Environmental impacts on clinical outcomes and the epigenome in patients with fibrotic interstitial lung disease

by

Gillian C. Goobie

MD, University of Calgary, 2014

FRCPC, Internal Medicine, University of Calgary, 2017

FRCPC, Respiratory Medicine, University of British Columbia, 2019

Submitted to the Graduate Faculty of the

Department of Human Genetics

School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Gillian C. Goobie

It was defended on

October 19, 2022

and approved by

John R. Shaffer, PhD, Assistant Professor of Human Genetics, School of Public Health, University of Pittsburgh

Ryan L. Minster, PhD, MSIS, Assistant Professor of Human Genetics, School of Public Health, University of Pittsburgh

Daniel J. Kass, MD, Associate Professor of Medicine, School of Medicine, University of Pittsburgh

S. Mehdi Nouraie, MD, PhD, Associate Professor of Medicine, School of Medicine, University of Pittsburgh

Dissertation Director: Yingze Zhang, PhD, Professor of Medicine and Human Genetics, Schools of Medicine and Public Health, University of Pittsburgh

Environmental impacts on clinical outcomes and the epigenome in patients with fibrotic interstitial lung disease

Gillian C. Goobie, PhD

University of Pittsburgh, 2022

Fibrotic interstitial lung diseases (fILDs) are a group of conditions characterized by lung scarring, functional limitation, and a high morbidity and mortality. Air pollution and socioeconomic disadvantage have been linked with adverse outcomes in idiopathic pulmonary fibrosis (IPF), the most common form of fILD, but these impacts have not been evaluated in large, geographically-diverse cohorts of patients with fILD. This dissertation sought to evaluate the impact of particulate matter with a diameter of ≥2.5um ($PM_{2.5}$) and neighborhood-level disadvantage on clinical outcomes, genomic, and epigenomic mechanisms of disease in a diverse cohort of patients with fILD.

Aim 1 demonstrated that patients with fILD who live in neighborhoods with greater disadvantage in the U.S., but not Canadian cohort, experience increased mortality and lower odds of receiving lung transplant. This work highlights how health disparities may be exacerbated by healthcare system structure, and raises questions of how environmental factors in disadvantaged neighborhoods contribute to these adverse outcomes. Aim 2 demonstrated that increased exposures to $PM_{2.5}$ and its anthropogenic constituents (particularly sulfate, nitrate, and ammonium) are associated with increased mortality and worse lung function in fILDs. Aim 3 demonstrated that $PM_{2.5}$ and its constituents can influence DNA methylation (DNAm) patterns and telomere length in patients with fILD, and that these molecular changes may mediate $PM_{2.5}$-mortality associations. We found that increased $PM_{2.5}$, sulfate, and ammonium exposures were associated with higher

global DNAm using an ELISA-based assay. Then, epigenome-wide association studies identified multiple CpGs associated with high exposures to $PM_{2.5}$. Most significant CpGs were found in analyses of sulfate, ammonium, and sea salt $PM_{2.5}$ constituents, highlighting the potentially greater mechanistic relevance for these components of the $PM_{2.5}$ mixture. Lastly, we found that higher exposure to $PM_{2.5}$ and anthropogenic constituents was associated with shorter telomere length, which mediates a portion of the $PM_{2.5}$-mortality association.

By highlighting environmental impacts on clinical outcomes and molecular mechanisms of disease, we have unveiled potential causal pathways for how such exposures contribute to the development and progression of fILDs. This research has direct impacts on public health by providing critical data to inform environmental health policies that protect vulnerable populations.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

%5mC – percentage 5-methylcytosine (measure of global DNA methylation)
ADI – area deprivation index
AE – acute exacerbation
AJRCCM – American Journal of Respiratory and Critical Care Medicine
ARF – attributable risk fraction
BC – black carbon
CARE-PF – Canadian Registry for Pulmonary Fibrosis
CIMD – Canadian Index of Multiple Deprivation
CpG – cytosine-guanine dinucleotide
CRTAP – cartilage-associated protein gene
CTD-ILD – connective tissue disease-related interstitial lung disease
CO – carbon monoxide
COPD – chronic obstructive pulmonary disease
D_LCO – diffusion capacity of the lung for carbon monoxide
DMP – differentially methylated probe
DMR – differentially methylated region
DNAm – deoxyribonucleic acid methylation
EC – elemental carbon
ECM – extracellular matrix
ELISA – enzyme-linked immunosorbent assay
EPA – United States Environmental Protection Agency
FDR – false discovery rate
fHP – fibrotic hypersensitivity pneumonitis
fILD – fibrotic interstitial lung disease
FVC – forced vital capacity
GSEA – gene set enrichment analysis
GWAS – genome wide association study
HAA – high attenuation abnormalities
HR – hazard ration
HP – hypersensitivity pneumonitis
ICD-9-CM – International Classification of Diseases, ninth revision, clinical modification
IIP – idiopathic interstitial pneumonia
ILA – interstitial lung abnormality
ILD – interstitial lung disease
IPF – idiopathic pulmonary fibrosis
IQR – interquartile range
MESA – Multi-ethnic study on atherosclerosis
miRNA – micro RNA
mRNA – messenger RNA
MUC5B – mucin 5B gene
NH_4^+ – ammonium
NO_2 – nitrogen dioxide
NO_3^- –nitrate
NO_x – nitrogen oxides

O_3 – ozone
OM – organic matter
OR – odds ratio
PFF – Pulmonary Fibrosis Foundation
PM_{10} – particulate matter with a diameter ≤10μm
$PM_{2.5}$ – particulate matter with a diameter ≤2.5μm
Q1 – quartile one
Q2 – quartile two
Q3 – quartile three
Q4 – quartile four
RA-ILD – rheumatoid arthritis-associated interstitial lung disease
SES – socioeconomic status
SNP – single nucleotide polymorphism
SO_2 – sulfur dioxide
SO_4^{2-} – sulfate
SS – sea salt
TL – telomere length
UBC – University of British Columbia
UCLA – University of California Los Angeles
UPitt – University of Pittsburgh
UPMC – University of Pittsburgh Medical Center
U.S. – United States
WGS – whole genome sequencing

Preface

This dissertation is the culmination of multiple years of effort by myself and countless hours of support and contributions from my supervisors, collaborators, and coauthors. Dr. Yingze Zhang has supported me from the beginning in pursuing the research that most interested me. She has taught me so much about genetics, scientific inquiry, and translational research, while also being a great confidant, friend, and someone to always discuss outdoor adventures with. Dr. Mehdi Nouraie has been an incredible mentor and teacher, helping me throughout every aspect of this work, tirelessly reading my endless R markdown documents, and always providing me with confidence in my abilities. Dr. Daniel Kass has been such a strong advocate for my work, providing clinical expertise and career mentorship that has been critical for my progression at this stage in my training. Dr. Chris Ryerson, my longstanding mentor from the University of British Columbia (UBC), encouraged me to pursue this path and has been a stalwart supporter of my work from the very beginning, emulating the type of researcher, mentor, and clinician I would like to be.

I also could not have accomplished any of this without the other collaborators and mentors that have contributed their time and insights to this work. This includes but is not limited to Drs. Chris Carlsten, James Fabisiak, Kevin Gibson, Xiaoyun Li, Jon Alder, Kerri Johannson, Kathleen Lindell, all other Canadian Registry for Pulmonary Fibrosis (CARE-PF) investigators, Pulmonary Fibrosis Foundation (PFF) Patient Registry investigators, Ms. Xiaoping Chen and Melinda Klesen, all the members of Dr. Zhang's Translational Research Core Laboratory, the University of Pittsburgh Division of Pulmonary, Allergy, and Critical Care Medicine, and of course all the wonderful patients and staff from the Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease. I would like to thank the UBC Clinician Investigator Program (CIP), including

Tessa Feuchuk and Dr. Sian Spacey, for their support during the first two years of my PhD, without which I would have been unable to pursue this incredible opportunity. I would also like to thank the University of Pittsburgh Department of Human Genetics for the excellent education and support throughout my PhD. Thank you to Drs. John Shaffer and Ryan Minster for being on my dissertation committee and for your exceptional teaching throughout these past three years.

Of course, I would have been unable to accomplish any of this without the love and support of my incredible friends and family. To my husband, Sean, and our sweet rescue dog Coal, our shared love, encouragement, and laughter has helped me to get to the end of this PhD. Sean, thank you for always listening to my rants, helping me to figure out coding nightmares, and for always adventuring with me. To my parents, sister, and dear friends, thank you for all for the time you've each spent listening to me throughout my PhD, willingly or otherwise, while I've navigated my way through my research during this crazy period of time. I'm blessed with the best support system in the world, and I can't thank each of you enough for all you've done to help me get here.

Because of the enduring support of my incredible mentors and supervisors, as well as my wonderful family and friends, I had the opportunity to work on important, novel, and highly impactful questions during my PhD that I hope will influence the lives of patients with fibrotic interstitial lung disease and environmental health policies.

1.0 Introduction

Fibrotic interstitial lung diseases (fILDs) are a group of conditions characterized by lung scarring, progressive shortness of breath, and a high morbidity and mortality.(Wong, Ryerson, & Guler, 2020) The most common form of fILD is idiopathic pulmonary fibrosis (IPF), which has a median survival of 3-5 years from the time of diagnosis.(Raghu et al., 2011, 2018, 2015) The term "idiopathic" implies that the etiology underlying the development of IPF is unknown, however substantial research efforts have unveiled genetic, occupational, and demographic risk factors that contribute to IPF incidence.(Wolters et al., 2018) Other forms of fILD like connective tissue disease-ILD (CTD-ILD) and fibrotic hypersensitivity pneumonitis (fHP) have more clear underlying etiologies, including inflammation in the pulmonary interstitium related to autoimmune disease or antigen exposure, respectively.(Wong et al., 2020) What remains unknown for most of these fILDs, however, is how environmental and sociodemographic factors like neighborhood disadvantage and air pollution exposure contribute to the development and progression of these devastating conditions. Furthermore, few studies have explored how such environmental exposures contribute to disease pathobiology through alteration of genomic and epigenomic factors like telomere length and DNA methylation (DNAm).

With new World Health Organization recommendations for annual exposures to airborne particulate matter with a mean diameter of less than 2.5μm ($PM_{2.5}$), over 99% of the world's population now live in areas where air quality recommendations are not met.("Air pollution," 2022) In 2015 alone, 4.2 million deaths (7.6% of annual global deaths) were attributed to $PM_{2.5}$.(Cohen et al., 2017) In North America, air pollution disproportionately affects individuals living in neighborhoods with greater socioeconomic disadvantage, potentially explaining a portion

of the staggering health disparities that exist in the United States (U.S.).(Hajat, Hsia, & O'Neill, 2015) Both air pollution and living in neighborhoods with greater disadvantage have serious adverse impacts on patients with chronic respiratory conditions like asthma and chronic obstructive pulmonary disease (COPD) (Alfano et al., 2018; Bowe, Xie, Yan, & Al-Aly, 2019; Sahni, Talwar, Khanijo, & Talwar, 2017; Sbihi, Koehoorn, Tamburic, & Brauer, 2017; To et al., 2016), however the effects of these factors on patients with fILDs has not been previously.

My PhD dissertation addresses a critical knowledge gap in our understanding of the demographic and environmental factors that contribute to the adverse clinical outcomes and pathobiologic mechanisms of disease in patients with fILD. This novel and collaborative multi-national project evaluated how neighborhood disadvantage and $PM_{2.5}$ exposures contributed to adverse clinical outcomes and epigenetic changes in patients with fILD.

The following sections will provide a background to introduce fILDs, neighborhood disadvantage, air pollution, and the impact of air pollution on the epigenome in this population. This includes a review manuscript (Section 1.3), which I first-authored, that summarizes the recent literature linking air pollution with adverse clinical outcomes in patients with fILD and outlines how epigenetic changes may mediate these associations. It further outlines the steps that can be taken to explore the epigenetic impacts of environmental exposures in this population, which provides context and justification for the third aim of my PhD.

Section 1.4 that follows will describe the central hypotheses and specific aims of my PhD. Subsequent sections include the primary aims and manuscripts that have been published related to these aims.

1.1 Global Burden of Interstitial Lung Diseases

The global burden of fILDs is increasing. It is estimated that over 6 million people worldwide suffer from some form of fILD or pulmonary sarcoidosis, with a 43% increase in potential years of life lost from fILDs between 2007-2017.(GBD 2017 Disease and Injury Incidence and Prevalence Collaborators, 2018a, 2018b) The most common form of fILD is IPF, which represents 10-40% of ILD,(Duchemann et al., 2017; Ford-Sahibzada, Johannson, Goobie, & Fell, 2017) and has a median survival of 3-5 years from the time of diagnosis.(Raghu et al., 2011) The healthcare burden of IPF is significant, with estimates of average yearly healthcare costs of nearly $60,000 USD per patient and approximately 7,000 non-transplant hospital admissions of patients with IPF per year at an average cost of $16,000 USD per hospital admission.(Mooney, Raimundo, Chang, & Broder, 2017; Raimundo et al., 2016) Similarly, patients with various forms of fILD have demonstrably reduced employment and workplace productivity when compared with age- and sex-matched controls, further emphasizing the substantial economic burden of these conditions.(Algamdi et al., 2019) Because of the high cost borne by patients with fILD, personal finances and other sociodemographic factors may influence the choices that patients make, including where they receive their medical care or if they pursue expensive therapies such as antifibrotics or lung transplantation.(Gaffney, Woolhander, Himmelstein, & Mccormick, 2018) As such, it is critical that we develop a better understanding of how access to care and other social and environmental factors impact outcomes in patients with ILD.

1.2 Neighborhood Disadvantage and Interstitial Lung Disease

Low socioeconomic status (SES) and higher neighborhood-level disadvantage have been shown to have substantial adverse effects on clinical outcomes in chronic respiratory diseases like asthma and COPD.(Gershon, Dolmage, Stephenson, & Jackson, 2012; Hee Cho et al., 2016; Sahni et al., 2017) Despite these findings, there has been little research evaluating the impact of neighborhood disadvantage or other sociodemographic factors on clinical outcomes in fILDs.(Sesé, Cavalin, Bernaudin, Maesano, & Nunes, 2020) One small study of 52 patients with rheumatoid arthritis-associated ILD (RA-ILD) found that the hazard ratio for mortality was nearly doubled for patients with low SES.(Koduri et al., 2010) Data from the Nationwide Inpatient Sample demonstrated that individuals with IPF who ranked in the lowest quartile of SES (estimated from residential zip code data) were less likely to receive lung transplantation.(Gaffney et al., 2018) Additionally, non-Hispanic Black and Hispanic patients with IPF have significantly higher mortality while wait-listed on lung transplantation lists in comparison to matched White patients with IPF, emphasizing how sociodemographic disparities may affect clinical outcomes in patients with fILD.(Lederer et al., 2006) These discrepancies persisted after adjustment for individual SES, indicating that other environmental factors such as access to care may mediate a component of these adverse outcomes. A major limitation of these studies is that the definitions of SES or neighborhood-level disadvantage are not consistent or comprehensive.

The first aim of my PhD evaluated whether patients with fILD who live in neighborhoods with greater levels of disadvantage experience more adverse clinical outcomes as compared to patients with fILD who live in less disadvantaged areas. This was the first study to explore the impact of neighborhood-level disadvantage on to clinical outcomes in patients with fILD.

1.3 Air Pollution Impacts on Clinical Outcomes and the Epigenome in fILDs

The introduction for the second and third aims of my PhD, exploring the impact of $PM_{2.5}$ pollution on clinical outcomes and the epigenome in patients with fILD, is captured in the following review manuscript.

1.3.1 *AJRCCM* Review Manuscript – "Air Pollution and Interstitial Lung Diseases: Defining Epigenomic Effects"

Cite: Goobie, GC, *et al*. Air Pollution and Interstitial Lung Diseases: Defining Epigenomic Effects. *Am J Respir Crit Care Med* 2020; 202(9):1217-1224. DOI: 10.1164/rccm.202003-0836PP. PMID: 32569479.(Goobie et al., 2020)

The American Journal of Respiratory and Critical Care Medicine is an official journal of the American Thoracic Society.

Air pollution and interstitial lung diseases: defining epigenomic effects

Gillian C. Goobie[1,2], Mehdi Nouraie[3], Yingze Zhang[1,3], Daniel J. Kass[3], Christopher J. Ryerson[4,5], Christopher Carlsten[4,5], Kerri A. Johannson[6]

1. Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, U.S.A.
2. Department of Medicine, University of British Columbia, Vancouver, B.C., Canada.
3. Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A.
4. Division of Respiratory Medicine, Department of Medicine, University of British Columbia, Vancouver, B.C., Canada.
5. Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, B.C., Canada.

6. Division of Respiratory Medicine, Department of Medicine, University of Calgary, Calgary, AB, Canada.

Correspondence to:
Gillian C. Goobie, MD FRCPC
Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh
130 DeSoto Street, Pittsburgh, PA, 15261
Email: goobiegc@upmc.edu; Phone: 412-624-3018; Fax: 412-624-3020

Author Contributions: GCG & KAJ are responsible for initial manuscript production. MN, YZ, DJK, CJR and CC provided critical commentary and initial review of the manuscript. MN provided insight about biostatistical and epidemiologic methods. CC and YZ provided insight about epigenetic changes from air pollution and epigenetic methods. DJK, CJR, and KAJ provided interstitial lung disease clinical research expertise.

Support: GCG and MN do not have current grant or industry support relevant to this publication. YZ is supported in part by AR076024 NIH grants. DJK is supported in part by HL126990 NIH grants, Boehringer-Ingelheim grants, and receives collaborative research funding from Regeneron Pharmaceuticals, outside of the submitted work. CJR reports grants and personal fees from Boehringer-Ingelheim and Hoffman-La Roche, and grants from Michael Smith Foundation for Health Research, outside of the submitted work. CC reports support from the Canada Research Chairs Program. KAJ reports personal fees from Boerhinger-Ingelheim, Hoffman-La Roche, Theravance, and Blade Therapeutics, and reports grant support from the CHEST Foundation, the Pulmonary Fibrosis Society of Calgary, UCB Biopharma SPRL, and personal fees from the Three Lakes Foundation.

Running Head: Air pollution impacts in interstitial lung disease.

Subject category: 3.05 (Epigenetics), 6.02 (Air Pollution: Mechanisms), 6.06 (Gene-Environment Interaction), 9.23 (Interstitial Lung Disease)

Body word count: 3323
Table count: 1
Figure count: 1

1.3.1.1 Abstract

Over the past decade, air pollution has been increasingly recognized as an environmental risk factor for the development and progression of common respiratory diseases. In interstitial lung disease, higher cumulative air pollution exposure is associated with increased incidence, rate of decline in lung function, number of acute exacerbations, and mortality. The molecular mechanisms underlying these adverse clinical outcomes remain largely unexplored. In this pulmonary perspective piece, we review the associations between air pollution and clinical outcomes in interstitial lung disease, our current state of knowledge about the epigenome in interstitial lung disease, and how epigenetic methods can be applied to evaluate the impact of air pollution on the epigenome of these patients. Increased knowledge of how environmental exposures modify the epigenome of patients with interstitial lung disease will help to identify biomarkers of exposure, strategies for prevention, and avenues for mitigation of the morbidity and mortality associated with these pollutants.

Abstract word count: 150

Keywords: interstitial lung diseases, idiopathic pulmonary fibrosis, air pollution, epigenetics.

1.3.1.2 Introduction

Air pollution is a massive global health problem, with over 90% of the world's population living in areas where daily exposures exceed the World Health Organization's air quality recommendations.(World Health Organization, 2018) Outdoor sources of particulate matter with a diameter <2.5μm ($PM_{2.5}$) caused between 4.2 and 8.9 million premature deaths in 2015 alone.(Burnett et al., 2018; Cohen et al., 2017) The burden of disability and mortality from air pollution exposure is disproportionately experienced by vulnerable populations and patients with chronic respiratory diseases.(Bowe et al., 2019; Cohen et al., 2017) The adverse effects of air pollution exposure are well-established in patients with chronic obstructive pulmonary disease (COPD) and asthma,(Sbihi et al., 2017; To et al., 2016) whereas the impact on patients with interstitial lung disease (ILD) remains poorly characterized.

Occupational and environmental exposures contribute to the development and progression of ILD, through mechanisms not yet fully understood.(Blanc et al., 2019) Several previous single-center studies have demonstrated that air pollution exposure is associated with increased incidence and adverse outcomes in idiopathic pulmonary fibrosis (IPF), the most common form of ILD.(Conti et al., 2018; Johannson et al., 2014, 2018a; Sesé et al., 2018; Winterbottom et al., 2018) Airborne pollutants have multiple deleterious physiologic effects in the lungs, such as triggering alterations to mucosal surfaces by overwhelming ciliary and macrophage clearance mechanisms, inducing oxidative stress, and by transiting toxic metals into the bloodstream.(Schraufnagel et al., 2019) One other mechanism whereby air pollution likely mediates adverse impacts in ILD and other diseases is through epigenetic modifications,(Alfano et al., 2018) referring to molecular mechanisms that regulate gene expression without changing nucleotide base sequences.

In this Pulmonary Perspective, we review the associations between air pollution exposure and adverse clinical outcomes in patients with ILD. Subsequently, we present an overview of the current understanding of the role of epigenetics in ILD. Lastly, we summarize how epigenetic methods can be adapted to explore how changes to the epigenome may mediate the adverse impacts of air pollution in ILD. Given the increasing global burden of ILD (GBD 2017 Disease and Injury Incidence and Prevalence Collaborators, 2018b) and recent increases in air pollution-related mortality across the United States (U.S.),(Clay & Muller, 2019) it is more important than ever to understand the molecular mechanisms relating air pollution exposure to ILD development and progression. Air pollution may play an important role in lung remodeling and fibrogenesis, such that targeting this environmental risk factor may help to reduce the development and progression of ILD. Research in this area will inform pathophysiology, identify opportunities to reduce adverse impacts in at-risk individuals, and may guide policymakers who institute regulations on emissions standards and pollution mitigation strategies.

1.3.1.3 Air Pollution and Clinical Impacts in Interstitial Lung Disease

1.3.1.3.1 Preclinical Disease and Interstitial Lung Disease Incidence

Air pollution exposure is increasingly recognized as a risk factor for the development and progression of ILD (**Table 1**). Patients at risk for the development of ILD may be incidentally identified by interstitial lung abnormalities (ILA) or high attenuation abnormalities (HAA) on computed tomography (CT) scans of the chest. These subclinical features are associated with increased likelihood of ILD diagnosis and mortality.(Araki et al., 2016; Podolanczuk et al., 2016) One study of healthy individuals from the Multi-Ethnic Study of Atherosclerosis (MESA) found that increased 10-year nitrogen oxides (NO_x) exposure was associated with higher odds of ILA

incidence.(Sack et al., 2017) Another study involving healthy individuals enrolled in the Framingham Heart Study found that increased 5-year elemental carbon (EC) exposure was associated with increased odds of ILA and ILA progression.(Rice et al., 2019) This preliminary evidence suggests cumulative air pollution exposures are linked to pre-clinical ILD. Further study is required to evaluate whether pollution modifies the risk of progression to ILD in patients with ILAs or HAAs, how underlying genetics influence these risks, and whether imaging studies can be used for early identification of subclinical ILD in high risk populations with significant environmental exposures.

Exposure to airborne pollutants may be associated with ILD incidence. One study investigated the impact of average daily exposure to particulate matter with a diameter <10μm (PM_{10}), nitrogen dioxide (NO_2), and ozone (O_3) on IPF incidence in Northern Italy.(Conti et al., 2018) In unadjusted models, increased NO_2 concentration was associated with IPF incidence during the cold season, although not statistically significant on multivariable analysis. No significant association was found between PM_{10} or O_3 exposure and IPF incidence. This was the first study to evaluate the impact of air pollution on ILD incidence, but it was limited by only evaluating a small geographic region with limited heterogeneity in air pollution exposures. This emphasizes the need to expand this type of methodology to larger and more heterogenous populations while considering more comprehensive multi-pollutant models including the criteria pollutants $PM_{2.5}$, PM_{10}, NO_2, sulfur dioxide (SO_2), carbon monoxide (CO), O_3, and lead simultaneously.

Another recent study in India found a significant association between increased city-wide levels of $PM_{2.5}$ and percentage of cases of hypersensitivity pneumonitis (HP) enrolled in that center's ILD registry.(Singh et al., 2019) The authors postulate that exposure to airborne pollutants

may impair mucociliary clearance, leading to antigen retention and initiation of the immunologic and fibrogenic pathways contributing to the development of HP. Although not a formal incidence study, these findings suggest that the impact of airborne pollutants on the development of non-IPF ILD warrants further study.

1.3.1.3.2 Effects on Established Interstitial Lung Disease

High air pollution exposure is recognized as a risk factor for adverse clinical outcomes and lung function decline in patients with IPF, although evidence in other ILDs is lacking. One single-center U.S. study of patients with IPF found that each 5μg/m^3 increase in six-year cumulative exposure to PM_{10} was associated with an additional 46mL decline in forced vital capacity (FVC) per year.(Winterbottom et al., 2018) Cumulative exposures to $PM_{2.5}$ were not associated with rate of decline in lung function. Another study using weekly spirometry in 25 patients with IPF demonstrated that higher weekly mean levels of NO_2, $PM_{2.5}$, and PM_{10} were each associated with lower mean FVC over the study period.(Johannson et al., 2018a) There was no association between airborne pollutant exposure and rate of decline in lung function, but this study was limited by small patient numbers and a finite follow-up duration. These physiological studies help to inform how air pollution contributes to disease progression and adverse clinical outcomes in ILD.

A French study of 192 patients with IPF is the only study to demonstrate a positive association between PM_{10} or $PM_{2.5}$ exposure and all-cause mortality.(Sesé et al., 2018) Increased average O_3 exposure was also associated with increased number of acute exacerbations (AEs), although no association was found between NO_2, $PM_{2.5}$, or PM_{10} and AEs. This positive association between air pollution exposure and AE rate was first described in a South Korean cohort of patients with IPF.(Johannson et al., 2014) The mean level, maximum level, and number of exceedances above accepted standards for O_3 and NO_2 over a six-week period preceding the

event was associated with increased incidence of AEs. Additional studies of these important clinical outcomes are needed to evaluate air pollution effects on these outcomes in patients with non-IPF ILD and to explore the molecular mechanisms underlying these effects in order to identify mitigation strategies.

Table 1 – Summary of selected clinical outcomes from air pollution exposure in patients with interstitial lung disease (ILD) and pre-clinical disease.

Population	Sample Size	Exposure Evaluated	Outcome Measured	Effect Size	Confidence Interval (CI)	P-value	Reference
Healthy U.S. subjects from MESA	671	40ppb increase in 10-year mean NO_x exposure	Odds of ILA	1.77	1.06, 2.95	**0.03**	Sack C. *et al.* 2017 (Sack et al., 2017)
	495		Percent increase in HAA per year	0.45%	-0.02%, 0.92%	0.06	
Healthy U.S. subjects from Framingham Heart Study	344	5-year EC exposure of 0.14μg/m^3	Odds of ILA	1.27	1.04, 1.55	NR	Rice M.B. *et al.* 2019 (Rice et al., 2019)
	709		Odds of ILA progression	1.33	1.00, 1.77	NR	
Incident cases of IPF in Northern Italy based on ICD-9-CM code 516.3	2093	10μg/m^3 increase in 5-year mean NO_2 exposure during cold season	Incidence rate of IPF	7.93%*	0.36%, 16.08%	NR	Conti S. *et al.* 2018 (Conti et al., 2018)
Newly diagnosed patients enrolled in ILD-India Registry	842	1μg/m^3 increase in mean annual $PM_{2.5}$	Odds of being diagnosed with HP over other forms of ILD	1.007	1.001, 1.013	**0.017**	Singh S. *et al.* 2019 (Singh et al., 2019)
IPF patients enrolled in French ILD COhort FIbrose (COFI)	192	10μg/m^3 increase in preceding 6-week mean O_3	Hazard ratio for AE event	1.47	1.13, 1.92	**0.005**	Sesé *et al.* 2018 (Sesé et al., 2018)
		10μg/m^3 increase in mean PM_{10} from inclusion to death, transplant, or censoring	Hazard ratio for mortality	2.01	1.07, 3.77	**0.03**	
		10μg/m^3 increase in mean $PM_{2.5}$ from inclusion to death, transplant, or censoring	Hazard ratio for mortality	7.93	2.93, 21.33	**<0.001**	
IPF patients enrolled in longitudinal ILD cohort in Seoul, South Korea	436	Increased mean O_3 over 6-week exposure period	Hazard ratio for AE event	1.57	1.09, 2.24	**0.01**	Johannson *et al.* 2014 (Johannson et al., 2014)
		Increased maximum O_3 over 6-week exposure period		1.42	1.11, 1.82	**0.01**	
		Increased number of exceedances above air quality standards for O_3 over 6-week exposure period		1.51	1.17, 1.94	**0.002**	
		Increased mean NO_2 over 6-week exposure period		1.41	1.04, 1.91	**0.03**	
		Increased maximum NO_2 over 6-week exposure period		1.27	1.01, 1.59	**0.04**	
		Increased number of exceedances above air quality standards for NO_2 over 6-week exposure period		1.20	1.10, 1.31	**<0.001**	
IPF patients seen at single U.S. center	135	5μg/m^3 increase in mean PM_{10} from enrollment to death, transplant, or censoring	Rate of decline in FVC	46mL/yr increased FVC decline	12mL/yr, 81mL/yr	NR	Winterbottom *et al.* 2018 (Winterbotto

							m et al., 2018)
IPF patients at single U.S. center given home spirometers	25	1ppb increase mean NO_2 over study period (up to 40 weeks)	Difference in mean FVC % predicted over study period (measured weekly)	-0.45%	-0.85%, -0.05%	**0.03**	Johannson *et al.* 2018 (Johannson et al., 2018a)
		1μg/m^3 increase mean $PM_{2.5}$ over study period (up to 40 weeks)		-0.45%	-0.84%, -0.07%	**0.02**	
		1μg/m^3 increase mean PM_{10} over study period (up to 40 weeks)		-0.57%	-0.92%, -0.21%	**0.003**	

Abbreviations: AE (acute exacerbation); FVC (forced vital capacity); EC (elemental carbon); HAA (high attenuation abnormalities); ICD-9-CM (International Classification of Diseases, ninth revision, clinical modification); ILA (interstitial lung abnormalities); ILD (interstitial lung disease); IPF (idiopathic pulmonary fibrosis); MESA (Multi-ethnic study on atherosclerosis); NO_x (nitrogen oxides); NR (not reported); ppb (parts per billion); U.S. (United States). *Unadjusted analysis. Results did not meet significance in multivariate analysis.

1.3.1.4 The Epigenome in Interstitial Lung Diseases

Despite the increasing body of literature linking exposure to multiple airborne pollutants with adverse outcomes in IPF, there remain critical knowledge gaps in the mechanisms underlying these relationships. Epigenetic mechanisms are prime candidates for evaluation given known alterations to the epigenome in patients with ILD and known epigenetic impacts of airborne pollutants in healthy individuals and in patients with other chronic respiratory diseases.(Alfano et al., 2018; Tzouvelekis & Kaminski, 2015; Yang et al., 2014) The most commonly studied epigenetic mechanisms are DNA methylation (DNAm), histone modifications, and non-coding RNAs, especially micro RNAs (miRNA). Epigenomic patterns are inherited between cells, but environmental exposures throughout a lifetime can significantly change one's epigenetic landscape.(Toraño, García, Luis Fernández-Morera, Niño-García, & Fernández, 2016) Most epigenetic factors have been studied in IPF,(Tzouvelekis & Kaminski, 2015) but the role air pollution plays in altering these factors remains unknown. To understand the epigenetic impacts of air pollution in ILD, it is important to first understand how the epigenome is altered in these patients.

Widespread alterations in DNAm patterns occur in lung tissue from patients with IPF in comparison to controls.(Yang et al., 2014) Some of these alterations occur near genes implicated in IPF pathogenesis, such as *TOLLIP*, *NOTCH1*, and *FBXO32*. Altered gene expression nearby these differentially methylated regions was found in these and other IPF-relevant genes, supporting the notion that changes in DNAm may mediate adverse mechanisms that contribute to IPF development and progression. Plasma cell-free DNAm patterns can be used with moderate specificity to distinguish between patients with fibrotic ILD and lung cancer or COPD,(Wielscher

et al., 2015) illustrating how DNAm patterns may represent novel diagnostic biomarkers in patients with ILD.

Histone modifications also have pathophysiologic relevance in pulmonary fibrosis. In bleomycin mouse models of pulmonary fibrosis and in IPF-derived fibroblasts, histone modifications are associated with alterations in apoptotic pathways.(Huang et al., 2013) Inhibiting histone deacetylase, which leads to alterations in histone modification and DNAm patterns, results in increased fibroblast apoptosis and prolonged survival in bleomycin-injured mice.(Sanders et al., 2014) These data support that resistance to apoptosis, which is thought to represent a major pathophysiologic mechanism in fibroproliferative diseases like IPF, is mediated in part by epigenetic changes.(Bagnato & Harari, 2015) Histone modifications in circulating nucleosomes have also been used to distinguish between serum from healthy subjects and patients with IPF,(Guiot et al., 2017) emphasizing its potential utility as a biomarker of disease.

Multiple previous studies, as reviewed elsewhere,(Tzouvelekis & Kaminski, 2015) have demonstrated that non-coding RNAs play an important role in ILD, through regulation of fibroblast proliferation, pro- and anti-fibrotic pathways, and as potential disease biomarkers and therapeutic targets.

1.3.1.5 Investigating the Epigenetic Effects of Air Pollution in Interstitial Lung Disease

Air pollution may contribute to ILD development and progression by altering the epigenome in ways that lead to upregulation of aberrant inflammatory or pro-fibrotic responses. Air pollution has been shown to impact each of the three main types of epigenetic mechanisms *in vitro*, *in vivo*, in healthy individuals, and in people with chronic diseases. The known mechanisms whereby air pollution modifies the epigenome are illustrated in the **Figure 1**.(Alfano et al., 2018) Little is known about the impact of air pollution on the epigenome of patients with ILD, and a

systematic approach is required to gain a comprehensive understanding of these relationships. DNAm is the most frequently evaluated epigenetic marker and would be a natural starting point, followed by studies evaluating histone modifications, non-coding RNA patterns, and gene-environment interactions. Research approaches range from *in vitro* studies of ILD-relevant cell types, to *in vivo* models of pulmonary fibrosis, to observational and experimental studies in patients with ILD.

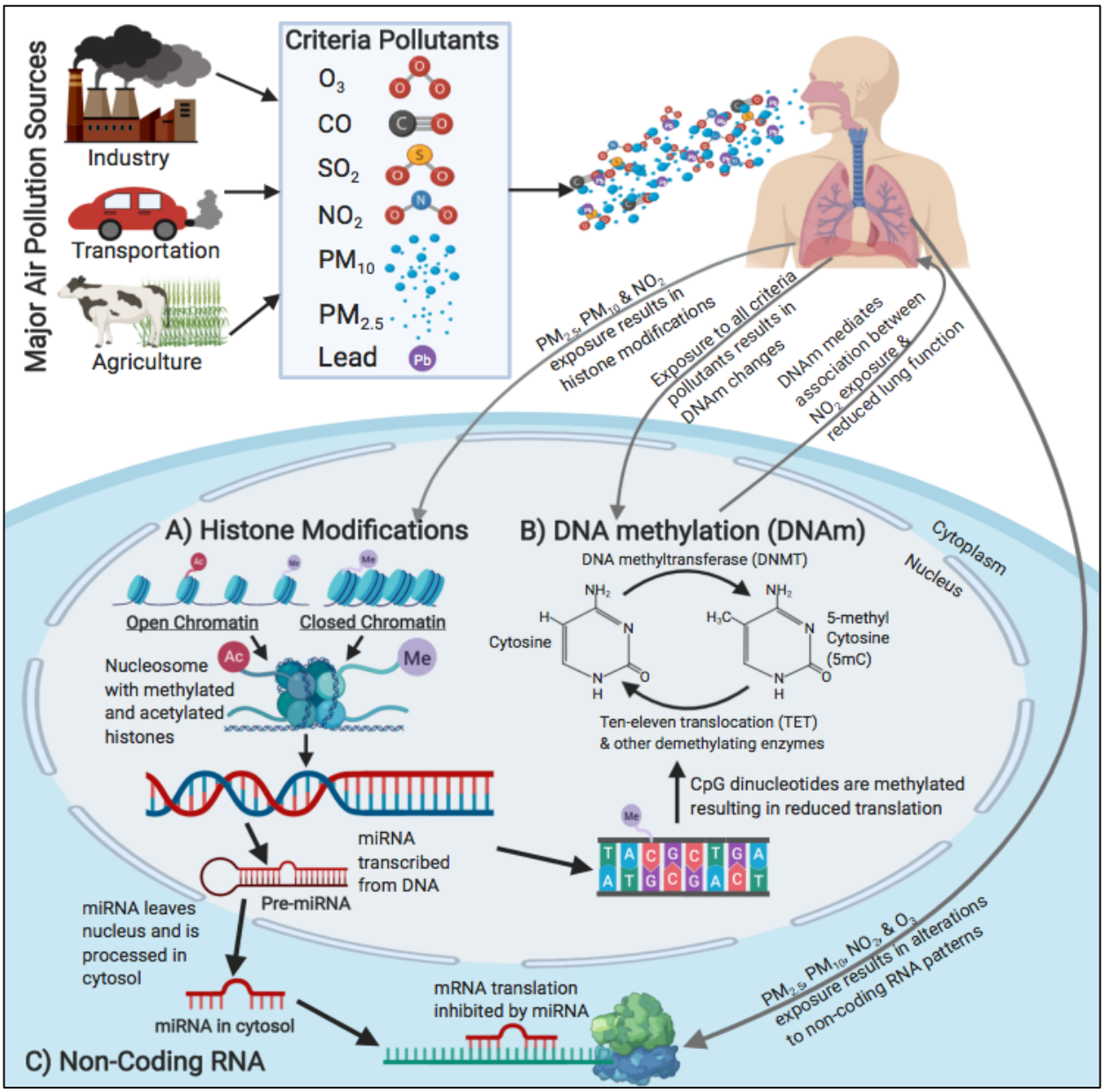


Figure 1 – Major air pollution sources, criteria air pollutants (ozone, carbon monoxide, sulfur dioxide, nitrogen dioxide, particulate matter and lead) and how these affect the three primary epigenetic mechanisms. A: Histone modifications (methylation and acetylation) regulate whether chromatin is open or closed and thus accessible to transcriptional machinery. **B:** DNA methylation (DNAm) at cytosine-guanine dinucleotide (CpG) sites generally reduces nearby gene expression. Alterations to DNAm have been shown to mediate the association between NO_2 exposure and reduced pulmonary function. **C:** The most studied non-coding RNA is micro RNA (miRNA), which influences gene expression post-transcriptionally by impairing messenger RNA (mRNA) translation. Figure made with BioRender.

1.3.1.5.1 In Vitro Methods

In vitro methods can be used to investigate how airborne pollutants contribute to epigenetic changes and disease mechanisms on a cellular level. A study using human bronchial epithelial cells found that $PM_{2.5}$ exposure resulted in globally reduced DNAm, site-specific histone modifications, shortened telomere length, and altered telomerase activity in a concentration and exposure-dependent manner, especially in cells derived from patients with COPD.(Leclercq et al., 2017) These telomere findings have been validated in human studies, with a recent meta-analysis finding that each $5\mu g/m^3$ increase in $PM_{2.5}$ exposure is associated with -0.03 (relative units) shorter telomeres.(Miri et al., 2019) This may have important implications given the pathophysiologic relevance of short telomeres in multiple forms of ILD.(Courtwright & El-Chemaly, 2019) Recent studies in nasal mucociliary epithelial cells have demonstrated changes in gene expression profiles in response to treatment with $PM_{2.5}$ organic extract.(Montgomery et al., 2020) Similar studies should be performed to delineate the transcriptomic and epigenomic responses to airborne pollutants in alveolar epithelial cells, fibroblasts and immune cells derived from normal controls and patients with ILD. This will help to clarify on a molecular level how air pollution triggers immune dysregulation and fibrogenesis.

1.3.1.5.2 In Vivo Methods

Model organisms, for example bleomycin lung-injured mice, can be used to evaluate how exposure to airborne pollutants affects disease pathophysiology *in vivo*. One study exposed rats to traffic-related air pollutants and demonstrated a dose- and time-responsive change in DNAm and histone modifications at multiple specific regions across the genome in both blood and lung tissue in molecular pathways relevant to chronic respiratory diseases.(R. Ding et al., 2017) The *in vivo* evaluation of air pollution impacts on fibrosis is limited by the lack of existing animal models of

pollutant-induced fibrosis. Aged or genetically modified model organisms could be used to better simulate patient characteristics of ILD.

1.3.1.5.3 Observational Genomic & Epigenomic Methods

Genomic and epigenomic patterns can be investigated at specific loci of known pathogenic relevance to ILD. For example, DNAm has been investigated at one key transcription factor of regulatory T-cells that is important in both IPF and asthma, Forkhead box transcription factor 3 (*FOXP3*).(Kotsianidis et al., 2009) Two studies have demonstrated altered DNAm at regulatory cytosine-guanine dinucleotide (CpG) sites of the *FOXP3* locus in relation to ambient exposures to polyaromatic hydrocarbons, NO_2, CO, and $PM_{2.5}$ in asthmatics.(Nadeau et al., 2010; Prunicki et al., 2018) Altered DNAm in these and other related pathways may play an important role in immune dysregulation that contributes to the development of ILD.

Genome-wide DNAm studies can incorporate an assessment of air pollution exposures and permit mediation analyses to deduce whether adverse outcomes are mediated by epigenetic changes. Mediation analysis considers the effect of a mediator variable that more clearly explains the relationship between exposure and outcome.(VanderWeele, 2016) A similar analysis was performed in healthy subjects where alterations in DNAm patterns were found to mediate NO_2-induced reductions in lung function.(de F.C. Lichtenfels et al., 2018) Other studies have evaluated genome-wide histone acetylation profiles in healthy individuals and noted alterations in the histone landscape in individuals with higher $PM_{2.5}$ exposure.(Liu et al., 2015) This approach could be adapted by performing global DNAm or histone modification analyses on ILD patients with high risk air pollution exposures, thereby identifying modifiable epigenetic regions that may lead to fibrogenesis or immune dysregulation. Previously performed genome-wide association studies can also be re-evaluated to investigate for the presence of any interactions between significant single

nucleotide polymorphisms (SNPs) and exposure to specific airborne pollutants, as was previously done for childhood asthma.(Gref et al., 2017) These approaches represent feasible extensions of prior research that can be performed using biologic samples collected from large ILD patient registries and biobanks.

1.3.1.5.4 Experimental Methods

Experimental approaches may provide evidence of causal associations between air pollution exposure, epigenetic changes, and outcomes in ILD. One approach is to evaluate how individual-level interventions, such as personal air quality monitors or air purifiers, modulate epigenomic responses and clinical outcomes in response to air pollution exposure.(Christopher Carlsten, Salvi, Wong, & Chung, 2020; Janjua, Powell, Atkinson, Stovold, & Fortescue, 2019) One case-control study in Beijing provided healthy subjects with personal air pollution monitors and found that increased PM_{10} exposure was associated with decreased histone H3 methylation.(Zheng et al., 2016) A randomized, double-blind crossover trial provided 36 healthy adults in Shanghai with air purifiers to lower personal indoor $PM_{2.5}$ exposure.(H. Li et al., 2018) Peripheral blood genome-wide DNAm was analysed before and after the air purifier intervention and significant alterations were noted at 49 CpG loci, with involvement of inflammatory, oxidative stress, cell survival, and apoptosis pathways. Similar analyses could be performed in patients with ILD, with repeated blood sampling during high and low pollution periods to assess for altered epigenomic patterns as a consequence of exposure. A novel systematic review is currently underway exploring the role of individual interventions aimed at reducing the adverse impacts of air pollution exposure in patients with chronic respiratory diseases.(Janjua et al., 2019) It is essential that healthcare providers remain appraised of interventions that can help protect vulnerable patient populations from the harmful impacts of air pollution exposure.

Experimental approaches have also been undertaken whereby patients with asthma are exposed to diesel exhaust followed by a bronchoscopy to investigate the impact on epigenetic patterns.(Clifford et al., 2017; Jiang, Jones, Sava, Kobor, & Carlsten, 2014; Rider et al., 2016) A double-blind, randomized crossover study exposed 13 asthmatics to diesel exhaust and found alterations in expression of multiple miRNAs, and subsequent downregulation of anti-oxidant pathways.(Yamamoto et al., 2013) These effects were attenuated by the addition of N-acetylcysteine, suggesting a potential role for this anti-oxidant in mitigating the effects of traffic-related air pollution in asthmatics. Similar evaluations of the impact of N-acetylcysteine on pollution-induced epigenetic changes in ILD would be useful given the potential efficacy of this drug in some patients with ILD.(Oldham et al., 2018) The role of other anti-oxidant therapies, such as B-vitamins, is also worth further study given their beneficial effect on air pollution-induced DNAm changes in CD4+ T-cells.(Zhong et al., 2017)

One crossover study aimed to determine if interactions exist between exposure to diesel exhaust and allergens.(Clifford et al., 2017) Similar interaction analyses could be performed in patients with ILD, looking at the interaction between smoking, sociodemographic factors, or occupational exposures and air pollution exposures on epigenetic patterns in these patients. These approaches will be essential in helping to elucidate the complex network of interactions that occur between the "exposome" and the epigenome in patients with ILD.(Vrijheid, 2014)

1.3.1.6 Limitations and Future Directions

Although our knowledge of the impact of air pollution and other environmental factors on the epigenome is rapidly expanding, there exist significant methodological and knowledge limitations. Previous studies have demonstrated low reproducibility of DNAm and other epigenetic patterns between different groups of patients with respiratory diseases.(Rider & Carlsten, 2019;

Tzouvelekis & Kaminski, 2015) This is likely due to different analysis methods, cell types, patient populations, and environmental exposures. Consistency in methods and correlation of epigenetic modifications with expression profiles and clinical outcomes in patients with ILD will be required. It is also imperative that we clarify the relationship between epigenetic patterns in peripheral blood and lung tissue, so that we may then explore how these patterns vary with air pollution exposure. Single cell DNAm sequencing & other single cell epigenetic techniques promise to address in part the impact of airborne pollutants on epigenetic patterns in ILD-relevant cell types.

Further research is needed to investigate the impact of air pollution on other forms of ILD, potentially considering fibrotic ILDs together given shared disease pathophysiology.(Cottin et al., 2019) Future studies should also consider utilizing more complex multi-pollutant analyses such as Bayesian Kernel Regression models to evaluate the effects of simultaneous exposures to multiple airborne pollutants.(Coull et al., 2015) This should include analyses of the major criteria pollutants ($PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, O_3, and lead), atmospheric heavy metals, and polychlorinated or polybrominated pollutants, as these have all been associated with adverse impacts on lung function and respiratory disease development.(Carpenter, Ma, & Lessner, 2008; S. Hansen et al., 2016)

Recent U.S. evidence suggests that $PM_{2.5}$-associated deaths most affect individuals living in neighborhoods with greater socioeconomic deprivation and non-Hispanic Black or African American populations.(Bowe et al., 2019) This demonstrates the concept of environmental justice and emphasizes the need to consider potential interactions or confounding by sociodemographic factors in air pollution research. Investigation of interaction effects between multiple airborne pollutants, other environmental exposures (e.g. cigarette smoke, allergens, socioeconomic factors), and genetic or epigenetic factors will paint a more detailed picture of how the “miasma” of airborne

pollutants contribute to disease pathophysiology.(Christopher Carlsten, 2018; Coull et al., 2015; M. G. Jones & Richeldi, 2014)

Future directions will also involve validating epigenetic modifications as biomarkers of air pollution exposure in patients with ILD. Given the potentially reversible nature of epigenetic modifications, these mechanisms have potential as prognostic and therapeutic targets to mitigate the adverse impacts of air pollution in ILD. Large ILD patient registries and biobanks should be further developed with plans for multi-national collaborative efforts aimed at elucidating the multi-omic effects of air pollution on patients with ILD. These cooperative efforts will facilitate novel avenues for diagnosis, monitoring progression, and disease prevention. These approaches need to be undertaken in conjunction with public health policies aimed at reducing global air pollution exposures.

1.3.1.7 Conclusions

The burden of ILD is increasing worldwide,(Cohen et al., 2017) yet there remain substantial knowledge gaps in our understanding of the environmental risk factors contributing to the development and progression of these conditions. Recent research indicates that exposure to airborne pollutants is associated with increased incidence and adverse clinical outcomes in IPF. These data are still sparse and need to be validated in larger multi-center cohorts, utilizing multi-pollutant models and longer time periods of assessment. Additionally, we need to investigate the role that air pollution and other environmental exposures play in non-IPF ILDs and the potential interactions between different exposures that contribute to disease development.

Given recent findings that air pollution levels have been increasing across the U.S. since 2016, resulting in an additional 9,700 premature deaths attributable to air pollution in 2018 alone,(Clay & Muller, 2019) it is imperative that researchers understand the biologic mechanisms

whereby airborne pollutants contribute to disease. Epigenetic modifications are a likely mechanism through which air pollution can interfere with normal physiologic functions. Exploring the impact of air pollution on the epigenome of patients with ILD will provide critical insights into how environmental factors contribute to the development and progression of these highly morbid conditions. Increased understanding of the genome-epigenome-environment interactions in patients with ILD and other chronic diseases may enable prevention and mitigation strategies aimed at reducing the disease burden associated with environmental pollution.(Chris Carlsten et al., 2014) Central Hypothesis and Specific Aims

1.4 Central Hypothesis

The central hypothesis and specific aims for this PhD dissertation are shown in **Figure 2**. Subsequently, **Figure 3** is the directed acyclic graph that shows the clinical outcomes of interest and the specific pathways for statistical analyses that are performed throughout this project.

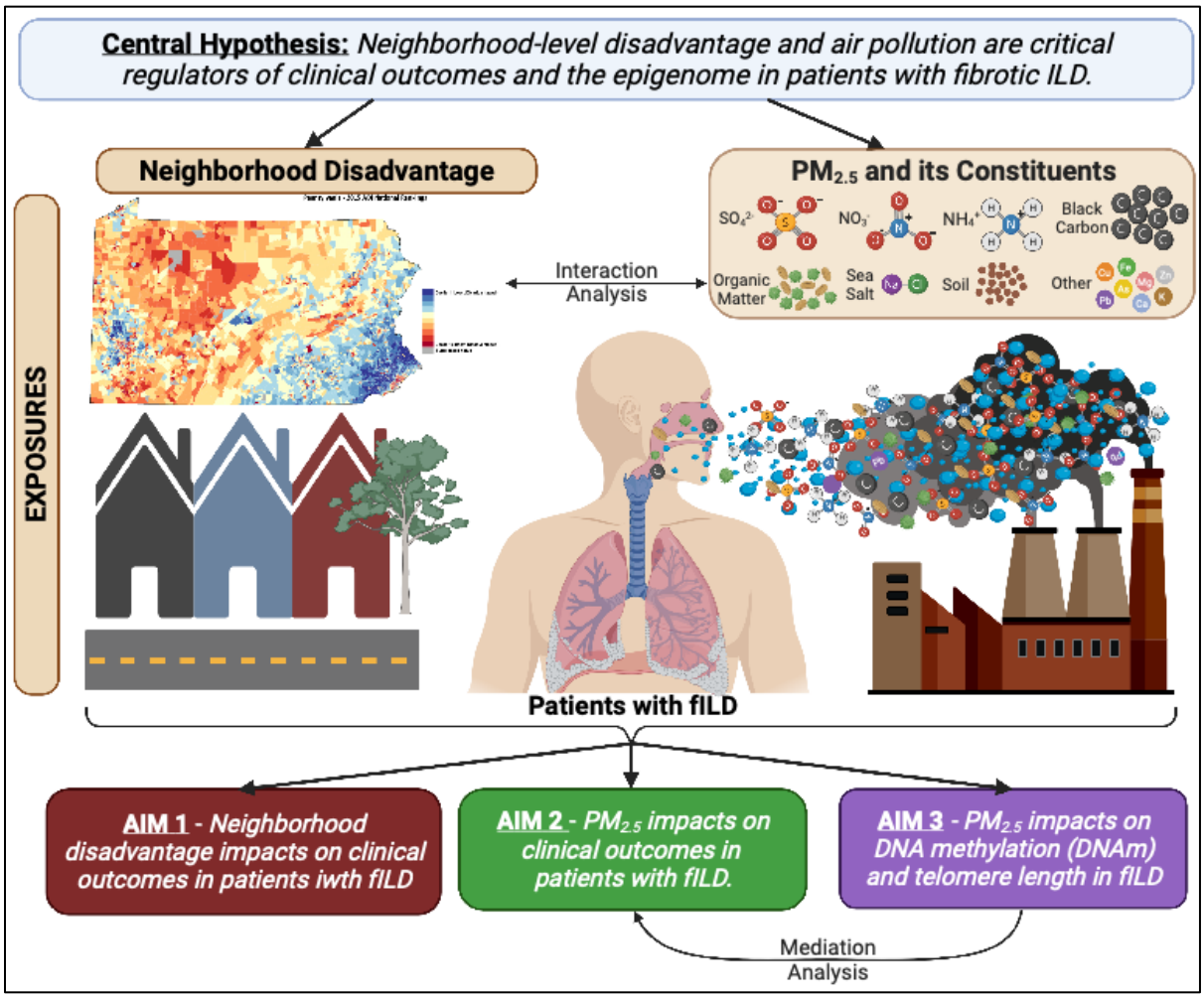


Figure 2 – Central hypothesis and specific aims for PhD dissertation.

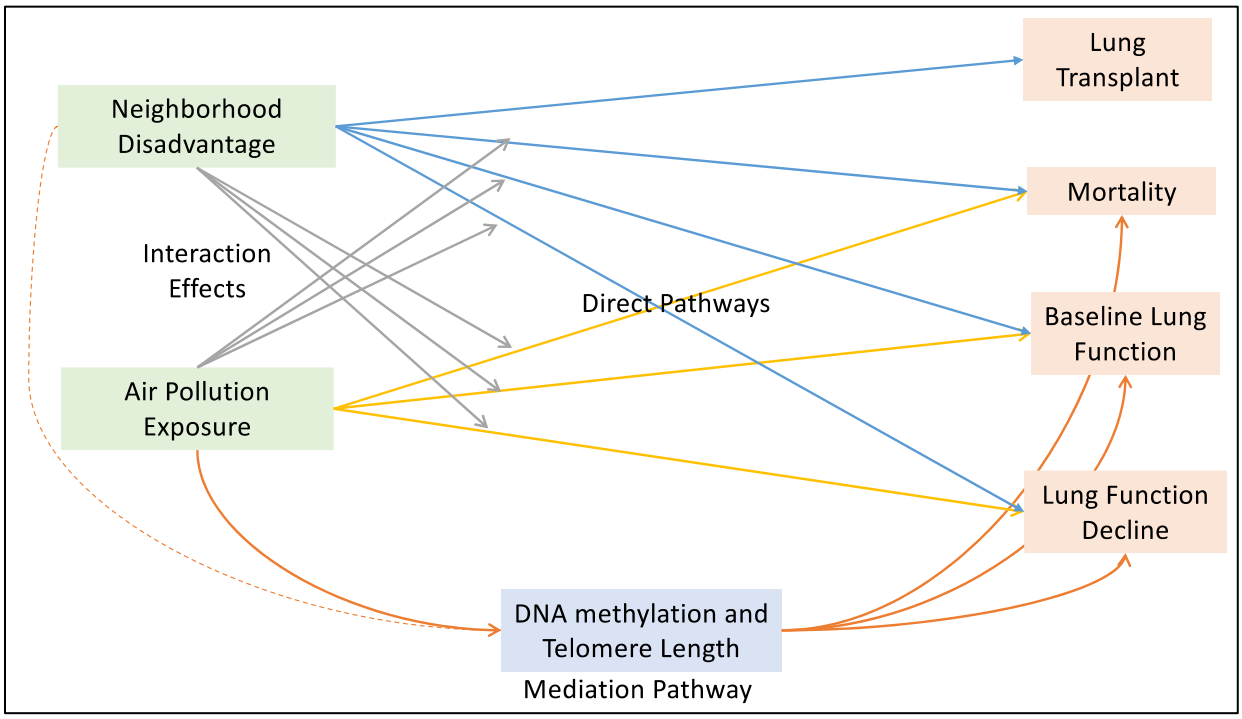


Figure 3 – Directed acyclic graph for primary exposures and outcomes for PhD project.

1.4.1 Aim 1 – Neighborhood Disadvantage Impacts on Patients with fILD

The first part of my PhD aims to determine the impact of neighborhood-level disadvantage on clinical outcomes in patients with fibrotic ILD. We evaluated the impact of neighborhood-level disadvantage as measured by the Area Deprivation Index (ADI) in U.S. patients and the Canadian Index of Multiple Deprivation (CIMD) in Canadian on the clinical outcomes of mortality, odds of receiving lung transplantation, baseline lung function, and rate of decline in lung function in patients with fILD. This work is the first of its kind to evaluate the impact of neighborhood-level disadvantage on clinical outcomes in patients with fILD. This work also draws important contrasts in the divergent impact of neighborhood disadvantage between a U.S. and Canadian cohort, where one of the critical differences is access to universal healthcare in Canada, but not the U.S. The results of Aim 1 have been published and are presented in *Section 2*.

1.4.2 Aim 2 – $PM_{2.5}$ Impacts on Patients with fILD

The second part of my PhD aims to determine the impact of exposure to $PM_{2.5}$ and its associated constituent components on clinical outcomes in patients with fibrotic ILD. We evaluated the association of 5-year exposures to $PM_{2.5}$ and its underlying constituent components (sulfate, SO_4^{2-}; nitrate, NO_3^-; ammonium, NH_4^+; black carbon, BC; organic matter, OM; sea salt, SS; and soil) with mortality, baseline lung function and rate of lung function decline in patients with fILD. The results of this work represent the largest and most geographically-diverse study of the impacts of air pollution in patients with fILD to date, and represents one of very few studies that has looked at air pollution impacts in a cohort of patients that includes non-IPF fILDs. This work also incorporates the findings from Aim 1 by adjusting for neighborhood disadvantage in analyses to address the potential confounding and intersecting relationships between pollution exposures and socioeconomic disadvantage. The results from Aim 2 have been published and are presented in *Section 3*.

1.4.3 Aim 3 – $PM_{2.5}$ Impacts on DNA Methylation and Telomere Length in fILD

The third part of my PhD aims to determine the impact of exposure to $PM_{2.5}$ and its constituents on DNA methylation (DNAm) patterns and telomere length in patients with fibrotic ILD. Aim 3.1 evaluates the association of $PM_{2.5}$ and constituent components with global DNAm in patients with IPF. In Aim 3.2, we evaluated the association of $PM_{2.5}$ and constituent components with genome-wide DNA methylation patterns and specifically investigated for alterations in regions of the genome known to be associated with fILD pathophysiology. In Aim 3.3, we evaluated the association of $PM_{2.5}$ and constituent components with leukocyte telomere length

(TL) in patients with fILD, given the established pathophysiologic relevance of short telomeres in these patients. In each of the sub-aims of Aim 3, we also investigated whether any significant clinical outcomes from Aim 2 were mediated by the epigenomic or genomic alterations found in each part of Aim 3.

2.0 Aim 1 – Neighborhood Disadvantage Impacts on Patients with fILDs

2.1 – Aim 1 Manuscript

2.1.1 *AJRCCM* Original Research Manuscript – "Neighborhood-level disadvantage impacts on patients with fibrotic interstitial lung disease"

Cite: Goobie, GC, *et al*. Neighborhood-Level Disadvantage Impacts on Patients with Fibrotic Interstitial Lung Disease. *Am J Respir Crit Care Med* 2022; 205(4):459-467. DOI: 10.1164/rccm.202109-2065OC. PMID: 34818133.(Goobie, Ryerson, et al., 2022)

The American Journal of Respiratory and Critical Care Medicine is an official journal of the American Thoracic Society.

The data supplement for this manuscript is provided in **Appendix A**.

Neighborhood-level disadvantage impacts on patients with fibrotic interstitial lung disease.

Gillian C. Goobie, MD[1,2,3]; Christopher J. Ryerson, MD, MAS[4,5]; Kerri A. Johannson, MD, MPH[6]; Erin Schikowski, MD[7]; Richard H. Zou, MD[8]; Nasreen Khalil, MD[4]; Veronica Marcoux, MD[9]; Deborah Assayag, MD[10]; Hélène Manganas, MD[11]; Jolene H. Fisher, MD[12]; Martin R.J. Kolb, MD, PhD[13]; Kevin F. Gibson, MD[2,8]; Daniel J. Kass, MD[2,8]; Yingze Zhang, PhD[8]; Kathleen O. Lindell, RN, PhD[2,14*]; S. Mehdi Nouraie, MD, PhD[8*]

1. Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, U.S.A. Email: goobiegc@upmc.edu.
2. Simmons Center for Interstitial Lung Disease, Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A.
3. Clinician Investigator Program, Department of Medicine, University of British Columbia, Vancouver, B.C., Canada.

4. Division of Respiratory Medicine, Department of Medicine, University of British Columbia, Vancouver, B.C., Canada. Email: chris.ryerson@hli.ubc.ca, nkhalil@mail.ubc.ca.
5. Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, B.C., Canada.
6. Division of Respiratory Medicine, Department of Medicine, University of Calgary, Calgary, AB, Canada. Email: kerri.johannson@albertahealthservices.ca.
7. Division of General Internal Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A. Email: schikowskiem@upmc.edu.
8. Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A. Emails: zour@upmc.edu, gibsonkf@upmc.edu, kassd2@upmc.edu, zhanyx@upmc.edu, nouraies@upmc.edu.
9. Division of Respirology, Critical Care and Sleep Medicine, College of Medicine, University of Saskatchewan, Saskatoon, SK, Canada. Email: vsm940@mail.usask.ca.
10. Division of Respiratory Medicine, Department of Medicine, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada. Email: deborah.assayag@mcgill.ca.
11. Département de Médecine, Centre Hospitalier de l'Université de Montréal, Montreal, QC, Canada. Email: helene.manganas.med@ssss.gouv.qc.ca.
12. Division of Respirology, Department of Medicine, University of Toronto, Toronto, ON, Canada. Email: jolene.fisher@uhn.ca.
13. Department of Medicine, Firestone Institute for Respiratory Health, The Research Institute of St. Joe's Hamilton, St. Joseph's Healthcare, McMaster University, Hamilton, ON, Canada. Email: kolbm@mcmaster.ca.
14. College of Nursing, Medical University of South Carolina, Charleston, SC, U.S.A. Email: lindellk@musc.edu.

*Denotes co-senior authorship

ORCID IDs: 0000-0001-7982-5635 (G.C.G.); 0000-0002-0168-5967 (R.H.Z.); 0000-0001-6597-7830 (D.J.K.); 0000-0001-6947-2901 (Y.Z.); 0000-0002-2927-0659 (K.O.L.); 0000-0001-7465-0581 (M.N.).

Correspondence to:
Gillian C. Goobie, MD FRCPC
Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh
130 DeSoto Street, Pittsburgh, PA, 15261
Email: goobiegc@upmc.edu; Phone: 412-624-3018; Fax: 412-624-3020

Author Contributions: GCG, KOL, and SMN are responsible for study development, analysis, and initial manuscript production. CJR, KAJ, KFG, YZ, and DJK provided critical commentary and initial review of the manuscript. SMN provided insight about biostatistical and epidemiologic methods. ES and RHZ provided code and raw data for analysis of outcomes in the U.S. cohort. NK, VM, DA, HM, JHF, and MRJK contributed to the development of the CARE-PF cohort. All authors had full access to the data analyzed in this study, accept responsibility for publication, and provided critical input on the final version of the submitted manuscript.

Sources of Support: The Canadian Registry for Pulmonary Fibrosis (CARE-PF) is sponsored by Boehringer Ingelheim Canada, but this entity has had no influence on the conceptualization, analysis, manuscript production, or any other aspects of this project. GCG receives research

funding and support through the Pulmonary Fibrosis Foundation (PFF) Scholars Award Program and the University of British Columbia Clinician Investigator Program. CJR, KAJ, DA, VM, and MRJK report personal fees and grants from Boehringer Ingelheim and Hoffman La Roche outside the submitted work. CJR also reports fees from AstraZeneca, Veracyte, and Cipla Ltd. KAJ also reports personal fees and grants from Theravance, Blade Therapeutics, Chest Foundation, University of Calgary School of Medicine, Three Lakes Foundation, University Hospital Foundation, Pliant Therapeutics, and sits on the data safety monitoring board of the PFOX trial. VM also reports grants from AstraZeneca. HM reports grants from and sits on the advisory board of Boehringer Ingelheim. DA also reports personal fees and grants from Novartis. HM also reports grants from BMS, Galapagos. JHF reports grants from the Canadian Pulmonary Fibrosis Foundation and the University of Toronto, as well as consulting fees for Boehringer Ingelheim and AstraZeneca. MRJK also reports personal fees and grants from Horizon, Cipla, Abbvie, Belerophon, Algernon, CSL Behring, Novartis, Pieris, sits on data safety monitoring boards for Covance and United Therapeutics, and serves as Chief Editor of the European Respiratory Journal. DJK is supported in part by HL126990 NIH grants, Boehringer-Ingelheim grants, and receives collaborative research funding from Regeneron Pharmaceuticals. YZ is supported in part by NIAMS and NIH R21 grants and a Roche grant. NK, ES, RHZ, KFG, KOL, and SMN report no competing interests.

Running head: Neighborhood disadvantage in fibrotic ILD.

Descriptor: 2.9 Racial, Ethnic, or Social Disparities in Lung Disease and Treatment

Body word count: 3164

This article has an online data supplement, which is accessible from this issue's table of content online at www.atsjournals.org

2.1.2 Abstract

Rationale: Fibrotic interstitial lung diseases (fILDs) represent a group of pathologic entities characterized by scarring of the lungs and high morbidity and mortality. Research investigating how socioeconomic and residential factors impact outcomes in patients with fILDs is lacking.

Objectives: To determine the association between neighborhood-level disadvantage and presentation severity, disease progression, lung transplant, and mortality in patients with fILD from the United States (U.S.) and Canada.

Methods: Multi-center, international, prospective cohort study of 4729 patients with fILD from one U.S. and eight Canadian ILD registry sites. Neighborhood-level disadvantage was measured by the area deprivation index (ADI) in the U.S. and the Canadian Index of Multiple Deprivation (CIMD) in Canada.

Measurements and Main Results: In the U.S., but not Canadian cohort, patients with fILD living in neighborhoods with the greatest disadvantage (top quartile) experience the highest risk of mortality (hazard ratio=1.51, p=0.002) and in subgroups of patients with idiopathic pulmonary fibrosis (IPF), the top quartile of disadvantage experienced the lowest odds of lung transplant (odds ratio=0.46, p=0.04). Greater disadvantage was associated with reduced baseline diffusion capacity for carbon monoxide (D_LCO) in both cohorts, but it was not associated with baseline forced vital capacity (FVC) or FVC or D_LCO decline in either cohort.

Conclusions: Patients with fILD who live in areas with greater neighborhood-level disadvantage in the U.S. experience higher mortality, and patients with IPF experience lower odds of lung transplant. These disparities are not seen in Canadian patients, which may indicate differences in access to care between the U.S. and Canada.

Keywords: interstitial lung disease, health equity, healthcare disparities, residence characteristics, social determinants of health.

2.1.3 At a Glance Commentary

Scientific knowledge on this subject: There has been limited previous investigation of the impact of socioeconomic and neighborhood-level factors on clinical outcomes in patients with fibrotic interstitial lung diseases (fILDs). Between January 16, 2020 and August 16, 2021, we searched the scientific literature in PubMed (with no date or language restrictions) for "ILD", "interstitial lung disease", "IPF", "idiopathic pulmonary fibrosis", "socioeconomic", and "neighborhood" in various combinations to identify articles focusing on the impact of individual and neighborhood-level socioeconomic disadvantage on patients with fILDs. We found two studies where socioeconomic status was associated with worse clinical outcomes in patients with idiopathic pulmonary fibrosis (IPF), but no studies evaluating the impact of socioeconomic factors on clinical outcomes in diverse cohorts of patients with fILD.

What this study adds to the field: This study provides the first evidence to support that patients with fILD living in neighborhoods greater disadvantage in the U.S. experience higher mortality, and in patients with a diagnosis of IPF, are less likely to receive lung transplantation as an end-stage therapy. These findings are not seen in the Canadian cohort of patients with fILD, where patients may have improved access to care in a universal healthcare system

2.1.4 Introduction

Fibrotic interstitial lung diseases (fILDs) represent a group of pathologic entities characterized by dyspnea, high morbidity, and early mortality. Treatments of fILD can include immunosuppressant and anti-fibrotic medications as well as non-pharmacologic measures.(George et al., 2020) In end-stage fILD, lung transplantation is one of the only life-prolonging interventions available.(Wong et al., 2020) Recent findings indicate that low income may predict worse survival in patients with idiopathic pulmonary fibrosis (IPF), the most common form of fILD.(Sese et al., 2021) Patients with IPF in the United States (U.S.) with low socioeconomic status (SES) are also less likely to receive lung transplantation, pulmonary rehabilitation, or diagnostic surgical lung biopsy.(Gaffney et al., 2018) The impact of residential, socioeconomic, and healthcare systems factors on clinical outcomes in patients with other forms of fILD remains largely unexplored.(Sesé et al., 2020)

Neighborhood-level disadvantage is a socioeconomic and residential factor associated with income, education level, food security, safety, and health behaviors that contributes to health disparities for minoritized, low-income, and other vulnerable communities.(Kind & Buckingham, 2018) The area deprivation index (ADI) is a measure of neighborhood-level disadvantage based on U.S. 2018 American Community Survey data that serves as a surrogate measure of SES.(Kind & Buckingham, 2018) The Canadian Index of Multiple Deprivation (CIMD)(*The Canadian Index of Multiple Deprivation: User Guide*, 2019) produces a similar score based on material and social disadvantage measured from 2016 Canadian Census data. Greater neighborhood-level disadvantage is associated with adverse outcomes in several chronic respiratory diseases,(Ejike et al., 2020; Galiatsatos et al., 2020; Oates et al., 2019) but the impact on patients with IPF and other forms of fILD has not been investigated.

The U.S. healthcare system is comprised of multiple separate public and private insurers that currently leaves 9% of Americans uninsured,("Congressional Research Service (CRS) Reports. U.S. Health Care Coverage and Spending.," 2021) resulting in both high costs and substantial inequities.(Dickman, Himmelstein, & Woolhandler, 2017; OECD, 2020) In contrast, Canadian provinces provide universal healthcare coverage to all Canadian citizens and permanent residents.("Government of Canada. Understand how health care works in Canada.," 2021) To our knowledge, this is the first study to evaluate the differential impact of neighborhood-level disadvantage on outcomes in vulnerable patient populations between the U.S. and Canadian healthcare systems. We hypothesized that greater neighborhood-level disadvantage, as quantified by the ADI or CIMD, would be associated with increased mortality, reduced odds of lung transplant, worse baseline lung function, and more rapid lung function decline in two well-defined cohorts of patients with fILD in the U.S. and Canada. Some of the results of these studies have previously been reported in the form of an abstract.(Goobie et al., 2021)

2.1.5 Methods

2.1.5.1 Study Populations

The U.S. cohort included patients with fILD prospectively enrolled in the University of Pittsburgh Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease at the University of Pittsburgh Medical Center (UPMC) Registry between 2000 to 2021. The Canadian cohort included patients with fILD prospectively enrolled in one of the eight Canadian Registry for Pulmonary Fibrosis (CARE-PF) sites between 2015 to 2021,(Ryerson et al., 2016) as well as some patients who were enrolled in preexisting single-center registries at participating sites prior to 2015.

Adult patients with fILD in either registry were eligible for inclusion (including IPF, chronic hypersensitivity pneumonitis, connective tissue disease-ILD/CTD-ILD, non-IPF idiopathic interstitial pneumonias, pneumoconiosis, and unclassifiable ILD). Non-fibrotic ILDs were excluded (including sarcoidosis, vasculitis, and pulmonary alveolar proteinosis). Date of diagnosis was defined as the first visit to the specialist ILD registry center.

Ethics approval was obtained from the University of Pittsburgh (STUDY20050209) and the University of British Columbia as the coordinating site for CARE-PF (#H19-01989 and #H20-01454).

2.1.5.2 Neighborhood-Level Disadvantage Determination

The most recent residential address was used to determine ADI or CIMD scores for each patient. ADI scores are measured on a national scale based on 2018 American Community Survey data and ranges from 1-100, with 1 reflecting the least and 100 reflecting the greatest disadvantage. CIMD scores are based on 2016 Canadian Census data and are measured on a continuous scale using a factor score approach for four dimensions of the index. The average factor score was used as our continuous exposure variable, with higher scores reflecting the greatest disadvantage. In our Canadian cohort, CIMD score ranged from -1.3 to 2.6. Analyses were performed with continuous and quartiled ADI or CIMD, wherein approximately equal numbers of patients with fILD were grouped into each quartile, with quartile 1 (Q1) reflecting the least disadvantage and quartile 4 (Q4) reflecting the greatest disadvantage (cut-offs in **Appendix Table 1**). Quartiled analyses were used to aid in comparability of the results between the two countries.

2.1.5.3 Clinical Outcomes

Survival, lung transplantation status, and lung function measurements (forced vital capacity, FVC, and diffusion capacity of the lung for carbon monoxide, D_LCO) were discerned from specialist ILD clinic records and electronic health records. Baseline FVC and D_LCO were defined as the first tests performed within 6 months of diagnosis. All FVC and D_LCO measurements obtained throughout the course of follow-up were used to determine the rate of change in lung function.

2.1.5.4 Statistical Analyses

Survival analyses were performed using Cox proportional hazards regression, considering time to death or lung transplant as a composite outcome. For sensitivity, competing hazards survival analyses, considering lung transplant a competing risk for death, were also performed and is reported in the Online Data Supplement. We had two modelling approaches, first (the "partially adjusted" models) adjusting for pre-specified covariates of age at diagnosis and sex, and second (the "fully adjusted" models) adjusting for the covariates of race, baseline FVC and D_LCO percent predicted, and smoking history, which are factors that may interact with neighborhood disadvantage.(Gaffney, Himmelstein, Christiani, & Woolhandler, 2021; Kawachi, Daniels, & Robinson, 2005) There was no imputation of missing covariates and observations where covariates were missing were dropped. Adjustments for sex, age, and baseline lung function were made as these factors predict mortality in IPF and other fILDs;(Ley et al., 2012; Ryerson et al., 2014) for race, given impacts on pulmonary function and mortality;(Adegunsoye et al., 2018) and for smoking, given impacts on fILD incidence, lung function, transplant referral, and potentially mortality.(Margaritopoulos, Vasarmidi, Jacob, Wells, & Antoniou, 2015)

The impact of continuous and quartiled ADI or CIMD on odds of lung transplant was evaluated using generalized binomial linear models, with adjustments for the same covariates as above.

Multivariable linear regression was used to evaluate the impact of neighborhood disadvantage on baseline FVC and D_LCO percent predicted. Linear mixed effects models with random intercept and slope were used to evaluate the association between ADI or CIMD and rate of change in FVC or D_LCO percent predicted. Lung function analyses were adjusted for sex, age at diagnosis, and smoking history, but not race as some registry sites produced race-adjusted estimates of percent predicted values.

Subgroup analyses for all outcomes were performed for patients with a specific fILD diagnosis of IPF. Analyses were performed using R (version 4.0.2, www.r-project.org).

2.1.6 Results

2.1.6.1 Baseline Patient Characteristics

There were 1372 U.S. and 3357 Canadian patients with fILD who met eligibility criteria. Baseline demographics, ADI or CIMD scores, baseline lung function, follow-up duration, and censoring outcomes are shown in **Table 2**, alongside the number of patients with missing covariates. These results are broken down by quartile in **Appendix Table 1**. State and province breakdown is shown in **Appendix Table 2**. In the U.S. cohort, the most common diagnosis was IPF (50%), patients had a median follow-up duration of 3.1 years, 196 (14%) received a lung transplant, and 683 (50%) died over the duration of follow-up. In the Canadian cohort, the most common diagnosis was CTD-ILD (38%), patients had a median follow-up duration of 3.2 years, 174 (5%) received a lung transplant, and 761 (23%) died over the duration of follow-up. 12% of

patients in the U.S. and 21% in the Canadian cohort had a self-reported non-White race, with these individuals representing a larger proportion of the patients in the highest quartile (Q4) of ADI or CIMD (17% and 31%, respectively; **Appendix Table 1**).

Table 2 – Patient characteristics by U.S. or Canadian cohort.

Patient Characteristics	U.S. Cohort	Canadian Cohort
	N=1372	N=3357
Age at diagnosis, median (IQR), years	66 (58, 73)	66 (57, 73)
Sex, n (%)		
Male	761 (55%)	1664 (50%)
Female	611 (45%)	1693 (50%)
Self-reported race, n (%)		
White	1209 (88%)	2658 (79%)
Black	56 (4%)	53 (2%)
Asian	5 (0.4%)	382 (11%)
Indigenous*	2 (0.1%)	85 (2%)
Pacific Islander	0 (0%)	25 (1%)
Unknown	100 (7%)	154 (5%)
Self-reported ethnicity, n (%)		
Not Hispanic	1148 (84%)	2772 (83%)
Hispanic	2 (0.1%)	68 (2%)
Unknown	222 (16%)	517 (15%)
Smoking history, n (%)		
Never	403 (29%)	1269 (38%)
Former	639 (47%)	1906 (57%)
Current	37 (3%)	172 (5%)
Unknown	293 (21%)	10 (0.3%)
ILD diagnostic group, n(%)		
Idiopathic pulmonary fibrosis	688 (50%)	918 (27%)
Connective tissue disease-ILD	290 (21%)	1283 (38%)
Fibrotic hypersensitivity pneumonitis	54 (4%)	257 (8%)
Pneumoconiosis	24 (2%)	28 (1%)
Non-IPF idiopathic interstitial pneumonia	65 (5%)	108 (3%)
Other ILD	50 (4%)	120 (4%)
Unclassifiable or not yet diagnosed	201 (14%)	643 (19%)
ADI or CIMD, median score (IQR)	61 (44, 78)	-0.03 (-0.35, 0.38)
Baseline FVC % Predicted, median (IQR), n (%)	66 (53, 81), 1079 (79%)	75 (61, 89), 2942 (88%)
Baseline D_LCO % Predicted, median (IQR), n (%)	49 (37, 63), 1008 (73%)	57 (44, 71), 2367 (71%)
Follow-up Duration, median (IQR), years	3.09 (1.21, 6.23)	3.18 (1.78, 5.15)
Cause of censoring, n (%)		
Death	683 (50%)	761 (23%)
Lung Transplantation	196 (14%)	174 (5%)
Lost to follow-up or censored by data extraction	493 (36%)	2422 (72%)

*Includes Native American, American Indian, Alaskan First Nations, & other Indigenous persons in the U.S.; First Nations, Métis, Inuit, and other Indigenous persons in Canada.

The distribution of the U.S. and Canadian cohort's ADI or CIMD was compared with the distribution across the primary referral states and provinces (**Appendix Figure 1**). Qualitatively, the U.S. cohort appeared to under-sample individuals from the most disadvantaged neighborhoods, but neither cohort was significantly different from the referral distribution when cohorts were divided into deciles for Chi-squared goodness of fit comparison of the distributions.

2.1.6.2 Association of ADI or CIMD with Survival

In partially and fully adjusted Cox models in the U.S. cohort, higher continuous ADI was associated with higher mortality (fully adjusted HR 1.006 per unit ADI, 95%CI 1.002-1.010, p=0.004), as was Q4 of ADI (fully adjusted HR 1.51, 95%CI 1.17-1.95, p=0.002) (**Figure 4A, Appendix Table 3**). There was no association between continuous CIMD score or Q4 and mortality in the full Canadian cohort adjusted for the same covariates (**Figure 4B, Appendix Table 3**). The mortality effect was maintained in subgroup analyses of patients with IPF in the partially and fully adjusted model for Q4 of ADI in the U.S. cohort, but not in the Canadian cohort (**Figure 4C/D, Appendix Table 3**). Both continuous ADI and CIMD met proportionality assumptions, but when plotted as a hazard function over time, both exert the greatest impact on mortality in the first two years following diagnosis (**Appendix Figure 2**). *Post-hoc* analyses were subsequently performed with cohorts split into deciles with approximately equal patient numbers. In the U.S. cohort, deciles 8 and above experienced the highest mortality, whereas there was no significant difference between the deciles in the Canadian cohort (**Appendix Figure 3**).

Competing hazards survival models were performed as sensitivity analyses, with a full at-risk table of outcomes shown in **Appendix Table 4**, results of the analyses in **Appendix Table 5**, and cumulative incidence curves in **Appendix Figure 4.** In the U.S. cohort, the findings were consistent with the Cox models, indicating that mortality was higher with increased neighborhood-

level disadvantage. No difference in transplant outcomes was noted in these models. In competing hazards models in the Canadian cohort, there was no association between higher CIMD score and mortality or transplant. In both cohorts, individuals of non-White race had lower probability of a transplant outcome (U.S. HR 0.28, 95%CI 0.11-0.71; Canadian HR 0.37, 95%CI 0.20-0.67).

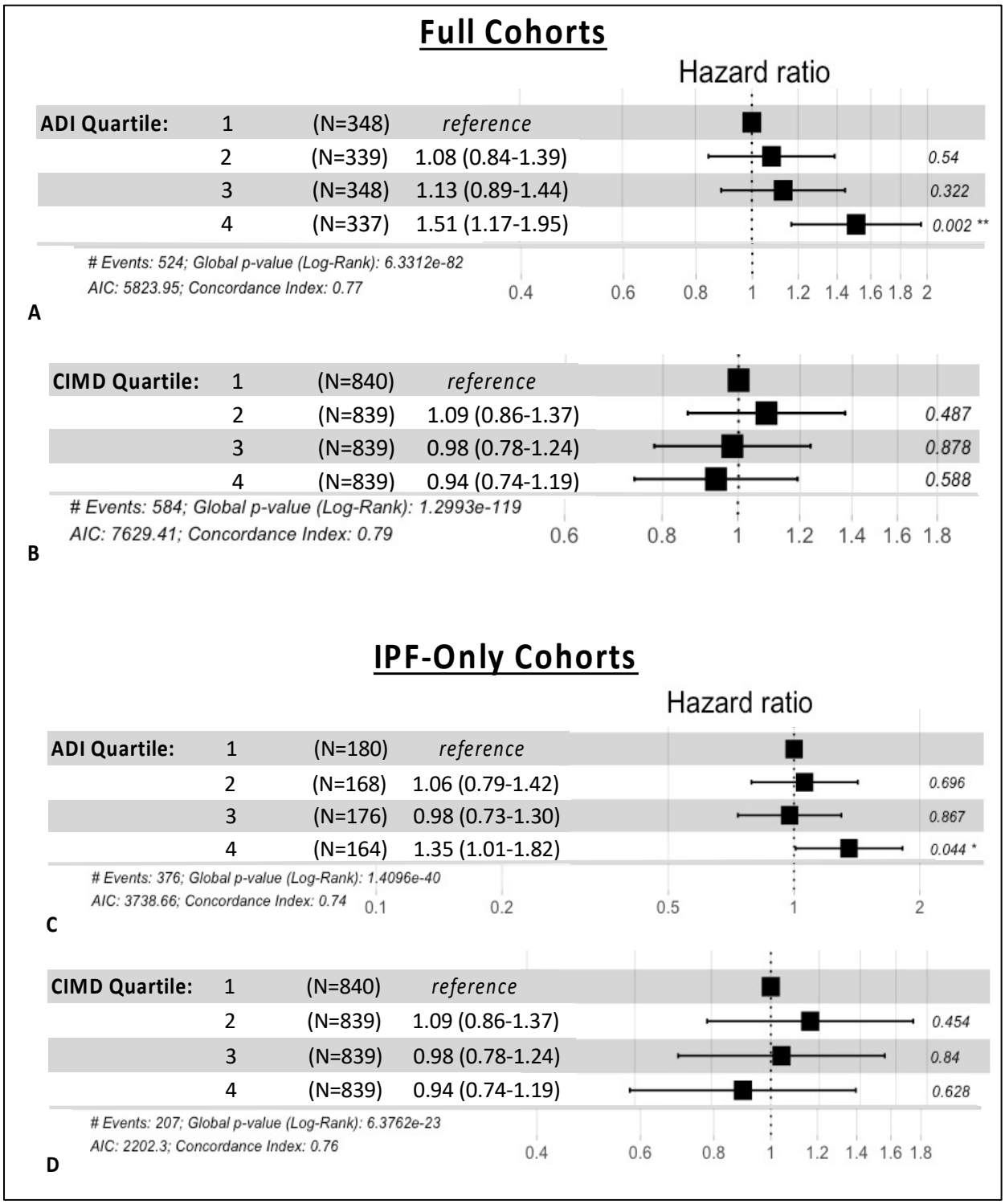


Figure 4 – Neighborhood disadvantage forest plots. Forest plots of effect estimate for Cox proportional hazards survival analysis complete models (covariates included are sex, age at diagnosis, smoking history, White or non-White race, baseline FVC, and baseline DLCO) in A) U.S. full cohort, B) Canadian full cohort, C) U.S. IPF-only cohort, and D) Canadian IPF-only cohort. ADI, area deprivation index; AIC, Akaike information criterion; CIMD, Canadian index of multiple deprivation; IPF, idiopathic pulmonary fibrosis.

2.1.6.3 Association of ADI or CIMD with Odds of Lung Transplant

Analyses were subsequently performed to evaluate the odds of lung transplant as an independent outcome from survival analyses. In baseline, partially adjusted, and fully adjusted generalized linear models, neither continuous ADI nor CIMD scores were associated with odds of lung transplant. There was no difference in odds of transplant between the quartiles in the U.S. cohort, but in the Canadian cohort, quartile 2 (Q2) was associated with higher odds of transplant compared to Q1 (OR 1.79, 95%CI 1.06-3.10, p=0.03) (**Table 3**, with complete model results shown in **Appendix Table 6**). In subgroup analyses of patients with IPF in the U.S. cohort, higher ADI (greater neighborhood disadvantage) was associated with lower odds of lung transplant (continuous OR 0.986 per unit ADI, 95%CI 0.975-0.996, p=0.01; Q4 OR 0.46, 95%CI 0.22-0.95, p=0.04) (**Figure 5, Table 3, Appendix Table 6**). In patients with IPF in the Canadian cohort, there was no association between continuous CIMD score or CIMD quartiles and odds of transplant.

Table 3 – Summary of the association between ADI or CIMD and odds of lung transplant. Results of fully adjusted models presented (adjusted for age at diagnosis, sex, smoking history, race, baseline FVC, and baseline D_LCO). P-values <0.05 are **bolded.**

U.S. Cohort				**Canadian Cohort**			
Odds of Lung Transplant Generalized Binomial Linear Models – Full Cohort							
	OR	**95% CI**	**P-value**		**OR**	**95% CI**	**P-value**
Continuous ADI (N = 797)				**Continuous CIMD (N = 2362)**			
Continuous ADI	0.995	0.985, 1.004	0.27	Continuous CIMD	1.01	0.68, 1.46	0.97
Quartiled ADI (N = 797)				**Quartiled CIMD (N = 2362)**			
Quartile 1	*reference*			Quartile 1	*reference*		
Quartile 2	0.95	0.53, 1.70	0.86	Quartile 2	1.79*	1.06, 3.10	**0.03**
Quartile 3	1.05	0.60, 1.86	0.86	Quartile 3	1.17	0.66, 2.08	0.60
Quartile 4	0.85	0.47, 1.55	0.60	Quartile 4	1.32	0.73, 2.38	0.36
Odds of Lung Transplant Generalized Binomial Linear Models – IPF-Only Cohort							
	OR	**95% CI**	**P-value**		**OR**	**95% CI**	**P-value**
Continuous ADI (N = 444)				**Continuous CIMD (N = 591)**			
Continuous ADI	0.986*	0.975, 0.996	**0.009**	Continuous CIMD	1.1	0.50, 1.98	0.98
Quartiled ADI (N = 444)				**Quartiled CIMD (N = 591)**			
Quartile 1	*reference*			Quartile 1	*reference*		
Quartile 2	1.18	0.59, 2.35	0.64	Quartile 2	2.31	0.96, 5.96	0.07
Quartile 3	0.83	0.42, 1.64	0.58	Quartile 3	1.49	0.57, 4.05	0.42
Quartile 4	0.46†	0.22, 0.95	**0.04**	Quartile 4	1.48	0.49, 4.40	0.48

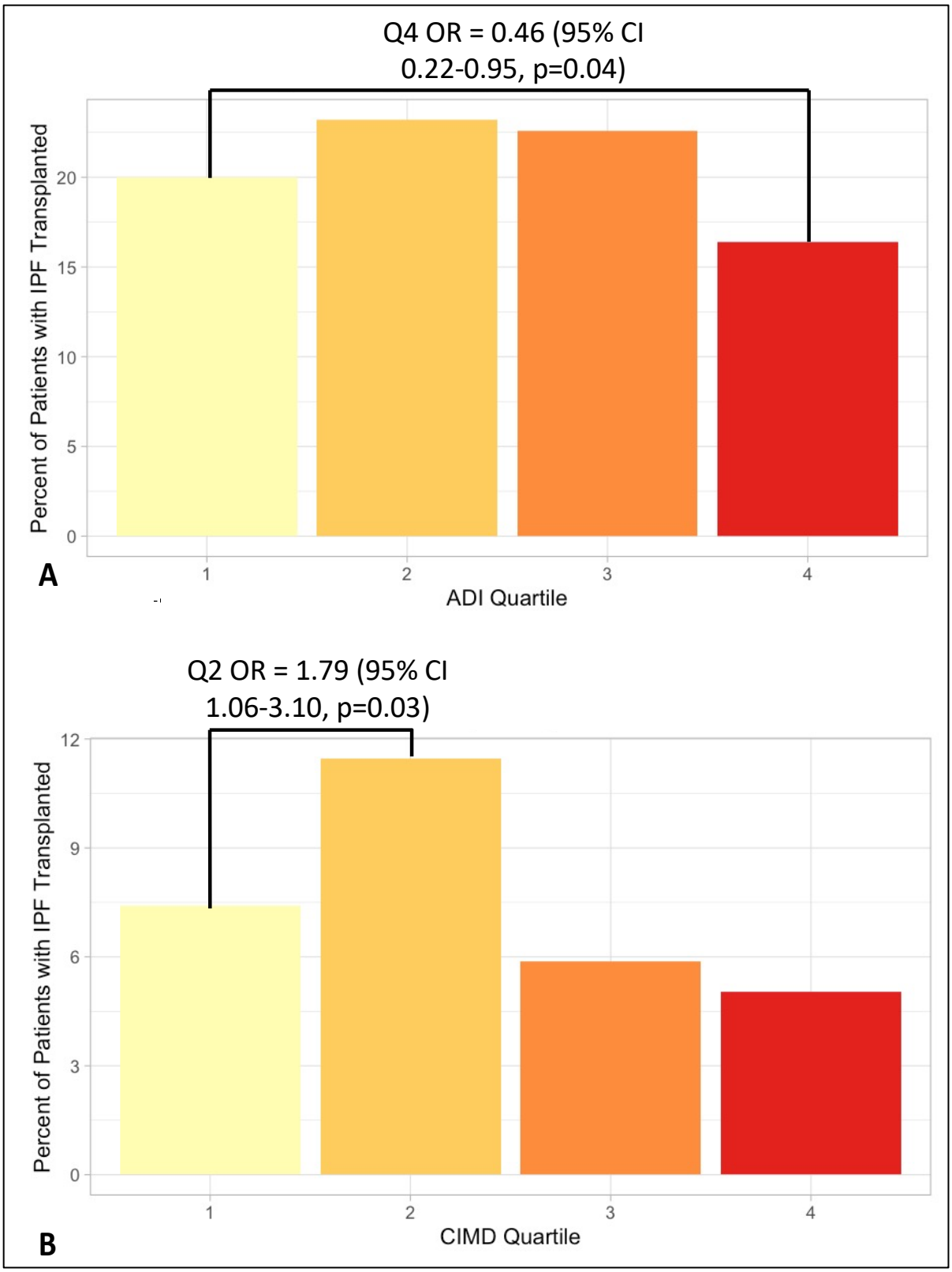


Figure 5 – Proportion of patients with IPF who received lung transplant. A) U.S. cohort. B) Canadian cohort. ADI, area deprivation index; CIMD, Canadian index of multiple deprivation; IPF, idiopathic pulmonary fibrosis.

2.1.6.4 Association of ADI or CIMD with Baseline Lung Function

In unadjusted and fully adjusted models in the U.S. cohort, neither continuous nor quartiled ADI was associated with differences in baseline FVC. In fully adjusted models in the Canadian cohort, higher CIMD was associated with lower baseline FVC (continuous β = -1.47, 95%CI -2.79 to -0.16, p=0.02; Q4 β = -2.10, 95%CI -4.10 to -0.10, p=0.04) (**Figure 6A-B**, **Appendix Table 7**). Subgroup analyses on patients with IPF did not demonstrate any differences in baseline FVC with increasing neighborhood disadvantage in either cohort.

In fully adjusted models in the U.S. cohort, higher continuous ADI was not associated with lower baseline D_LCO (continuous β = -0.05, 95%CI -0.11 to 0.01, p=0.12), but Q4 was associated with reduced D_LCO compared to Q1 (Q4 β = -4.32, 95%CI -8.08 to -0.55, p=0.02). In fully adjusted models in the Canadian cohort, higher continuous CIMD score was associated with lower baseline D_LCO (continuous β = -3.23, 95%CI -4.70 to -1.77, p<0.001). Additionally, quartiles 2-4 of CIMD were each associated with lower baseline D_LCO compared to Q1 (Q4 β = -4.57, 95%CI -6.80 to -2.34, p<0.001) (**Figure 6C-D, Appendix Table 8**). Subgroup analyses on patients with IPF did not demonstrate any differences in baseline D_LCO with increasing neighborhood disadvantage in either cohort.

2.1.6.5 Association of ADI or CIMD with Longitudinal Lung Function

In all models for both full and IPF-only U.S. and Canadian cohorts, neither continuous nor quartiled ADI or CIMD were associated with the rate of FVC decline (**Appendix Table 9**). In all models in the full and IPF-only U.S. cohort, neither continuous nor quartiled ADI were associated with the rate of D_LCO decline. In the full Canadian cohort, living in CIMD quartile 3 (Q3) was associated with less D_LCO decline, but continuous CIMD score was not associated with D_LCO

decline (**Appendix Table 10**). There was no association between rate of D_LCO decline and CIMD in the IPF-only Canadian subgroup.

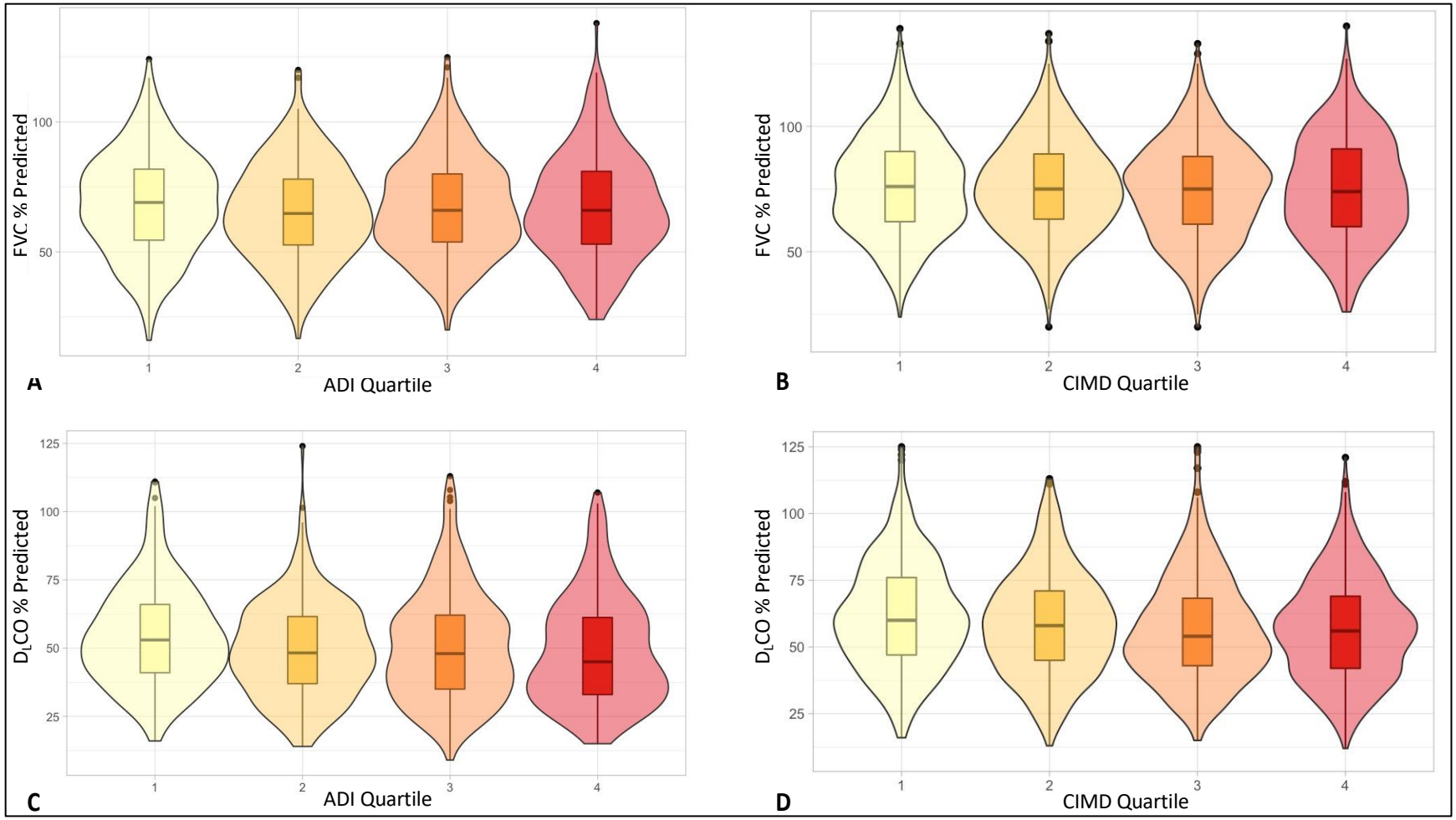


Figure 6 – Baseline lung function by neighborhood disadvantage quartile. Baseline percent predicted forced vital capacity (FVC) by neighborhood disadvantage quartile in A) U.S. and B) Canadian cohort, and diffusion capacity for carbon monoxide (D_LCO) by neighborhood disadvantage quartile in C) U.S. and D) Canadian cohort. ADI, area deprivation index; CIMD, Canadian index of multiple deprivation; D_LCO, diffusion capacity for carbon monoxide; FVC, forced vital capacity.

2.1.7 Discussion

Neighborhood-level disadvantage is linked to adverse health outcomes in several chronic diseases,(Galiatsatos et al., 2020; Kind et al., 2014; Oates et al., 2019) but the impact of this residential socioeconomic factor has not been evaluated in patients with fILD. This study found that in a U.S., but not Canadian cohort, that greater disadvantage increased the risk of mortality and, in patients with IPF, reduced the odds of lung transplant. This study highlights clinical disparities in a vulnerable patient population that occurs between two countries with and without a universal healthcare system.

To our knowledge, this is the first study to demonstrate an association between socioeconomic factors and mortality in a diverse cohort of patients with fILD.(Sese et al., 2021) The increased mortality with higher ADI in the U.S. cohort may be due to reduced access to specialist or allied health services or increased exposure to environmental hazards like air pollution.(Bowe et al., 2019; Kind & Buckingham, 2018) Recent evidence indicates that patients with IPF and low socioeconomic status experience increased mortality and are exposed to higher levels of particulate matter with a diameter of ≤2.5μm ($PM_{2.5}$),(Sese et al., 2021) which is a separate risk factor for mortality in IPF.(Sesé et al., 2018) Increased exposure and burden of mortality related to $PM_{2.5}$ is associated with higher ADI across the U.S.,(Bowe et al., 2019) but similar inequities related to neighborhood disadvantage have not been demonstrated in Canada.

The U.S. cohort experienced higher mortality overall, but this is likely related to the higher proportion of patients with IPF, the most aggressive form of fILD, in the U.S. cohort compared to the Canadian cohort. Additionally, UPMC is a tertiary transplant referral center, and thus may care for more severe patients than some of the Canadian subspecialty ILD centers in CARE-PF.

Our findings that increasing ADI is associated with lower odds of receiving a lung transplant in patients with IPF is consistent with prior research. A previous U.S. study found that patients with IPF in the lowest quartile of SES were less likely to undergo lung transplantation.(Gaffney et al., 2018) The lack of association between greater neighborhood disadvantage and reduced lung transplantation in the Canadian cohort may reflect healthcare-associated differences in access to these services between the U.S. and Canada. In a non-universal healthcare system, as in the U.S., patients with greater disadvantage may be less frequently referred for transplant consideration than in Canada, where transplantation-related care would be covered under the universal healthcare system. Disparities in mortality and transplant outcomes between the U.S. and Canada are consistent with previous literature demonstrating increased mortality and lower proportions of lung transplantation in U.S. patients with cystic fibrosis compared to Canadian patients.(Stephenson et al., 2017)

Although race was not the primary exposure of interest in our study, we found that individuals of non-White race were more heavily represented in the highest neighborhood disadvantage quartiles. Our lung transplant analyses indicated that individuals of non-White race had a lower probability of a lung transplant outcome in both cohorts (see Online Data Supplement **Appendix Tables 5** and **6**). Our competing hazards analyses highlight some limitations of traditional Cox proportional hazards survival analyses that consider death and transplant as composite outcomes. Our Cox models and previous studies have suggested lower ILD mortality in individuals of non-White race,(Adegunsoye et al., 2018) but our competing hazards models indicate that this may be related to lower rates of transplantation, and not lower mortality. The interpretation of these results is limited by the minimal racial diversity in our U.S. cohort, which restricted our ability to evaluate refined racial categories as a factor in health disparities. Factors

including historical oppression, marginalization, and systemic racism may contribute to racial inequities in access to interventions such as lung transplantation.(Boyd, Lindo, Weeks, & Mclemore, 2020) Our findings indicate that neighborhood-level disadvantage, race, and access to lung transplantation are complex and likely intersecting factors requiring further evaluation in more diverse populations of patients with fILD. Furthermore, clinicians and public health professionals should advocate for policies that ensure equity in access to transplantation across all socioeconomic, racial, and ethnic groups.

Our study has several limitations. We did not have data on occupational exposures or individual SES, so were unable to control for these factors. We were unable to consider ILD medication use in our study, but it is plausible that individuals living in more disadvantaged areas may have less access to expensive therapeutics like anti-fibrotics or immunosuppressants, which can slow the rate of disease progression.(Behr et al., 2021; Flaherty et al., 2019; King et al., 2014; Richeldi et al., 2014; Tashkin et al., 2006, 2016) The ADI and CIMD are both composite scores of multiple census-level metrics indicating material and social disadvantage in the U.S. and Canada, respectively. Despite their similarities, the construction of the scores is different, thus limiting the ability to directly compare results for continuous scores. As such, quartiled analyses were performed to provide consistency between the cohorts. We did not have data on insurance coverage for the participants in the U.S. cohort, although it is assumed that most had some form of coverage at the time of registry enrollment. All patients over 65 years of age should also be covered by Medicare. Indigenous persons in Canada, who are most highly represented in Saskatchewan, Manitoba, and Northern territories, often experience the most disadvantage and are the least likely to receive specialist care.(Auditor General of Canada, 2015; Kolahdooz, Nader, Yi, & Sharma, 2015) CARE-PF's minimal sampling of these regions may lead to an underestimation

of the impact of neighborhood disadvantage in our Canadian cohort. Furthermore, we are limited by only evaluating a single-center U.S. cohort with relatively limited geographic and racial diversity. This work would benefit greatly from analysis of the impact of neighborhood disadvantage on larger and more diverse cohorts of patients with fILD. Lastly, we were unable to consider ADI or CIMD as time-varying covariates as these indices are not reproduced on a yearly basis and our registries did not capture multiple residential addresses.

2.1.8 Conclusions

Fibrotic interstitial lung diseases (fILDs) are devastating conditions with high morbidity and mortality. Environmental and demographic factors substantially impact incidence and mortality of fILD, but the impact of socioeconomic factors on outcomes has remained largely unexplored. This is the first study to demonstrate an association between neighborhood-level disadvantage and mortality in patients with fILD and is also the first to demonstrate differences in fILD outcomes between universal and non-universal healthcare systems, although this is an area that requires further study. We found that neighborhood-level disadvantage impacts lung transplantation in U.S. patients with IPF, with a suggestion that race also impacts lung transplant outcomes in both the U.S. and Canadian cohorts of patients with fILD. Together, these findings indicate that substantial health disparities exist in the fILD patient population, most predominately in the U.S. These disparities may be mitigated by improving access to healthcare, by ensuring equitable consideration of patients for lung transplantation, and by instituting poverty-reducing interventions on a neighborhood-level.

2.1.9 Acknowledgements

We would like to thank the fILD patients from the Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease and each of the CARE-PF registry sites for their time and commitment to clinical research. We would also like to thank all other contributors to CARE-PF who are not listed as authors on this manuscript, including Drs. Alyson W. Wong, Mohsen Sadatsafavi, Julie Morisset, Shane Shapera, Charlene D. Fell, Gerard Cox, Teresa To, Andrea S. Gershon, Nathan Hamby, Andrew J. Halayko, and Pearce G. Wilcox.

2.1.10 Data Sharing

Because of the restrictions of the consents signed by participants in the Simmons Center for ILD Registry and CARE-PF, individual participant data will not be made available.

3.0 Aim 2 – $PM_{2.5}$ Impacts on Patients with fILDs

3.1 – Aim 2 Manuscript

3.1.1 *JAMA Internal Medicine* Original Research Manuscript – "Association of particulate matter exposure with lung function and mortality in fibrotic interstitial lung disease: A multinational cohort study"

Cite: Goobie, GC, *et al*. Association of particulate matter exposure with lung function and mortality in fibrotic interstitial lung disease: A multinational cohort study. *JAMA Intern Med* 2022 Oct 17: e224696. DOI: 10.1001/jamainternmed.2022.4696. PMID: 36251286.

The data supplement for this manuscript is provided in **Appendix B**.

Association of particulate matter exposure with lung function and mortality in fibrotic interstitial lung disease: A multinational cohort study

Gillian C. Goobie, MD[1,2,3]; Christopher Carlsten, MD MPH[4,5]; Kerri A. Johannson, MD, MPH[6]; Nasreen Khalil, MD[4]; Veronica Marcoux, MD[7]; Deborah Assayag, MD[8]; Hélène Manganas, MD[9]; Jolene H. Fisher, MD[10]; Martin R.J. Kolb, MD, PhD[11]; Kathleen O. Lindell, RN, PhD[2,12] James P. Fabisiak, PhD[13]; Xiaoping Chen, MS[2,14]; Kevin F. Gibson, MD[2,14]; Yingze Zhang, PhD[14]; Daniel J. Kass, MD[2,14]; Christopher J. Ryerson, MD, MAS*[4,15]; S. Mehdi Nouraie, MD, PhD*[14]

1. Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, U.S.A. Email: goobiegc@upmc.edu.
2. Simmons Center for Interstitial Lung Disease, Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A.
3. Clinician Investigator Program, Department of Medicine, University of British Columbia, Vancouver, B.C., Canada.
4. Division of Respiratory Medicine, Department of Medicine, University of British Columbia, Vancouver, B.C., Canada. Email: carlsten@mail.ubc.ca, chris.ryerson@hli.ubc.ca, nkhalil@mail.ubc.ca.
5. Air Pollution Exposure Laboratory, Vancouver Coastal Health Research Institute, Vancouver, B.C., Canada.
6. Division of Respiratory Medicine, Department of Medicine, University of Calgary, Calgary, AB, Canada. Email: kerri.johannson@ahs.ca.
7. Division of Respirology, Critical Care and Sleep Medicine, College of Medicine, University of Saskatchewan, Saskatoon, SK, Canada. Email: vsm940@mail.usask.ca.
8. Division of Respiratory Medicine, Department of Medicine, McGill University, Montreal, QC, Canada. Email: deborah.assayag@mcgill.ca.
9. Département de Médecine, Centre Hospitalier de l'Université de Montréal, Montreal, QC, Canada. Email: helene.manganas.med@ssss.gouv.qc.ca.
10. Division of Respirology, Department of Medicine, University of Toronto, Toronto, ON, Canada. Email: jolene.fisher@uhn.ca.
11. Department of Medicine, Firestone Institute for Respiratory Health, The Research Institute of St. Joe's Hamilton, St. Joseph's Healthcare, McMaster University, Hamilton, ON, Canada. Email: kolbm@mcmaster.ca.
12. College of Nursing, Medical University of South Carolina, Charleston, SC, U.S.A. Email: lindellk@musc.edu.
13. Department of Environmental and Occupational Health, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, U.S.A. Email: fabs@pitt.edu.
14. Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A. Emails: chenx9@upmc.edu, gibsonkf@upmc.edu, kassd2@upmc.edu, zhanyx@upmc.edu, nouraies@upmc.edu.
15. Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, B.C., Canada.

*Denotes co-senior authorship.

ORCID IDs: 0000-0001-7982-5635 (G.C.G.); 0000-0003-1205-5511 (K.A.J.); 0000-0003-0283-8985(J.P.F.); 0000-0001-6597-7830 (D.J.K.); 0000-0001-6947-2901 (Y.Z.); 0000-0002-2927-0659 (K.O.L.); 0000-0001-7465-0581 (S.M.N.).

Correspondence to:
Gillian C. Goobie, MD FRCPC
Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh
130 DeSoto Street, Pittsburgh, PA, 15261
Email: goobiegc@upmc.edu; Phone: 412-624-3018; Fax: 412-624-3020

Author Contributions: All authors meet criteria for authorship as recommended by the International Committee of Medical Journal Editors (ICMJE). GCG, CJR, and MN are responsible for study development, analysis, and initial manuscript production. CC, KAJ, KOL, JPF, KFG, YZ, and DJK provided critical commentary and initial review of the manuscript. MN provided insight about biostatistical and epidemiologic methods. NK, VM, DA, HM, JHF, and MRJK contributed to the development of the CARE-PF cohort. XC provided raw clinical data for analysis of outcomes in the Simmons cohort. All authors had full access to the data analyzed in this study, accept responsibility for publication, and provided critical input on the final version of the submitted manuscript.

Declaration of Interests: This was an independent, investigator-initiated study supported by Boehringer Ingelheim Pharmaceuticals, Inc. (BIPI). BIPI had no role in the design, analysis, or interpretation of the results in this study; BIPI was given the opportunity to review the manuscript for medical and scientific accuracy as it relates to BIPI substances, as well as intellectual property considerations. GCG receives research funding and support through the Pulmonary Fibrosis Foundation (PFF) Scholars Award Program and the University of British Columbia Clinician Investigator Program. CC is supported by the Canada Research Chair program. CJR, DA, and MRJK report personal fees and grants from Boehringer Ingelheim and Hoffman La Roche outside the submitted work. KAJ reports personal fees, non-financial support and other from Boehringer-Ingelheim and the Three Lakes Foundation, personal fees from Hoffman-La Roche Ltd., and grants from the University Hospital Foundation, the University of Calgary Cumming School of Medicine, and the Pulmonary Fibrosis Society of Calgary. VM reports grants and personal fees from Boehringer Ingelheim Canada, Hoffman La-Roche Ltd., and Astra Zeneca, and grants from the University of Saskatchewan and the Royal University Hospital Foundation. MRJK also reports personal fees and grants from GSK, Gilead, Actelion, Respivert, Genoa, Alkermes, Pharmaxis, Prometric, Indalo and Third Pole. HM reports grants from Hoffman La Roche, Galapagos and BMS and personal fees and research grants from Boehringer Ingelheim. DA also reports personal fees and grants from Novartis. JHF reports grants from the Canadian Pulmonary Fibrosis Foundation and personal fees from Boehringer Ingelheim and Astra Zeneca, outside of the submitted work. JPF reports funding from the Heinz Endowments, Pittsburgh, PA (#E6820). DJK is supported in part by AR060780, HL133232, UL1 TR001857, AR076024 NIH grants, Boehringer-Ingelheim grants, and receives collaborative research funding from Regeneron Pharmaceuticals, outside of the submitted work. YZ is supported in part by AR076024 NIH grants. NK, XC, and KOL report no competing interests. SMN receives grant support from Boehringer Ingelheim U.S.A.

Body word count: 3102/3000
Table count: 2 (13 supplemental)
Figure count: 3 (5 supplemental)

Keywords: air pollution; particulate matter; interstitial lung disease; idiopathic pulmonary fibrosis.

3.1.2 Key Points

Question: How do particulate matter with a diameter ≤2.5um ($PM_{2.5}$) and its constituents impact clinical outcomes in patients with fibrotic interstitial lung disease (fILD)?

Findings: $PM_{2.5}$ varied greatly across North America, with Western Pennsylvania patients exposed to the most $PM_{2.5}$ and its human-derived constituents (sulfate, nitrate, ammonium). Increasing exposure to $PM_{2.5}$ constituent mixtures was consistently associated with worse baseline lung function, more rapid progression, and increased mortality, with sulfate, nitrate, and ammonium driving associations.

Meaning: $PM_{2.5}$ exposures adversely impact fILD baseline severity, progression, and mortality in patients with fILD, with human-derived constituents of $PM_{2.5}$ pollution appearing to drive this risk.

3.1.3 Abstract

Importance: Particulate matter with a diameter ≤2.5μm ($PM_{2.5}$) adversely impacts patients with idiopathic pulmonary fibrosis, but associations in other fibrotic interstitial lung diseases (fILD) and contributions of $PM_{2.5}$ composition to adverse outcomes remains unclear.

Objective: To investigate the impact of $PM_{2.5}$ exposure on mortality and lung function in fILD.

Design: Multi-center, international, prospective cohort study.

Setting: Simmons Center fILD Registry at the University of Pittsburgh, forty-two sites of Pulmonary Fibrosis Foundation (PFF) Registry, and eight sites of Canadian Registry for Pulmonary Fibrosis (CARE-PF).

Participants: 6683 patients with fILD (1424 Simmons, 1870 PFF, 3389 CARE-PF).

Exposures: $PM_{2.5}$ and constituent exposures were estimated using hybrid models, combining satellite-derived aerosol optical depth with chemical transport models and ground-based $PM_{2.5}$ measurements.

Main Outcomes and Measures: Multivariable linear regression tested associations of exposures 5-years pre-enrollment with baseline forced vital capacity (FVC) and diffusion capacity for carbon monoxide (D_LCO). Multivariable Cox models tested associations of exposures in the 5-years pre-censoring with mortality and linear mixed models with lung function decline. Multi-constituent analyses were performed using quantile-based g-computation. Cohort effect estimates were meta-analyzed. Models adjusted for age, sex, smoking history, race, a socioeconomic variable, and site (PFF and CARE-PF).

Results: Median follow-up across the three cohorts was 2.9 years (IQR=1.5-4.5) with death in 28% and lung transplantation in 10% of patients. $PM_{2.5}$ exposure ≥8μg/m^3 was associated with

a hazard ratio (HR) for mortality of 4.40 (95%CI 3.51-5.51) in Simmons, 1.71 (95%CI 1.32-2.21) in PFF, and 1.45 (95%CI 1.18-1.79) in CARE-PF. Increasing sulfate, nitrate, and ammonium $PM_{2.5}$ constituent exposures were associated with increased mortality across all cohorts, and multi-constituent models demonstrated that these constituents tended to exert the most harmful impacts on mortality and baseline lung function. Meta-analyses revealed consistent harmful impacts of sulfate and ammonium on mortality and rate of FVC and D_LCO decline, and harmful impacts of increasing levels of $PM_{2.5}$ multi-constituent mixture on all outcomes.

Conclusions and Relevance: $PM_{2.5}$ exposure is associated with baseline severity, disease progression and mortality in patients with fILD. The most harmful effects appear driven by sulfate, ammonium, and nitrate constituents, highlighting the need for reductions in human-derived sources of pollution.

3.1.4 Introduction

Fibrotic interstitial lung diseases (fILDs) are a group of pulmonary conditions characterized by dyspnea, radiographic pulmonary fibrosis, and a high morbidity and mortality.(Cottin et al., 2018; Raghu et al., 2018) Idiopathic pulmonary fibrosis (IPF) is the most common and severe form of fILD, whose etiology remains incompletely understood.(Raghu et al., 2018) Air pollution adversely impacts IPF development and progression;(Conti et al., 2018; Johannson et al., 2018b; Sesé et al., 2018; Winterbottom et al., 2018) however, the impact of particulate matter with a diameter of ≤2.5μm ($PM_{2.5}$) on outcomes in patients with diverse forms of fibrotic interstitial lung disease (fILD) remains unclear.(Goobie et al., 2020) Furthermore, the contribution of specific $PM_{2.5}$ constituents to these outcomes has never been explored. Given the

severity and complex etiology of fILDs, there exists an urgent need to understand how environmental factors contribute to these diseases.

$PM_{2.5}$, which is made up of a miasma of fine airborne particles that exist in the atmosphere alongside gaseous pollutants, is responsible for 4.2-8.9 million premature deaths annually.(Burnett et al., 2018; Cohen et al., 2017) Satellite-derived hybrid models can estimate ambient $PM_{2.5}$ levels across the globe, with recent approaches enabling the speciation of $PM_{2.5}$ into constituent components, including sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic matter (OM), sea salt (SS), and soil.(Hammer et al., 2020; Van Donkelaar, Martin, Li, & Burnett, 2019) Constituents primarily derived from anthropogenic (i.e. human-derived) sources include SO_4^{2-}, NO_3^-, and NH_4^+. Complex reactions of gaseous emissions from fossil-fuel combustion, industrial activities (e.g. steel production), and agriculture result in the formation of ammonium-sulfate, ammonium-nitrate, and acidic sulfate and nitrate particles,(Hewitt, 2002; Plautz, 2018; H. Zhang, Hu, Kleeman, & Ying, 2014) with acidic particles exerting some of the most adverse impacts on human health.(Gwynn, Burnett, & Thurston, 2000; Spengler, Koutrakis, Dockery, Raizenne, & Speizer, 1996) BC is derived from multiple sources including fossil fuel and wood combustion from anthropogenic and natural sources.(Forbes, Raison, & Skjemstad, 2006) SS, soil, and OM constituents are components of normal atmospheric composition.(Philip et al., 2014)

We sought to evaluate the impact of exposure to $PM_{2.5}$ and its constituents on outcomes in patients with fILD using a multi-national cohort of patients with fILD from across the United States (U.S.) and Canada. We hypothesized that higher $PM_{2.5}$ and constituents related to anthropogenic sources (SO_4^{2-}, NO_3^-, NH_4^+, BC) would be associated with lower baseline lung function, more rapid lung function decline, and increased mortality. This work reflects the largest,

most geographically-diverse evaluation of the impacts of air pollution in patients with fILD to date. It is also the first study to evaluate the contribution of $PM_{2.5}$ constituents to outcomes in this population.

3.1.5 Methods

3.1.5.1 Study Populations

Adult patients with fILD whose diagnoses were made by specialist ILD physicians according to current clinical practice guidelines and best available evidence were eligible for inclusion.(Raghu et al., 2018, 2020) Non-fibrotic ILDs were excluded. The first U.S. cohort ("Simmons") included patients with fILD prospectively enrolled in the University of Pittsburgh Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease Registry between 2000 to 2021. The second U.S. cohort ("PFF") included patients with fILD prospectively enrolled in one of 42 Pulmonary Fibrosis Foundation (PFF) registry sites between 2016 to 2021.(B. R. Wang et al., 2020) Overlapping subjects in Simmons and PFF were excluded from the PFF cohort. The Canadian cohort ("CARE-PF") included patients with fILD prospectively enrolled in one of eight Canadian Registry for Pulmonary Fibrosis sites between 2015 to 2021,(Ryerson et al., 2016) and patients previously enrolled in single-center registries at CARE-PF sites prior to 2015.

Ethics approval was obtained from the University of Pittsburgh (STUDY20050209, STUDY21030226) and the University of British Columbia (#H19-01989 and #H20-01454).

3.1.5.2 Demographics, Residential Data, and Clinical Outcomes

Patient demographics (age at enrollment, sex, smoking history, self-reported race), most recent residential address (or 5-digit zip code for PFF cohort), specific fILD diagnosis, and lung

function (height- and weight-adjusted percent predicted forced vital capacity (FVC) and diffusion capacity of the lung for carbon monoxide (D_LCO)) were discerned from electronic health records (EHRs). Simmons data was in part extracted using the University of Pittsburgh Health Record Research Request Service.(Visweswaran et al., 2022) Baseline FVC and D_LCO were defined as the first tests performed within 6 months of enrollment. All FVC and D_LCO measurements obtained throughout follow-up were collected. Date of death and lung transplantation were confirmed through periodic extraction of these data from EHRs by each site's registry managers. Patients were considered "lost to follow-up" if they had not died, received a lung transplant, or had their registry record updated within 1 year of the censorship date (January 27, 2021 for Simmons; March 23, 2021 for CARE-PF). Details on losses to follow-up were not made available by the PFF Registry (censor date July 15, 2021).

The most recent residential address was geocoded into latitude and longitude coordinates using ArcGIS. PFF patients were assigned the centroid coordinates of their 5-digit zip code. Residential location was used to determine a socioeconomic variable, calculating the area deprivation index (ADI) for Simmons, the Canadian Index of Multiple Deprivation (CIMD) for CARE-PF,(Goobie, Ryerson, et al., 2022; Kind & Buckingham, 2018; *The Canadian Index of Multiple Deprivation: User Guide*, 2019) and percent of 5-digit zip below the poverty level for PFF (based on U.S. Census data).(U.S. Census Bureau, n.d.)

3.1.5.3 Particulate Matter and Constituent Component Exposure Determination

Estimates of monthly average total $PM_{2.5}$ and constituent (SO_4^{2-}, NO_3^-, NH_4^+, BC, OM, SS, soil) mass ($\mu g/m^3$) were acquired from the Atmospheric Composition Analysis Group online repository from 2000-2018 for $PM_{2.5}$ and 2000-2017 for constituents.(Hammer et al., 2020; Van Donkelaar et al., 2019) These data provide estimates for $PM_{2.5}$ and constituents at 0.01°x0.01°

(~1.1km^2) across North America based on satellite-derived aerosol optical depth measurements combined with chemical transport models and ground-based measurements. Total $PM_{2.5}$ mass, SO_4^{2-}, NO_3^-, NH_4^+, BC, OM, SS, and soil were ten-fold cross-validated with ground-based measurements, demonstrating an R^2 (and root mean squared error) values of 0.70 (1.6), 0.96 (0.3), 0.90 (0.3), 0.86 (0.2), 0.59 (0.1), 0.57 (0.8), 0.80 (0.1), and 0.60 (0.2), respectively.(Van Donkelaar et al., 2019) Residential coordinates were matched to nearest coordinates of pollutant data using the "ncdf4" package in R. Average exposures for each patient were determined for 5-years pre-censoring (censoring defined as the time of death, lung transplant, or cessation of follow-up) for mortality and lung function decline analyses. Average exposures were determined for 5-years pre-enrollment for baseline lung function analyses, and as a sensitivity analysis, for lung function decline analyses.

$PM_{2.5}$ exposures were evaluated continuously and as dichotomized low vs high exposures ($<$ or $\geq 8 \mu g/m^3$) based on American Thoracic Society (ATS) recommendations for yearly average $PM_{2.5}$ exposures.(Cromar, Gladson, Hicks, Marsh, & Ewart, 2021) Constituents were dichotomized based on their median value across the three cohorts.

3.1.5.4 Statistical Analyses

Survival analyses were performed using Cox proportional hazards regressions, considering death and lung transplant as a composite outcome. Assumptions were checked with Schoenfeld residuals. Spline models were constructed to evaluate the hazard ratio (HR) for mortality across different ranges of $PM_{2.5}$ and constituents across each cohort. Associations of $PM_{2.5}$ and constituents with baseline FVC and D_LCO were evaluated using multivariable linear regression. Linear mixed effects models with random intercepts and slopes for each patient were used to evaluate associations with rate of change in FVC or D_LCO.

Adjusted models included covariates of age at enrollment, sex, smoking history, race, ADI for Simmons, percent below poverty for PFF, CIMD for CARE-PF, and site (PFF and CARE-PF only). Adjustments for age, sex, baseline lung function, and neighborhood-level disadvantage were made as these factors predict mortality in fILDs;(Goobie, Ryerson, et al., 2022; Ley et al., 2012) for race, given impacts on pulmonary function and mortality;(Adegunsoye et al., 2018) for smoking, given impacts on lung function and potentially mortality;(Margaritopoulos et al., 2015) and for site to control for site-specific effects.

Cohort-specific attributable risk fractions (ARF) were calculated to determine the proportion of mortality attributable to $PM_{2.5}$ or anthropogenic constituent exposure, using the following formula:

$$Cohort\ Attributable\ Risk\ Fraction = \frac{P_e(HR - 1)}{1 + P_e(HR - 1)}$$

where P_e is the prevalence of high pollutant exposure in the cohort, and HR is the hazard ratio for high exposure in the fully-adjusted models.

Multi-constituent analyses of mortality and baseline lung function outcomes were performed using quantile-based g-computation with a linear additive approach for the addition of each $PM_{2.5}$ constituent, as has been previously employed with this exposure-matching approach.(Zhao et al., 2022)

A random effects meta-analysis of effect estimates across the three cohorts was performed for all primary outcomes, with I^2 values for heterogeneity reported.

Subgroup and sensitivity analyses are provided in the *Online Material*. This includes subgroup results for patients with IPF, a sensitivity analysis of mortality impacts pre- and post-2015 (where 2015 is the median year of enrollment across the three cohorts) to account for time-varying confounding, and a sensitivity analysis of $PM_{2.5}$ mortality impacts in 5-year pre-censoring

averages of warm (April-September) versus cold (October-March) month exposures to account for seasonality effects.

Analyses were performed using R (version 4.0.2, www.r-project.org).

3.1.6 Results

3.1.6.1 Baseline Patient Characteristics and Pollutant Exposures

Eligibility was met by 1424 Simmons, 1870 PFF, and 3389 CARE-PF patients with fILD, who were followed for a median of 3.1, 2.5, and 3.2 years, respectively. Baseline characteristics are shown in **Table 4**, with exposure breakdowns by site in **Appendix Table 11.** The most common diagnosis in Simmons and PFF was IPF, versus CTD-ILD in CARE-PF. Patient characteristics by low vs high $PM_{2.5}$ exposures (**Appendix Table 12**) demonstrate that higher proportions of non-White patients lived in high compared to low exposure areas (13% vs 8% in Simmons and PFF; 22% vs 12% in CARE-PF).

Table 4 – Patient demographics by cohort.

Patient Characteristics	Simmons Cohort	PFF Cohort	CARE-PF Cohort
	N=1424	N=1870	N=3389
$PM_{2.5}$ in 5yrs pre-enrollment, median (IQR), µg/m^3	11.4 (9.8, 13.6)	9.1 (7.9, 10.2)	6.2 (5.2, 8.1)
$PM_{2.5}$ in 5yrs pre-censoring, median (IQR), µg/m^3	9.4 (7.8, 11.4)	7.9 (7.0, 8.8)	6.2 (5.3, 7.3)
Age at enrollment, median (IQR), years	66 (58, 73)	68 (61, 73)	66 (57, 73)
Male sex, n (%)	795 (56%)	1186 (63%)	1672 (49%)
Self-reported race, n (%)			
White	1258 (88%)	1669 (89%)	2682 (79%)
Black	56 (4%)	96 (5%)	53 (2%)
Asian	5 (0.4%)	48 (3%)	382 (11%)
Indigenous[a]	2 (0.1%)	3 (0.2%)	88 (3%)
Pacific Islander	0 (0%)	3 (0.2%)	26 (1%)
Unknown	103 (7%)	51 (3%)	158 (4%)
Smoking history, n (%)			
Never	413 (29%)	779 (42%)	1279 (38%)
Former	664 (47%)	"Ever"	1922 (57%)
Current	38 (2%)	1091 (58%)	176 (5%)
Unknown	309 (21%)	0 (0%)	0 (0%)
fILD diagnostic group, n(%)			

Idiopathic pulmonary fibrosis	716 (50%)	1202 (64%)	924 (27%)
Connective tissue disease-ILD	300 (21%)	310 (17%)	1298 (38%)
Fibrotic hypersensitivity pneumonitis	55 (4%)	152 (8%)	259 (8%)
Pneumoconiosis	26 (2%)	0 (0%)	28 (1%)
Non-IPF idiopathic interstitial pneumonia	68 (5%)	144 (8%)	109 (3%)
Other fILD[b]	50 (3%)	0 (0%)	121 (4%)
Unclassifiable or not yet diagnosed	209 (15%)	62 (3%)	650 (19%)
Urbanicity			
Metropolitan (>50,000 people)	1078 (76%)	1611 (86%)	2333 (69%)
Micropolitan (10,000-50,000 people)	223 (16%)	138 (7%)	597 (18%)
Rural (<10,000 people)	122 (8%)	120 (7%)	459 (13%)
Neighborhood Disadvantage (ADI for Simmons or CIMD for CARE-PF), median score (IQR)	62 (44, 78)	N/A	-0.02 (-0.35, 0.38)
Percent of 5-digit zip below poverty line, median (IQR)	N/A	9% (6-14%)	N/A
Baseline FVC % Predicted, median (IQR)[c]	66 (53, 81)	67 (55, 80)	75 (62, 89)
Baseline D_LCO % Predicted, median (IQR)[d]	49 (37, 63)	40 (31, 51)	57 (44, 71)
Follow-up Duration, median (IQR), years	3.1 (1.2, 6.3)	2.5 (1.4, 3.5)	3.2 (1.8, 5.2)
Cause of censoring, n (%)			
Death	707 (50%)	429 (23%)	765 (23%)
Lung Transplantation	201 (14%)	258 (14%)	176 (5%)
Lost to follow-up (no registry update for >1 year)	181 (13%)	N/A	23 (0.7%)
Censored	335 (23%)	1183 (63%)	2425 (72%)

[a] – Includes Native American, American Indian, Alaskan First Nations, & other Indigenous persons in the U.S.; First Nations, Métis, Inuit, and other Indigenous persons in Canada.
[b] – Includes drug-, radiation-, aspiration-, or acute lung injury-induced fILD.
[c] – FVC was available for 76% of patients enrolled in Simmons, 91% of PFF, and 88% of CARE-PF.
[d] – D_LCO was available for 70% of patients enrolled in Simmons, 85% of PFF, and 70% of CARE-PF
Abbreviations: ADI, area deprivation index; CIMD, Canadian Index of Multiple Deprivation; CARE-PF, Canadian Registry for Pulmonary Fibrosis; D_LCO, diffusion capacity of lung for carbon monoxide; fILD, fibrotic interstitial lung disease; FVC, forced vital capacity; IPF, idiopathic pulmonary fibrosis; IQR, interquartile range; $PM_{2.5}$, particulate matter with a diameter of 2.5μm or less; PFF, Pulmonary Fibrosis Foundation.

Figure 7 shows the geographic distribution of $PM_{2.5}$ across North America, site locations, and the median breakdown of constituents across the three cohorts. **Appendix Figure 5** shows cohort-specific correlations of $PM_{2.5}$ constituents in the 5-years pre-censoring . Simmons patients experienced the highest exposures to $PM_{2.5}$, SO_4^{2-}, NO_3^-, NH_4^+, and BC, followed by PFF, then CARE-PF. $PM_{2.5}$ exposures in the 5-year pre-enrollment and 5-year pre-censoring were highly correlated (combined cohorts r=0.84).

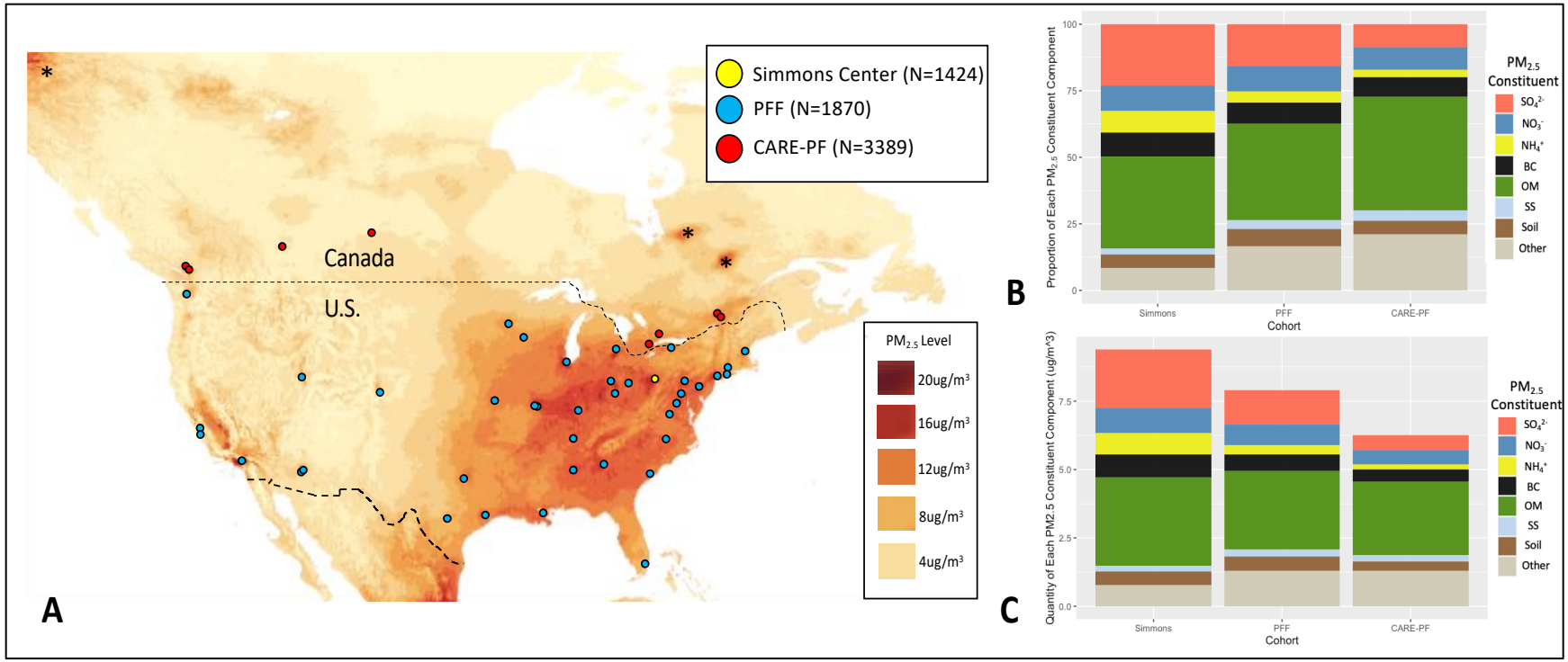
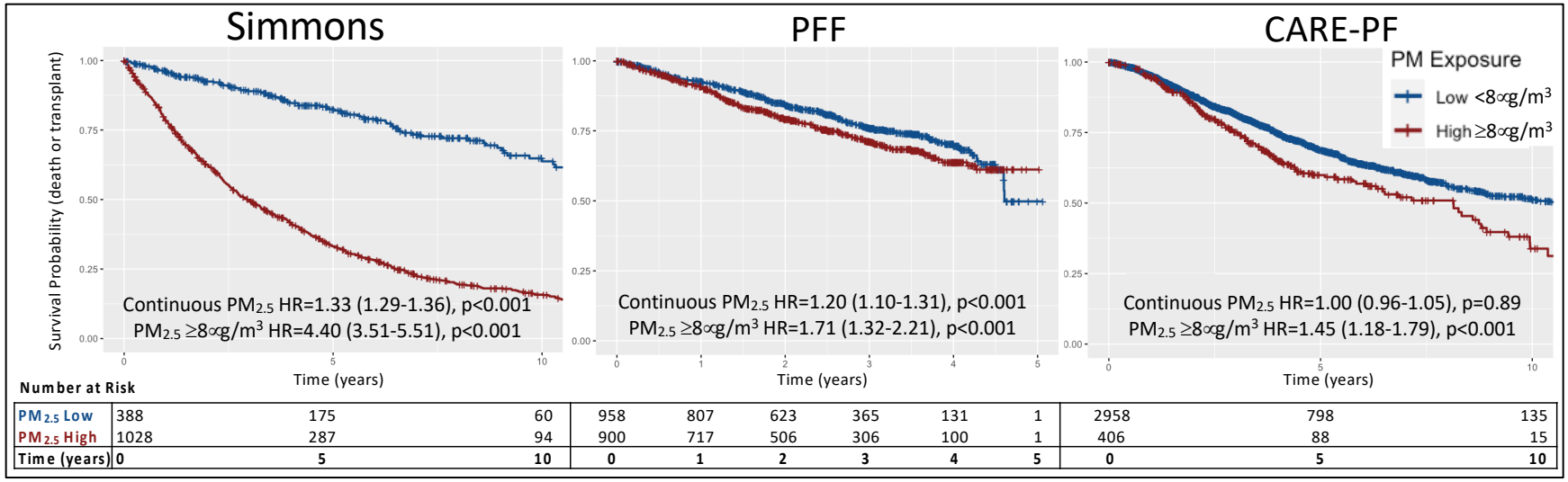


Figure 7 – **$PM_{2.5}$ distribution and constituent composition across three cohorts.** A) Average of satellite-derived $PM_{2.5}$ level ($\mu g/m^3$) across North America for a representative year (2005) with Simmons Center for ILD referral center in yellow, PFF registry referral centers in blue, and CARE-PF ILD referral centers in red. $PM_{2.5}$ and constituent component estimates are accurate to 0.01°by 0.01° (approximately 1.1km^2 at the equator), with average monthly estimates available across North America from 2000-2018. **B)** Proportion of median total $PM_{2.5}$ mass in 5-years pre-censoring that each constituent component makes up in each cohort. **C)** Total mass of median exposure to $PM_{2.5}$ in 5-years pre-censoring broken down by each constituent component (measured in $\mu g/m^3$) across each cohort. *Hot spots noted in Quebec and Alaska (denoted by *) reflect wildfires that occurred in those locations in 2005, highlighting how high $PM_{2.5}$ levels during such exceptional events can drive up yearly averages of exposures in these remote, rural regions.

3.1.6.2 Association of $PM_{2.5}$ and Constituent Components with Survival

In all cohorts, 5-year pre-censoring $PM_{2.5}$ exposures $\geq 8 \mu g/m^3$ were associated with increased mortality, with the highest effect size in Simmons (**Figure 2**, **Appendix Table 13**). Continuous models demonstrate similar findings, although the effect is not significant for CARE-PF, with spline models indicating a potentially non-linear association between total $PM_{2.5}$ mass and mortality in CARE-PF (**Appendix Figure 6**). Meta-analysis indicates high heterogeneity between the cohorts (I^2=98%) but supports that increasing $PM_{2.5}$ is associated with mortality (HR=1.18, 95%CI=1.02-1.37, p=0.03) (**Appendix Table 14**).



Number at Risk

PM2.5 Low	388	175	60	958	807	623	365	131	1	2958	798	135
PM2.5 High	1028	287	94	900	717	506	306	100	1	406	88	15
Time (years)	0	5	10	0	1	2	3	4	5	0	5	10

Figure 8 – Survival by low versus high (< or ≥8ug/m^3) $PM_{2.5}$ exposures in 5-years pre-censoring. Kaplan-Meier survival curves for associations of exposures to $PM_{2.5}$ total mass in the 5-years pre-censoring, where death and transplant are considered composite outcomes. Hazard ratios (HR) reported for dichotomized and continuous models are adjusted for age at enrollment, sex, race, smoking history, a socioeconomic variable, and site (PFF and CARE-PF only).

Analyses broken down by $PM_{2.5}$ constituents (**Table 5**, **Appendix Table 13**) demonstrate strong associations between higher SO_4^{2-}, NO_3^-, and NH_4^+ in the 5-years pre-censoring and mortality across all cohorts (**Appendix Figure 7**). Effects of other constituents (BC, OM, SS, soil) were less consistent (**Appendix Table 13**). Spline models of continuous HRs for total $PM_{2.5}$ mass and each constituent are shown in **Appendix Figure 6.** Multi-constituent models demonstrate consistent impacts of increasing $PM_{2.5}$ constituent mixture on mortality (meta-analysis HR=2.30 per 1-quantile increase in mixture, 95%CI=2.11-2.50, $p<0.001$, $I^2=25\%$), with SO_4^{2-} and NH_4^+ contributing the most harm in all cohorts (**Figure 9, Appendix Table 14, Appendix Table 15**).

Table 5 – Results from adjusted models for primary outcomes of mortality (death and transplant considered composite outcomes), baseline forced vital capacity (FVC) and diffusion capacity of the lung for carbon monoxide (D_LCO), and rate of decline in FVC and D_LCO. Single pollutant effect estimates are per each 1μg/m^3 increase in $PM_{2.5}$ or constituent. Multi-constituent effect estimates are per 1 quantile increase in the $PM_{2.5}$ mixture of all constituents (SO_4^{2-}, NO_3^-, NH_4^+, BC, OM, SS, and soil). Analyses are adjusted for age at enrollment, sex, smoking history, race, a socioeconomic variable (ADI for Simmons, percent of 5-digit zip below poverty in PFF, CIMD for CARE-PF), and site (in PFF and CARE-PF). Exposure period for mortality and lung function decline models are the average of monthly exposures in the 5-years pre-censoring, while baseline lung function exposure periods are for the average of monthly exposures in the 5-years pre-enrollment. Unadjusted models and models for other $PM_{2.5}$ constituents are reported in the *Online Material*. Significant associations are **bolded**.

Outcome	Simmons			PFF			CARE-PF			Meta-Analysis			
	HR/β (95% CI)	p	n	HR/β (95% CI)	p	n	HR/β (95% CI)	p	n	HR/β (95% CI)	p	I^2	n
Mortality													
Total $PM_{2.5}$	**1.33 (1.29-1.36)**	**<0.001**	1372	**1.20 (1.10-1.31)**	**<0.001**	1832	1.00 (0.96-1.05)	0.89	3353	**1.18 (1.02-1.37)**	**0.03**	98%	6557
SO_4^{2-}	**1.79 (1.70-1.89)**	**<0.001**		**132.19 (78.12-223.70)**	**<0.001**		**2.26 (2.05-2.48)**	**<0.001**		8.02 (0.52-122.63)	0.13	99%	
NO_3^-	**3.59 (3.10-4.16)**	**<0.001**		**2.48 (1.74-3.53)**	**<0.001**		**6.26 (4.16-9.42)**	**<0.001**		**3.78 (2.30-6.20)**	**<0.001**	82%	
NH_4^+	**4.31 (3.83-4.86)**	**<0.001**		**903.17 (408.40-1998.00)**	**<0.001**		**36.22 (27.32-48.03)**	**<0.001**		**50.99 (2.46-1056.64)**	**0.01**	99%	
Multi-Constituent	**2.19 (1.93-2.48)**	**<0.001**		**2.76 (2.15-3.54)**	**<0.001**		**2.30 (2.02-2.61)**	**<0.001**		**2.30 (2.11-2.50)**	**<0.001**	25%	
Baseline FVC													
Total $PM_{2.5}$	**-0.98 (-1.45 to -0.50)**	**<0.001**	1048	0.20 (-0.40 to 0.79)	0.52	1672	-0.07 (-0.59 to 0.46)	0.80	2958	-0.30 (-1.00 to 0.41)	0.41	82%	5678
SO_4^{2-}	**-1.85 (-2.73 to -0.98)**	**<0.001**		-1.17 (-2.72 to 0.38)	0.14		1.38 (-1.17 to 3.93)	0.29		-0.90 (-2.52 to 0.73)	0.28	65%	
NO_3^-	**-4.13 (-7.61 to -0.65)**	**0.02**		1.11 (-1.21 to 3.42)	0.35		1.77 (-1.47 to 5.01)	0.28		-0.29 (-3.78 to 3.20)	0.87	73%	
NH_4^+	**-4.80 (-7.09 to -2.52)**	**<0.001**		-1.83 (-4.74 to 1.08)	0.22		**4.87 (0.49 to 9.25)**	**0.03**		-0.85 (-6.28 to 4.58)	0.76	87%	
Multi-Constituent	**-4.44 (-6.20 to -2.69)**	**<0.001**		-1.61 (-3.66 to 0.44)	0.12		**-3.75 (-5.13 to -2.37)**	**<0.001**		**-3.38 (-4.88 to -1.87)**	**<0.001**	56%	
Baseline D_LCO													
Total $PM_{2.5}$	-0.13 (-0.63 to 0.36)	0.60	978	**-0.86 (-1.42 to -0.31)**	**0.002**	1547	0.006 (-0.54 to 0.56)	0.98	2383	-0.32 (-0.84 to 0.19)	0.22	64%	4908
SO_4^{2-}	-0.23 (-1.14 to 0.69)	0.62		**-3.76 (-5.19 to -2.32)**	**<0.001**		-0.03 (-2.83 to 2.78)	0.98		-1.43 (-3.87 to 1.02)	0.25	88%	
NO_3^-	0.49 (-3.12 to 4.08)	0.79		-0.51 (-2.66 to 1.63)	0.64		0.23 (-3.16 to 3.62)	0.89		-0.14 (-1.76 to 1.47)	0.86	0%	
NH_4^+	-0.25 (-2.63 to 2.13)	0.83		**-6.54 (-9.23 to -3.85)**	**<0.001**		2.84 (-1.89 to 7.57)	0.24		-1.54 (-6.88 to 3.80)	0.57	88%	

Multi-Constituent	**-4.14 (-5.96 to -2.33)**	**<0.001**		**-2.40 (-4.31 to -0.48)**	**0.01**		**-4.02 (-5.47 to -2.57)**	**<0.001**		**-3.64 (-4.61 to -2.66)**	**<0.001**	9%	
FVC Decline													
Total $PM_{2.5}$	**-0.40 (-0.53 to -0.27)**	**<0.001**	1055	0.007 (-0.28 to 0.30)	0.96	1153	-0.01 (-0.13 to 0.11)	0.86	2959	-0.15 (-0.42 to 0.12)	0.29	90%	5167
SO_4^{2-}	**-0.88 (-1.13 to -0.64)**	**<0.001**		**-3.39 (-5.37 to -1.40)**	**<0.001**		**-3.73 (-4.95 to -2.52)**	**<0.001**		**-2.53 (-4.45 to -0.62)**	**0.01**	92%	
NO_3^-	**-2.82 (-3.90 to -1.74)**	**<0.001**		-0.89 (-2.21 to 0.42)	0.18		**-1.35 (-2.40 to -0.29)**	**0.01**		**-1.72 (-2.86 to -0.58)**	**0.003**	67%	
NH_4^+	**-2.16 (-2.73 to -1.58)**	**<0.001**		**-7.04 (-10.41 to -3.68)**	**<0.001**		**-9.05 (-11.19 to -6.91)**	**<0.001**		**-5.93 (-10.18 to -1.69)**	**0.006**	95%	
D_LCO Decline													
Total $PM_{2.5}$	**-0.28 (-0.42 to -0.15)**	**<0.001**	1013	0.09 (-0.24 to 0.43)	0.58	1090	0.08 (-0.04 to 0.21)	0.19	2775	-0.05 (-0.31 to 0.21)	0.70	88%	4878
SO_4^{2-}	**-0.67 (-0.93 to -0.41)**	**<0.001**		**-2.93 (-5.28 to -0.58)**	**0.02**		**-3.29 (-4.64 to -1.95)**	**<0.001**		**-2.12 (-3.93 to -0.30)**	**0.02**	88%	
NO_3^-	**-2.61 (-3.80 to -1.43)**	**<0.001**		-0.23 (-1.77 to 1.34)	0.79		-0.66 (-179 to 0.46)	0.25		-1.21 (-2.66 to 0.24)	0.10	75%	
NH_4^+	**-1.74 (-2.35 to -1.12)**	**<0.001**		-4.04 (-8.15 to 0.07)	0.05		**-8.42 (-10.78 to -6.06)**	**<0.001**		**-4.66 (-8.77 to -0.54)**	**0.03**	93%	

Abbreviations: ADI, area deprivation index; CIMD, Canadian Index of Multiple Deprivation; CARE-PF, Canadian Registry for Pulmonary Fibrosis; D_LCO, diffusion capacity of lung for carbon monoxide; FVC, forced vital capacity; HR, hazard ratio; IQR, interquartile range; NH_4^+, ammonium; NO_3^-, nitrate; $PM_{2.5}$, particulate matter with a diameter ≤2.5μm; PFF, Pulmonary Fibrosis Foundation; SO_4^{2-}, sulfate.

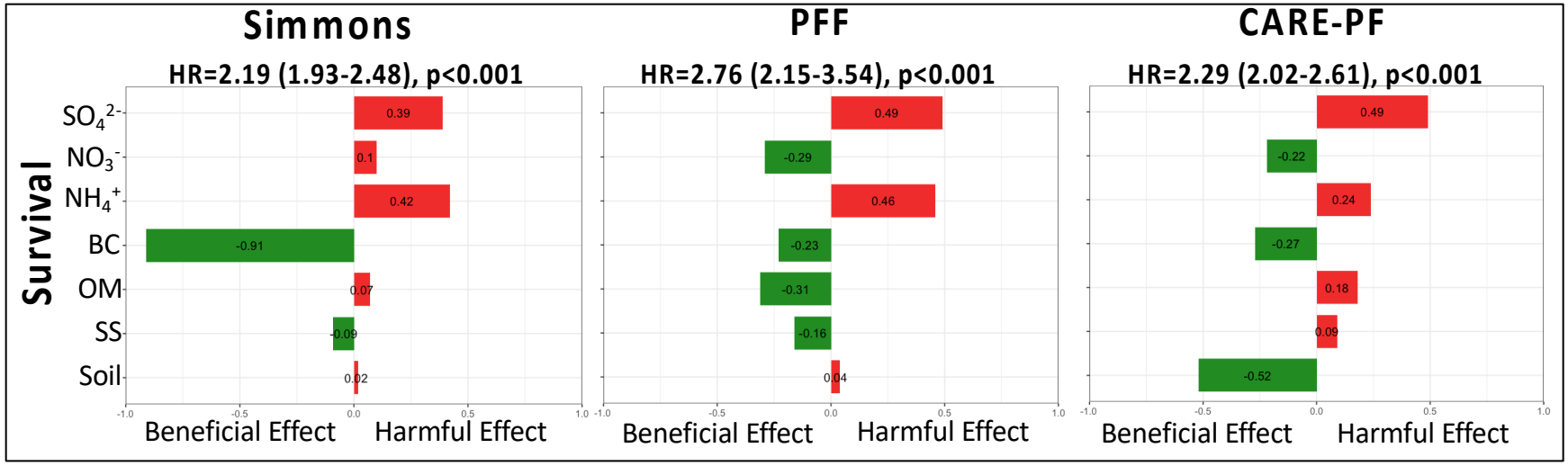


Figure 9 – Tornado plots of $PM_{2.5}$ constituent impacts on mortality in multi-pollutant models. Results are reported from adjusted quantile-based g-computation Cox proportional hazards survival models where 5-year pre-censoring estimates for SO_4^{2-}, NO_3^-, NH_4^+, BC, OM, SS, and soil were included. All models were adjusted for age at enrollment, sex, race, smoking history, a socioeconomic variable, and site (for PFF and CARE-PF). The weight of effect in a direction is displayed over the bars of each plot with bars representing harmful effects displayed in red and bars representing protective effects in green. The sum of all positive weights equals 1 and all negative weights equals -1 (i.e. cannot directly compare effect size between positive and negative weights). The HR (95% CI) and p-value for a 1-quantile increase in the overall mixture is reported above each plot.

Attributable risk fractions (ARFs) were calculated for dichotomized total $PM_{2.5}$ mass, and SO_4^{2-}, NO_3^-, and NH_4^+. The ARF for $PM_{2.5}$ ≥8μg/m^3 was 0.71 in Simmons, 0.26 in PFF, and 0.05 in CARE-PF indicating that if high exposures to $PM_{2.5}$ were removed, that 71% of the premature mortality in the Simmons cohort could be avoided, as compared to only 5% in CARE-PF. The attributable fractions for anthropogenic constituents were greatest in Simmons, followed by PFF, then CARE-PF (**Appendix Table 16**). SO_4^{2-} and NH_4^+ carried the highest risk burdens.

Subgroup analyses in patients with IPF show generally consistent effects (**Appendix Table 17**). Effect sizes varied between pre- and post-2015 year of enrollment subgroups indicating some time variability, but directionality remained consistent (**Appendix Table 18**). In Simmons and CARE-PF, the HR associated with increasing $PM_{2.5}$ is higher for cold compared to warm months, whereas the opposite is seen for PFF, indicating regional variability in seasonal effects (**Appendix Table 19**).

3.1.6.3 Association of $PM_{2.5}$ and Constituent Components with Baseline Lung Function

In adjusted Simmons models, a 1μg/m^3 increase in 5-year pre-enrollment $PM_{2.5}$ was associated with a 0.98% lower percent-predicted baseline FVC (95% confidence interval (CI) -1.45 to -0.50, p<0.001), but was not significant in PFF or CARE-PF (**Table 5, Appendix Table 20**). Multi-constituent analyses in Simmons and CARE-PF demonstrated that increased exposures to the $PM_{2.5}$ constituent mixture were associated with lower baseline FVC (**Appendix Figure 8**), with SO_4^{2-} and NH_4^+ consistently demonstrating harmful effects. Meta-analyses indicate that across the three cohorts a 1-quantile increase in the constituent mixture is associated with a 3.38% lower percent-predicted baseline FVC (**Table 5, Appendix Table 14, Appendix Table 15**).

Total $PM_{2.5}$ in the 5-years pre-enrollment was only associated with lower baseline D_LCO in PFF (β=-0.86, 95%CI -1.42 to -0.31, p=0.002), as were SO_4^{2-} and NH_4^+ constituents (**Table 5, Appendix Table 21**). Multi-constituent models indicate consistent negative impacts of increasing $PM_{2.5}$ mixture on baseline D_LCO (**Appendix Figure 9**, **Table 5, Appendix Table 21**), again with SO_4^{2-} and NH_4^+ consistently demonstrating harmful effects. Meta-analysis indicates that each 1-quantile increase in constituent mixture is associated with a 3.64% lower baseline percent-predicted D_LCO (95%CI -4.61 to -2.66, p<0.001, I^2=9%).

3.1.6.4 Association of $PM_{2.5}$ and Constituent Components with Lung Function Decline

We evaluated the rate of FVC decline in 74% of Simmons, 62% of PFF, and 87% of CARE-PF patients. Each 1μg/m^3 increase in 5-year pre-censoring $PM_{2.5}$ exposures in Simmons was associated with an additional 0.4% decline in FVC percent predicted/year (95%CI -0.53 to -0.27, p<0.001), but this effect was not seen in PFF or CARE-PF (**Table 5, Appendix Table 22**). Higher exposures to SO_4^{2-}, NO_3^-, and NH_4^+ were associated with more rapid decline in meta-analysis (**Table 5**, **Appendix Table 14**).

We evaluated the rate of D_LCO decline in 71% of Simmons, 58% of PFF, and 82% of CARE-PF patients. Each 1μg/m^3 increase in 5-year pre-censoring $PM_{2.5}$ exposures in Simmons was associated with an additional 0.28% decline in D_LCO percent predicted/year (95%CI -0.42 to -0.15, p<0.001), but this effect was not seen in PFF or CARE-PF (**Table 5, Appendix Table 23**). Higher exposures to SO_4^{2-} and NH_4^+ were associated with more rapid decline in meta-analysis (**Table 5**, **Appendix Table 14**).

3.1.7 Discussion

This study of 6683 patients with fILD from across North America demonstrates that $PM_{2.5}$ and its constituents, primarily SO_4^{2-}, NO_3^-, and NH_4^+, adversely impact mortality, baseline severity, and disease progression in patients with fILD. Differences in mortality and lung function impacts between the three cohorts demonstrates how $PM_{2.5}$ constituents related to industry and human activities contribute most significantly to the adverse impacts of $PM_{2.5}$ in patients with fILD.

In this geographically and diagnostically diverse cohort, we demonstrate an association between high $PM_{2.5}$ exposure and mortality. This is most pronounced in the Simmons cohort, which has the highest burden of heavy industry-associated $PM_{2.5}$ constituents: SO_4^{2-}, NO_3^-, and NH_4^+. We demonstrate consistency in the mortality impact attributable to SO_4^{2-}, NO_3^-, and NH_4^+ across all three cohorts, highlighting how these constituents may be primary drivers of $PM_{2.5}$-associated mortality in fILDs. Our findings are consistent with recent literature that demonstrates increased all-cause mortality associated with $PM_{2.5}$ constituents SO_4^{2-} and NO_3^-, as compared with constituents like soil or OM.(Kazemiparkouhi et al., 2022; C. Wang et al., 2022) Recent work also indicates that higher NH_4^+ is associated with increased ILD incidence in patients with rheumatoid arthritis, indicating that this constituent may have pathophysiologic relevance to both the development and progression of fILDs.(Zhao et al., 2022)

Attributable risk fractions (ARFs) illustrate how $PM_{2.5}$'s contribution to fILD mortality varies substantially depending on the mass and constituent makeup of $PM_{2.5}$ in a region. An ARF can exceed 100% because of complex interactions between social, environmental, and biologic risk factors, indicating the need to interpret these findings with caution,(Levine, 2007) but the Simmons attributable risk fraction of 0.71 for patients exposed to high $PM_{2.5}$ ($\geq 8 \mu g/m^3$) implies

that 71% of this group's premature mortality could be averted if the exposure did not occur. This metric is most useful for weighing relative burdens across cohorts, implying that the mortality burden attributable to high $PM_{2.5}$ in the Simmons cohort is ~2.7 times greater than the PFF cohort and ~14 times greater than the CARE-PF cohort. Policy interventions that reduce $PM_{2.5}$ total mass to below ATS standards, with specific targeting of anthropogenic sources of emissions, may have the greatest impact in reducing mortality in this vulnerable group.

While there was consistency in the impact of high $PM_{2.5}$ exposure ($\geq 8 \mu g/m^3$) on mortality and baseline FVC across all three cohorts, there were inconsistencies in some constituent analyses, baseline D_LCO models, and lung function decline analyses. Inconsistencies in lung function decline analyses may relate to the higher proportions of patients with IPF in Simmons and PFF, whereas CARE-PF has a larger proportion of connective tissue disease-ILD (CTD-ILD), wherein patients may experience a more indolent disease course.(Ryerson et al., 2014) $PM_{2.5}$ constituent distribution also varied greatly, with Simmons demonstrating the highest proportion of deleterious constituents from industrial activities, including SO_4^{2-} and NH_4^+, whereas CARE-PF has higher proportions of OM and other unmeasured constituents, which are frequently derived from biomass burning and non-anthropogenic sources.(Philip et al., 2014) Similarly, CARE-PF patients experienced lower total exposures to $PM_{2.5}$ and constituent components as compared to Simmons or PFF patients. In Simmons, this is likely related to the earlier establishment of the ILD registry alongside higher historical pollution exposures in Western Pennsylvania where steel, coal, and other metal industries predominate.(Dutzik, Group, Barber, Research, & Center, 2019) The relatively minimal exposure of CARE-PF patients to SO_4^{2-} and NH_4^+ may blunt the significance of associations between total $PM_{2.5}$ mass and mortality.

While this study is strengthened by its large, geographically- and diagnostically-diverse multi-national cohort of patients with fILD, it is not without limitations. Pollution exposures were estimated at a patient's most recent residential address, but this approach does not account for mobility or changes in address (data for which was unavailable), leading to some risk of exposure misclassification. This study also evaluated pollution exposures over a pre-specified period of 5 years pre-censoring (or 5 years pre-enrollment for baseline lung function analyses), and future work is needed to determine the best at-risk period in these patients. Given that this is not a population-based study, analyses of exposures impacts on clinical outcomes may suffer from selection bias as not all patients may have access to subspecialist ILD care. Additionally, patients lost to follow-up represented >10% of the Simmons cohort and were not available for the PFF cohort, which may impact mortality estimates if these losses were not truly "random" as was assumed. Further work is also needed to understand the biologic mechanisms underpinning the association between $PM_{2.5}$, its constituent components, and adverse clinical outcomes in this population.(Goobie et al., 2020) Lastly, extensions of this work should employ strategies to source-apportion the $PM_{2.5}$ constituents SO_4^{2-}, NO_3^-, and NH_4^+ that patients are exposed to, thereby enabling more stringent regulatory oversight of the human sources of emissions from which these constituents are derived.

3.1.8 Conclusions

This study represents the largest and most geographically-diverse evaluation of the impacts of $PM_{2.5}$ pollution on patients with diverse forms of fILD, and the first to evaluate the impact of specific $PM_{2.5}$ constituents on outcomes in this population. We found that exposures to total $PM_{2.5}$ mass $\geq 8 \mu g/m^3$ were consistently associated with increased mortality, worse baseline lung function,

and more rapid lung function decline. Multi-constituent models demonstrated that SO_4^{2-}, NO_3^-, and NH_4^+ constituents, which are primary by-products of industrial and transportation activities, are the main contributors to $PM_{2.5}$-associated adverse outcomes in patients with fILD. This work unveils new paradigms in our understanding of the importance of $PM_{2.5}$ composition to health outcomes, indicating a strong need to evaluate constituent-specific effects in other diseases. These findings are of critical importance to the development of policies aimed at protecting vulnerable populations, such as patients with fILD, as they demonstrate how anthropogenic sources of pollution may contribute more significantly to disease morbidity and mortality.

3.1.9 Acknowledgements

We would like to thank the fILD patients from the Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease and each of the CARE-PF registry sites for their time and commitment to clinical research. We thank all patients who participated in the Pulmonary Fibrosis Foundation (PFF) Registry. We also thank the investigators and other staff at participating PFF Care Centers for providing clinical data, and the Pulmonary Fibrosis Foundation, which established and has maintained the PFF Registry since 2016, and lastly, the many generous donors of the PFF Registry.

3.1.10 Data Sharing

Because of the restrictions of the consents signed by participants in the Simmons Center for ILD Registry, PFF, and CARE-PF, individual participant data will not be made available. The code repository for this manuscript, containing code for exposure matching, statistical analysis,

and figure creation is available from <https://github.com/gcgoobie/PM2.5_ClinicalOutcomes>. Further information regarding registry protocols, consent forms, or other specific data can be made available upon request by contacting the corresponding author, Dr. Gillian Goobie (goobiegc@upmc.edu or gcgoobie@alumni.ubc.ca).

4.0 Aim 3 – $PM_{2.5}$ Impacts on DNA Methylation and Telomere Length in fILDs

4.1 – Aim 3.1 Global DNA Methylation Manuscript

4.1.1 Original Research Manuscript under review with *Environmental Pollution* – "$PM_{2.5}$ and constituent component impacts on global DNA methylation in patients with idiopathic pulmonary fibrosis"

The following manuscript is under review with *Environmental Pollution* and the following version is available as a preprint on SSRN at: http://ssrn.com/abstract=4204633

The supplementary data for this manuscript is available in **Appendix C**.

$PM_{2.5}$ and constituent component impacts on global DNA methylation in patients with idiopathic pulmonary fibrosis

Gillian C. Goobie, MD[1,2,3]; Xiaoyun Li, MD[4]; Christopher J. Ryerson, MD, MAS[5,6]; Christopher Carlsten, MD MPH[5,7]; Kerri A. Johannson, MD, MPH[8]; James P. Fabisiak, PhD[9]; Kathleen O. Lindell, RN, PhD[2,10]; Xiaoping Chen, MS[2,4]; Kevin F. Gibson, MD[2,4]; Daniel J. Kass, MD[2,4]; S. Mehdi Nouraie, MD, PhD[4]; Yingze Zhang, PhD[1,2,4]

1. Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, U.S.A. Email: goobiegc@upmc.edu.
2. Simmons Center for Interstitial Lung Disease, Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A.
3. Clinician Investigator Program, Department of Medicine, University of British Columbia, Vancouver, B.C., Canada.
4. Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A. Emails: lix11@upmc.edu; chenx9@upmc.edu; gibsonkf@upmc.edu, kassd2@upmc.edu, zhanyx@upmc.edu, nouraies@upmc.edu.

5. Division of Respiratory Medicine, Department of Medicine, University of British Columbia, Vancouver, B.C., Canada. Email: carlsten@mail.ubc.ca, chris.ryerson@hli.ubc.ca.
6. Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, B.C., Canada.
7. Air Pollution Exposure Laboratory, Vancouver Coastal Health Research Institute, Vancouver, B.C., Canada.
8. Division of Respiratory Medicine, Department of Medicine, University of Calgary, Calgary, AB, Canada. Email: kerri.johnnson@albertahealthservices.ca.
9. Department of Environmental and Occupational Health, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, U.S.A. Email: fabs@pitt.edu.
10. College of Nursing, Medical University of South Carolina, Charleston, SC, U.S.A. Email: lindellk@musc.edu.

ORCID IDs: 0000-0001-7982-5635 (G.C.G.); 0000-0003-1205-5511 (K.A.J.); 0000-0003-0283-8985(J.P.F.); 0000-0001-6597-7830 (D.J.K.); 0000-0001-7465-0581 (S.M.N.); 0000-0001-6947-2901 (Y.Z.).

Correspondence to:
Gillian C. Goobie, MD FRCPC
Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh
130 DeSoto Street, Pittsburgh, PA, 15261
Email: goobiegc@upmc.edu; Phone: 412-624-3018; Fax: 412-624-3020

Author Contributions: All authors meet criteria for authorship as recommended by the International Committee of Medical Journal Editors (ICMJE). GCG, YZ, and MN are responsible for study development, analysis, and initial manuscript production. XL, CJR, CC, KAJ, JPF, KOL, XC, KFG, and DJK provided critical commentary and initial review of the manuscript. MN provided insight about biostatistical and epidemiologic methods. XC provided raw clinical data for analysis of outcomes in the Simmons cohort. All authors had full access to the data analyzed in this study, accept responsibility for publication, and provided critical input on the final version of the submitted manuscript.

Declaration of Interests: This was an independent, investigator-initiated study supported by the Pulmonary Fibrosis Foundation (PFF). PFF had no role in the design, analysis, or interpretation of the results in this study; PFF was given the opportunity to review the manuscript for medical and scientific accuracy. GCG receives research funding and support through the Pulmonary Fibrosis Foundation (PFF) Scholars Award Program and the University of British Columbia Clinician Investigator Program. XL, KOL, and XC report no competing interests. CC is supported by the Canada Research Chair program. CJR reports personal fees and grants from Boehringer Ingelheim and Hoffman La Roche outside the submitted work. KAJ reports personal fees, non-financial support and other from Boehringer-Ingelheim and the Three Lakes Foundation, personal fees from Hoffman-La Roche Ltd., and grants from the University Hospital Foundation, the University of Calgary Cumming School of Medicine, and the Pulmonary Fibrosis Society of Calgary. JPF reports funding from the Heinz Endowments, Pittsburgh, PA (#E6820). DJK is supported in part by AR060780, HL133232, UL1 TR001857, AR076024 NIH grants, Boehringer-Ingelheim grants,

and receives collaborative research funding from Regeneron Pharmaceuticals, outside of the submitted work. SMN receives grant support from Boehringer Ingelheim USA. YZ is supported in part by an AR076024 NIH grant.

Body word count (no specified limit): 3453
Table count (no specified limit)**: 3 (+ 4 supplemental)**
Figure count (no specified limit)**: 3 (+ graphical abstract + 2 supplemental)**

Keywords: air pollution; particulate matter; interstitial lung disease; idiopathic pulmonary fibrosis; DNA methylation; epigenetics.

Highlights (3-5 bullet points, max 85 characters each):

- Increased $PM_{2.5}$ exposures associated with higher global DNA methylation %
- Increased SO_4, NO_3, NH_4, and black carbon associated with higher DNA methylation %
- Global DNA methylation mediates 2-5% of $PM_{2.5}$-mortality association in fibrotic ILD

4.1.1.1 Graphical Abstract

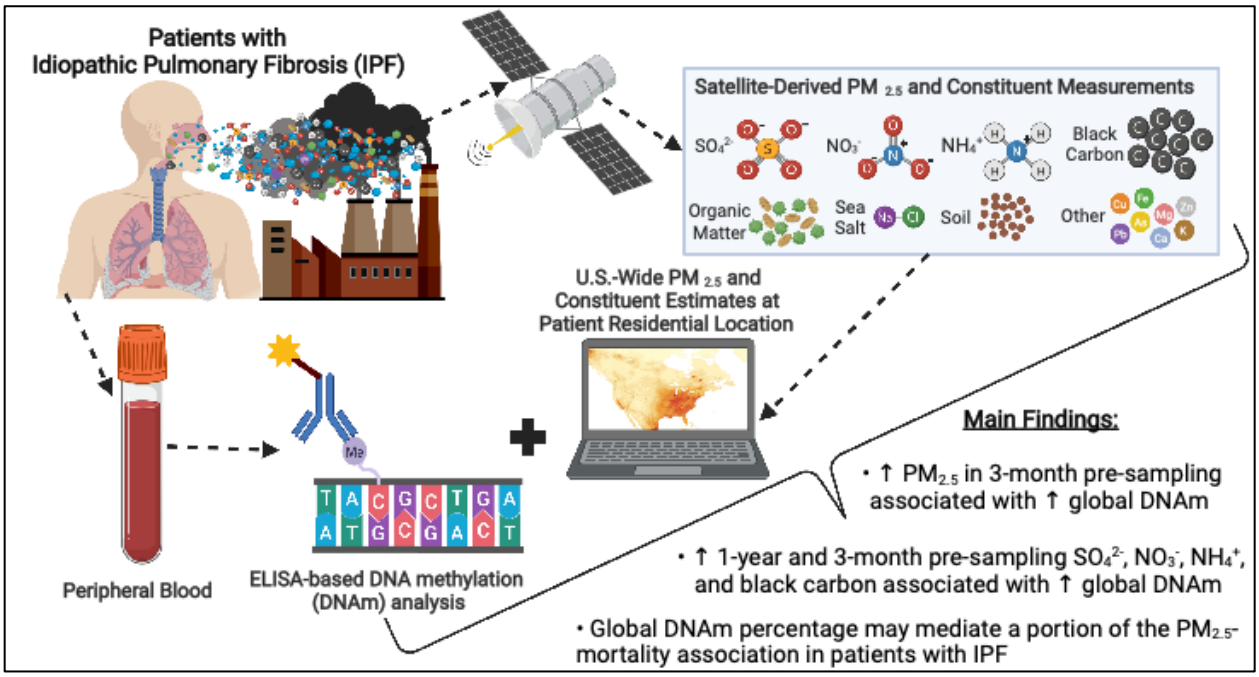


Figure 10 – Graphical absract for $PM_{2.5}$ association with global DNA methylation manuscript.

4.1.1.2 Abstract

Background: Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive interstitial lung disease (ILD) whose outcomes are worsened with air pollution exposures. DNA methylation (DNAm) patterns are altered in lungs and blood from patients with IPF, but the relationship between air pollution exposures and DNAm patterns in IPF remains unexplored.

Objective: To evaluate the association of $PM_{2.5}$ and constituent components with global DNAm in patients with IPF.

Methods: Patients with IPF enrolled in either the University of Pittsburgh Simmons Center for ILD Registry or the U.S.-wide Pulmonary Fibrosis Foundation (PFF) Patient Registry with peripheral blood DNA samples were included. Average of monthly exposures to $PM_{2.5}$ and constituents over 1-year and 3-months pre-blood collection were matched to patient residential coordinates using satellite-derived hybrid models. Global DNAm percentage (%5mC) was

determined using the ELISA-based MethylFlash assay. Associations of pollutants with %5mC were assessed using beta-regression, Cox models for mortality, and linear regression for baseline lung function. Mediation proportion was determined for models where pollutant-mortality and pollutant-%5mC associations were significant.

Results: This study included 313 Simmons and 746 PFF patients with IPF. Higher $PM_{2.5}$ 3-month exposures prior to blood collection were associated with higher %5mC in Simmons (β=0.02, 95%CI 0.0003-0.05), with trends in the same direction in the 1-year period in both cohorts. Higher exposures to sulfate, nitrate, ammonium, and black carbon constituents were associated with higher %5mC in multiple models. Percent 5mC was not associated with IPF mortality or lung function, but mediates between 2-5% of the association of $PM_{2.5}$, sulfate, or ammonium with mortality.

Conclusions: Higher global DNAm may be a novel biomarker for increased $PM_{2.5}$ and anthropogenic constituent exposure in patients with IPF. Mechanistic research is needed to determine if DNAm has pathogenic relevance in mediating associations between pollutants and mortality in IPF.

4.1.1.3 Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive lung disease of unknown etiology, with a median survival of three to five years from the time of diagnosis.(Raghu et al., 2011) Exposures to particulate matter with a diameter of 2.5um or less ($PM_{2.5}$) and other airborne pollutants contribute to the incidence,(Conti et al., 2018; Singh et al., 2019) progression,(Dales, Blanco-Vidal, & Cakmak, 2020; Johannson et al., 2014, 2018a; Winterbottom et al., 2018) and mortality(Goobie, Carlsten, Johannson, Marcoux, et al., 2022; Sesé et al., 2018) of IPF and other forms of interstitial lung disease (ILD). Despite this increasing body of epidemiologic evidence,

our mechanistic understanding of how air pollution contributes to IPF pathophysiology is lacking.(Goobie et al., 2020)

DNA methylation (DNAm) patterns in the blood and lungs are modified in response to air pollution exposure in healthy and chronic lung disease populations.(Jiang et al., 2014; Rider & Carlsten, 2019) DNAm patterns are also altered in the blood and lung tissue of patients with IPF compared with controls, with differentially methylated regions (DMRs) in lung tissue influencing downstream gene expression at IPF-relevant loci.(Guiot et al., 2017; Wielscher et al., 2015; Yang et al., 2014) Enzyme-linked immunosorbent assay (ELISA)-based methods enable the rapid and sensitive colorimetric measurement of global DNAm status (i.e. the percentage of cytosine residues that are methylated, %5mC) with small quantities of DNA.(Kurdyukov & Bullock, 2016; "MethylFlash Global DNA Methylation (5-mC) ELISA Easy Kit (Colorimetric) | EpiGentek," n.d.) Using this method, mice chronically exposed to high levels of $PM_{2.5}$ experienced global hypomethylation in blood and lung tissues.(Z. Li et al., 2019) In contrast, individuals living in high-exposure air pollution regions in India had higher %5mC by this method compared to individuals living in low-exposure regions.(Mishra et al., 2021) No previous studies have evaluated the relevance of global DNAm status in IPF and the potential association of this epigenetic biomarker with exposures to ambient $PM_{2.5}$ pollution.

Using a geographically diverse cohort of patients with IPF from one well-characterised single-center cohort from Western Pennsylvania and thirty-nine tertiary ILD referral sites from a United States (U.S.)-wide registry, we sought to describe global DNAm status in patients with IPF and to determine whether global DNAm serves as a useful biomarker of acute and chronic $PM_{2.5}$ and constituent component exposures in these patients. We hypothesized that increased exposures to $PM_{2.5}$ and primarily anthropogenically-derived constituent components (sulfate/SO_4^{2-},

nitrate/NO_3^-, and ammonium/NH_4^+) would be associated with alterations in %5mC. Exploratory analyses evaluated the association of %5mC with baseline lung function and mortality in patients with IPF and to determine whether %5mC mediates a portion of the association between $PM_{2.5}$ or constituent component exposures and mortality.

4.1.1.4 Methods

4.1.1.4.1 Study Population and Clinical Data

Adults with a diagnosis of IPF made by a specialist ILD clinician according to clinical practice guidelines at either the Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease (Simmons cohort) or one of the Pulmonary Fibrosis Foundation Patient Registry sites (PFF cohort) were eligible.(Raghu et al., 2018) Only patients with peripheral blood samples taken during registry enrollment were included, with the first DNA samples collected during enrollment used if the DNA yield was of sufficient quantity and quality. Simmons patients were enrolled in the Simmons Center ILD Registry between 2000-2021, while patients from one of the 39 PFF Patient Registry sites were enrolled between 2016-2021 (cohort and site breakdown in **Appendix Table 24**).(B. R. Wang et al., 2020) The University of Pittsburgh (UPitt), while a site in the PFF Patient Registry, was excluded from the PFF cohort to prevent overlap of patients between Simmons and PFF.

Demographic, residential, and clinical data was obtained from electronic health records for all patients. The UPitt Health Record Research Request (R3) Service was used to extract additional Simmons cohort data on race and initial encounter dates.(Visweswaran et al., 2022) Baseline lung function analyses included the first measurement of percent predicted values for forced vital

capacity (FVC) and diffusion capacity of the lung for carbon monoxide (D_LCO) in the 6 months before or after sample collection.

Ethics approval was obtained from UPitt (STUDY20030223, STUDY19040326, STUDY21030226), including for use of PFF data, and for all PFF sites individually.

4.1.1.4.2 $PM_{2.5}$ and Constituent Component Exposure Estimation

Full residential address was available for Simmons patients, enabling geocoding into precise latitude and longitude coordinates. 5-digit zip code was the most granular residential data available for PFF patients. Latitude and longitude coordinates were assigned to the centroid of each PFF patient's zip code using the "zipcodeR" package in R (version 4.0.2. www.r-project.org).(Rozzi, 2021) Coordinates from each cohort were then matched to the nearest coordinates for $PM_{2.5}$ and constituent component measurements from the Atmospheric Composition Analysis Group online data repository.(Van Donkelaar et al., 2019) These data provide average monthly estimates of total $PM_{2.5}$ mass and $PM_{2.5}$ constituent composition including sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic matter (OM), sea salt (SS), and soil from 2000-2018 for total $PM_{2.5}$ and 2000-2017 for constituents. Estimates are resolved to a geographic area of approximately 1.1km^2 using the "ncdf4" package in R.(Pierce, 2021)

Average of monthly exposures to $PM_{2.5}$ and constituents were calculated for multiple pre-blood sampling periods (**Appendix Table 26**). Subsequently, one long-term (average of all monthly values in the 1-year pre-sampling) and one short-term (average of all monthly values in the 3-months pre-sampling) period were selected for primary analyses.

4.1.1.4.3 DNA Isolation

Genomic DNA was isolated from peripheral whole blood samples taken from patients enrolled in the Simmons cohort using the Puregene DNA isolation kit or QiaAmp blood DNA isolation kit (Qiagen) and from the PFF cohort using the DNeasy Blood and Tissue Kit (Qiagen), all according to the manufacturer's instructions. PFF samples were transferred to UPitt for simultaneous analysis alongside the Simmons samples. Prior to analysis, all DNA samples were stored at -20°C (Simmons) and -80°C (PFF). DNA concentration and purity was assessed using the Infinite M200 Pro NanoQuant Plate Reader. Samples with an A260/A280 absorbance ratio outside of the 1.6-2.0 range were excluded.

4.1.1.4.4 Global DNA Methylation (DNAm) Assay

Global %5mC was determined using the ELISA-based MethylFlash™ Global DNAm Easy Kit (Epigentek Group Inc., Farmingdale, NY, USA). Each sample was evaluated in duplicate with 100ng of DNA, which is bound to the high-DNA affinity strip wells, followed by the addition of a 5mC antibody, which binds to methylated DNA. Colorimetric detection of 5mC antibody binding is determined by measuring absorbance at 450nm using the Infinite M200 Pro Plate Reader, and the mean measurement of sample duplicates was used to calculate %5mC, according to the manufacturer's instructions, which was expressed as percent of total DNA bases.("MethylFlash Global DNA Methylation (5-mC) ELISA Easy Kit (Colorimetric) | EpiGentek," n.d.)

4.1.1.4.5 Statistical Analysis

The Simmons and PFF cohorts were analyzed separately for all outcomes, serving as a derivation and validation cohort, respectively, for the evaluation of %5mC as a biomarker of $PM_{2.5}$ and constituent exposures. Beta regression models were constructed to evaluate the association of

pollutant exposures with %5mC levels. Adjusted models controlled for age at diagnosis, sex, smoking history, and site (PFF only). Age, sex, and smoking history are well-established contributors to DNAm patterns, thus requiring adjustment in multivariable models.(M. J. Jones, Goodman, & Kobor, 2015; Martin & Fry, 2018; Solomon et al., 2022)

Cox proportional hazards models were used to evaluate associations of %5mC or pollutants with survival, considering death or lung transplantation as a composite outcome. Linearity and proportionality assumptions were tested. "Partially-adjusted" models adjust for age at diagnosis, sex, and smoking history, while "fully-adjusted" models additionally adjust for race, baseline FVC, baseline D_LCO, and site (PFF only). Adjustments were made for age, sex, and baseline lung function as these factors are well established contributors to mortality in patients with IPF,(Ley, Collard, & King, 2011; Ryerson et al., 2014) for smoking history given impacts on DNAm and potentially IPF mortality,(Margaritopoulos et al., 2015; Martin & Fry, 2018) for race given impacts on pulmonary function and mortality,(Adegunsoye et al., 2018) and for site in the PFF cohort to control for site-specific differences in management practices.

Multivariable linear regression models were used to evaluate the association of %5mC with baseline FVC and D_LCO. "Partially-adjusted" models controlled for age at diagnosis, sex, and smoking history, while "fully-adjusted" models additionally controlled for race and site (PFF only).

Mediation analyses were performed to evaluate the proportion of associations between pollutants and mortality that were mediated by %5mC (visual depiction of mediation shown in **Appendix Figure 11**). Analyses were only performed for models where associations of pollutants with %5mC and associations of pollutants with mortality outcomes were significant. Mediation proportion was determined by the following formula:

$$Mediation\ proportion = 1 - (\frac{coefficient\ of\ model\ with\ pollutant + \%5mC}{coefficient\ of\ model\ with\ pollutant\ alone})$$

All analyses were performed using R (version 4.0.2., www.r-project.org).

4.1.1.5 Results

4.1.1.5.1 Cohort Characteristics and Pollutant Exposures

In both cohorts, the median (interquartile range (IQR)) time between registry enrollment and sample collection was 0.0 (0.0-0.0) years, indicating the most samples were collected on the date of enrollment. Patient characteristics for each cohort are shown in **Table 6**. Age at diagnosis was similar between the two cohorts, with both being predominately male, White, having a former or current history of smoking, and living in a metropolitan area. Baseline FVC was similar between the two cohorts, but D_LCO was lower in PFF. Follow-up durations were similar, and most patients in both cohorts had sample collection at the time of registry enrollment. More patients in Simmons experienced a terminal outcome of death or lung transplantation, primarily due to enrollment of patients since 2000 in the Simmons Registry, compared to since 2015 in PFF.

Table 6 – Patient demographics by cohort.

Patient Characteristics	Simmons Cohort	PFF Cohort
	N=313	N=746
Age at diagnosis, median (IQR), years	68 (61-74)	69 (64-74)
Male sex, n (%)	213 (68%)	559 (75%)
Self-reported race, n (%)		
White	289 (92%)	706 (94%)
Black	8 (3%)	7 (1%)
Asian	0 (0%)	20 (3%)
Indigenous[a]	1 (0.3%)	0 (0%)
Unknown	15 (5%)	13 (2%)
Smoking history, n (%)		
Never	86 (28%)	267 (36%)
Former	7 (2%)	"Ever"
Current	207 (66%)	479 (64%)
Unknown	13 (4%)	0 (0%)
Urbanicity		

Metropolitan (>50,000 people)	237 (76%)	627 (84%)
Micropolitan (10,000-50,000 people)	50 (16%)	63 (9%)
Rural (<10,000 people)	26 (8%)	55 (7%)
Baseline FVC % Predicted, median (IQR)[b]	65 (53-79)	67 (55-80)
Baseline D_LCO % Predicted, median (IQR)[c]	46 (35-60)	40 (30-49)
Follow-up Duration, median (IQR), years	2.7 (1.1-5.2)	2.4 (1.3-3.3)
Cause of censoring, n (%)		
Death	187 (60%)	213 (29%)
Lung Transplantation	82 (26%)	104 (14%)
Lost to follow-up or censored by data extraction	44 (14%)	429 (57%)
$PM_{2.5}$ Exposure (ug/m^3), median (IQR)		
1 year pre-sampling	11.2 (9.5-13.4)	7.6 (6.7-8.3)
3 months pre-sampling	10.4 (8.8-12.7)	7.4 (6.4-8.5)
Percent DNA methylation by MethylFlash, median (IQR)	0.15 (0.08-0.24)	0.10 (0.08-0.12)

[a] – Includes Native American, American Indian, Alaskan First Nations, & other Indigenous persons.
[b] – FVC was available for 97% of patients enrolled in Simmons, 91% of PFF.
[c] – D_LCO was available for 91% of patients enrolled in Simmons, 86% of PFF.
D_LCO, diffusion capacity of lung for carbon monoxide; FVC, forced vital capacity; IPF, idiopathic pulmonary fibrosis; IQR, interquartile range; $PM_{2.5}$, particulate matter with a diameter of 2.5μm or less; PFF, Pulmonary Fibrosis Foundation.

$PM_{2.5}$ and constituent exposures were generally higher in Simmons compared to PFF over all periods (**Table 6, Appendix Table 26**). One-year pre-sampling median exposures to $PM_{2.5}$ total mass, broken down by constituent composition, are shown in **Figure 11**. The only PFF site with higher total $PM_{2.5}$ exposures than Simmons was the University of California Los Angeles (UCLA), with Simmons median 1 year pre-sampling exposures of $12.0ug/m^3$ (interquartile range/IQR 10.1-$13.9ug/m^3$) and UCLA exposures of $12.2ug/m^3$ (IQR 8.7-$14.3ug/m^3$). Exposure to primarily anthropogenic $PM_{2.5}$ constituents, notably SO_4^{2-} and NH_4^+, was highest in Simmons, whereas NO_3^- was highest in UCLA, reflecting differences in pollutant emission sources between sites. Sea salt (SS) composition was greatest in coastal sites of PFF, including Stanford University and the University of California San Francisco/UCSF (both in San Francisco Bay Area) and the University of Miami. Organic matter (OM) made up the largest fraction of $PM_{2.5}$ at most sites, reflecting particulates largely derived from biomass burning.(Philip et al., 2014; Y. Wang et al., 2022)

Global DNAm fraction (%5mC) was highest in the Simmons cohort (median 0.15%, IQR 0.08-0.24%), followed by the University of Pennsylvania (UPenn) site of PFF (median 0.13%,

0.10-0.14%). **Figure 12** depicts the distributions of %5mC by cohort and site, listing PFF sites in order from highest to lowest median %5mC. **Appendix Figure 12** presents a scatterplot matrix with correlations between %5mC, $PM_{2.5}$, and its constituents.

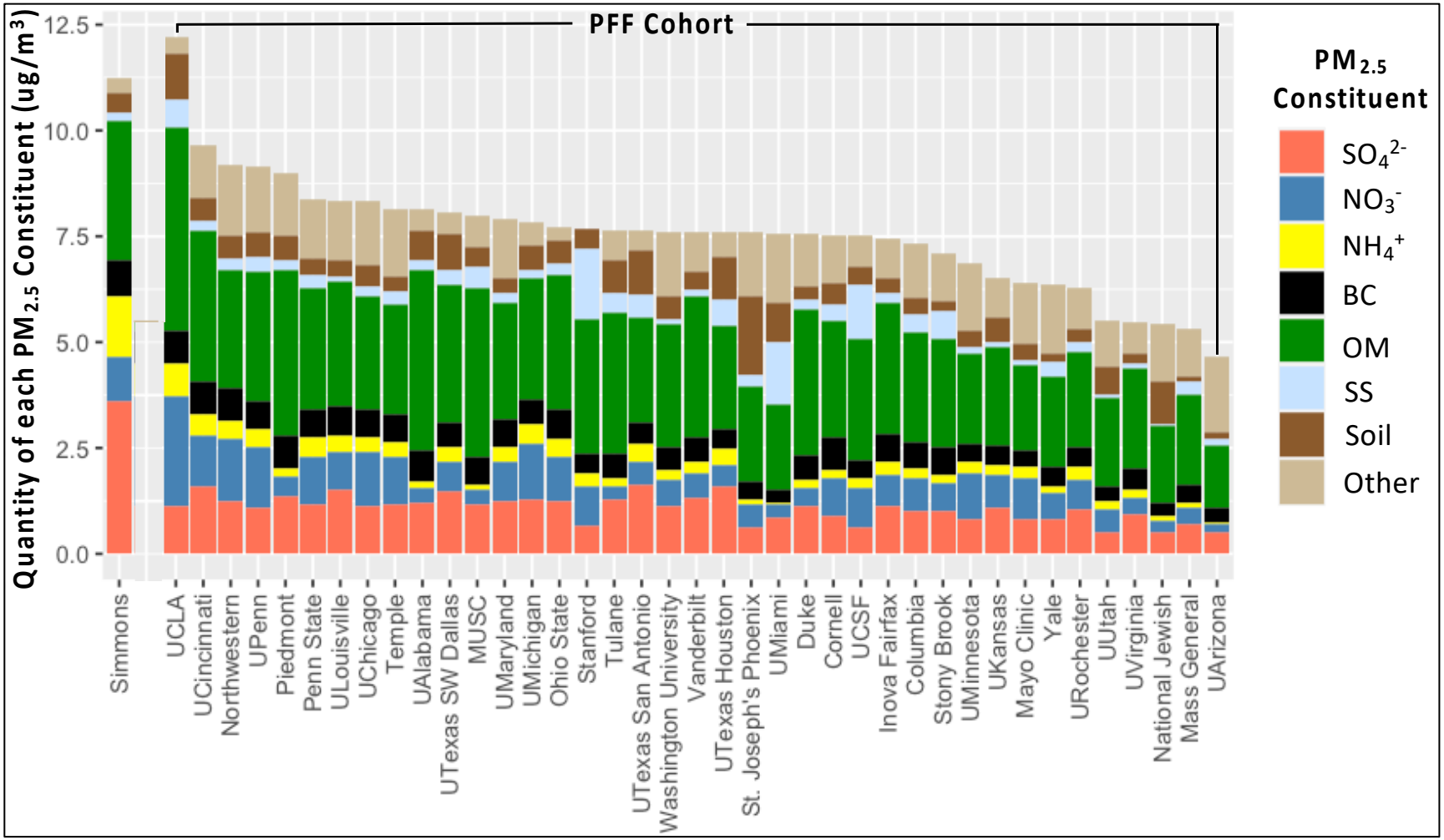


Figure 11 – $PM_{2.5}$ total mass and constituent distribution by cohort and site. Distribution of total $PM_{2.5}$ mass, broken down into median exposures to each constituent component across all individuals at that cohort or PFF site* in the 1-year pre-sampling period. PFF sites are ordered from left to right by highest to lowest median $PM_{2.5}$ total mass exposures in 1-year pre-sampling period. Abbreviations: BC, black carbon; NH_4^+, ammonium; NO_3^-, nitrate; OM, organic matter; PFF, Pulmonary Fibrosis Foundation; $PM_{2.5}$, particulate matter with a diameter ≤2.5um; SO_4^{2-}, sulfate; SS, sea salt.

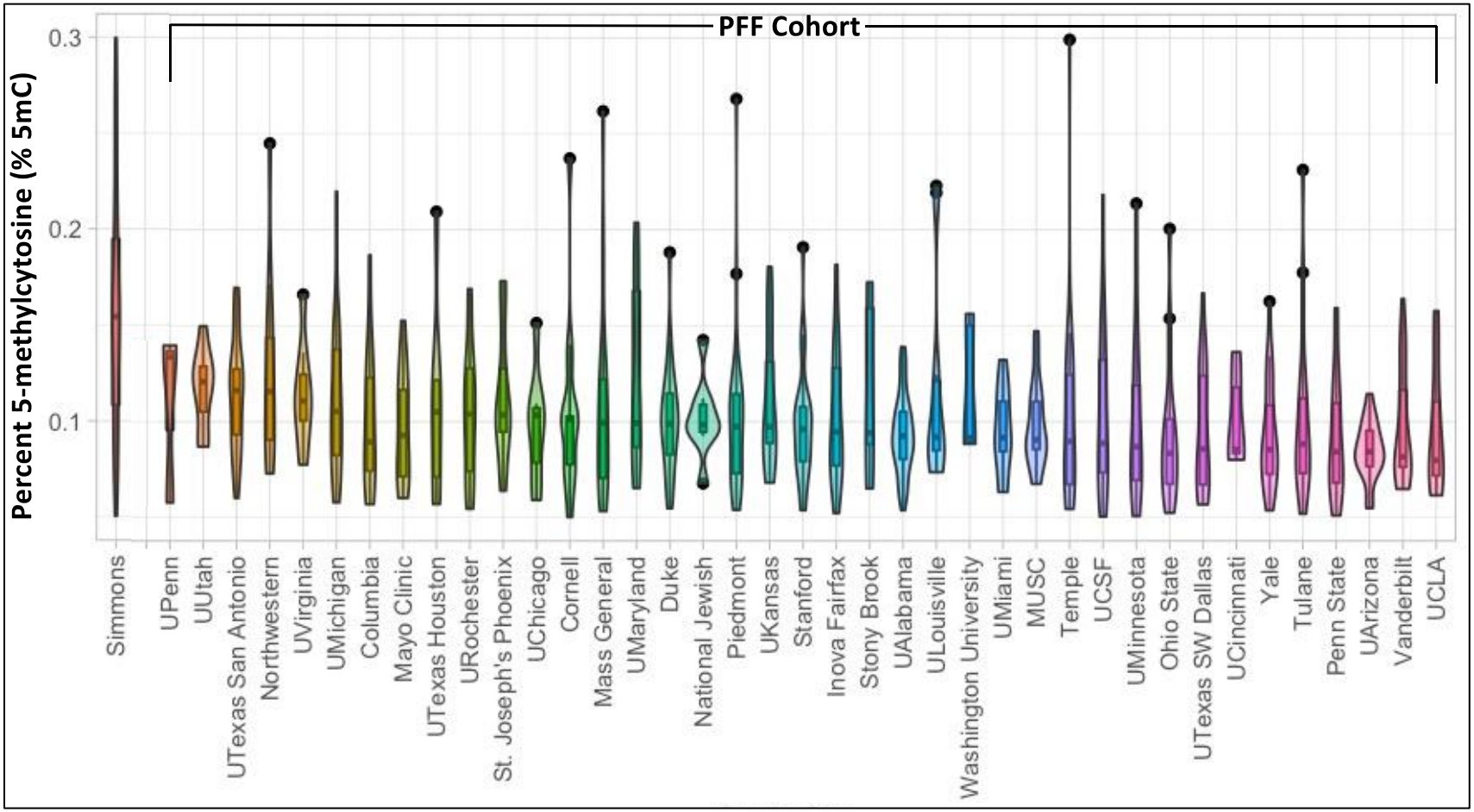


Figure 12 – Percent DNA methylation at 5-methylcytosine residues (%-5mC) by study cohort and PFF site. Percent 5mC is determined by ELISA-based MethylFlash assay. PFF sites are organized from left to right by highest to lowest median %5mC. The dots present in some violin plots reflect outliers within each site. Fill color is different for each site.

4.1.1.5.2 Pollutant Associations with Global DNAm Percentage (%5mC)

Associations of $PM_{2.5}$ or constituents with %5mC in unadjusted and adjusted models are reported in **Appendix Table 27** with adjusted model results shown in **Figure 13**. In fully-adjusted models in the Simmons cohort, higher $PM_{2.5}$ in the 3-months pre-sampling (**Panel 13B**) was associated with higher %5mC (β=0.02, 95%CI 0.0003 to 0.05, p=0.047), with a consistent direction and magnitude of effect in the 1-year pre-sampling period in both cohorts (**Panel 13A**) that did not reach statistical significance thresholds (Simmons β=0.03, 95%CI -0.004 to 0.06, p=0.09; PFF β=0.03, 95%CI -0.007 to 0.06, p=0.12). Associations in unadjusted models for both periods in the Simmons cohort were statistically significant and positively correlated.

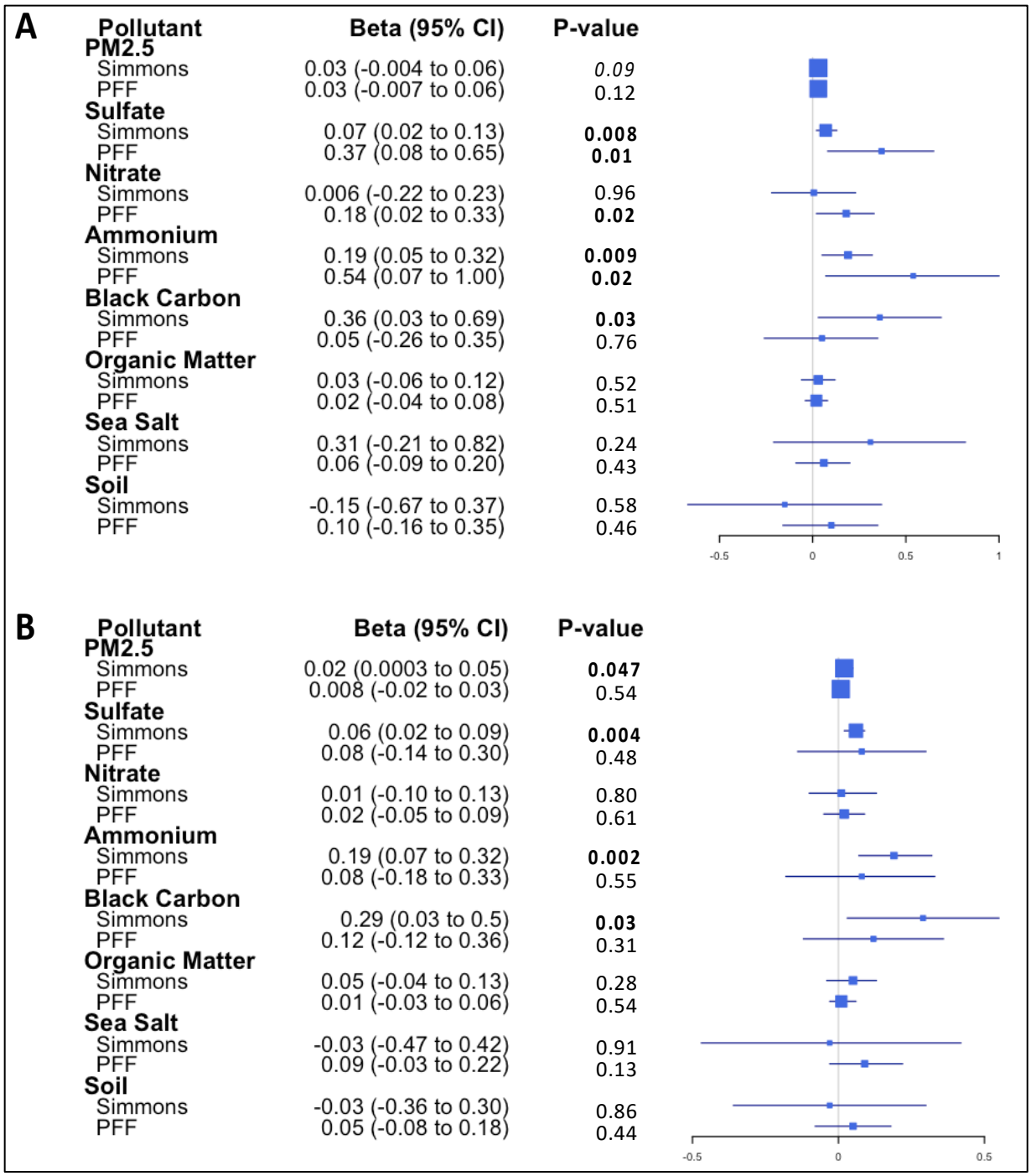


Figure 13 – Association between $PM_{2.5}$ or constituent component and percent global DNAm. Results shown are from beta regression models adjusting for age at diagnosis, sex, smoking history, and in PFF cohort only, study site. A) 1-year pre-sampling models and B) 3-months pre-sampling models. Unadjusted model results are reported in Table S2. Box size reflects the width of confidence intervals.

Constituent-specific analyses demonstrated that higher SO_4^{2-} and NH_4^+ exposures were associated with higher %5mC in all models in the Simmons cohort and in the adjusted 1-year pre-

sampling model in the PFF cohort. Higher 1-year pre-sampling NO_3^- was associated with higher %5mC in the PFF cohort alone. All models in the Simmons cohort demonstrated positive associations between BC and %5mC, but this was not replicated in PFF. OM, SS, and soil exposures were not associated with differences in %5mC in any models. In 3-month models, higher SO_4^{2-}, NH_4^+, and BC exposures in the Simmons cohort alone were associated with higher %5mC.

4.1.1.5.3 Global DNAm Fraction (%5mC) Associations with Clinical Outcomes

Global %5mC was not associated with all-cause mortality, baseline FVC, or baseline D_LCO in either cohort (**Table 7**). Unadjusted and partially-adjusted mortality analyses in the Simmons cohort had a direction of effect that suggested higher %5mC may be associated with increased mortality, but this effect size decreased substantially when baseline FVC and D_LCO were added to the fully-adjusted model. Consistently, the direction of effect on baseline FVC is negative and nears statistical significance thresholds in the Simmons cohort, suggesting that patients with higher %5mC may have lower baseline FVC (i.e. more impaired pulmonary function at presentation).

Table 7 – DNA methylation (DNAm) association with clinical outcomes. Partially-adjusted mortality models control for age at diagnosis, sex, and smoking history, and fully-adjusted models additionally control for race and site (in PFF only). Partially-adjusted models for baseline forced vital capacity (FVC) and diffusion capacity of the lung for carbon monoxide (D_LCO) models control for age at diagnosis and sex, while fully-adjusted models additionally control for race, smoking history, and site (in PFF only).

Model	**Simmons Cohort**			**PFF Cohort**		
Mortality Analyses	**HR (95% CI)**	**p-value**	**n**	**HR (95% CI)**	**p-value**	**n**
Unadjusted	1.43 (0.86-2.38)	0.17	313	0.86 (0.38-1.95)	0.72	746
Partially-Adjusted	1.39 (0.81-2.39)	0.23	313	0.84 (0.39-1.79)	0.64	746
Fully-Adjusted	1.18 (0.59-2.36)	0.65	286	1.29 (0.64-2.59)	0.47	637
Baseline FVC Analyses	**β (95% CI)**	**p-value**	**n**	**β (95% CI)**	**p-value**	**n**
Unadjusted	-5.49 (-15.47 to 4.48)	0.28	305	-1.94 (-7.72 to 3.84)	0.51	678
Partially-Adjusted	-6.22 (-16.08 to 3.63)	0.21	305	-1.35 (-7.06 to 4.36)	0.64	678
Fully-Adjusted	-6.92 (-16.50 to 2.65)	0.16	305	-1.38 (-7.18 to 4.41)	0.64	678
Baseline D_LCO Analyses	**β (95% CI)**	**p-value**	**n**	**β (95% CI)**	**p-value**	**n**
Unadjusted	-3.30 (-15.52 to 8.92)	0.60	286	1.30 (-4.05 to 6.65)	0.63	638

Partially-Adjusted	-5.17 (-17.11 to 6.77)	0.40	286	1.30 (-4.05 to 6.66)	0.63	638
Fully-Adjusted	-4.71 (-16.69 to 7.27)	0.44	286	1.90 (-3.42 to 7.22)	0.48	638

4.1.1.5.4 Pollutant Associations with Mortality and Mediation by %5mC

Associations of 1-year and 3-month pre-sampling exposures to $PM_{2.5}$ and constituents with mortality are shown in **Appendix Table 28**. Higher $PM_{2.5}$ over both exposure periods was associated with increased mortality in unadjusted and partially adjusted models in the Simmons cohort and fully-adjusted models in the PFF cohort (Simmons partially-adjusted 1-year exposure hazard ratio/HR=1.08, 95%CI=1.03-1.13, p=0.003; PFF fully-adjusted 1-year exposure HR=1.16, 95%CI=1.003-1.13, p=0.045). Increased SO_4^{2-} and NH_4^+ exposures were associated with increased mortality in both 1-year and 3-month periods in Simmons unadjusted and partially-adjusted models and in fully-adjusted models in PFF. Increased NO_3^- was associated with increased mortality in unadjusted and partially-adjusted 1-year Simmons models and fully-adjusted 1-year and 3-month PFF models. Increased BC and OM were also associated with increased mortality in the fully-adjusted 1-year and 3-month models in PFF alone.

Table 8 reports the mediation proportion of %5mC for the association between pollutants and mortality. Only one analysis from PFF was assessed for mediation proportion (fully-adjusted SO_4^{2-} 1-year pre-sampling models), whereas unadjusted and partially-adjusted 1-year and 3-month Simmons models for $PM_{2.5}$, SO_4^{2-}, and NH_4^+ were assessed. Percent 5mC appears to mediate 2-3% of the association between $PM_{2.5}$ and mortality, 2-4% of the association between SO_4^{2-} and mortality, and 3-5% of the association between NH_4^+ and mortality.

Table 8 – Proportion of mediation by 5-methylcytosine (5mC) of the association between $PM_{2.5}$ or constituent components and mortality. Partially-adjusted models control for age at diagnosis, sex, and smoking history. Fully-adjusted models additionally control for race, baseline forced vital capacity (FVC), baseline diffusion capacity of the lung for carbon monoxide (D_LCO), and site (in PFF only). Mediation proportion is presented only for analyses where the association of $PM_{2.5}$ or constituent component with percent 5mC is significant and where the association of $PM_{2.5}$ or constituent component and mortality is significant. Significant associations are bolded.

Cohort and Exposure Period	Coefficient	HR	95%CI	p	n	Mediation Proportion
Simmons $PM_{2.5}$ 1yr Pre-Sampling						
Unadjusted	0.077	**1.080**	**1.03-1.13**	**0.002**	313	**3%**
Unadjusted + 5mC	0.074	**1.077**	**1.03-1.13**	**0.002**	313	
Partially-adjusted	0.074	**1.077**	**1.03-1.13**	**0.003**	313	**2%**
Partially-adjusted + 5mC	0.073	**1.075**	**1.02-1.13**	**0.004**	313	
Simmons $PM_{2.5}$ 3mo Pre-Sampling						
Unadjusted	0.057	**1.059**	**1.02-1.10**	**0.002**	307	**3%**
Unadjusted + 5mC	0.056	**1.057**	**1.02-1.10**	**0.002**	307	
Partially-adjusted	0.053	**1.055**	**1.02-1.09**	**0.004**	307	**2%**
Partially-adjusted + 5mC	0.052	**1.054**	**1.02-1.09**	**0.004**	307	
Simmons SO_4^{2-} 1yr Pre-Sampling						
Unadjusted	0.151	**1.163**	**1.07-1.27**	**<0.001**	310	**3%**
Unadjusted + 5mC	0.146	**1.158**	**1.06-1.26**	**0.001**	310	
Partially-adjusted	0.150	**1.162**	**1.07-1.27**	**<0.001**	310	**3%**
Partially-adjusted + 5mC	0.147	**1.158**	**1.06-1.26**	**0.001**	310	
PFF SO_4^{2-} 1yr Pre-Sampling						
Fully-adjusted	1.341	**3.824**	**1.22-12.02**	**0.02**	633	**2%**
Fully-adjusted + 5mC	1.318	**3.737**	**1.18-11.83**	**0.02**	633	
Simmons SO_4^{2-} 3mo Pre-Sampling						
Unadjusted	0.090	**1.094**	**1.03-1.16**	**0.003**	297	**4%**
Unadjusted + 5mC	0.086	**1.090**	**1.03-1.16**	**0.005**	297	
Partially-adjusted	0.079	**1.083**	**1.02-1.15**	**0.01**	297	**4%**
Partially-adjusted + 5mC	0.076	**1.079**	**1.02-1.15**	**0.01**	297	
Simmons NH_4^+ 1yr Pre-Sampling						
Unadjusted	0.368	**1.445**	**1.15-1.81**	**0.001**	310	**3%**
Unadjusted + 5mC	0.356	**1.427**	**1.14-1.79**	**0.002**	310	
Partially-adjusted	0.363	**1.438**	**1.15-1.80**	**0.002**	310	**3%**
Partially-adjusted + 5mC	0.353	**1.423**	**1.13-1.79**	**0.002**	310	
Simmons NH_4^+ 3mo Pre-Sampling						
Unadjusted	0.313	**1.368**	**1.12-1.68**	**0.003**	297	**5%**
Unadjusted + 5mC	0.297	**1.350**	**1.10-1.65**	**0.005**	297	
Partially-adjusted	0.282	**1.326**	**1.08-1.63**	**0.007**	297	**5%**
Partially-adjusted + 5mC	0.268	**1.307**	**1.06-1.61**	**0.01**	297	

4.1.1.6 Discussion

In this novel study, we found that %5mC levels were generally low in patients with IPF and that increased exposure to $PM_{2.5}$ in the three months prior to peripheral blood sample acquisition was associated with higher %5mC levels. We also found that this positive association is most pronounced for primarily anthropogenic constituents, including SO_4^{2-}, NO_3^-, NH_4^+, and

BC, suggesting that these largely human-derived sources of pollution may be driving epigenetic dysregulation in patients with IPF. Furthermore, we found that %5mC may mediate a small but potentially-significant portion of the association of $PM_{2.5}$, SO_4^{2-}, and NH_4^+ with increased mortality, indicating the need for further mechanistic studies to explore the role of DNAm in IPF pathophysiology.

We found that the %5mC measured in peripheral blood samples taken from patients with IPF was generally low in comparison to previous general population studies using the same ELISA-based global DNAm analysis methods. For example, the median %5mC was 0.15% in Simmons, and 0.10% in PFF, as compared to 0.32% in a general population study where patients had a median age of 55, indicating a state of relative hypomethylation in our population.(Tellez-Plaza et al., 2014) Multiple factors may contribute to this, including the older age of our patients,(M. J. Jones et al., 2015) frailty,(Bellizzi et al., 2012) or potentially higher lifetime burdens of occupational or environmental exposures that were not captured in this study.

To our knowledge, this is the first study to evaluate the association of $PM_{2.5}$ composition with %5mC in a vulnerable disease population. Potential pathophysiologic relevance for these primarily anthropogenic $PM_{2.5}$ constituents is suggested by the association of %5mC with exposures to SO_4^{2-} and NH_4^+, and to a lesser extent NO_3^- and BC. Similarly, exposures to these constituents were associated with the greatest adverse impact on survival in this population.(Goobie, Carlsten, Johannson, Marcoux, et al., 2022) This is of particular importance given that $PM_{2.5}$ composition varied substantially depending on cohort and site. Exposures to SO_4^{2-} and NH_4^+ constituents, which are primarily derived from industrial activities and agriculture,(Plautz, 2018; H. Zhang et al., 2014) were highest in the Simmons cohort, whereas exposures to NO_3^-, which is primarily derived from fossil fuel combustion,(H. Zhang et al., 2014)

were highest in the UCLA site. Our findings suggest that patients in these regions may be at the highest risk of experiencing $PM_{2.5}$-associated epigenetic alterations and adverse clinical outcomes. Public health interventions aimed at reducing $PM_{2.5}$ pollution from anthropogenic sources that predominately emit these particulate components may provide the greatest benefits for this vulnerable population group.

What remains unclear is whether total $PM_{2.5}$ mass and the specific constituents SO_4^{2-}, NH_4^+, NO_3^-, and BC directly impact DNAm regulation. $PM_{2.5}$ can induce the formation of reactive oxygen species (ROS) and trigger inflammatory cytokine cascades, which may lead to dysregulation of DNAm maintenance mechanisms.(Rider & Carlsten, 2019) While no studies have investigated whether ammonium sulfate and ammonium nitrate particles exert adverse toxicity in vitro,(Park et al., 2018) these particles do interact with other chemicals in the atmosphere to produce compounds like sulfuric and nitric acid.(Weber, Guo, Russell, & Nenes, 2016) Such acidic compounds induce airway and alveolar epithelial damage,(Treon, Dutra, Cappel, Sigmon, & Younker, 1950) which may trigger inflammatory pathways and ROS that contribute to DNAm alterations. Furthermore, these particles can increase the solubility of toxic metals that exist in particulate matter, which may enable their cellular uptake and facilitate toxicity.(Park et al., 2018)

Many health effect studies consider the effects of $PM_{2.5}$ as a whole, however $PM_{2.5}$ needs to be considered as a complex mixture of chemical species whose relative contributions total particulate exposure varies markedly over time and space. Combustion processes produce particulate soot (or BC) directly into the air. In addition, gaseous pollutants like SO_2 and NO_x can transform into secondary particulates through complex chemical reactions involving multiple species including formation of sulfuric and nitric acid, which ultimately contribute to $PM_{2.5}$ as SO_4^{2-} and NO_3^- particles.(Foltescu et al., 1996; Hewitt, 2002) Thus, it is often unclear as to which

types of $PM_{2.5}$ contribute the most to adverse health events. We cannot definitively point to which specific chemical forms of $PM_{2.5}$ (e.g. H_2SO_4, HNO_3, salts of ammonium sulfate/nitrate) possess the most biological activity for initiating epigenetic changes as significant autocorrelation exists between pollutants. However, the greater association of DNA methylation with secondary $PM_{2.5}$ components compared to primary forms of $PM_{2.5}$ should stimulate additional work on understanding the role of atmospheric chemical transformations in $PM_{2.5}$ biological effects. Some toxicological and epidemiological studies have suggested sulfate and nitrate containing aerosols possess less adverse potential than total $PM_{2.5}$ mass,(Reiss et al., 2007) however, the role of acidic aerosols have been highlighted by others as important offending agents.(Gwynn et al., 2000; Spengler et al., 1996; Thurston, Ito, Lippmann, & Hayes, 1989)

While several studies have shown that higher exposures to particulate pollution *in utero* and during childhood are associated with global DNA hypomethylation, specifically at long interspersed nuclear elements-1 (LINE-1),(Breton et al., 2016; Cai et al., 2017) few studies have evaluated the impact of pollution exposures on global DNAm status in elderly chronic disease populations. The direction and magnitude of effect of pollution on global %5mC may vary depending on age and disease state of the population studied, although most studies have suggested that higher pollution exposures are associated with global DNA hypomethylation,(Plusquin et al., 2017; Rider & Carlsten, 2019) which contrasts with our findings. Furthermore, pollution exposure can cause hypomethylation at some CpG loci, while simultaneously inducing hypermethylation at others, with further variations across cell types.(Honkova et al., 2022; Rider & Carlsten, 2019) Our study was unable to include control subjects as non-diseased individuals are not captured in the Simmons or PFF registries. Future investigations into the impact of $PM_{2.5}$ and constituent exposures on matched healthy populations compared with IPF will be useful to determine if these

exposures differentially impact individuals with and without ILD. More detailed analyses of DNAm patterns using locus-specific bisulfite sequencing techniques will enable a more nuanced exploration of differential methylation patterns at IPF-relevant loci.

This study has several limitations. The measurement of global DNAm using an ELISA-based method is a crude approach for evaluating this highly variable epigenetic factor, which has highly specific variations at 5'-cytosine-phosphate-guanine-3' (CpG) sites across the genome and different cell types. Nonetheless, global DNAm status can serve as a useful biomarker of aging, smoking, and other environmental exposures, highlighting its potential utility as a biomarker of ambient pollution exposures.(M. J. Jones et al., 2015) PFF $PM_{2.5}$ exposure estimations were determined from 5-digit zip codes, which has the potential to result in a greater degree of exposure misclassification as compared to our Simmons cohort where full residential addresses were available. Additionally, samples from the PFF cohort are more recent with patients experiencing generally lower $PM_{2.5}$ exposures, which may impede our ability to detect significant impacts of $PM_{2.5}$ on DNAm if these effects are most predominate at higher exposure levels. Sample size was somewhat small in the Simmons cohort, resulting in limitations in our ability to detect significant effects in the %5mC and clinical outcomes analyses. Future studies with increased patient numbers and highly precise exposure estimates (as was feasible in the Simmons cohort) may demonstrate significant associations between higher %5mC and mortality or baseline FVC. Lastly, this study only considered $PM_{2.5}$ and its associated constituent components for impacts on %5mC. Future studies should consider evaluating the impact of gaseous pollutants and particulate matter of other size proportions (e.g. PM_{10} and ultrafine particulate matter) on global DNAm status.

4.1.1.7 Conclusions

Using a precise satellite-derived method of air pollution exposure estimation, this represents the first study to evaluate the association of $PM_{2.5}$ and constituent component exposure with global DNAm status in peripheral blood samples from patients with IPF. These findings indicate that %5mC may serve as a biomarker of short- and long-term exposures to ambient pollution in this population, where $PM_{2.5}$ and its anthropogenically-derived constituents have been shown to adversely impact disease severity and mortality.(Goobie, Carlsten, Johannson, Marcoux, et al., 2022; Sesé et al., 2018) Future research is needed to determine if patients who have molecular evidence of high $PM_{2.5}$ exposures derive greater benefit from avoidance of ongoing high pollution exposures or from treatment with anti-fibrotic or future disease-modifying therapies.

4.1.1.8 Acknowledgements

We would like to thank the fILD patients from the Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease for their time and commitment to clinical research. We thank all patients who participated in the Pulmonary Fibrosis Foundation (PFF) Patient Registry. We also thank the investigators and other staff at participating PFF Care Centers for providing clinical data, and the Pulmonary Fibrosis Foundation, which established and has maintained the PFF Patient Registry since 2016, and lastly, the many generous donors of the PFF Patient Registry. We would also like to thank Diane Lavsa for her work quantifying the DNA samples for this project. BioRender was used in the development of the graphical abstract for this manuscript.

4.1.1.9 Data Sharing

The code for all analyses performed in this study is available via: https://github.com/gcgoobie/PM2.5_MethylFlash and for the exposure matching pipeline via:

https://github.com/gcgoobie/PM2.5_ClinicalOutcomes/tree/main/PM2.5andConstituent_Matching

4.2 Aim 3.2 – $PM_{2.5}$ Impacts on Epigenome-Wide DNA Methylation in fILDs

4.2.1 Introduction

Imperative to our understanding of the effects of air pollution in patients with fILD is an assessment of the molecular mechanisms that contribute to these adverse clinical outcomes. One mechanism whereby air pollution may contribute to disease development and deterioration is through changes to an individual's epigenome, particularly in the form of altered DNA methylation (DNAm). Air pollution is associated with alterations in DNAm in healthy individuals and in patients with asthma, COPD, and lung cancer.(Alfano et al., 2018; de F.C. Lichtenfels et al., 2018; Peng et al., 2019; Rider & Carlsten, 2019) DNAm changes in lung tissue from patients with IPF are associated with altered expression of IPF-related genes, including *NOTCH1* and *TOLLIP*.(Yang et al., 2014). Alterations in DNAm patterns have been noted in multiple studies comparing IPF patients to controls,(Rabinovich et al., 2012; Sanders et al., 2012; Yang et al., 2014) however, no studies have evaluated the impact of air pollution on epigenome-wide DNAm in patients with fILD.

This study (Aim 3.2 of my PhD thesis) aimed evaluate how air pollution modifies global and locus-specific DNAm patterns in patients with fILD. We hypothesized that patients would demonstrate alterations in methylation at cytosine-phosphate-guanine (CpG) dinucleotide sites that are involved in fibro-inflammatory pathways of relevance to the development and progression of fILDs. We further anticipate that these epigenetic changes may mediate a portion of the adverse

impact of high $PM_{2.5}$ and constituent exposures on mortality and lung function in patients with fILD (as demonstrated in Aim 2).

4.2.2 Methods

4.2.2.1 Study Population and Clinical Data

Adults with a fILD diagnosis made by a specialist ILD clinician according to clinical practice guidelines at either the Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease (Simmons cohort) or the University of British Columbia site of the Canadian Registry for Pulmonary Fibrosis (CARE-PF cohort) were eligible.(Ryerson et al., 2016) Only patients with peripheral blood samples taken during registry enrollment were included, with the first DNA samples collected during enrollment used if the DNA yield was of sufficient quantity and quality. Simmons patients were enrolled in the Simmons Center ILD Registry between 2000-2021, while patients from CARE-PF were enrolled between 2015-2021.

Demographic, residential, and clinical data was obtained from electronic health records for all patients. The UPitt Health Record Research Request (R3) Service was used to extract additional Simmons cohort data on race and initial encounter dates.(Visweswaran et al., 2022) Baseline lung function analyses included the first measurement of percent predicted values for forced vital capacity (FVC) and diffusion capacity of the lung for carbon monoxide (D_LCO) in the 6 months before or after sample collection.

Residential location was used to determine neighborhood-level disadvantage, calculating the area deprivation index (ADI) for Simmons and the Canadian Index of Multiple Deprivation (CIMD) for CARE-PF.(Goobie, Ryerson, et al., 2022; Kind & Buckingham, 2018; *The Canadian Index of Multiple Deprivation: User Guide*, 2019)

Ethics approval was obtained from the University of Pittsburgh (STUDY20050209, STUDY21030226) and the University of British Columbia (#H19-01989 and #H20-01454).

4.2.2.2 $PM_{2.5}$ and Constituent Component Exposure Estimation

Full residential address was available for all patients, enabling geocoding into precise latitude and longitude coordinates, which were then matched to the nearest coordinates for $PM_{2.5}$ and constituent component measurements from the Atmospheric Composition Analysis Group online data repository.(Van Donkelaar et al., 2019) These data provide average monthly estimates of total $PM_{2.5}$ mass from 2000-2018 and $PM_{2.5}$ constituent composition including sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic matter (OM), sea salt (SS), and soil from 2000-2017. Estimates are resolved to a geographic area of approximately 1.1km^2 using the "ncdf4" package in R.(Pierce, 2021) Average of monthly exposures to $PM_{2.5}$ and constituents were calculated for the average of all monthly values in the 5-years pre-sampling and the average of all monthly values in the 3-months pre-sampling.

4.2.2.3 DNA Isolation

Genomic DNA was isolated from peripheral whole blood samples taken from patients enrolled in the Simmons cohort using the Puregene DNA isolation kit or QiaAmp blood DNA isolation kit (Qiagen) and from the CARE-PF cohort using the DNeasy Blood and Tissue Kit (Qiagen), all according to the manufacturer's instructions. CARE-PF samples were transferred to UPitt for simultaneous analysis alongside the Simmons samples. Prior to analysis, all DNA samples were stored at -20°C. DNA concentration and purity was assessed using the Infinite M200 Pro NanoQuant Plate Reader. Samples with an A260/A280 absorbance ratio outside of the 1.7-2.0 range were excluded.

4.2.2.4 DNA Methylation Measurement & Data Processing

DNA fragments were bisulfite converted using the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA, USA), which enables distinction between methylated versus unmethylated cytosines. Genome-wide DNA methylation was then analyzed using the Infinium MethylationEPIC BeadChip Array, which evaluates DNAm levels at >850,000 CpG sites across the genome.("Infinium MethylationEPIC Data Sheet | Illumina," 2019)

A multi-package approach to processing and analyzing raw DNAm data produced from the Infinium MethylationEPIC array was implemented.(Maksimovic, Phipson, & Oshlack, 2022) First, raw .idat files for each sample underwent pre-processing, quality control, filtering, and analysis using R statistical computing environment (version 4.2.1, www.r-project.org). Using the *minfi* package,(K. D. Hansen et al., 2022) samples with a mean detection p-value >0.05 were excluded (n=2). The detection p-value compares the total unmethylated and methylated signal at each CpG location and compares it to the background signal inferred from negative control probes, such that p-values >0.05 or >0.01 indicate a failed probe where the signal from the probe is not significantly different from the background.(Wilhelm-Benartzi et al., 2013)

All samples are from blood, and thus it was not expected there would be major global differences in methylation patterns between the samples. This enabled normalization using a subset quantile procedure, which helps to mitigate the impacts of artifacts in the data that are the result of differences between Type-I and Type-II methylation probes and other technical sources of variation.(Fortin et al., 2014; Touleimat & Tost, 2012)

Next, for probe filtering, we removed any CpG probes whose detection p-value was >0.01 on one or more samples. We also removed sex chromosome probes given that our study includes patients of male and female sex. Next, we removed probes that had a single nucleotide

polymorphism (SNP) at the CpG site, as this can result in impaired hybridization and detection of methylation for individuals with a variant allele at that locus.(Wilhelm-Benartzi et al., 2013) Subsequently, we removed cross-reactive probes that do not uniquely hybridize to a single CpG site on the MethylationEPIC array.(McCartney et al., 2016)

Following normalization and filtering steps, we calculated the M-values and Beta-values (β-values) at each CpG probe. β-values are calculated with the following equation in the *minfi* package:

$$\beta = \frac{M}{(M + U + \alpha)}$$

where M is the methylated signal intensity at a CpG probe, U is the unmethylated signal intensity, and α is a constant, usually set to 100, that is used as an offset to avoid dividing M by small values. The β-value essentially reflects the proportion of methylation at a single CpG locus. The M-value is related to the β-value through log transformation, and is more versatile for use in statistical analyses. It is calculated through the following equation in the *minfi* package:

$$Mvalue = log_2 \frac{M}{U}$$

Following calculation of β- and M-values, we calculated the average β-value for each individual. This was simply the arithmetic mean β-value across all retained CpGs for each sample.

4.2.2.5 Statistical Analysis

Multivariable beta regression was used to evaluate the association of pollutants with average β-value across all retained CpG probes (after normalization and filtering steps completed). Results of adjusted models are reported, where analyses include covariates of age at diagnosis,

sex, race, smoking history, and neighborhood-level disadvantage score. Both cohorts were evaluated separately.

M-values were used to detect differentially methylated probes (DMPs) between cohorts and between low vs high exposures to $PM_{2.5}$ and constituents using the *limma* package.(Smyth et al., 2019) For combined cohort analyses, a $PM_{2.5}$ cut-point of $8\mu g/m^3$ based on the American Thoracic Society recommended annual standard was used.(Cromar et al., 2021) For combined analyses, constituent dichotomized cut-points were based on the median pollutant exposure across the two pooled cohorts, whereas for cohort-specific analyses, $PM_{2.5}$ and constituent cut-points were based on the median exposure within that cohort. First, a contrasts matrix was developed for the primary comparisons and covariates, and then a linear model was used to evaluate for univariate associations between methylation status at each CpG locus and predictors or covariates of interest. Subsequently, this linear model fit is run through the empirical Bayes procedure of *limma*, which moderates standard errors towards a common value, computing moderated t-statistics, F-statistics, and log odds of differential methylation. The result is a list of significantly up- and down-methylated CpGs, from which an output table is generated where the top CpGs are annotated to the human genome build 19 through the package *IlluminaHumanMethylationEPICanno.ilm10b2.hg19*. P-values were adjusted according to the Benjamini & Hochberg procedure whereby a false discovery rate (FDR) threshold of 0.05 was determined to be acceptable.(Benjamini, 1995)

Differentially-methylated region (DMR) analysis will later be performed using the *DMRcate* package.(Peters et al., 2015) These analyses are not reported in this dissertation.

Fixed effects meta-analysis of cohort-specific DMP analyses, weighted by sample size, was performed using METAL.(Willer, Li, & Abecasis, 2010) Raw p-values were input into the

METAL pipeline and a FDR threshold of 0.05 was subsequently applied using the "FDR" function of the *fuzzySim* R package on the resulting meta-analysis p-values.(Barbosa, n.d.)

Gene set enrichment analysis (GSEA) was performed using the "GOmeth" function in the *missMethyl* package in R.(Maksimovic, Oshlack, & Phipson, 2021) This approach adjusts for biases due to genes with more annotated CpG sites being more likely to be identified as differentially methylated, as well as biases related to the 10% of CpGs that are annotated to more than one gene. Significantly enriched gene sets are identified by taking the list of significant differentially-methylated CpG probes (DMPs) for an analysis and mapping them to the gene annotation associated with *IlluminaHumanMethylationEPICanno.ilm10b2.hg19*, followed by testing for enrichment of GO terms or KEGG pathways with a Wallenius' non-central hypergeometric test. FDR thresholds of 0.05 were used.(Benjamini, 1995) This procedure was performed for the cohort-specific DMP analyses for the Simmons and CARE-PF cohorts for the 5-year and 3-month pre-sampling exposure periods.

The top CpG in the Simmons cohort and meta-analysis was evaluated for associations with clinical outcomes of mortality using Cox proportional hazards regression and baseline FVC and D_LCO using multivariable linear regression. Analyses were adjusted for age at diagnosis, sex, smoking history, race, IPF vs non-IPF diagnosis, and disadvantage score. Associations of $PM_{2.5}$ and constituents with this top CpG were individually tested using multivariable linear models adjusted for the same covariates to confirm the findings and magnitude of effect detected in the DMP analyses above. Subsequently, Cox proportional hazards models were constructed that first excluded, then included the top CpG, in addition to the covariates mentioned, to determine the proportion of the pollutant-mortality association that was mediated by the top CpG. Mediation proportion was determined by the following formula:

$$Mediation\ proportion = 1 - (\frac{coefficient\ of\ model\ with\ pollutant + top\ CpG}{coefficient\ of\ model\ with\ pollutant\ alone})$$

All analyses were performed in R (R statistical computing environment, version 4.2.1).

4.2.3 Results

4.2.3.1 Sample Processing, Quality Control, and Probe Filtering

Two samples with a mean detection p-value >0.05 were removed from further analysis, and one with a mean detection p-value just below 0.05 was retained (**Figure 14**). Subsequently, normalization was performed. The raw and normalized density plots of beta values for the full cohort are shown in **Figure 15,** illustrating how this procedure reduces the between sample variation in β-values.

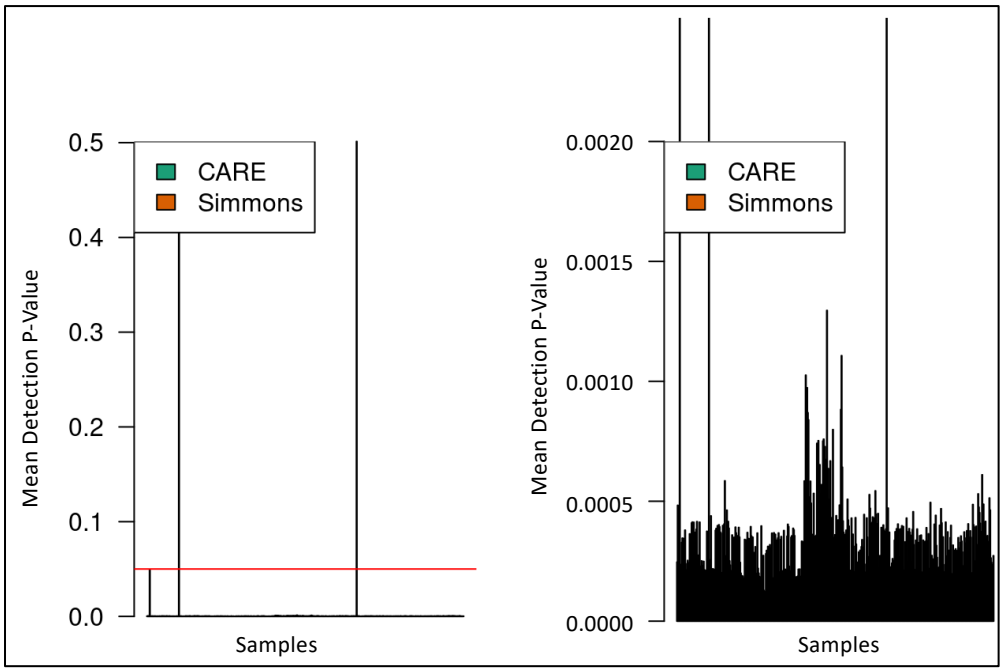


Figure 14 – Mean detection p-value plots. Plots of all 478 samples assayed on Infinium MethylationEPIC array, with the y-axis demonstrating the mean detection p-value across all CpGs for each sample with the horizontal bar on the left plot reflecting the detection p-value cutoff. There are two samples that exceed that cutoff (which were

excluded from further analyses) and one where the mean detection p-value sits right at the cutoff of 0.05 (which was included). The plot on the right is zoomed in to show the mean detection p-values for the majority of samples.

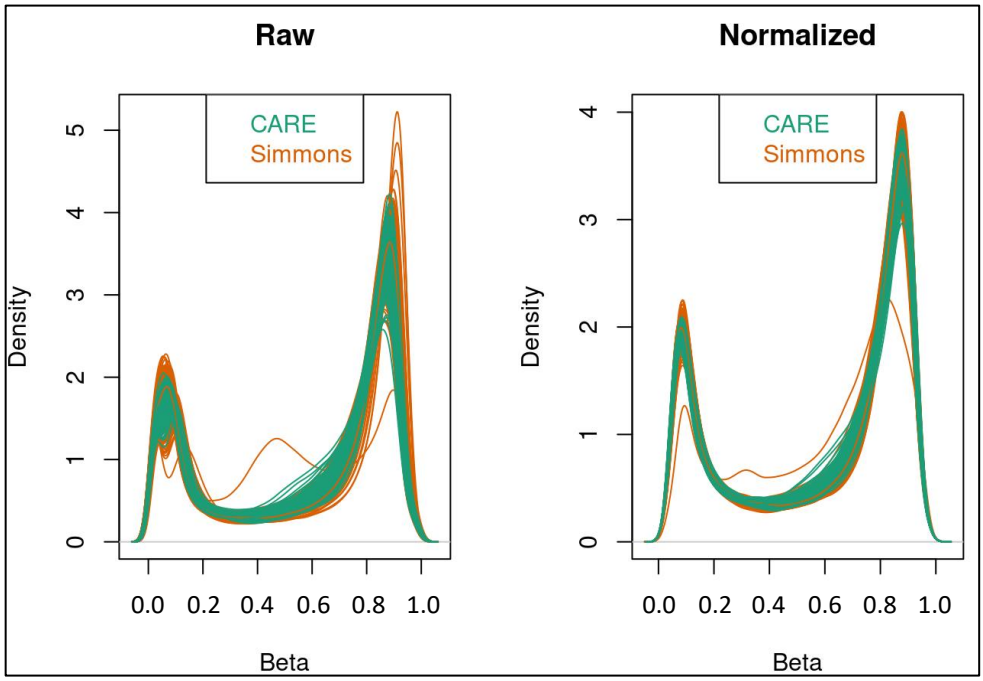


Figure 15 – Density plot of raw vs quantile normalized samples. There is a line for each of the 476 samples that passed detection p-value thresholds, and samples are normalized against background signal using a subset quantile approach. One sample from Simmons continues to have a slightly abnormal distribution despite normalization, although the rest of the samples appear to have most probes clustering around β-values of 0 or 1, reflecting completely unmethylated and methylated CpGs, respectively, which is the pattern we would expect.

Prior to filtering, β-values were calculated for 865,859 CpG probes. Filtering out probes whose detection p-value was >0.01 on one or more samples resulted in a loss of 142,052 probes. Removal of probes on sex chromosomes additionally resulted in the loss of 15,656 probes. Another 22,750 probes were removed as they coincided with single nucleotide polymorphisms, which may interfere with probe hybridization. Lastly, 36,348 cross-reactive probes were removed. This left 649,053 probes to analyze.

4.2.3.2 Cohort Characteristics

Table 9 demonstrates the characteristics of patients whose samples were included from the two cohorts included in this epigenome-wide chip-based DNAm study.

Table 9 – Patient demographics by cohort.

Patient Characteristics	Simmons Cohort	CARE-PF Cohort
	N=306	N=170
Age at diagnosis, median (IQR), years	65 (58-72)	66 (59-73)
Male sex, n (%)	170 (56%)	89 (52%)
Self-reported race, n (%)		
White	275 (90%)	129 (76%)
Black	16 (5%)	0 (0%)
Asian	1 (0.3%)	25 (15%)
Indigenous[a]	1 (0.3%)	8 (5%)
Pacific Islander	0 (0%)	6 (4%)
Unknown	13 (4%)	2 (1%)
Smoking history, n (%)		
Never	73 (24%)	50 (29%)
Former	161 (53%)	113 (66%)
Current	6 (2%)	7 (4%)
Unknown	66 (22%)	0 (0%)
Specific fILD Diagnosis		
Idiopathic pulmonary fibrosis	172 (56%)	69 (41%)
Connective tissue disease-ILD	92 (30%)	33 (19%)
Fibrotic hypersensitivity pneumonitis	14 (5%)	22 (13%)
Non-IPF idiopathic interstitial pneumonia	3 (1%)	0 (0%)
Other fILD[b]	3 (1%)	0 (0%)
Unclassifiable ILD	22 (7%)	46 (27%)
Baseline FVC % Predicted, median (IQR)[c]	65 (53-78)	77 (66-91)
Baseline D_LCO % Predicted, median (IQR)[d]	46 (34-60)	52 (40-63)
Follow-up Duration, median (IQR), years	3.8 (1.6-7.1)	4.1 (2.5-5.6)
Cause of censoring, n (%)		
Death	173 (57%)	42 (25%)
Lung Transplantation	55 (18%)	13 (8%)
Lost to follow-up or censored by data extraction	78 (25%)	115 (68%)
$PM_{2.5}$ Exposure (ug/m^3), median (IQR)		
5 years pre-sampling	12.1 (10.1-14.2)	5.1 (4.5-5.8)
3 months pre-sampling	10.5 (8.8-12.7)	4.4 (3.7-5.8)

[a] – Includes Native American, American Indian, Alaskan First Nations, & other Indigenous persons.
[b] – Includes drug-, radiation-, aspiration-, or acute lung injury-induced fILD.
[c] – FVC was available for 96% of patients enrolled in Simmons, 94% of CARE-PF.
[d] – D_LCO was available for 92% of patients enrolled in Simmons, 79% of CARE-PF.
CARE-PF, Canadian Registry for Pulmonary Fibrosis; D_LCO, diffusion capacity of lung for carbon monoxide; fILD; fibrotic interstitial lung disease; FVC, forced vital capacity; IPF, idiopathic pulmonary fibrosis; IQR, interquartile range; $PM_{2.5}$, particulate matter with a diameter of 2.5μm or less.

Distributions of $PM_{2.5}$ and constituents between the two cohorts are demonstrated in **Figure 16**. This highlights how there was very little overlap in 5-year pre-censoring exposures to $PM_{2.5}$, SO_4^{2-}, NO_3^-, and NH_4^+ between the two cohorts, with Simmons patients consistently experiencing higher exposures than CARE-PF patients.

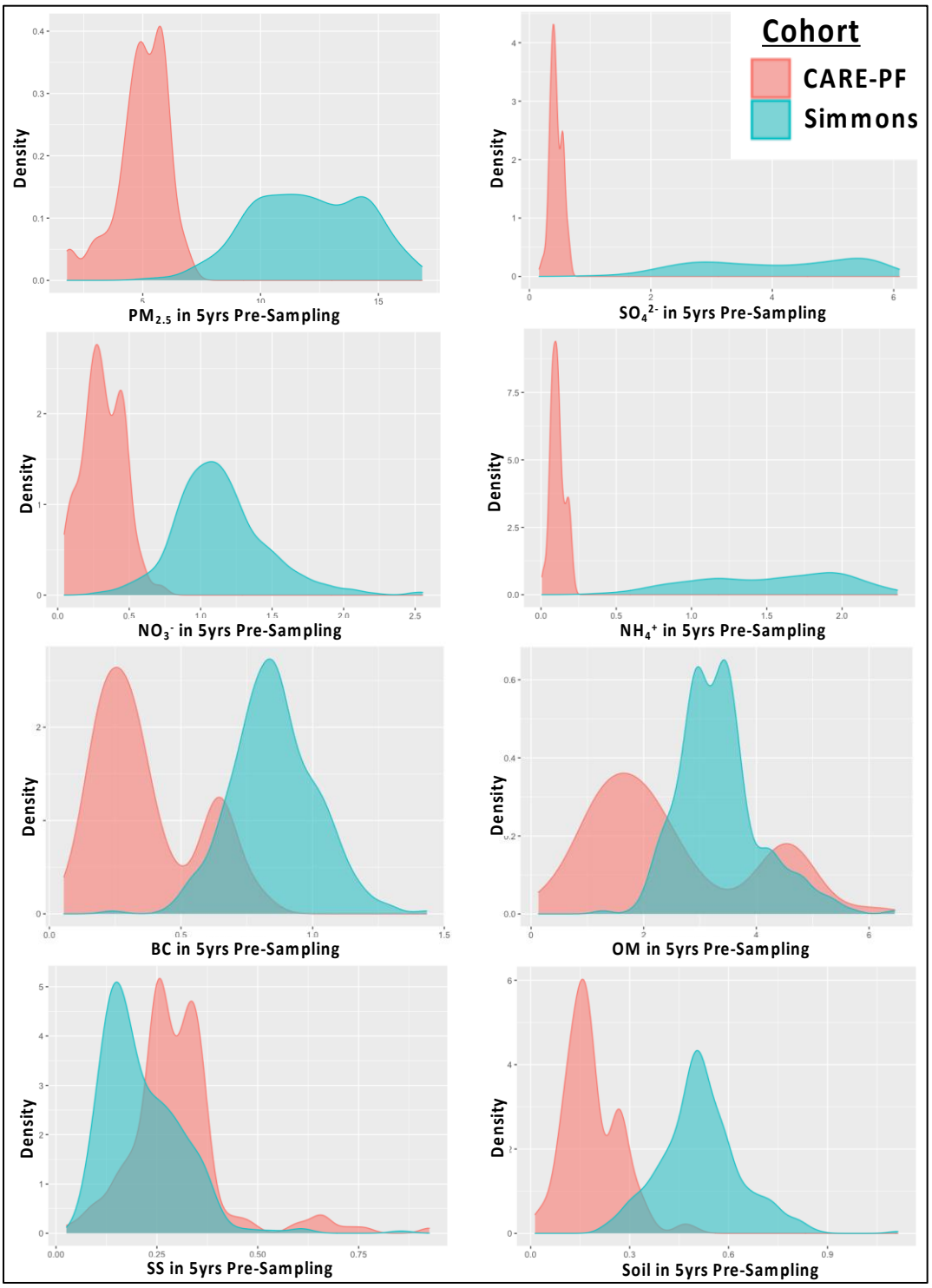


Figure 16 – Distribution of $PM_{2.5}$ and constituent components between Simmons and CARE-PF. Y-axes reflect the density of patients within each cohort having a certain level of pollutant exposure and x-axes reflect the pollutant mass exposures in $\mu g/m^3$. Simmons has higher range of exposures to $PM_{2.5}$, SO_4^{2-}, NO_3^-, NH_4^+, BC, and Soil.

4.2.3.3 Average Beta Values Across Genome Association with $PM_{2.5}$ and Constituent Exposures

Average β-value across all retained CpGs was slightly higher in the CARE-PF cohort as compared to the Simmons cohort (CARE-PF average β-value median=57.2%, IQR=57.1-57.3%; Simmons median=57.1%, IQR-57.1-57.2%). Association of pollutant exposures with average β-value was assessed as in Aim 3.1 to evaluate for consistency with ELISA-based global DNAm findings. This analysis found, contrary to Aim 3.1 that there was no association of higher $PM_{2.5}$ or constituent exposures with average β-value in the Simmons cohort. However, in the CARE-PF cohort, there was a significant association of higher SO_4^{2-}, NO_3^-, BC, and OM with higher average β-value (**Figure 17**). For example, the β-value of 0.00758 for SO_4^{2-} in CARE-PF indicates that for each 1ug/m^3 increase in SO_4^{2-} exposure in the 5-years pre-sampling, there is a 0.76% increase in average β-value. These findings indicate some directional consistency with Aim 3.1, although patients from CARE-PF were not analyzed for global DNAm using the ELISA-based platform.

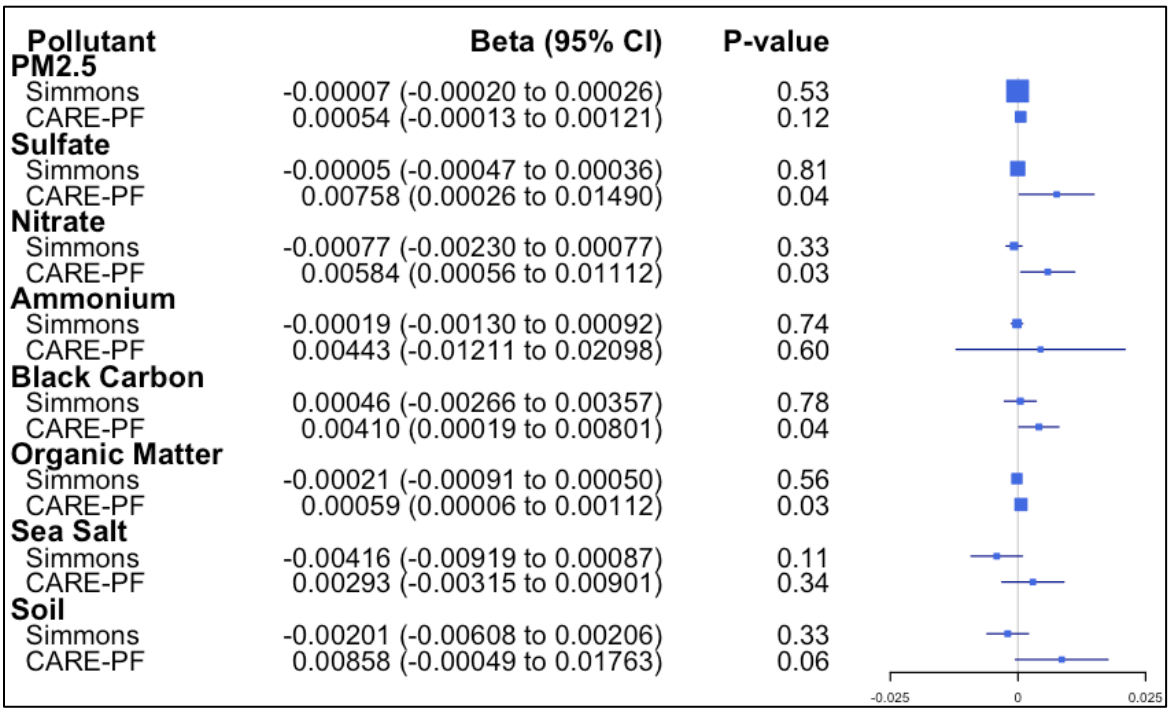


Figure 17 – $PM_{2.5}$ and constituent impact on average β-value across all probes in epigenome-wide DNAm study.

4.2.3.4 Differentially Methylated Probes (DMPs) by Cohort

To consider the difference in methylation patterns by cohort, we evaluated for DMPs between cohorts. Approximately 247,000 CpGs were differentially-methylated between CARE-PF and Simmons cohort patients in analyses adjusted for age at diagnosis, sex, race, smoking history, and IPF vs non-IPF diagnosis. The top 50 DMPs from this analysis are shown in **Table 10**.

Table 10. – Top 50 differentially-methylated probes between Simmons and CARE-PF cohorts in adjusted models. Analyses adjusted for age at diagnosis, sex, smoking history, race, and IPF vs non-IPF diagnosis.

CpG	Chr	Position	Strand	Gene	UCSC Gene Region Feature	Adjusted P-value	logFC
cg00365642	chr2	8456058	-	*LINC00299*	Body	8.01E-23	0.48
cg01826636	chr22	23540280	+	*BCR*	Body	8.01E-23	-0.31
cg18535415	chr1	200983238	+		Body	1.84E-22	-0.41
cg02039171	chr14	23588162	+	*CEBPE*	1stExon	9.06E-22	-0.32
cg04082721	chr16	85815851	-	*U1*	Body	9.06E-22	-0.48
cg22979041	chr5	134787907	+	*TIFAB*	1stExon; 5'UTR	9.06E-22	0.29
cg06275281	chr8	66865816	+			9.06E-22	-0.69
cg18868509	chr1	36390633	+	*EIF2C1*		9.06E-22	0.38
cg11671363	chr5	148810177	+	*MIR145*	TSS200; Body	9.51E-22	-0.62

cg12672818	chr19	4518822	-	*PLIN4*	TSS1500	9.51E-22	-0.31
cg07245150	chr21	44484449	+		Body	1.12E-21	0.33
cg13747899	chr6	33171911	+	*SLC39A7*	TSS1500; 3'UTR	1.68E-21	-0.34
cg01618660	chr9	100882376	+	*TRIM14*	TSS1500	2.59E-21	-0.22
cg24455884	chr10	64216590	-		Body	3.55E-21	0.40
cg03884100	chr17	44248641	-	*KANSL1*	Body	3.55E-21	-0.66
cg07561747	chr1	200983313	-		Body	3.55E-21	-0.57
cg16920725	chr7	37269786	-		Body	3.55E-21	-0.75
cg10948353	chr9	79087332	+	*GCNT1*	5'UTR	3.55E-21	0.41
cg04919592	chr7	2607232	+	*IQCE*	Body	3.55E-21	-0.69
cg11181843	chr12	110816408	+	*ANAPC7*	Body	3.55E-21	-0.80
cg12740667	chr6	13480235	-	*GFOD1*	Body; 5'UTR	3.55E-21	-0.38
cg08961248	chr1	45026634	-	*RNF220*	Body	3.68E-21	0.35
cg22196167	chr8	96133344	+			3.68E-21	0.28
cg13394762	chr14	103871452	-	*MARK3*	Body	3.68E-21	-0.76
cg01502428	chr17	76850266	+	*TIMP2*	3'UTR	4.33E-21	-0.67
cg04068159	chr3	195973977	+		Body	4.33E-21	-0.72
cg00705730	chr2	106438120	-	*NCK2*	5'UTR	4.41E-21	-0.83
cg06264019	chr20	62884301	+			4.41E-21	0.30
cg11958949	chr6	151701920	-	*ZBTB2*	5'UTR	4.41E-21	0.43
cg04289740	chr14	20959034	-			4.41E-21	-0.72
cg03865101	chr12	1883663	+	*ADIPOR2*	Body	4.41E-21	-0.75
cg19517136	chr11	63259705	-	*HRASLS5*	TSS1500	4.43E-21	-0.64
cg27479052	chr1	192953366	+			5.24E-21	-0.62
cg17175527	chr12	109059806	-	*CORO1C*	Body	5.24E-21	-0.66
cg12621097	chr9	88151643	-			5.54E-21	-0.68
cg10098414	chr4	89446385	+	*PIGY*	TSS1500	5.91E-21	-0.72
cg06746829	chr1	17424524	-		Body	5.97E-21	-0.54
cg25074794	chr10	45958885	-		Body	6.04E-21	-0.68
cg19772161	chr6	3746570	+	*PXDC1*	Body	6.58E-21	-0.46
cg03385871	chr18	46311648	-		Body	6.73E-21	0.49
cg26808656	chr1	95189559	-	*RP11-86H7.1*	Body	6.73E-21	-0.43
cg25739938	chr2	9610621	-	*CPSF3*	Body	7.21E-21	-0.78
cg14393004	chr2	219222941	-	*C2orf62*	Body; TSS1500	7.21E-21	-0.52
cg25731731	chr13	41019891	-		Body	7.21E-21	0.38
cg15325734	chr1	9203356	-			7.21E-21	-0.27
cg15135286	chr2	33359281	+	*LTBP1*	Body; TSS1500	7.21E-21	-0.68
cg23516981	chr5	14356625	+		Body	7.32E-21	0.38
cg24235882	chr4	54928822	-		Body	7.53E-21	-0.57
cg14054880	chr3	9433589	-	*RP11-58B17.1*	Body	7.81E-21	-0.78
cg01796438	chr3	11312864	-	*ATG7*	TSS1500	7.94E-21	-0.56

4.2.3.5 Differentially Methylated Probes (DMPs) by $PM_{2.5}$ and Constituent Exposures – Full Cohort

DMPs from adjusted analyses (including covariates of age at diagnosis, sex, race, smoking history, IPF vs non-IPF diagnosis, and cohort) between low versus high $PM_{2.5}$ exposures (< or ≥8μg/m^3) and all constituents are shown in **Table 11**. Initial analyses demonstrate 17

differentially-methylated CpGs in the $PM_{2.5}$ analysis, 3669 for SO_4^{2-}, 0 for NO_3^-, 2978 for NH_4^+, 0 for BC, 1 for OM, 44,659 for SS, and 0 for soil. Twenty-five of the top fifty CpGs (50%) are shared in analyses of DMPs between low vs high SO_4^{2-} and NH_4^+ in the 5-years pre-sampling period, indicating similar methylation pathways may be impacted by these pollutants.

Table 11 – Significant differentially-methylated probes between low vs high PM2.5 and constituents in the 5-yrs pre-sampling (< or ≥8ug/m3) in full cohort adjusted models. Reporting in table limited to top 50 CpGs for pollutants with >50 signficantly differentially-methylated probes. Analyses adjusted for cohort, age at diagnosis, sex, smoking history, race, and IPF vs non-IPF diagnosis. CpGs that are **bolded** reflect peak differentially methylated probes that are shared between more than one pollutant.

CpG	Chr	Position	Strand	Gene	UCSC Gene Region Feature	Adjusted P-value	logFC
Significant Probes for low vs high $PM_{2.5}$ (< or ≥8ug/m^3)							
cg18307928	chr13	114085650	+		Body	0.0003	0.57
cg06929120	chr1	56530823	-			0.0005	-0.88
cg10515414	chr1	15650268	+		Body	0.0008	0.54
cg06100457	chr5	180236921	+	*MGAT1*	TSS200; 5'UTR	0.005	0.77
cg08003732	chr12	7023752	+	*ENO2*	1stExon; 5'UTR	0.007	0.76
cg09134205	chr7	154003215	-	*DPP6*	Body	0.01	0.53
cg03713570	chr16	30789484	+	*RP11-2C24.6*		0.01	-0.41
cg19901956	chr11	77921274	-	*USP35*	Body	0.01	-0.36
cg05098233	chr16	475948	+	*RAB11FIP3*	1stExon; 5'UTR	0.01	-0.57
cg03800969	chr6	159082213	+	*SYTL3*	5'UTR	0.01	-0.92
cg21048542	chr6	80714164	+	*TTK*	TSS200	0.01	0.79
cg02945060	chr5	171469660	-	*STK10*	3'UTR	0.01	-0.43
cg22892539	chr5	1667258	-			0.04	1.03
cg15059239	chr5	121187571	+	*FTMT*	TSS200	0.04	0.81
cg10406526	chr19	1051079	-		Body	0.04	0.55
cg13018448	chr6	33385440	+	*CUTA*	Body; 1stExon	0.04	-0.46
cg03414202	chr12	128850432	-		Body	0.045	0.94
Significant Probes for low vs high SO_4^{2-} (< or ≥2.89ug/m^3)							
cg00002743	chr2	86742476	-	*U6*	Body	1.43E-07	0.35
cg25354716	chr3	33159475	-	*CRTAP*	Body	1.43E-07	0.37
cg07637658	chr19	45887042	-		Body	2.24E-06	0.34
cg12578536	chr5	43003251	+			2.24E-06	0.43
cg08811509	chr9	95640366	-	*ZNF484; ANKRD19P*	TSS200;TSS200	4.08E-06	-0.21
cg23446514	chr15	66797150	+	*SNORD18A; ZWILCH; RPL4*	TSS1500; 1stExon; 5'UTR	5.70E-06	-0.38
cg26508775	chr20	55050494	-	*snoU13*	Body	6.11E-06	0.44
cg01727346	chr22	45596797	-	*KIAA0930*	Body	6.94E-06	0.24
cg02896970	chr16	87460610	-		Body	1.04E-05	0.15
cg24664079	chr7	43908858	+	*MRPS24*	Body	1.04E-05	-0.14
cg00796866	chr16	88744522	+	*SNAI3; RP5-1142A6.3*	Body; 3'UTR	1.04E-05	0.23
cg13149902	chr3	113325172	-	*SIDT1*	Body	2.39E-05	0.32
cg25503565	chr17	38264529	-			2.55E-05	-0.24
cg10016770	chr1	244805996	-			2.71E-05	0.13
cg20659463	chr18	9102542	+	*NDUFV2*	TSS200	3.06E-05	-0.34
cg11867686	chr11	67169368	+	*PPP1CA*	5'UTR; 1stExon	3.36E-05	-0.31

cg07162498	chr10	47900640	-	*FAM21B*	Body	3.36E-05	0.22
cg02562419	chr6	159049626	+		Body	3.36E-05	0.15
cg13286500	chr7	75854009	+	*SRRM3*	5'UTR	5.45E-05	-0.20
cg25665417	chr3	50327690	-	*IFRD2*	Body	5.63E-05	0.20
cg16486145	chr10	33890129	+	*RP11-476F14.1*		5.63E-05	0.22
cg21298511	chr1	201476053	+	*RP11-134G8.7; CSRP1*	TSS200; 5'UTR	5.63E-05	-0.41
cg10300729	chr10	65284527	+	*REEP3*	Body	5.63E-05	0.25
cg13409259	chr22	41682119	+	*RANGAP1*	1stExon; TSS200; 5'UTR	5.63E-05	-0.29
cg01304461	chr12	10103711	+	*AC091814.3; CLEC12A*	TSS1500	5.88E-05	-0.36
cg08699803	chr11	74585894	-	*RP11-147I3.1*	Body	5.88E-05	0.22
cg18031226	chr19	57835157	-	*ZNF543*	Body	6.34E-05	0.25
cg20628191	chr1	20265229	-			6.34E-05	0.30
cg10964394	chr22	41176719	-	*SLC25A17*	Body	6.34E-05	0.21
cg21261121	chr20	43380263	-	*RP4-781B1.2*	3'UTR	6.34E-05	0.16
cg25531473	chr3	52810463	-	*ITIH1*	TSS1500	6.34E-05	0.28
cg19847283	chr19	20349250	-			6.49E-05	-0.28
cg12900931	chr11	3400491	+	*ZNF195*	TSS200	6.49E-05	-0.21
cg20866393	chr17	21952612	+			6.49E-05	0.21
cg16681977	chr1	11332969	+	*UBIAD1*	TSS1500	7.06E-05	-0.19
cg24828071	chr12	123128962	+	*HCAR1*		7.06E-05	0.14
cg10816414	chr5	141416597	-			7.06E-05	0.18
cg06305048	chr19	53466199	-	*ZNF321P; ZNF816*	TSS200	7.23E-05	-0.16
cg18720285	chr11	15096341	+	*CALCB*	Body	7.31E-05	-0.20
cg21400851	chr17	73056681	-		Body	7.43E-05	0.29
cg25013378	chr19	1569576	+			7.56E-05	0.23
cg01340089	chr12	22199200	-	*CMAS*	1stExon; 5'UTR	7.56E-05	-0.26
cg16508522	chr3	141319423	-	*RASA2*	Body	8.18E-05	0.16
cg27223543	chr1	45098125	+		Body	8.38E-05	-0.22
cg26430167	chr5	87984943	+	*CTC-467M3.1*		8.42E-05	-0.19
cg12232041	chr15	22491284	-			9.14E-05	-0.34
cg12438413	chr17	38510643	+	*RARA*	Body	9.24E-05	0.21
cg00086995	chr17	65027382	+	*AC005544.1; CACNG4*	3'UTR	9.35E-05	0.20
cg08008065	chr1	62673941	-		Body	9.44E-05	0.27
cg01309005	chr12	121517574	+			9.44E-05	0.24
Significant Probes for low vs high NH_4^+ (< or ≥1.13ug/m^3)							
cg25354716	chr3	33159475	-	*CRTAP*	Body	6.11E-15	0.39
cg07637658	chr19	45887042	-		Body	3.69E-12	0.34
cg23446514	chr15	66797150	+	*SNORD18A; ZWILCH; RPL4*	TSS1500; 1stExon; 5'UTR	1.35E-11	-0.38
cg00002743	chr2	86742476	-	*U6*	Body	4.77E-11	0.32
cg12578536	chr5	43003251	+			3.07E-10	0.40
cg08811509	chr9	95640366	-	*ZNF484; ANKRD19P*	TSS200	3.58E-10	-0.20
cg26508775	chr20	55050494	-	*snoU13*	Body	5.77E-10	0.41
cg10300729	chr10	65284527	+	*REEP3*	Body	8.48E-10	0.26
cg19847283	chr19	20349250	-			1.18E-09	-0.28
cg12667999	chr17	47901743	-	*KAT7*	Body	1.43E-09	0.29
cg25013316	chr1	175483605	+	*TNR*	5'UTR	2.18E-09	-0.37
cg10816414	chr5	141416597	-			2.35E-09	0.18
cg01727346	chr22	45596797	-	*KIAA0930*	Body	2.60E-09	0.22
cg00081062	chr1	76209564	-		Body	3.41E-09	0.27
cg02896970	chr16	87460610	-		Body	3.50E-09	0.14
ch.4.113910337F	chr4	113690888	+			3.91E-09	-0.22

cg10582639	chr11	121526840	-			4.96E-09	-0.17
cg04378177	chr9	100505109	-	*RP11-546O6.4*		5.06E-09	0.15
cg01340089	chr12	22199200	-	*CMAS*	1stExon; 5'UTR	5.52E-09	-0.26
cg26430167	chr5	87984943	+	*CTC-467M3.1*		6.27E-09	-0.19
cg00796866	chr16	88744522	+	*SNAI3; RP5-1142A6.3*	Body; 3'UTR	6.41E-09	0.20
cg21298511	chr1	201476053	+	*RP11-134G8.7; CSRP1*	TSS200; 5'UTR	6.51E-09	-0.40
cg14201314	chr16	34984295	-	*RP11-352B15.1; RN5S415*	Body	6.54E-09	0.16
cg19257979	chr20	37101387	-	*RALGAPB*	TSS200	6.79E-09	-0.28
cg01304461	chr12	10103711	+	*AC091814.3; CLEC12A*	TSS1500	6.80E-09	-0.34
cg06305048	chr19	53466199	-	*ZNF321P; ZNF816*	TSS200	6.96E-09	-0.16
cg13409259	chr22	41682119	+	*RANGAP1*	1stExon; TSS200; 5'UTR	9.49E-09	-0.28
cg26080286	chr14	39459310	+			9.90E-09	-0.29
cg00746783	chr21	46977243	+			1.00E-08	0.16
cg18720285	chr11	15096341	+	*CALCB*	Body	1.04E-08	-0.19
cg27348461	chr8	105678889	+	*RP11-127H5.1*		1.12E-08	0.17
cg20659463	chr18	9102542	+	*NDUFV2*	TSS200	1.23E-08	-0.31
cg07162498	chr10	47900640	-	*FAM21B*	Body	1.53E-08	0.20
cg06686702	chr19	46684843	-			1.80E-08	-0.28
cg16921868	chr3	128967831	-	*COPG1*	TSS1500	1.88E-08	-0.20
cg27611827	chr17	41392569	+			2.11E-08	-0.27
ch.12.4961748 9F	chr12	51331222	+			2.15E-08	-0.15
cg08008065	chr1	62673941	-		Body;	2.20E-08	0.26
cg16127415	chr7	105925141	+	*NAMPT*	Body	2.21E-08	-0.24
cg12930714	chr14	24835006	-	*NFATC4*	TSS1500	2.24E-08	-0.19
cg01576531	chr8	29206508	+	*DUSP4*	TSS200;Body	2.33E-08	-0.35
cg14808759	chr19	40854069	+	*PLD3; C19orf47*	TSS1500; 5'UTR	2.42E-08	-0.35
cg07867325	chr3	73159953	-	*U2*		2.63E-08	0.38
cg24828071	chr12	123128962	+	*HCAR1*		2.74E-08	0.13
cg11496113	chr5	34627766	-			2.75E-08	0.20
cg05969296	chr11	8932792	-	*AKIP1*	5'UTR; TSS200; TSS1500	2.87E-08	-0.27
cg13197283	chr7	152433782	-			2.88E-08	0.24
cg23967848	chr2	71556215	+	*ZNF638*		3.04E-08	0.19
cg25665417	chr3	50327690	-	*IFRD2*	Body	3.25E-08	0.18
cg08876558	chr11	59394462	-	*AP000442.1; AP000442.1*		3.25E-08	0.36
Significant Probes for low vs high OM (< or ≥3.03ug/m^3)							
cg21824017	chr2	15804594	+			0.035	0.16
Significant Probes for low vs high SS (< or ≥0.24ug/m^3)							
cg25849390	chr7	56128170	+	*CCT6A; SNORA15*	Body	4.77E-11	0.42
cg22932101	chr14	71815816	+			3.83E-10	0.23
cg11898358	chr3	172325619	+			3.83E-10	0.22
cg11009716	chr3	52888489	-		Body	3.83E-10	0.29
cg07835232	chr7	140302680	+	*DENND2A*	TSS1500	5.15E-10	0.18
cg01145910	chr1	5729401	+	*RP11-154H17.1*		1.26E-09	0.29
cg26118675	chr2	96256841	-	*TRIM43*	TSS1500	1.27E-09	0.41
cg23319427	chr10	45684486	+	*U6*		8.17E-09	0.26
cg10685380	chr5	158745090	-	*RNU4ATAC2P*	Body	1.61E-08	0.31
cg11284631	chr5	38870649	+		Body	2.29E-08	0.19
cg20034552	chr4	70696320	-			2.87E-08	0.24
cg07637658	chr19	45887042	-		Body	2.99E-08	0.26
cg24207610	chr19	41632920	-		Body	2.99E-08	0.29

cg00727630	chr20	50721366	+	*ZFP64*	Body	3.07E-08	-0.28
cg18642567	chr14	21755798	+	*RPGRIP1*	TSS1500	3.07E-08	0.34
cg18057710	chr1	197705078	+	*DENND1B*	5'UTR; Body	4.15E-08	0.24
cg24839008	chr3	13335779	+			4.58E-08	0.19
cg17830540	chr3	41923177	-		Body	4.58E-08	0.30
cg12720152	chr14	91087476	-	*TTC7B*	Body	4.58E-08	0.22
cg23207011	chr1	96884428	+			4.88E-08	0.36
cg06747358	chr2	135870040	+	*RAB3GAP1*	Body	4.88E-08	0.30
cg26459588	chr11	11838030	-			4.88E-08	0.26
cg12590791	chr6	89931408	-	*GABRR1*	5'UTR	4.88E-08	0.21
cg03528558	chr20	49991083	-			4.88E-08	0.14
cg20834311	chr2	164123027	-			4.88E-08	0.27
cg04386839	chr1	158534143	+	*OR6P1*	TSS1500	4.88E-08	0.21
cg15152958	chr3	9369039	-	*RP11-380O24.1*		4.89E-08	0.22
cg12198254	chr3	11521788	+		Body	4.89E-08	0.19
cg10585698	chr12	113795049	-	*PLBD2; PLBD2*	TSS1500	5.07E-08	0.21
cg04193083	chr17	41323562	+	*BRCA1; NBR1*	5'UTR	5.46E-08	-0.26
cg13000555	chr12	104460455	+		Body	5.52E-08	0.21
cg00470768	chr15	41285147	+	*INO80*	Body	5.52E-08	0.27
cg08006046	chr6	11106504	+	*ERVFRD-1; ERVFRD-1*	Body; 5'UTR	5.52E-08	0.18
cg20673919	chr11	125702313	-	*PATE4*	TSS1500	5.52E-08	0.21
cg00945341	chr19	22694413	+			6.78E-08	0.19
cg25312694	chr7	17500466	+	*AC019117.2; AC019117.1*	Body	7.10E-08	0.12
cg11853096	chr1	50332874	-	*AGBL4*	Body	7.10E-08	0.28
cg05935240	chr7	43713241	-	*C7orf44*	Body; 5'UTR	7.10E-08	0.17
cg12323760	chr1	40544720	-	*PPT1*	Body	7.10E-08	0.18
cg11136041	chr1	158111350	+	*RP11-404O13.1*		7.10E-08	0.29
cg25595449	chr12	74431418	+	*RP11-711C17.2*		7.46E-08	0.21
cg09150840	chr22	19304193	+			8.38E-08	0.24
cg07455406	chr14	21077527	+	*AL163195.3*		8.76E-08	0.23
cg27408765	chr12	63359135	-			8.76E-08	0.18
cg18593556	chr20	47965469	-			8.76E-08	0.22
cg04248127	chr4	71226047	-	*SMR3A*	TSS1500	8.76E-08	0.27
cg08131547	chr19	9691569	+	*ZNF121*	5'UTR	1.00E-07	0.29
cg18098774	chr2	107084802	+	*RGPD3*	TSS200	1.02E-07	0.07
cg09950479	chr7	80610724	+			1.05E-07	0.20
cg17214388	chr19	3431061	-		Body	1.15E-07	0.17

4.2.3.6 Differentially Methylated Probes (DMPs) by $PM_{2.5}$ and Constituent Exposures – Cohort-Specific

Given substantial differences in the distribution of $PM_{2.5}$ and constituent components between the Simmons and CARE-PF cohort, with very few patients from CARE-PF in the high exposure group for any pollutant and very few from Simmons in the low exposure groups (as demonstrated in **Figure 16**), we elected to perform cohort-specific DMP analysis, followed by meta-analysis. **Table 12** demonstrates the top differentially-methylated CpGs for all constituents,

with the top 50 listed for analyses with >50 significant CpGs. For $PM_{2.5}$, there were 256 DMPs; 338 for SO_4^{2-}; 1 for NO_3^-; 28 for NH_4^+; 1065 for SS; and 1 for Soil. There were no significantly DMPs for BC or OM. The top CpG for $PM_{2.5}$, SO_4^{2-}, NH_4^+, and SS analyses in the 5-years pre-sampling period was cg25354716, which is annotated to the body of the gene *CRTAP*. There were shared top CpGs with $PM_{2.5}$ or other constituents in 35/50 (70%) of the $PM_{2.5}$ analysis, 36/50 (72%) for SO_4^{2-}, 1/1 (100%) for NO_3^-, 17/28 (61%) for NH_4^+, and 9/50 (18%) for SS, indicating multiple shared pathways of involvement for $PM_{2.5}$ and several of its constituents.

Table 12 – Significant differentially-methylated probes between low vs high PM2.5 and constituents in the 5-yrs pre-sampling in Simmons cohort-alone adjusted models. Reporting in table limited to top 50 CpGs for pollutants with >50 signficantly differentially-methylated probes. Analyses adjusted for age at diagnosis, sex, smoking history, race, and IPF vs non-IPF diagnosis. CpGs that are **bolded** reflect peak differentially methylated probes that are shared between more than one pollutant.

CpG	Chr	Position	Strand	Gene	UCSC Gene Region Feature	Adjusted P-value	logFC
Significant Probes for low vs high $PM_{2.5}$ (< or ≥12.07μg/m^3)							
cg25354716	chr3	33159475	-	*CRTAP*	Body	3.80E-06	0.32
cg07637658	chr19	45887042	-		Body	1.80E-05	0.31
cg19782158	chr8	36641197	-	*KCNU1*	TSS1500	1.07E-04	0.21
cg18924184	chr14	107257615	+			1.04E-03	0.13
cg06686702	chr19	46684843	-			1.68E-03	-0.26
cg12578536	chr5	43003251	+			1.68E-03	0.32
cg03846240	chr6	35723021	-			2.23E-03	0.18
cg26060971	chr3	52407402	+	*DNAH1*	5'UTR	2.23E-03	-0.24
cg08811509	chr9	95640366	-	*ZNF484; ANKRD19P*	TSS200	2.23E-03	-0.17
cg14201314	chr16	34984295	-	*RP11-352B15.1; RN5S415*	Body	2.23E-03	0.13
cg27457921	chr17	43976811	-	*MAPT*	5'UTR	3.15E-03	0.19
cg17598704	chr4	148885518	-	*ARHGAP10*	5'UTR	3.15E-03	0.10
cg11496113	chr5	34627766	-			3.15E-03	0.17
cg24277705	chr10	89603720	+	*CFL1P1*	Body	3.42E-03	0.17
cg03270074	chr2	95870287	-			3.52E-03	0.15
cg15967169	chr6	30167836	+	*TRIM26*	5'UTR	3.62E-03	0.18
cg13821031	chr9	116035939	+	*CDC26*	5'UTR	3.71E-03	0.25
cg08543976	chr17	19699755	+	*ULK2*	Body	3.94E-03	0.18
cg08954646	chr6	31495960	-	*MCCD1*	TSS1500	3.94E-03	0.16
cg09131901	chr1	45980096	-		Body	3.94E-03	0.09
cg09209493	chr12	20124555	+	*AC087237.1; RP11-405A12.2*		4.19E-03	0.12
cg22487204	chr1	156435421	-	*MEF2D*	3'UTR	4.64E-03	0.10
cg09502141	chr9	4700338	-		Body	4.70E-03	0.21
cg12406406	chr2	139654951	+		Body	4.70E-03	0.14
cg19116668	chr7	99932089	-		Body	4.77E-03	-0.20
cg13000555	chr12	104460455	+		Body	5.34E-03	0.19
cg15303300	chr1	163844956	-			6.05E-03	-0.23

cg21261121	chr20	43380263	-	*RP4-781B1.2*	3'UTR	6.05E-03	0.13
cg01304461	chr12	10103711	+	*AC091814.3; CLEC12A*	TSS1500	6.05E-03	-0.26
cg22506343	chr3	127173672	-			6.05E-03	0.12
cg08109808	chr17	78029642	+		Body	6.50E-03	0.14
cg26508775	chr20	55050494	-	*snoU13*	Body	6.63E-03	0.31
cg07162498	chr10	47900640	-	*FAM21B*	Body	6.63E-03	0.17
cg22549556	chr8	70042274	+			7.27E-03	0.16
cg00796866	chr16	88744522	+	*SNAI3; RP5-1142A6.3*	Body; 3'UTR	7.27E-03	0.16
cg20628191	chr1	20265229	-			7.81E-03	0.22
cg22637307	chr2	68590998	-	*AC015969.3; PLEK*	TSS1500	8.85E-03	0.14
cg07835232	chr7	140302680	+	*DENND2A*	TSS1500	9.06E-03	0.15
cg26080286	chr14	39459310	+			9.06E-03	-0.22
cg10297223	chr3	148414649	-	*AGTR1*	TSS1500	9.30E-03	0.23
cg05014601	chr6	119378161	+	*Y_RNA; FAM184A*	Body; 5'UTR	9.53E-03	0.21
cg12894234	chr5	149125770	-	*PPARGC1B*	Body	9.71E-03	0.15
cg05526438	chr5	174350112	+	*CTC-281M20.1*		9.71E-03	0.16
cg01901468	chr8	128250972	-			9.71E-03	0.17
cg12667999	chr17	47901743	-	*KAT7*	Body	1.01E-02	0.21
cg21400851	chr17	73056681	-		Body	1.01E-02	0.22
cg24713063	chr20	35459146	-	*SOGA1*	Body	1.01E-02	0.09
cg15635761	chr13	45411234	+			1.01E-02	0.11
cg10709925	chr3	149768629	-	*PFN2*		1.04E-02	0.08
cg07057218	chr9	6086163	-			1.08E-02	0.10
Significant Probes for low vs high SO_4^{2-} (< or ≥4.17μg/m^3)							
cg25354716	chr3	33159475	-	*CRTAP*	5'UTR	1.98E-07	0.35
cg07637658	chr19	45887042	-			7.56E-07	0.33
cg08811509	chr9	95640366	-	*ZNF484; ANKRD19P*	TSS200; 5'UTR	1.59E-04	-0.19
cg12578536	chr5	43003251	+			1.59E-04	0.35
cg19782158	chr8	36641197	-	*KCNU1*	TSS1500	6.29E-04	0.20
cg06686702	chr19	46684843	-			6.29E-04	-0.27
cg12667999	chr17	47901743	-	*KAT7*	5'UTR; TSS200	6.64E-04	0.24
cg03270074	chr2	95870287	-			6.64E-04	0.16
cg12406406	chr2	139654951	+			7.05E-04	0.15
cg24277705	chr10	89603720	+	*CFL1P1*	5'UTR	9.73E-04	0.18
cg07162498	chr10	47900640	-	*FAM21B*	5'UTR	9.92E-04	0.18
cg26508775	chr20	55050494	-	*C20orf43; snoU13*	5'UTR; TSS1500	9.92E-04	0.34
cg11496113	chr5	34627766	-			1.08E-03	0.18
cg10274208	chr13	41351044	+			1.10E-03	-0.32
cg26060971	chr3	52407402	+	*DNAH1*	5'UTR	1.13E-03	-0.25
cg03846240	chr6	35723021	-			1.86E-03	0.18
cg14201314	chr16	34984295	-	*RP11-352B15.1*	5'UTR; TSS200	2.10E-03	0.13
cg12894234	chr5	149125770	-	*PPARGC1B*	5'UTR	2.10E-03	0.16
cg26634055	chr16	72042068	+	*DHODH*	TSS1500	2.17E-03	0.14
cg01901468	chr8	128250972	-			2.26E-03	0.19
cg13821031	chr9	116035939	+	*CDC26*	5'UTR; 3'UTR	2.42E-03	0.26
cg09834444	chr16	89937472	+	*SPIRE2*	3'UTR	2.83E-03	0.16
cg17598704	chr4	148885518	-	*ARHGAP10*	5'UTR	2.83E-03	0.10
cg21337074	chr16	57515799	-	*DOK4*	5'UTR; TSS200	2.83E-03	0.18
cg20628191	chr1	20265229	-			2.83E-03	0.24
cg21261121	chr20	43380263	-	*RP4-781B1.2*	TSS200	2.83E-03	0.13
cg19116668	chr7	99932089	-			2.83E-03	-0.21
cg24434232	chr7	73695326	+			2.90E-03	0.17
cg06576783	chr6	53572893	-			3.09E-03	0.18

cg27457921	chr17	43976811	-	*MAPT*	5'UTR	3.09E-03	0.18
cg22637307	chr2	68590998	-	*AC015969.3; PLEK*	3'UTR; TSS1500	3.19E-03	0.15
cg12120326	chr22	30237309	-			3.19E-03	0.17
cg00796866	chr16	88744522	+	*SNAI3; RP5-1142A6.3*	3'UTR; 5'UTR	3.19E-03	0.17
cg03929077	chr10	51844517	-	*FAM21A*	5'UTR; 5'UTR	3.38E-03	0.16
cg21118749	chr6	153042148	+			3.38E-03	-0.14
cg01993027	chr1	29487000	-	*SRSF4*	3'UTR; TSS1500; 1stExon; TSS200	4.22E-03	0.22
cg18924184	chr14	107257615	+			4.34E-03	0.12
cg14776910	chr1	92355253	-	*TGFBR3*	5'UTR	4.37E-03	0.21
cg21400851	chr17	73056681	-			4.37E-03	0.23
cg01304461	chr12	10103711	+	*AC091814.3; CLEC12A*	3'UTR; TSS1500	4.37E-03	-0.26
cg08954646	chr6	31495960	-	*MCCD1*	TSS1500	4.37E-03	0.16
cg08543976	chr17	19699755	+	*ULK2*	TSS200; TSS1500	5.45E-03	0.17
cg09561417	chr17	18223010	+			5.45E-03	0.10
cg26080286	chr14	39459310	+			5.45E-03	-0.22
cg24030675	chr11	108093402	+	*ATM; NPAT*	TSS1500; 1stExon; 5'UTR; TSS200	5.45E-03	-0.31
cg08282960	chr11	111955637	-	*C11orf57; TIMM8B*	3'UTR	5.45E-03	0.29
ch.4.2387433R	chr4	129608123	+			5.45E-03	-0.18
cg02224047	chr7	101883734	-			6.14E-03	0.18
cg04951822	chr12	113345598	+	*RP1-71H24.1*	3'UTR	6.54E-03	-0.24
cg23446514	chr15	66797150	+	*RPL4; SNORD18A; ZWILCH*	3'UTR; TSS1500; TSS200; 1stExon; 5'UTR	6.73E-03	-0.27
Significant Probes for low vs high NO_3^- (< or ≥1.11µg/m^3)							
cg15967169	chr6	30167836	+	*TRIM26*	3'UTR; 5'UTR	3.75E-03	0.19
Significant Probes for low vs high NH_4^+ (< or ≥1.61µg/m^3)							
cg25354716	chr3	33159475	-	*CRTAP*	5'UTR	2.03E-04	0.29
cg07637658	chr19	45887042	-			2.71E-04	0.29
cg19782158	chr8	36641197	-	*KCNU1*	TSS1500	1.33E-03	0.19
cg12578536	chr5	43003251	+			2.91E-03	0.32
cg01901468	chr8	128250972	-			8.99E-03	0.18
cg12406406	chr2	139654951	+			2.55E-02	0.13
cg23446514	chr15	66797150	+	*SNORD18A; ZWILCH; RPL4*	TSS1500; TSS200; 1stExon; 5'UTR	2.55E-02	-0.26
cg03270074	chr2	95870287	-			2.55E-02	0.14
cg08811509	chr9	95640366	-	*ZNF484; ANKRD19P*	TSS200; 5'UTR	2.55E-02	-0.15
cg17598704	chr4	148885518	-			2.55E-02	0.10
cg26508775	chr20	55050494	-	*snoU13*	TSS1500	2.55E-02	0.30
cg06400595	chr12	56551642	-	*RP11-603J24.14; MYL6; RP11-603J24.18; MYL6B*	3'UTR; TSS1500; 5'UTR; 3'UTR	2.60E-02	-0.36
cg06515159	chr21	34400659	+	*OLIG2; AP000282.2*	1stExon; 3'UTR	2.65E-02	-0.21
cg14201314	chr16	34984295	-	*RP11-352B15.1; RN5S415*	5'UTR; TSS200	2.82E-02	0.12
cg18924184	chr14	107257615	+			2.95E-02	0.11
cg12667999	chr17	47901743	-	*KAT7*	5'UTR	2.95E-02	0.20
cg06712335	chr16	2028325	-	*TBL3*	ExonBnd	2.95E-02	0.15
cg15967169	chr6	30167836	+	*TRIM26*	5'UTR; 3'UTR	3.16E-02	0.17
cg08954646	chr6	31495960	-	*MCCD1*	TSS1500	3.46E-02	0.14
cg14348828	chr7	142985245	-	*AC073342.12; CASP2*	TSS1500	3.46E-02	-0.29
cg11374452	chr9	112951179	-			3.46E-02	-0.16
cg09834444	chr16	89937472	+	*SPIRE2*	3'UTR	3.46E-02	0.14
cg09228876	chr3	42738590	+	*HHATL*	ExonBnd	3.46E-02	0.19

cg27037944	chr9	89873967	+	*SNORA26*	TSS1500	3.46E-02	-0.15
cg03846240	chr6	35723021	-			3.46E-02	0.15
ch.4.2387433R	chr4	129608123	+			4.29E-02	-0.17
Significant Probes for low vs high SS (< or ≥0.19μg/m^3)							
cg25354716	chr3	33159475	-	*CRTAP*	Body	3.01E-07	0.32
cg07637658	chr19	45887042	-		Body	3.01E-07	0.32
cg00934735	chr5	180582587	-	*OR2V2*	1stExon	1.18E-05	0.25
cg07835232	chr7	140302680	+	*DENND2A*	TSS1500	3.72E-05	0.18
cg13000555	chr12	104460455	+		Body	5.81E-05	0.22
cg24702040	chr1	233501790	+		Body	5.81E-05	0.12
cg12406406	chr2	139654951	+		Body	7.44E-05	0.15
cg03860054	chr16	55691102	-	*SLC6A2*	Body	1.15E-04	0.24
cg12198254	chr3	11521788	+		Body	1.15E-04	0.20
cg08879913	chr10	58130574	-			1.15E-04	0.18
cg12253859	chr10	96748928	-	*CYP2C9*	3'UTR	1.15E-04	0.20
cg24030675	chr11	108093402	+	*ATM; NPAT*	TSS200	1.76E-04	-0.34
cg01339959	chr2	184125517	-			1.76E-04	0.21
cg23562228	chr19	49657666	-	*HRC*	1stExon	1.76E-04	0.11
cg07162498	chr10	47900640	-	*FAM21B*	Body	1.76E-04	0.18
cg21261121	chr20	43380263	-	*RP4-781B1.2*	3'UTR	1.76E-04	0.14
cg18098774	chr2	107084802	+	*RGPD3*	TSS200	1.76E-04	0.07
cg11496113	chr5	34627766	-			1.76E-04	0.17
cg13300098	chr9	33036573	+	*DNAJA1*	ExonBnd; Body	1.78E-04	0.28
cg22142205	chr17	80132787	+		Body	1.78E-04	0.22
cg27408765	chr12	63359135	-			1.78E-04	0.17
cg15091333	chr7	100908800	+			1.78E-04	0.27
cg04193083	chr17	41323562	+	*BRCA1; NBR1*	5'UTR	1.78E-04	-0.27
cg23679920	chr12	42376057	-			1.86E-04	0.21
cg25849390	chr7	56128170	+	*CCT6A; SNORA15*	Body	1.86E-04	0.36
cg22932101	chr14	71815816	+			1.86E-04	0.20
cg17214388	chr19	3431061	-		Body	2.29E-04	0.18
cg15152958	chr3	9369039	-	*RP11-380O24.1*		2.29E-04	0.22
cg00281333	chr1	40203613	-	*PPIE*	TSS1500	2.29E-04	0.18
cg01123250	chr2	210673545	+		Body	2.76E-04	0.25
cg24800930	chr14	24454176	-	*DHRS4L2*	5'UTR	2.76E-04	0.25
cg20577728	chr15	23034331	+	*NIPA2*	5'UTR; 1stExon	3.25E-04	-0.22
cg22549556	chr8	70042274	+			3.27E-04	0.17
cg06880365	chr19	36870044	+	*ZFP14*		4.12E-04	-0.33
cg14445171	chr9	131419910	+	*WDR34*	TSS1500	4.12E-04	0.20
cg22778132	chr4	186317907	-	*LRP2BP; ANKRD37*	1stExon; 5'UTR	4.22E-04	-0.24
cg11342415	chr10	1034284	+	*GTPBP4; AL359878.1*	TSS200	4.22E-04	-0.34
cg26784412	chr6	33088710	-		Body	4.33E-04	0.19
cg15613100	chr5	72804620	+	*AC099522.1*		4.33E-04	0.23
cg04135907	chr9	95645293	-	*RP11-526D8.7; ANKRD19P*	Body	4.33E-04	0.26
cg23071864	chr10	3171130	-	*PFKP*	Body	4.33E-04	0.14
cg16952272	chr9	15422596	+	*SNAPC3*	TSS200	4.33E-04	-0.29
cg10705570	chr15	20498417	-			4.40E-04	0.25
cg11590508	chr16	31883967	+	*ZNF267*	TSS1500	4.40E-04	0.18
cg19782158	chr8	36641197	-	*KCNU1*	TSS1500	4.49E-04	0.18
cg27184249	chr1	38010230	-	*SNIP1*	Body	4.91E-04	0.20
cg08876558	chr11	59394462	-	*AP000442.1*		4.91E-04	0.31

cg11009716	chr3	52888489	-		Body	5.00E-04	0.24
cg17380870	chr3	195808602	-	*TFRC*	5'UTR	5.00E-04	-0.21
cg08173730	chr3	72144570	+	*RP11-398A8.3*	Body	5.00E-04	0.18
Significant Probes for low vs high Soil (< or ≥0.51μg/m^3)							
cg04670072	chr7	45243745	+			4.20E-03	0.30

In the CARE-PF cohort, there were no DMPs that passed the FDR of 0.05, however **Table 13** displays the top 10 CpGs for each pollutant for reference.

Table 13 – Significant differentially-methylated probes between low vs high $PM_{2.5}$ and constituents in the 5-yrs pre-sampling in CARE-PF cohort-alone adjusted models. Analyses adjusted for age at diagnosis, sex, smoking history, race, and IPF vs non-IPF diagnosis.

CpG	Chr	Position	Strand	Gene	UCSC Gene Region Feature	Adjusted P-value	logFC
Top 10 Probes for low vs high $PM_{2.5}$ (< or ≥5.07μg/m^3)							
cg15425530	chr17	2415361	+	*METTL16*	TSS200	0.91	0.19
cg11132921	chr8	66560891	+	*MTFR1*	5'UTR	0.91	-0.28
cg11495351	chr10	98110991	+		Body	0.91	0.24
cg23882683	chr14	90975893	+			0.91	-0.25
cg13567450	chr21	41876012	-		Body	0.91	0.17
cg13228355	chr20	53223230	+		Body	0.91	-0.26
cg15695181	chr4	7071161	-	*GRPEL1*	TSS1500	0.91	0.26
cg09049982	chr20	32950073	+	*ITCH*	TSS1500	0.91	0.14
cg23959518	chr6	110011686	+	*FIG4; AKD1*	5'UTR; TSS1500	0.91	0.15
cg04087237	chr2	152955765	+	*CACNB4; AC079790.2*	TSS1500; TSS200	0.91	-0.21
Top 10 Probes for low vs high SO_4^{2-} (< or ≥0.42μg/m^3)							
cg10974128	chr10	93393509	-	*PPP1R3C*	TSS1500	0.26	-0.28
cg13389502	chr17	1961440	-		Body	0.31	-0.31
cg04146405	chr20	60953667	+			0.31	0.15
cg12068244	chr4	130455581	-			0.64	0.28
cg01856455	chr22	22289520	-	*PPM1F*	Body	0.68	0.12
cg11376470	chr1	57046577	-	*PPAP2B*	TSS1500	0.76	0.13
cg05934012	chr1	1533852	-	*C1orf233*		0.76	0.17
cg02948125	chr17	7560317	+	*ATP1B2*	3'UTR	0.76	-0.24
cg02696415	chr4	40994566	+		Body	0.76	-0.25
cg02955268	chr19	2729359	-	*AC006538.1*		0.76	0.14
Top 10 Probes for low vs high NO_3^- (< or ≥0.31μg/m^3)							
cg16793483	chr5	179751165	+	*GFPT2*	Body	0.39	-0.34
cg19648923	chr6	31543266	-	*TNF*	TSS200	0.39	-0.35
cg15236240	chr10	79781362	-	*POLR3A*	Body	0.39	0.15
cg01856455	chr22	22289520	-	*PPM1F*	Body	0.39	0.13
cg03634145	chr6	484044	-			0.39	-0.29
cg09525260	chr15	84523205	+		Body	0.39	-0.27
cg23882683	chr14	90975893	+			0.39	-0.26
cg02862835	chr12	110011242	+	*MMAB; MVK*	1stExon; TSS1500	0.39	0.15
cg01001508	chr6	34090014	-		Body; 5'UTR	0.39	-0.14
cg11753051	chr22	21311214	+	*XXbac-B135H6.15*	TSS200	0.39	-0.20
Top 10 Probes for low vs high NH_4^+ (< or ≥0.11μg/m^3)							
cg11817057	chr6	29924202	-			0.21	0.21

cg14008399	chr1	160012129	+	*RP11-226L15.4*	Body	0.28	0.17
cg17775944	chr14	96152671	+	*TCL1B; RP11-1070N10.6*	TSS200	0.28	-0.25
cg05864191	chr14	76448933	-	*TGFB3*	TSS1500	0.28	-0.18
cg02205657	chr10	115480704	-	*CASP7*	Body	0.53	0.18
cg21568453	chr2	58284130	-		Body; 5'UTR	0.56	0.31
cg11753929	chr6	43484254	+	*POLR1C*	TSS1500; Body	0.56	-0.18
cg07773095	chr21	33784558	+	*FAM176C*	TSS200	0.56	0.38
cg15676719	chr1	167598521	+	*RP3-455J7.4; RCSD1*	TSS1500	0.56	-0.19
cg13647486	chr5	138274641	-			0.56	0.20
Top 10 Probes for low vs high BC (< or ≥0.31μg/m^3)							
cg16010827	chr1	200708413	+	*CAMSAP2*	TSS1500	0.67	0.20
cg23983436	chr11	842963	-	*TSPAN4*	TSS1500; 1stExon; 5'UTR	0.67	0.39
cg15175266	chr5	22854704	+	*CDH12*	TSS1500	0.67	0.27
cg10623290	chr1	1912185	-	*C1orf222*	Body	0.67	0.27
cg15885814	chr21	43204246	-			0.67	-0.30
cg18750249	chr2	7499111	-			0.67	-0.28
cg16021998	chr16	19896851	+	*GPRC5B*	TSS1500; 1stExon	0.67	-0.49
cg12662206	chr5	121412927	-	*LOX*	TSS200; Body	0.67	-0.27
cg13209613	chr17	44849789	-		Body	0.67	-0.20
cg13802490	chr1	3649890	-	*TP73*	3'UTR	0.67	0.13
Top 10 Probes for low vs high OM (< or ≥2.03μg/m^3)							
cg16010827	chr1	200708413	+	*CAMSAP2*	TSS1500	0.32	0.20
cg17043230	chr12	3427729	+	*RP5-1063M23.2*		0.72	-0.41
cg15885814	chr21	43204246	-			0.72	-0.31
cg15175266	chr5	22854704	+	*CDH12*	TSS1500	0.72	0.26
cg23879460	chr3	10806569	-	*AC018495.2*	TSS1500	0.72	0.15
cg23983436	chr11	842963	-	*TSPAN4; POLR2L*	TSS1500; 1stExon; 5'UTR	0.72	0.38
cg10623290	chr1	1912185	-	*C1orf222*	Body	0.72	0.27
cg07636653	chr11	91834809	-			0.72	0.19
cg14074328	chr11	84646972	-	*AP000857.3*	Body	0.72	0.16
cg03105222	chr20	30778399	-	*TSPY26P*	TSS1500	0.72	-0.22
Top 10 Probes for low vs high SS (< or ≥0.28μg/m^3)							
cg16713743	chr21	34397135	+	*OLIG2; AP000282.2*	TSS1500	0.26	0.28
cg09578113	chr7	47420152	-	*TNS3*	Body	0.30	0.14
cg02695704	chr16	1250494	+		Body	0.30	0.35
cg09291095	chr5	15794192	-	*FBXL7*	Body; 5'UTR	0.38	-0.28
cg20982606	chr12	109535391	+	*UNG*	TSS1500; TSS200	0.85	-0.22
cg09195319	chr1	213031561	-	*FLVCR1; FLVCR1-AS1*	TSS200	1.00	0.36
cg07384708	chr19	372718	+	*THEG*	Body	1.00	-0.31
cg02405461	chr3	112740175	+	*C3orf17; RP11-572M11.44*		1.00	0.16
cg00669777	chr8	95565739	+	*RP11-267M23.4; KIAA1429*	TSS200	1.00	-0.12
cg06269372	chr10	79321319	+		Body	1.00	0.12
Top 10 Probes for low vs high Soil (< or ≥0.17μg/m^3)							
cg13241977	chr19	31745676	-			0.15	-0.20
cg03920301	chr11	79167349	-			0.15	0.27
cg14849022	chr4	141181563	+	*SCOC*	5'UTR	0.15	-0.35
cg25772769	chr9	19409284	+	*ACER2*	Body	0.15	0.21
cg19243826	chr19	14584926	-		Body	0.15	-0.20
cg27058239	chr3	197816796	+			0.15	-0.52
cg19473377	chr2	38891148	+			0.15	-0.28

cg10321393	chr1	15905387	-	*DNAJC16*	Body	0.15	0.21
cg26848594	chr1	84843639	-	*UOX*	Body	0.15	0.14
cg22227719	chr7	128337339	+	*RN5S243; RN5S242*		0.15	-0.24
cg13241977	chr19	31745676	-			0.15	-0.20

4.2.3.7 Meta-Analysis of Individual Cohort Results

Results from individual cohort DMP analyses for Simmons and CARE-PF were combined in meta-analyses that were performed using METAL.(Willer et al., 2010) **Table 14** demonstrates the top differentially-methylated CpGs on meta-analysis of the two cohorts for all constituents, with the top 50 listed for analyses with >50 significant CpGs. For 5-year pre-sampling exposures to $PM_{2.5}$, there were 65 DMPs, 296 for SO_4^{2-},0 for NO_3^-, 0 for NH_4^+ (although 3 CpGs were marginal with FDR-adjusted p-values <0.1), 0 for BC, 0 for OM (although 1 CpG was marginal with FDR-adjusted p-value <0.1), 363 for SS; and 0 for Soil. There were shared top CpGs with $PM_{2.5}$ or other constituents in 23/50 (46%) of the $PM_{2.5}$ analysis, 22/50 (44%) for SO_4^{2-}, and 8/50 (16%) for SS, indicating multiple shared pathways of involvement for $PM_{2.5}$ and several of its constituents. In the 5-yr pre-sampling period for low vs high $PM_{2.5}$, SO_4^{2-}, and SS, the top CpG was again cg25354716, which is annotated to body of the *CRTAP* gene. Another finding was the significance of cg22186557 as 47th and 67th most differentially methylated probe for the $PM_{2.5}$ and SO_4^{2-} analyses. Cg22186557 is annotated to the body of the *MUC5B* gene, which is a gene for which the promoter SNP rs37505950 has been identified as one of the most important variants influencing the phenotype and progression of IPF and other fILDs.(Peljto et al., 2013; Y. Zhang, 2017; Y. Zhang, Noth, Garcia, & Kaminski, 2011) Although the direction of effect of higher levels of pollutants on CpG loci was consistent for all the top 50 CpGs highlighted in **Table 13**, the heterogeneity as indicated by the I^2 was substantial (range 0-91%).

Table 14 – Meta-analysis of significant differentially-methylated probes in the two cohorts between low vs high $PM_{2.5}$ and constituents in the 5-yrs pre-sampling. Dichotomized cut-point used in each cohort was based on the median pollutant exposure in that cohort. Reporting in table limited to top 50 CpGs for pollutants with >50 signficantly differentially-methylated probes. Analyses adjusted for age at diagnosis, sex, smoking history, race, and IPF vs non-IPF diagnosis. CpGs that are **bolded** reflect peak differentially methylated probes that are shared between more than one pollutant.

CpG	Chr	Position	Strand	Gene	UCSC Gene Region Feature	Meta-Analysis Adjusted P-value	Meta-Analysis Direction of Effect	I^2 (%)
Significant Probes for low vs high $PM_{2.5}$ (< or ≥12.07μg/m³ in Simmons; < or ≥5.07μg/m³ in CARE-PF)								
cg25354716	chr3	33159475	-	*CRTAP*	Body	0.0009	++	91
cg07637658	chr19	45887042	-		Body	0.0014	++	90
cg06686702	chr19	46684843	-			0.0025	--	70
cg27457921	chr17	43976811	-	*MAPT*	5'UTR	0.0036	++	60
cg14201314	chr16	34984295	-	*RP11-352B15.1; RN5S415*	Body	0.0037	++	69
cg08543976	chr17	19699755	+	*ULK2*	Body	0.0059	++	61
cg13000555	chr12	104460455	+		Body	0.0059	++	52
cg19782158	chr8	36641197	-	*KCNU1*	TSS1500	0.0059	++	90
cg13826499	chr19	3235128	+	*CELF5*	Body	0.0072	++	0
cg24277705	chr10	89603720	+	*CFL1P1*	Body	0.0118	++	75
cg18924184	chr14	107257615	+			0.0122	++	87
cg24713063	chr20	35459146	-	*SOGA1*	Body	0.0122	++	44
cg08811509	chr9	95640366	-	*ZNF484; ANKRD19P*	TSS200	0.0170	--	83
cg15347108	chr12	25340920	+	*CASC1*	Body	0.0170	++	0
cg03846240	chr6	35723021	-			0.0170	++	84
cg26060971	chr3	52407402	+		Body	0.0197	--	84
cg21261121	chr20	43380263	-	*RP4-781B1.2*	3'UTR	0.0202	++	72
cg19116668	chr7	99932089	-		Body	0.0202	--	76
cg15635761	chr13	45411234	+			0.0202	++	60
cg26207631	chr1	247900912	-	*OR14K1; RP11-634B7.4*		0.0202	++	0
cg01123250	chr2	210673545	+		Body	0.0226	++	0
cg09782746	chr12	132553998	+		Body	0.0226	++	48
cg21400851	chr17	73056681	-		Body	0.0226	++	64
cg15967169	chr6	30167836	+	*TRIM26*	5'UTR	0.0255	++	82
cg08109808	chr17	78029642	+		Body	0.0255	++	75
cg20696644	chr3	46118582	-			0.0255	++	57
cg22487204	chr1	156435421	-	*MEF2D*	3'UTR	0.0255	++	80
cg00519320	chr16	75020429	+	*WDR59*	TSS1500	0.0260	++	0
cg17598704	chr4	148885518	-		Body	0.0260	++	85
cg19837214	chr12	133402551	-	*GOLGA3; RP11-46H11.10*	5'UTR	0.0260	++	28
cg05956803	chr5	137666464	+		Body	0.0272	++	0
cg13821031	chr9	116035939	+	*CDC26*	5'UTR	0.0312	++	84
cg05441198	chr13	61237593	-			0.0342	++	0
cg05844625	chr6	29976071	-	*HLA-J; ZNRD1-AS1*	Body	0.0342	++	0
cg08954646	chr6	31495960	-	*MCCD1*	TSS1500	0.0370	++	84
cg06515159	chr21	34400659	+	*OLIG2; AP000282.2*	3'UTR	0.0370	--	0
cg00281333	chr1	40203613	-	*PPIE*	TSS1500	0.0370	++	0

cg23322811	chr3	72259457	+	*RP11-433A10.2*		0.0370	++	19
cg15578948	chr1	1689012	+	*NADK*	Body	0.0370	++	0
cg08913530	chr10	96988478	+	*C10orf129; RP11-310E22.4*	Body	0.0370	++	55
cg08478283	chr1	117076803	-	*RP4-655J12.5*	Body	0.0370	++	0
cg01650137	chr4	153534702	+			0.0370	--	0
cg01381374	chr7	93474158	-	*GNGT1*		0.0371	++	24
cg10297223	chr3	148414649	-	*AGTR1*	TSS1500	0.0378	++	77
cg05509352	chr5	142092851	-			0.0378	--	51
cg09502141	chr9	4700338	-		Body	0.0393	++	84
cg22186557	chr11	1260163	+	*MUC5B*	Body	0.0407	++	67
cg11496113	chr5	34627766	-			0.0410	++	87
cg25348895	chr1	158434995	+	*OR10K1*	TSS1500	0.0424	++	0
cg01304461	chr12	10103711	+	*AC091814.3; CLEC12A*	TSS1500	0.0431	--	82
Significant Probes for low vs high SO_4^{2-} (< or ≥4.17μg/m^3 in Simmons; < or ≥0.42μg/m^3 in CARE-PF)								
cg25354716	chr3	33159475	-	*CRTAP*	Body	3.33E-07	++	84
cg07637658	chr19	45887042	-		Body	1.41E-06	++	83
cg14201314	chr16	34984295	-	*RP11-352B15.1; RN5S415*	Body	2.81E-04	++	0
cg06686702	chr19	46684843	-			5.98E-04	--	70
cg27457921	chr17	43976811	-	*MAPT*	5'UTR	1.42E-03	++	23
cg01901468	chr8	128250972	-			1.55E-03	++	50
cg26508775	chr20	55050494	-	*snoU13*	Body	1.75E-03	++	72
cg03270074	chr2	95870287	-			1.75E-03	++	78
cg12667999	chr17	47901743	-	*KAT7*	Body	1.75E-03	++	79
cg01123250	chr2	210673545	+		Body;Body	2.45E-03	++	0
cg24699002	chr3	45632441	+	*AC099539.1; LIMD1*		2.64E-03	++	0
cg17598704	chr4	148885518	-		Body	3.11E-03	++	62
cg24277705	chr10	89603720	+	*CFL1P1*	Body	3.11E-03	++	80
cg05441198	chr13	61237593	-			3.43E-03	++	0
cg26732754	chr10	134255605	+	*RP11-432J24.3*		3.81E-03	++	0
cg19782158	chr8	36641197	-	*KCNU1*	TSS1500	3.81E-03	++	85
cg23657686	chr11	10565836	-	*MRVI1-AS1*		4.17E-03	++	0
cg08811509	chr9	95640366	-	*ZNF484; ANKRD19P*	TSS200	4.46E-03	--	90
cg19116668	chr7	99932089	-		Body	4.46E-03	--	69
cg25881328	chr8	1809920	+		Body	4.46E-03	++	39
cg05956803	chr5	137666464	+		Body	4.46E-03	++	0
cg08811958	chr20	43349788	+	*RP11-445H22.4*	Body	4.46E-03	++	0
cg16854097	chr16	81643846	-	*CMIP*	TSS1500; Body	4.46E-03	++	23
cg00519320	chr16	75020429	+	*WDR59*	TSS1500	4.46E-03	++	0
cg12578536	chr5	43003251	+			4.46E-03	++	90
cg17671846	chr21	45341224	+	*U6; AGPAT3*	5'UTR	4.46E-03	++	0
cg16174683	chr11	31195609	+	*DCDC1*		4.65E-03	++	0
cg00281333	chr1	40203613	-	*PPIE*	TSS1500	5.72E-03	++	0
cg22637307	chr2	68590998	-	*AC015969.3;PLEK*	TSS1500	5.90E-03	++	72
cg21261121	chr20	43380263	-	*RP4-781B1.2*	3'UTR	6.00E-03	++	75
cg11135772	chr14	55526209	+	*MAPK1IP1L*	5'UTR	6.06E-03	++	0
cg00919534	chr3	38474585	-			6.06E-03	++	54
cg17753475	chr2	180477963	-		Body	6.06E-03	++	0
cg12894234	chr5	149125770	-	*PPARGC1B*	Body	6.06E-03	++	80

cg05509352	chr5	142092851	-			6.94E-03	--	42
cg27395226	chr7	104994322	-		Body	8.00E-03	--	0
cg11496113	chr5	34627766	-			8.00E-03	++	85
cg18924184	chr14	107257615	+			8.00E-03	++	72
cg11021810	chr14	89720620	-		Body	8.00E-03	++	0
cg07690796	chr12	125301012	+	*SCARB1*	Body	8.00E-03	++	0
cg19837214	chr12	133402551	-	*GOLGA3; RP11-46H11.10*	5'UTR	8.15E-03	++	63
cg12120326	chr22	30237309	-			8.15E-03	++	76
cg08543976	chr17	19699755	+	*ULK2*	Body	8.28E-03	++	69
cg26634055	chr16	72042068	+	*DHODH*	TSS1500	8.28E-03	++	82
cg06576783	chr6	53572893	-			8.53E-03	++	78
cg10600967	chr4	152196952	-	*PRSS48*	TSS1500	8.53E-03	++	0
cg15967169	chr6	30167836	+	*TRIM26*	5'UTR	8.53E-03	++	37
cg09489435	chr15	89561491	+			8.53E-03	++	62
cg02224047	chr7	101883734	-		Body	8.63E-03	++	68
cg13821031	chr9	116035939	+	*CDC26*	5'UTR	8.87E-03	++	82
Significant Probes for low vs high SS (< or ≥0.19μg/m³ in Simmons; < or ≥0.28μg/m³ in CARE-PF)								
cg25354716	chr3	33159475	-	*CRTAP*	Body	1.76E-06	++	86
cg23562228	chr19	49657666	-	*HRC*	1stExon	5.44E-05	++	42
cg00934735	chr5	180582587	-	*OR2V2*	1stExon	5.44E-05	++	84
cg07637658	chr19	45887042	-		Body	5.69E-05	++	92
cg26784412	chr6	33088710	-		Body	6.97E-05	++	0
cg11342415	chr10	1034284	+	*GTPBP4; AL359878.1*	TSS200	5.04E-04	--	54
cg01339959	chr2	184125517	-			6.71E-04	++	77
cg11590508	chr16	31883967	+	*ZNF267*	TSS1500	7.86E-04	++	60
cg16196201	chr6	52615988	+		Body	8.86E-04	++	0
cg01764438	chr16	9170298	+			1.36E-03	++	63
cg14238081	chr14	65438743	+	*RAB15*	1stExon	1.36E-03	--	0
cg07835232	chr7	140302680	+	*DENND2A*	TSS1500	1.36E-03	++	90
cg12697603	chr12	122751251	-	*VPS33A; RP11-512M8.5*	TSS200	1.99E-03	--	63
cg13374897	chr20	30613140	-		Body	2.26E-03	++	60
cg00281333	chr1	40203613	-	*PPIE*	TSS1500	2.34E-03	++	81
cg06197074	chr1	51788770	-	*TTC39A*	TSS1500; Body	2.45E-03	++	7
cg11412935	chr1	92950086	-	*GFI1*	5'UTR; TSS1500	2.45E-03	--	68
cg14445171	chr9	131419910	+	*WDR34*	TSS1500	2.45E-03	++	78
cg11284631	chr5	38870649	+		Body	2.45E-03	++	73
cg22932101	chr14	71815816	+			2.45E-03	++	84
cg13000555	chr12	104460455	+		Body	2.45E-03	++	90
cg16952272	chr9	15422596	+	*SNAPC3*	TSS200	3.60E-03	--	80
cg15091333	chr7	100908800	+			3.60E-03	++	86
cg06880365	chr19	36870044	+	*ZFP14*		3.60E-03	--	81
cg24800930	chr14	24454176	-	*DHRS4L2*	5'UTR	3.60E-03	++	83
cg20577728	chr15	23034331	+	*NIPA2*	5'UTR; 1stExon	3.60E-03	--	83
cg26406571	chr11	2422425	+	*TSSC4*	TSS1500	4.22E-03	--	0
cg00058786	chr12	49365465	+	*WNT10B*	5'UTR; 1stExon	4.61E-03	++	0
cg17214388	chr19	3431061	-		Body	4.61E-03	++	86
cg12198254	chr3	11521788	+		Body	4.61E-03	++	90

cg01123250	chr2	210673545	+		Body	4.61E-03	++	85
cg12406406	chr2	139654951	+		Body	4.68E-03	++	91
cg23679920	chr12	42376057	-			4.74E-03	++	87
cg22488975	chr1	184724009	+	*EDEM3*	1stExon;5'UTR	4.74E-03	--	0
cg01145910	chr1	5729401	+	*RP11-154H17.1*		4.74E-03	++	77
cg21261121	chr20	43380263	-	*RP4-781B1.2*	3'UTR	4.74E-03	++	88
cg17598704	chr4	148885518	-		Body	4.74E-03	++	77
cg27184249	chr1	38010230	-	*SNIP1*	Body	4.74E-03	++	82
cg11496113	chr5	34627766	-			4.76E-03	++	88
cg21091378	chr19	11215861	-		Body	4.78E-03	++	21
cg15613100	chr5	72804620	+	*AC099522.1*		4.78E-03	++	83
cg03860054	chr16	55691102	-	*SLC6A2*	Body	4.78E-03	++	90
cg24048921	chr22	41074275	-	*MCHR1*	TSS1500	4.78E-03	++	74
cg12253859	chr10	96748928	-	*CYP2C9*	3'UTR	4.78E-03	++	90
cg13701509	chr5	170833685	-	*NPM1*	Body; 3'UTR	4.93E-03	++	81
cg19115530	chr1	150334619	+	*RPRD2*		5.71E-03	++	74
cg02568557	chr11	129288930	-		Body	5.79E-03	++	74
cg04281845	chr10	10215481	-	*TCEB1P3*		5.87E-03	++	81
cg27223727	chr1	161228496	-	*PCP4L1*	TSS200	5.98E-03	--	6
cg21393713	chr7	69064801	+	*AUTS2*	1stExon	6.31E-03	--	80

A Manhattan plot and QQ-plot of the meta-analysis results for the 5-year pre-sampling DMP analysis for the Simmons and CARE-PF cohorts is shown in **Figure 18**. The genomic inflation for the analysis is high with a λ=1.46, although this finding is in keeping with high genomic inflation demonstrated in other epigenome-wide association studies.(van Iterson, van Zwet, & Heijmans, 2017)

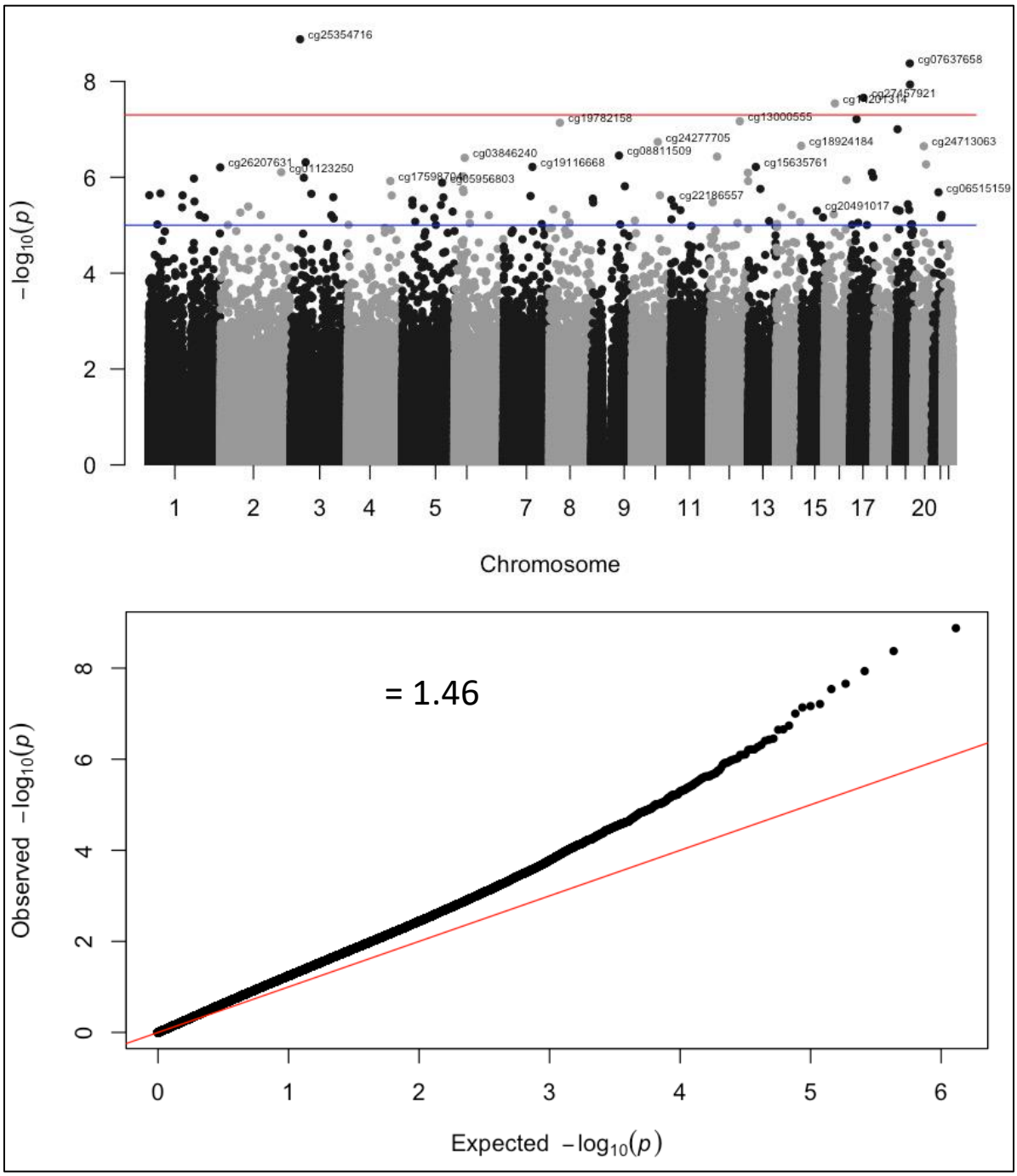


Figure 18 – Manhattan plot and QQ plot for $PM_{2.5}$ in 5-yr pre-sampling meta-analysis of cohort-specific DMP analyses. The top 20 most highly significant CpGs are annotated on the Manhattan plot in the top panel. Lambda score (i.e. genomic inflation) is reported on the QQ-plot on the bottom panel.

4.2.3.8 Gene Set Enrichment Analysis (GSEA)

Using the *missMethyl* package in R, we performed GSEA for each cohort-specific DMP analysis where significant CpGs were identified. In the Simmons cohort, this analysis was performed for $PM_{2.5}$, SO_4^{2-}, NO_3^-, NH_4^+, and SS in the 5-years pre-sampling, as well as $PM_{2.5}$, SO_4^{2-}, NH_4^+, BC, and OM in the3-months pre-sampling. In the CARE-PF cohort, only OM in the 3-months pre-sampling had significant CpGs, so this was also analyzed. None of the GSEAs identified significantly-enriched GO terms at a FDR of <0.05 using the "gometh" function in the *missMethyl* package for any of the above analyses.

4.2.3.9 Clinical Outcomes Analysis of cg25354716

Confirming the findings from the DMP analysis, we found that higher exposures to $PM_{2.5}$, SO_4^{2-}, NO_3^-, NH_4^+, OM, SS, and soil were associated with lower β-value at cg25354716 (**Table 15**, **Figure 19**). Multivariable linear regression models adjusted for age at diagnosis, sex, smoking history, race, IPF vs non-IPF diagnosis, and disadvantage score demonstrated that for each 1μg/m^3 increase in $PM_{2.5}$, there is a 0.01 unit decrease in cg25354716 β-value (95% CI-0.013 to -0.007, p<0.001). There were no significant associations between pollutant exposures and β-value at cg25354716 in the CARE-PF cohort, although the direction and magnitude of effect was consistent for all pollutants except NH_4^+ with SO_4^{2-} and SS nearing significance.

Table 15 – Association of pollutants with β-value at cg25454716. Results of models adjusting for age at diagnosis, sex, smoking history, race, IPF vs non-IPF diagnosis, and neighborhood disadvantage score are reported. Significant results are **bolded.**

Pollutant	Cohort	Linear Model β-value	95% CI	p-value
$PM_{2.5}$	Simmons	**-0.010**	**-0.013 to -0.007**	**<0.001**
$PM_{2.5}$	CARE-PF	-0.001	-0.007 to 0.006	0.86
SO_4^{2-}	Simmons	**-0.023**	**-0.028 to -0.018**	**<0.001**
SO_4^{2-}	CARE-PF	-0.054	-0.121 to 0.014	0.12
NO_3^-	Simmons	**-0.025**	**-0.044 to -0.006**	**0.01**
NO_3^-	CARE-PF	-0.018	-0.066 to 0.031	0.47
NH_4^+	Simmons	**-0.055**	**-0.068 to -0.042**	**<0.001**

NH_4^+	CARE-PF	0.047	-0.106 to 0.200	0.55
BC	Simmons	-0.020	-0.060 to 0.019	0.31
BC	CARE-PF	-0.018	-0.054 to 0.018	0.33
OM	Simmons	**-0.019**	**-0.028 to -0.010**	**<0.001**
OM	CARE-PF	-0.003	-0.008 to 0.002	0.28
SS	Simmons	**-0.198**	**-0.260 to -0.137**	**<0.001**
SS	CARE-PF	-0.048	-0.105 to 0.008	0.09
Soil	Simmons	**-0.070**	**-0.121 to -0.019**	**0.007**
Soil	CARE-PF	-0.033	-0.115 to 0.048	0.42

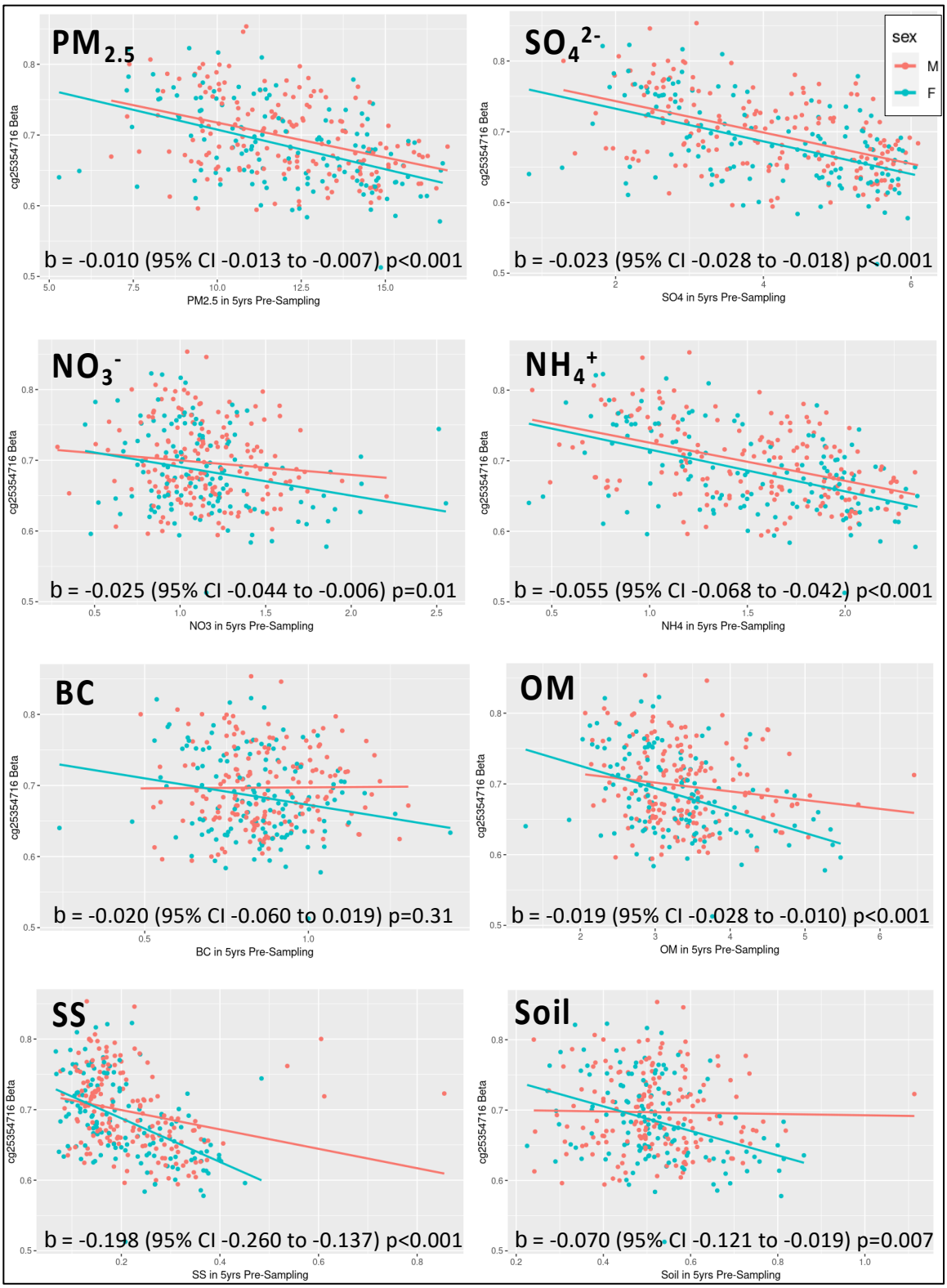


Figure 19 – Scatterplots demonstrating correlation of β-value at cg25354716 with pollutant exposures in 5 years pre-sampling (in μg/m^3). β reported on panel reflects the change in cg25354716 β-value per 1μg/m^3 increase in a pollutant in models adjusted for age at diagnosis, sex, smoking history, race, IPF vs non-IPF diagnosis, and disadvantage score. Higher exposures to $PM_{2.5}$, SO_4^{2-}, NO_3^-, NH_4^+, OM, SS, and soil were all associated with lower β-value at cg25354716.

We then evaluated the association of cg25354716 with mortality, baseline FVC, and baseline D_LCO. Cox proportional hazards models adjusted for age at diagnosis, sex, smoking history, race, IPF vs non-IPF diagnosis, and neighborhood disadvantage score in the Simmons cohort demonstrated that increased β-value at cg25354716 was associated with lower mortality (HR=0.03, 95% CI 0.003-0.38, p=0.006) with spline model of effect shown in **Figure 20**. This HR can be interpreted as, if the β-value at cg25354716 goes from completely unmethylated (β-value=0) to completely methylated (β-value=1) that the risk of mortality is 3% of the risk in the completely unmethylated patient. Given that the β-values at cg25354716 range from 0.51-0.85, we do not expect patients to go through this entire range of methylation. There was no significant association of cg25354716 β-value with mortality in the CARE-PF cohort (adjusted model HR=0.32, 95% CI 0.0004-255.27, p=0.74).

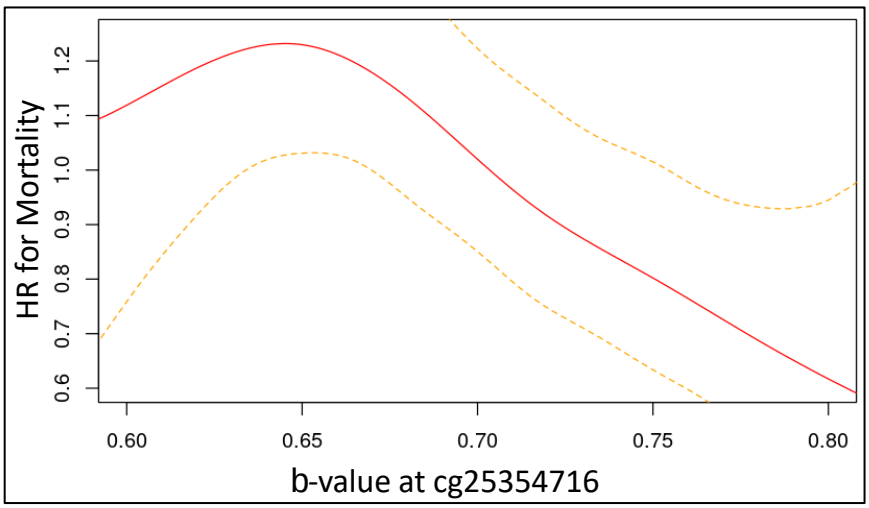


Figure 20 – Spline of adjusted model for association of cg25354716 β-value with mortality. Hazard ratio (HR) for mortality shown on the y-axis versus cg25354716 β-value on the x-axis, demonstrating that as β-value at this locus increases, the HR for mortality decreases. Models adjusted for age at diagnosis, sex, smoking history, race, IPF vs non-IPF diagnosis, and neighborhood disadvantage score.

Multivariable linear regression models adjusting for the same covariates demonstrated that higher methylation at cg25354716 was associated with lower baseline FVC and D_LCO in the Simmons cohort (**Figure 21**). Models indicate that if cg25354716 went from completely unmethylated (β-value=0) to completely methylated (β-value=1) that the baseline FVC would be 87% predicted units higher (95% CI 51-123%, p<0.001) and the baseline D_LCO would be 71% predicted units higher (95% CI 36-107%, p<0.001). There was no significant association between cg25354716 β-value and baseline FVC or D_LCO in the CARE-PF cohort (FVC β = -18%, 95% CI -88% to -52%, p=0.61; D_LCO β = 2%, 95% CI -65% to 70%, p=0.95).

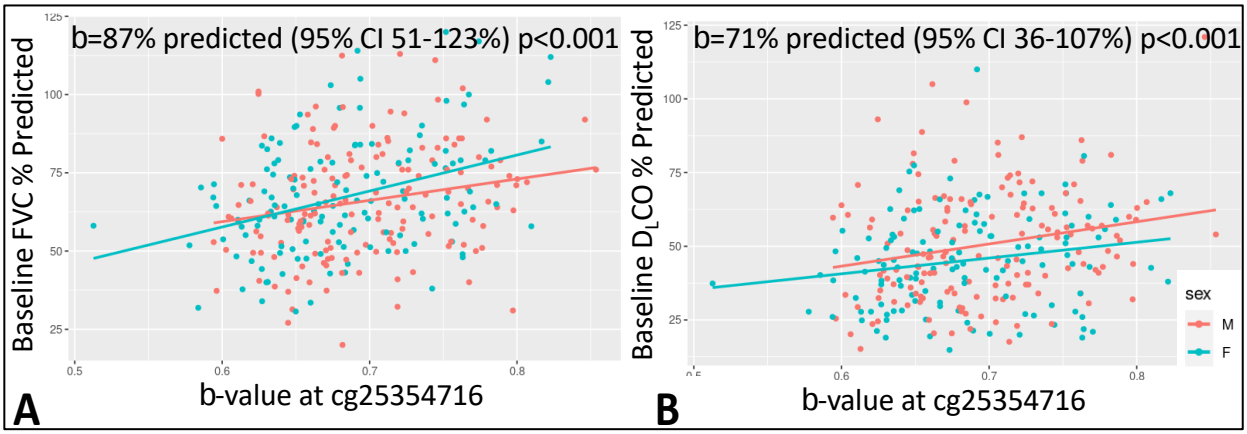


Figure 21 – Scatterplots of baseline lung function vs β-value at cg25354716. β reported above each graph reflects the effect estimate from adjusted multivariable regression models evaluating the association of β-value at cg25354716 with baseline FVC (panel A) or D_LCO (panel B). Models adjusted for age at diagnosis, sex, smoking history, race, IPF vs non-IPF diagnosis, and neighborhood disadvantage score.

Given the significant associations between pollutants and β-value at cg25354716 as well as between β-value at cg25354716 and clinical outcomes in the Simmons cohort, we met the criteria necessary to perform mediation analyses to determine the proportion of the pollutant-mortality association that is mediated by cg25354716 methylation. **Table 16** demonstrates the mediation proportion for pollutants where the 5-year pre-sampling pollutant exposure-mortality association was significant in Cox proportional hazards models adjusted for age at diagnosis, sex,

smoking history, race, IPF vs non-IPF diagnosis, and neighborhood disadvantage score. Given the lack of significant associations in the CARE-PF cohort, only Simmons mediation fractions are reported.

Table 16 – Mediation proportion by cg25354716 of relationship between pollutants and mortality in Simmons cohort. Results from Cox proportional hazard models with adjustments for age at diagnosis, sex, smoking history, race, IPF vs non-IPF diagnosis, and neighborhood disadvantage score. Significant effects are **bolded.**

Model	HR	95% CI	p-value	Mediation Proportion
$PM_{2.5}$ adjusted model without cg25354716	**1.11**	**1.05-1.18**	**<0.001**	0.25
$PM_{2.5}$ adjusted model with cg25354716	**1.08**	**1.01-1.16**	**0.02**	
SO_4^{2-} adjusted model without cg25354716	**1.24**	**1.11-1.38**	**<0.001**	0.19
SO_4^{2-} adjusted model with cg25354716	**1.19**	**1.04-1.35**	**0.009**	
NH_4^+ adjusted model without cg25354716	**1.68**	**1.24-2.26**	**<0.001**	0.25
NH_4^+ adjusted model with cg25354716	**1.47**	**1.05-2.06**	**0.02**	
SS adjusted model without cg25354716	**3.50**	**1.09-11.23**	**0.04**	0.40
SS adjusted model with cg25354716	2.12	0.57-7.83	0.26	

4.2.4 Discussion

This work demonstrates that $PM_{2.5}$ and constituent exposures are associated with alterations in locus-specific DNAm in patients with fILD. The highest numbers of significant DMPs were detected for the $PM_{2.5}$ constituents of SO_4^{2-}, NH_4^+, and sea salt (SS), moreso than total $PM_{2.5}$ mass alone. This implies that these constituents may have more mechanistic influence on DNAm than total $PM_{2.5}$, which is consistent with our clinical and global DNAm findings in Aim 2 and Aim 3.1, where these constituents were also shown to have the most impact on outcomes. These findings further support that it is critical to evaluate the impact of $PM_{2.5}$ composition on both clinical and molecular outcomes, as individual constituents can have highly variable associations with different mechanistic pathways.

Meta-analysis of results between Simmons and CARE-PF cohorts was necessary given the largely non-overlapping ranges of $PM_{2.5}$ and its constituents between the cohorts. This great

difference in exposures likely contributed to the high heterogeneity (I^2) seen in the meta-analysis results. Paired with the consistency in the direction of effect between the two cohorts for the top CpGs, this high heterogeneity indicates that the effect size of these pollutants on DNAm may vary at different exposure levels.

Differential methylation was noted in multiple analyses for cg25354716, which is annotated to the *CRTAP* gene. We further demonstrated that higher exposures to $PM_{2.5}$, SO_4^{2-}, NO_3^-, NH_4^+, OM, SS, and soil was associated with lower methylation at this locus and that high methylation at this locus was protective against mortality and lower baseline lung function in these patients. Subsequently, we demonstrated that cg25454716 mediates between 19-40% of the relationship between pollutants and mortality, indicating this as an important novel pathway for investigation of the relationship between pollution exposure and mortality in patients with fILD. This locus is of particular interest given its annotation to *CRTAP* (i.e. cartilage-associated protein), which is critical to the process of collagen chain trimerization and subsequently extracellular matrix (ECM) formation.("CRTAP Gene - GeneCards," 2022) ECM formation and regulation has long been implicated in the pathophysiology of fILDs, although no studies have previously identified *CRTAP* as a gene of interest in this disease.(Upagupta, Shimbori, Alsilmi, & Kolb, 2018) Autosomal recessive deficiency of *CRTAP* is associated with congenital osteogenesis imperfecta, which has been shown to result in excessive skeletal transforming growth factor β (TGF-β) activity and downstream signaling.(Grafe et al., 2014) This is relevant to patients with fILD, given the wealth of literature linking excessive TGF-β activity with the development and progression of pulmonary fibrosis.(Fernandez & Eickelberg, 2012) Future analyses should evaluate the impact of *CRTAP* and DNAm at cg25354716 on TGF-β signalling in *in vitro* cell lines relevant to fILDs,

such as fibroblasts. This is under consideration for subsequent analysis in Dr. Zhang's lab to complement this work.

Findings of significant differential methylation at cg22186557 in DMP meta-analysis of $PM_{2.5}$ and SO_4^{2-}, which is annotated to *MUC5B,* is also of particular interest to patients with fILD given the association of the *MUC5B* promoter SNP rs37505950 with fILD phenotypes and progression.(Peljto et al., 2013; Y. Zhang, 2017; Y. Zhang et al., 2011) Subsequent analyses will be performed to investigate the impact of this CpG on clinical outcomes, as was done for cg53254716. The significance of this CpG, however, highlights how pollution exposures may have impacts at well-established ILD-relevant loci, and sets the stage for future analyses that will investigate the impact of pollution exposures on DNAm at CpG loci within 1Mb of SNPs previously-established to be associated with fILD phenotypes.(Allen et al., 2017; Noth et al., 2013)

4.2.4.1 Limitations and Next Steps

This study reflects the first to evaluate the association of $PM_{2.5}$ with locus-specific DNAm changes in peripheral blood samples from patients with fILD, with the goal of identifying potential causal pathways whereby $PM_{2.5}$ and its constituents exert adverse clinical impacts on patients with fILD. While this work is highly novel, the present analyses are limited by several factors, some which will be addressed prior to publication of this work. For simplicity of initial analyses, dichotomized cut-points of $PM_{2.5}$ and constituent exposures were used, but future analyses will explore the association of continuous pollutants with locus-specific DNAm using these data. We plan to further expand these results by evaluating for differentially-methylated regions (DMRs) as well as DMPs within 1Mb of SNPs previously identified from genome wide association studies (GWAS) to be associated with the phenotype of IPF or other fILDs. This targeted evaluation has

previously been employed in DNAm analyses performed on samples from patients with IPF.(Borie et al., 2022) Additionally, these findings may be limited by the analysis of DNAm in patients with multiple forms of fILD, who may have some disease-intrinsic methylation differences. Although adjusting for the covariate of IPF vs non-IPF diagnosis should control for some of these disease-specific differences, there may still be disease-specific residual confounding. As such, future analyses will include subgroup analyses of patients with IPF and CTD-ILD, which are the two fILD diagnoses with the highest number of patients included in this analysis. We will also be performing a sensitivity analysis with adjustments for cell-type heterogeneity, as variations in blood cell composition may contribute to bias and some of the genomic inflation that was identified.(Houseman et al., 2012) Lastly, this analysis was performed in blood samples from patients with fILD. Although the primary tissue involved in patients with fILD is the lungs, findings in blood as still relevant given that this is the primary accessible tissue for molecular evaluation of these patients considering the high morbidity and mortality of lung biopsy.(Fisher et al., 2019) Future work in my postdoctoral fellowship will evaluate the concordance between peripheral blood DNAm patterns and lung tissue DNAm patterns in patients with fILD for whom lung explant tissue is available.

4.2.5 Conclusions

This is the first work to demonstrate associations of $PM_{2.5}$ and its constituents with altered DNAm patterns in patients with fILD. We identified a highly significant CpG, cg25354716, annotated to the *CRTAP* gene, as the top locus in multiple analyses, demonstrating this CpG to mediate a portion of the pollutant-mortality association. This reflects an important avenue for

future exploration, given the role of *CRTAP* in ECM formation and maintenance, which is critical to the pathogenesis of fILDs. These findings are relevant as they provide evidence for potential causality linking ambient pollution with molecular dysregulation in fILDs. Such findings are critical to determinations of causality, which are used to refine air quality regulations enforced by the U.S. Environmental Protection Agency (EPA) and other international regulatory bodies. Future analyses will expand on this work to solidify our molecular understanding of how airborne pollutants exert their adverse impacts in this vulnerable patient population.

4.3 Aim 3.3 – $PM_{2.5}$ Impacts on Telomere Length in fILDs

4.3.1 Introduction

Patients with idiopathic pulmonary fibrosis (IPF) have shorter telomeres than age-matched controls, even in patients without inherited telomere-related mutations.(Alder et al., 2008) Short telomeres are associated with more rapid progression and shorter survival in multiple forms of fILD, including IPF.(Courtwright & El-Chemaly, 2019) Higher exposures to particulate matter with a diameter ≤2.5μm ($PM_{2.5}$) is also associated with increased mortality in patients with fILD, but the mechanisms underlying this association remain unclear.(Goobie, Carlsten, Johannson, Marcoux, et al., 2022; Sesé et al., 2018)

Some studies have demonstrated that increased exposures to airborne pollutants are associated with shorter telomeres, although the data remains highly variable depending on the pollutant and population studied.(Miri et al., 2019) Few studies have evaluated the impact of air pollution on TL in older, potentially more vulnerable populations. One recent Spanish study did

not find any association between $PM_{2.5}$ exposures and quantitative polymerase chain reaction (qPCR)-measured TL in 280 patients with fILD, although this study was limited in being from a single center where $PM_{2.5}$ exposures were not matched to a specific residential location or anchored to patient-relevant date (e.g. date of blood sampling).(Shull, Planas-Cerezales, Lara Compte, Perona, & Molina-Molina, 2022) Given the limitations of the literature to date and the established relationships of $PM_{2.5}$ and telomere length (TL) with mortality, we postulated that $PM_{2.5}$ may induce telomere shortening and that shortened TL may mediate a portion of the $PM_{2.5}$-mortality association in patients with fILD.

4.3.2 Methods

4.3.2.1 Study Population and Clinical Data

Adult patients with a diagnosis of fILD (primarily IPF) enrolled in the Dorothy P. and Richard P. Simmons Center for ILD Registry at the University of Pittsburgh (UPitt) were eligible for enrollment. Only patients who had blood samples collected and DNA extracted at or near the time of registry enrollment, with subsequent whole genome sequencing (WGS) and TelSeq analysis were included. Patients with non-fibrotic forms of ILD and sarcoidosis were excluded.

Electronic health records and specialist ILD clinic visit documentation was used to obtain demographic, residential, and clinical information for all included patients. UPitt Health Record Research Request (R3) Service was used to extract additional Simmons cohort data on race and initial encounter dates.(Visweswaran et al., 2022) Baseline forced vital capacity (FVC) and diffusion capacity of the lung for carbon monoxide (D_LCO) was defined as the earliest measurements of these values within six months of registry enrollment.

Ethics approval was obtained from UPitt (STUDY20030223, STUDY19040326).

4.3.2.2 $PM_{2.5}$ and Constituent Component Exposure Estimation

$PM_{2.5}$ exposure matching was performed using satellite-derived hybrid models accessed via the Atmospheric Composition Analysis Group,("Atmospheric Composition Analysis Group » Surface PM2.5," 2020; Hammer et al., 2020; Van Donkelaar et al., 2019) whereby residential address at time of registry enrollment was converted into latitude and longitude coordinates, which where matched to the closest coordinates from the hybrid dataset. Average of monthly exposures to $PM_{2.5}$ and its constituents, sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic matter (OM), sea salt (SS), and soil, were estimated at each patient's most recent residential location for the short-term period of 3-months pre-sampling and the long-term period of 5-years pre-sampling. The granularity of this hybrid approach enables exposure estimates to be resolved to a geographic area of approximately 1.1km^2 using the *ncdf4* package in R.(Pierce, 2021)

4.3.2.3 Telomere Length (TL) Estimation

TL for patients in the Simmons cohort was estimated using a bioinformatic TelSeq approach from whole genome sequencing (WGS) data.(Z. Ding, Mangino, Aviv, Spector, & Durbin, 2014) WGS was performed on Simmons samples of DNA that was isolated from blood using the Illumnia HiSeq technology. TelSeq was then performed on the WGS data for each patient to estimate TL based on the average number of repeats of the telomeric nucleotide sequence (TTAGGG) for the DNA within each sample. These results were then age-corrected, resulting in an age-corrected TL estimate that was used as the outcome variable for statistical analyses evaluating the association of pollutants with TL.

4.3.2.4 Statistical Analysis

The association of $PM_{2.5}$ or constituent components with age-corrected TL was evaluated using multivariable linear regression in unadjusted models and models adjusted for sex, smoking history, and race. Multi-pollutant analysis using quantile-based g-computation with a linear additive approach was performed to determine which of the seven matched $PM_{2.5}$ constituents exert the most substantial effects on TL.(Zhao et al., 2022)

Cox proportional hazards models were used to evaluate associations of TL or pollutants with mortality, adjusting for sex, smoking, race, and age at fILD diagnosis (in pollutant-mortality models only). Spline models were constructed to evaluate for non-linearity in the relationship between TL and mortality. A *post-hoc* exploratory Cox analysis was performed whereby patients were grouped into low versus high pollutant exposures (based on median cohort exposures) and quartiles of age-corrected TL (with equal numbers of patients in each cohort) to explore whether pollutant exposures and TL interact to influence mortality.

Mediation analyses were performed for models where both the pollutant-TL relationship and the pollutant-mortality relationship were significant. The proportion of the pollutant-mortality relationship that was mediated by TL was determined using a traditional mediation framework where age-adjusted TL was sequentially added to significant pollutant-mortality models, from which the mediation proportion was calculated using the following formula:

$$Mediation\ proportion = 1 - (\frac{coefficient\ of\ model\ with\ pollutant + \%5mC}{coefficient\ of\ model\ with\ pollutant\ alone})$$

As a sensitivity analysis, a causal mediation framework was employed to evaluate the proportion of the pollutant-mortality relationship that was mediated by TL using the *mediation* package in R.(Tingley, Yamamoto, Hirose, Keele, & Princeton, 2014) Given the statistical constraints of the *mediation* package, Cox models were changed to accelerated failure time models.

Analyses were performed using R (version 4.2.1, www.r-project.org).

4.3.3 Results

4.3.3.1 Cohort Characteristics and Pollutant Exposures

Cohort characteristics are shown in **Table 14**. TL analysis was performed on 318 patients with fILD from the Simmons Center Registry, of which 212 (67%) were male, the median age at diagnosis was 69 (interquartile range 62-74), and 283 (89%) were White. Median $PM_{2.5}$ and constituent exposures were higher in the 5-year pre-sampling period compared to the 3-month period, although the top range of exposures were generally highest in the 3-month exposures reflecting shorter-term peaks.

Table 17 – Characteristics of patients with telomere length analysis performed.

Cohort Characteristics (n=318)	n (%) or median (IQR)
Sex	
Male	212 (67%)
Female	106 (33%)
Race	
White	283 (89%)
Black	7 (2%)
Indigenous	1 (0.3%)
Unknown	27 (8%)
Smoking History	
Never	94 (30%)
Former	205 (64%)
Current	8 (3%)
Unknown	11 (3%)
fILD Diagnostic Group	
Idiopathic pulmonary fibrosis (IPF)	303 (95%)
Connective tissue disease-ILD (CTD-ILD)	1 (0.3%)
Other idiopathic interstitial pneumonia (IIP)	2 (0.6%)
Unclassifiable or Other ILD	12 (4%)
Area deprivation index (ADI)	60 (42, 76)
Age at diagnosis (years)	67 (62, 74)
Baseline forced vital capacity (FVC) % pred	63 (51, 78)
Baseline diffusion capacity (D_LCO) % pred	45 (34, 59)
Time to censoring (years)	2.5 (1.0, 4.6)
Cause of censoring	
Death	211 (66%)
Lung Transplant	66 (21%)

Lost to follow-up or censored due to data extraction	41 (13%)
Pollutant exposures in 5 years pre-sampling (μg/m^3)	
Particulate matter with diameter ≤2.5μm ($PM_{2.5}$)	11.3 (9.7, 13.7)
Sulfate (SO_4^{2-})	3.8 (2.6, 5.0)
Nitrate (NO_3^-)	1.0 (0.9, 1.3)
Ammonium (NH_4^+)	1.5 (0.9, 1.8)
Black carbon (BC)	0.9 (0.7, 1.0)
Organic matter (OM)	3.3 (2.8, 3.7)
Sea salt (SS)	0.2 (0.1, 0.3)
Soil	0.5 (0.4, 0.6)
Pollutant exposures in 3 months pre-sampling (μg/m^3)	
Particulate matter with diameter ≤2.5μm ($PM_{2.5}$)	10.5 (8.8, 12.8)
Sulfate (SO_4^{2-})	3.0 (2.3, 4.0)
Nitrate (NO_3^-)	1.0 (0.6, 1.6)
Ammonium (NH_4^+)	1.3 (0.9, 1.8)
Black carbon (BC)	0.8 (0.7, 1.0)
Organic matter (OM)	3.2 (2.7, 3.8)
Sea salt (SS)	0.2 (0.1, 0.3)
Soil	0.5 (0.3, 0.6)

4.3.3.2 Pollutant Associations with Age-Corrected Telomere Length (TL)

Patients exposed to higher $PM_{2.5}$ in the 5-year pre-sampling period were found to have shorter age-corrected TL in both continuous and quartiled adjusted models (**Figure 19**). There was no significant association of $PM_{2.5}$ in 3-months pre-sampling with age-corrected TL. Results of linear models adjusted for sex, smoking history, and race of pollutant associations with age-corrected TL are shown in **Table 15**. Higher exposures to SO_4^{2-}, NO_3^-, and NH_4^+ in the 5-years pre-sampling as well as NO_3^- in the 3-months pre-sampling were associated with shorter age-corrected TL.

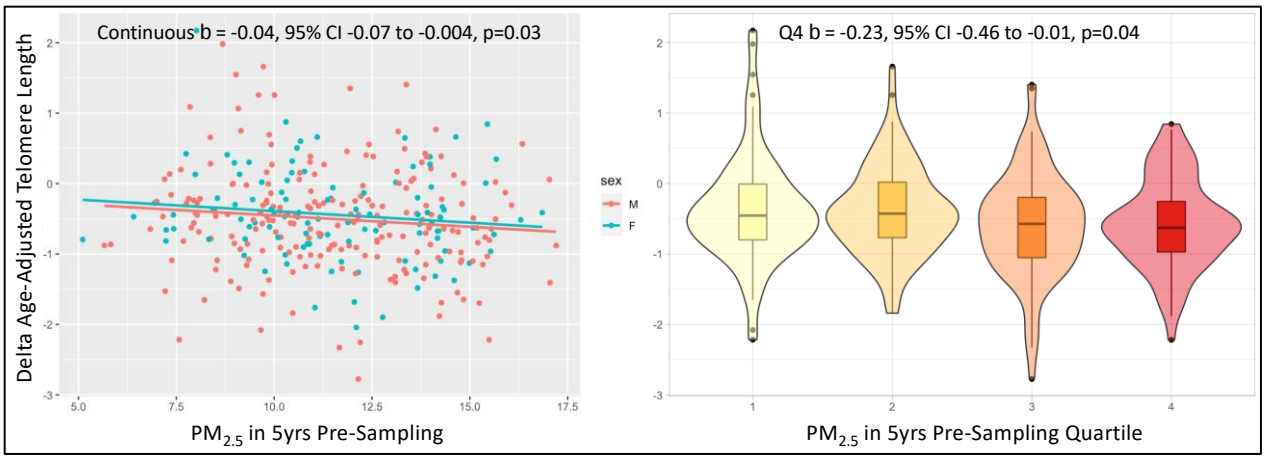


Figure 22 – Association of total $PM_{2.5}$ mass with age-adjusted telomere length. A) Scatterplot of continuous $PM_{2.5}$ exposure in 5-years pre-sampling against age-adjusted telomere length with results from linear model adjusted for sex, smoking history, and race reported. B) Violin plot of quartiled $PM_{2.5}$ exposure in 5-years pre-sampling against age-adjusted telomere length with results for quartile 4 (highest $PM_{2.5}$ exposure quartile) from linear model adjusted for sex, smoking history, and race reported.

Table 18 – Multivariable linear regression models for association of pollutants with age-corrected TL. Adjusted models include covariates of sex, smoking, and race. β-value signifies the change in age-adjusted TL per 1μg/m^3 increase in a pollutant, such that pollutants with a greater range of exposures tend to have smaller effect sizes per unit. Significant associations are **bolded** and marginal are *italicized*.

Model	β-value	95% CI	p	n
$PM_{2.5}$ in 5 years Pre-Sampling	**-0.04**	**-0.06, -0.004**	**0.03**	318
$PM_{2.5}$ in 3 months Pre-Sampling	-0.005	-0.03, 0.02	0.67	313
SO_4^{2-} in 5 years Pre-Sampling	**-0.06**	**-0.12, -0.003**	**0.04**	316
SO_4^2 in 3 months Pre-Sampling	-0.01	-0.05, 0.03	0.61	304
NO_3^- in 5 years Pre-Sampling	**-0.25**	**-0.45, -0.05**	**0.01**	316
NO_3^- in 3 months Pre-Sampling	**-0.12**	**-0.23, -0.01**	**0.03**	304
NH_4^+ in 5 years Pre-Sampling	**-0.18**	**-0.33, -0.03**	**0.02**	316
NH_4^+ in 3 months Pre-Sampling	-0.09	-0.22, 0.04	0.17	304
BC in 5 years Pre-Sampling	0.04	-0.31, 0.39	0.82	316
BC in 3 months Pre-Sampling	0.08	-0.20, 0.36	0.59	304
OM in 5 years Pre-Sampling	-0.07	-0.17, 0.02	0.14	316
OM in 3 months Pre-Sampling	-0.03	-0.11, 0.06	0.51	304
SS in 5 years Pre-Sampling	-0.20	-0.77, 0.36	0.48	316
SS in 3 months Pre-Sampling	-0.26	-0.76, 0.24	0.31	304
Soil in 5 years Pre-Sampling	-0.12	-0.68, 0.44	0.67	316
Soil in 3 months Pre-Sampling	0.07	-0.27, 0.42	0.69	304

Multi-pollutant analyses demonstrated similar direction and magnitude of effect of increased $PM_{2.5}$ mixture associations with age-corrected TL, but this did not reach significance thresholds (β per 1-quartile increase in $PM_{2.5}$ mixture = -0.09, 95%CI -0.22 to 0.02, p=0.11). The sum of negative effects on TL outweighed the sum of positive effects, with NH_4^+, followed by

OM, NO_3^-, and SS, respectively having the strongest negative associations with TL. Conversely, SO_4^{2-}, followed by BC and soil appeared to have positive associations with TL.

4.3.3.3 Telomere Length Associations with Mortality

Short TL was associated with increased mortality in models adjusted for sex, age, smoking, and race in the Simmons cohort (HR=1.41, 95%CI 1.07-1.86, p=0.02), consistent with previous findings (**Figure 20**).(Courtwright & El-Chemaly, 2019) Conversely, age-corrected TL as a continuous variable was not associated with mortality (HR=0.89, 95%CI 0.74-1.08, p=0.24). Spline models demonstrated a U-shaped variable HR curve, suggesting that patients with both low and high TL may experience increased mortality.

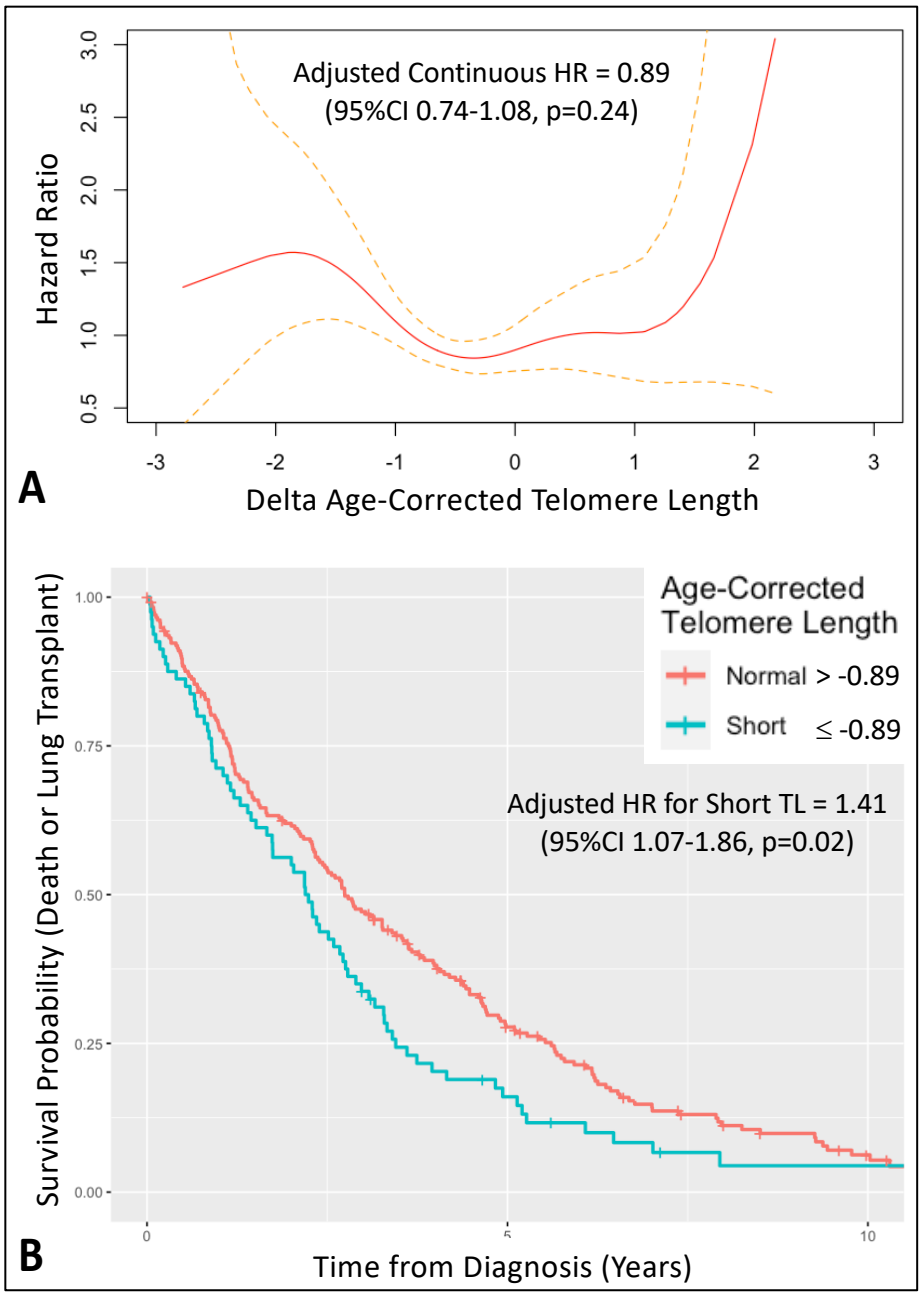


Figure 23 – Association of age-corrected TL with survival in Simmons cohort. A) Spline model of smoothed hazard ratio (HR) for death or transplant with inreasing age-corrected telomere length (TL) demonstrating that risk of death appears increased in age-corrected TL <-0.5 and >-1.5. HR reported is for Cox models of continuous age-corrected TL, adjusting for sex, age at diagnosis, smoking history, and race, demonstrating no significant linear association between TL and mortality. B) Kaplan-meier survival curve demonstrating association of short versus normal TL with mortality where "Normal" refers to top three quartiles of age-corrected telomere length (TL) and "Short" refers to bottom quartile. Hazard ratio (HR) reported is for Cox models of short telomere length associations with mortality in models adjusted for sex, age at diagnosis, smoking history, and race.

4.3.3.4 Pollutant Associations with Mortality

Increasing $PM_{2.5}$, SO_4^{2-}, and NH_4^+ in the 5-years and 3-months pre-sampling, and marginally increasing NO_3^-, BC, and OM in the 5-years pre-sampling were associated with increased mortality in models adjusted for age at diagnosis, sex, smoking, and race (**Table 16**). Findings were consistent with those demonstrated in previously published results from this cohort.(Goobie, Carlsten, Johannson, Khalil, et al., 2022)

Table 19 - Cox proportional hazards models for association of pollutants with mortality. Adjusted models include covariates of age at diagnosis, sex, smoking, and race. Hazard ratio (HR) signifies the mortality risk associated with a 1μg/m^3 increase in a pollutant, such that pollutants with a greater range of exposures tend to have smaller effect sizes per unit.Significant associations are **bolded** and marginal are *italicized*.

Model	HR	95% CI	p	n
$PM_{2.5}$ in 5 years Pre-Sampling	**1.13**	**1.07-1.18**	**<0.001**	316
$PM_{2.5}$ in 3 months Pre-Sampling	**1.08**	**1.05-1.12**	**<0.001**	311
SO_4^{2-} in 5 years Pre-Sampling	**1.26**	**1.15-1.38**	**<0.001**	314
SO_4^2 in 3 months Pre-Sampling	**1.13**	**1.07-1.20**	**<0.001**	302
NO_3^- in 5 years Pre-Sampling	*1.33*	*0.97-1.81*	*0.07*	314
NO_3^- in 3 months Pre-Sampling	0.98	0.83-1.14	0.75	302
NH_4^+ in 5 years Pre-Sampling	**1.69**	**1.33-2.14**	**<0.001**	314
NH_4^+ in 3 months Pre-Sampling	**1.46**	**1.21-1.77**	**<0.001**	302
BC in 5 years Pre-Sampling	*1.66*	*0.99-2.80*	*0.06*	314
BC in 3 months Pre-Sampling	1.37	0.92-2.05	0.13	302
OM in 5 years Pre-Sampling	*1.14*	*0.99-1.32*	*0.07*	314
OM in 3 months Pre-Sampling	1.11	0.98-1.27	0.11	302
SS in 5 years Pre-Sampling	0.74	0.30-1.87	0.53	314
SS in 3 months Pre-Sampling	0.98	0.43-2.25	0.97	302
Soil in 5 years Pre-Sampling	1.36	0.60-3.10	0.47	314
Soil in 3 months Pre-Sampling	0.87	0.51-1.50	0.62	302

4.3.3.5 Telomere Length Mediation of Pollutant-Mortality Associations

Using both a traditional and a causal mediation approach, the proportion of the pollutant-mortality relationship was calculated for pollutants where there was a significant association between the pollutant and TL as well as the pollutant and mortality. TL is estimated to mediate between 1-3% of the relationship of both $PM_{2.5}$ and SO_4^{2-} with mortality, 8-22% of the relationship of NO_3^- with mortality, and 2-5% of the relationship of NH_4^+ with mortality (**Table 17**). Causal mediation analysis, which required survival models to be changed from Cox to accelerated failure

time models, demonstrated significance for the role of short TL in mediating the NH_4^+-mortality relationship.

Table 20 – Proportion of pollutant-mortality relationships mediated by age-corrected telomere length (TL). Mediation analyses only performed when pollutant-TL and pollutant-mortality associations were significant or marginal. All survival models are for the 5-year pre-sampling period and are adjusted for age at diagnosis, sex, smoking, and race. Hazard ratio (HR) signifies the mortality risk associated with a 1μg/m^3 increase in a pollutant, such that pollutants with a greater range of exposures tend to have smaller effect sizes per unit. Significant associations are **bolded** and marginal are *italicized*.

Model	HR	95% CI	p	Traditional Mediation Proportion	Causal Mediation Analysis			n
					Proportion	95% CI	p	
$PM_{2.5}$	**1.13**	**1.07-1.18**	**<0.001**					
$PM_{2.5}$ + Continuous TL	**1.12**	**1.06-1.18**	**<0.001**	0.02	0.02	-0.05 to 0.11	0.48	316
$PM_{2.5}$ + Short TL	**1.12**	**1.07-1.18**	**<0.001**	0.03	0.01	-0.10 to 0.15	0.69	
SO_4^{2-}	**1.26**	**1.15-1.38**	**<0.001**					
SO_4^{2-} + Continuous TL	**1.25**	**1.14-1.37**	**<0.001**	0.01	0.02	-0.04 to 0.10	0.48	314
SO_4^{2-} + Short TL	**1.25**	**1.14-1.37**	**<0.001**	0.03	0.03	-0.04 to 0.13	0.39	
NO_3^-	*1.33*	*0.97-1.81*	*0.07*					
NO_3^- + Continuous TL	1.29	0.95-1.77	0.11	0.09	0.08	-0.32 to 1.08	0.42	314
NO_3^- + Short TL	1.25	0.91-1.71	0.16	0.22	0.14	-1.03 to 1.30	0.12	
NH_4^+	**1.69**	**1.33-2.14**	**<0.001**					
NH_4^+ + Continuous TL	**1.67**	**1.31-2.13**	**<0.001**	0.02	0.02	-0.06 to 0.11	0.59	314
NH_4^+ + Short TL	**1.64**	**1.29-2.09**	**<0.001**	0.05	**0.04**	**0.0005 to 0.12**	**0.04**	

4.3.3.6 Pollutant-Telomere Length (TL) Interactions in Mortality Analyses

Exploratory analysis of pollutant-TL interactions in Cox survival models demonstrated that patients with high $PM_{2.5}$ exposure (above median Simmons exposure of 11.3μg/m^3) and the shortest age-corrected TL (TL quartile 1/Q1) had significantly increased mortality compared to patients with low $PM_{2.5}$ exposure (<11.3μg/m^3) and short telomeres (TL Q1/high $PM_{2.5}$ exposure HR=1.98, 95%CI 1.21-3.24, p=0.006). Although not significant, patients with the longest TL (TL quartile 4/Q4) and high $PM_{2.5}$ exposures also had greater HR for mortality compared to TL Q1 patients with low exposures (HR=1.50, 95%CI 087-2.58, p=0.15)(**Figure 21**).

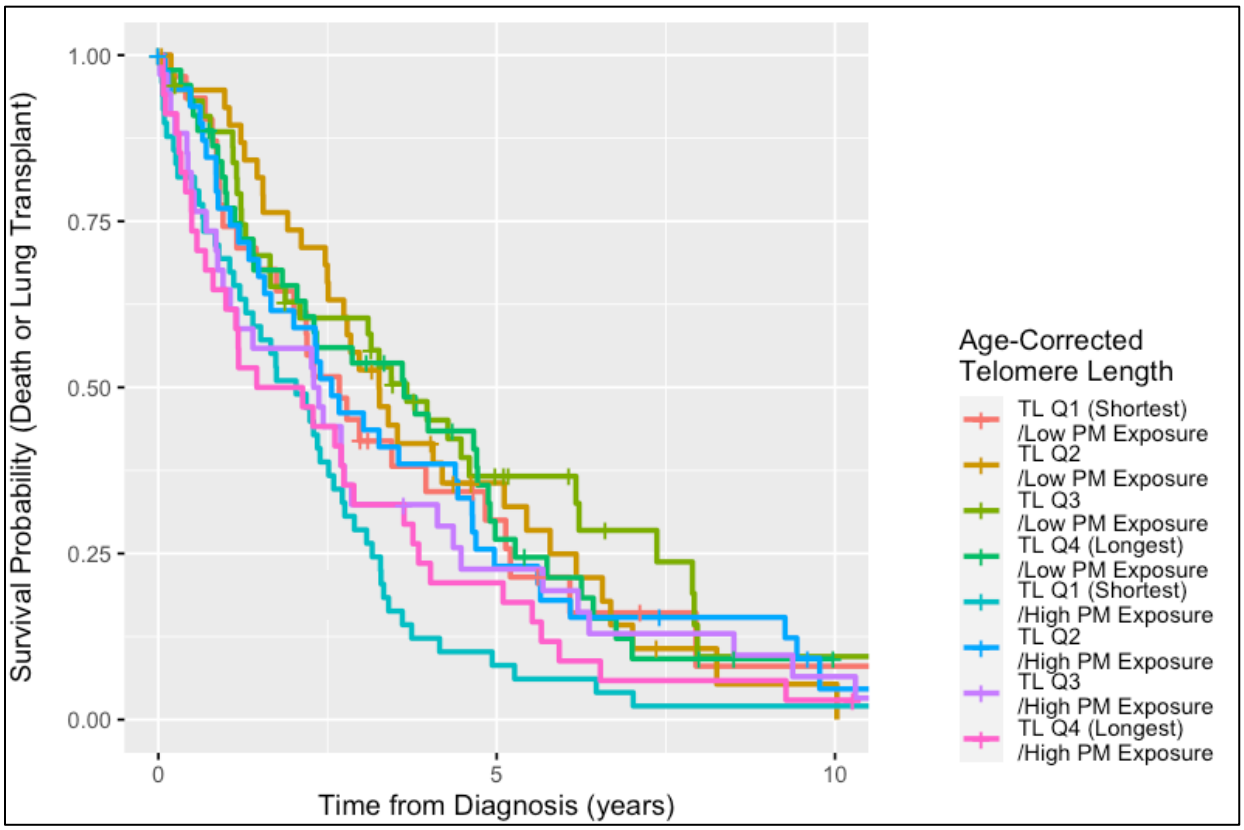


Figure 24 – Kaplan-Meier survival curve of telomere length (TL) quartile and $PM_{2.5}$ interaction. The lowest curves are seen with TL quartile 1(Q1) and quartile 4 (Q4) with high $PM_{2.5}$ exposures (≥11.3μg/m^3).

4.3.4 Discussion

This study represents the first evaluation of the association of $PM_{2.5}$ constituent component associations with telomere length in a cohort of patients with fILD, serving to unveil novel insights about the environmental pathophysiology of these diseases. We found that higher long-term (5-year pre-sampling) exposures to $PM_{2.5}$, SO_4^{2-}, NO_3^-, and NH_4^+ was associated with shorter age-corrected TL. We further validated previous findings demonstrating the associations of short TL and these pollutants with increased mortality in this population. Subsequently, through traditional and causal mediation approaches, we demonstrated that TL mediates between 1-22% of pollutant-mortality associations depending on the pollutant and modeling structure used. We also identified novel insights about the non-linear relationship between TL and mortality, revealing through our

interaction analyses that both short and long TL may interact with $PM_{2.5}$ to increase mortality risk. These findings may help to explain some of the discrepancies in the literature, where different studies have yielded opposite directions of effect for pollution impacts on TL.(Miri et al., 2019) Given that both $PM_{2.5}$ exposures and TL are associated with mortality in patients with fILD, these data provide evidence for a possible causal pathway linking air pollution to adverse outcomes in this vulnerable population.

Short telomeres below the 10^{th} percentile are found in up to 25% of patients with sporadic IPF, with up to 10% of these patients having identifiable telomere-related mutations.(Cronkhite et al., 2008) Short telomeres have also been demonstrated in patients with non-IPD fILDs, including rheumatoid arthritis-ILD, fibrotic hypersensitivity pneumonitis (fHP), and pleuroparenchymal fibroelastosis.(Courtwright & El-Chemaly, 2019; Ley et al., 2017) Longitudinal studies of patients with telomere-related mutations and severely shortened telomeres indicate that patients do not typically develop radiographic evidence or symptoms of fILDs until their 40s to 60s at the earliest.(Newton et al., 2016) These findings support that short TL precedes and contributes to fILD development, rather than the presence of a fILD leading to the development of short TL. Furthermore, telomere-related mutations demonstrate incomplete penetrance, whereby not all individuals with a telomere-associated mutation will develop a fILD.(Courtwright & El-Chemaly, 2019) This underscores the potential importance of environmental exposures like air pollution for triggering the development of telomere-related disease such as fILDs.

The present work is limited by this being a single-center evaluation of the association of pollutants with TL, however future work will involve evaluating the impact of $PM_{2.5}$ on TL in a cohort of patients with fILD from the University of British Columbia (UBC) site of CARE-PF and a separate cohort from the University of Chicago. Our analysis of TL was also performed on

leukocyte DNA, not DNA from lung cell types, although given the higher turnover rate and involvement of inflammatory cells in fILD pathophysiology, blood leukocytes may still reflect the most appropriate cell type to study pollution effects on TL in this population. We were only able to consider the most recent residential address for patients, and we used this to assign exposures to patients at the time of blood sampling, which has the potential to introduce exposure misclassification if people moved during their disease. Our cohort was also limited in its racial diversity, illustrating the need to translate these findings to cohorts more representative of general populations. Lastly, we did not adjust for the presence of telomere-related mutations in this analysis, but given our access to WGS data, we will evaluate for the presence of these telomere-related mutations and perform subgroup analyses excluding those with known mutations.

Future analyses should be performed to determine how pollutants influence clinical outcomes and fILD development in patients with short and long telomeres, given our preliminary findings indicating more harmful impacts of pollution in both extremes of TL. We will be expanding our work to address telomere responses to pollution exposures in future *in vitro* and *in vivo* extensions of this work. Future work should also evaluate the cause-specific mortality of patients with low versus high TL in these cohorts, given recent evidence that patients with long TL may be particularly susceptible to malignancies.(Chun-on et al., 2022) These findings will help to elucidate the mechanistic pathways whereby pollution may exert differential impacts on TL and disease development.

5.0 Conclusions

Together, the body of work presented in this PhD thesis reflects the largest and most geographically-diverse evaluation of the impact of neighborhood disadvantage and air pollution on clinical outcomes in patients with fILD to date. It is one of the few studies in the fILD epidemiologic literature to include all patients with fILD, and not just those with IPF. This is critically important given the high morbidity and mortality of all fILDs as well as the increasing global burden of these conditions.(Ma et al., 2022) This work also represents some of the first research to evaluate the underlying mechanisms explaining the associations between air pollution exposures and adverse clinical outcomes in patients with fILD.(Goobie et al., 2020) Such mechanistic evidence is critical to providing evidence of causal linkages between environmental exposures and adverse health outcomes, which is an essential step in informing environmental health policy. This project has highlighted novel pathways for further investigation that have the potential for targeting of future therapeutics to mitigate the harmful impacts of pollution exposure.

5.1 Summary of Results

Aim 1 was the first study of its kind to demonstrate an association of neighborhood-level disadvantage with mortality, odds of receiving a lung transplant, and baseline lung function in patients with fILD. The discrepancies in influence of neighborhood disadvantage on fILD outcomes between the U.S. and Canada were starkly apparent in this work, highlighting the potential contribution of healthcare system differences to these findings.(Gaffney & Podolanczuk,

2022) This work also highlighted the need to evaluate the contribution of environmental exposures to these disparities, given the well-established history of environmental injustice, especially in the U.S.(Bowe et al., 2019)

Aim 2 reflects the largest and most geographically-diverse evaluation of the impact of air pollution on patients with fILD to date. Our satellite-derived hybrid exposure estimation approach used in this study reflects an improvement from previous exposure estimation modalities used in much of the previous fILD and air pollution literature, given its high spatial resolution and ability to specify $PM_{2.5}$ constituent composition.(Van Donkelaar et al., 2019) In this study, we found that high exposures to $PM_{2.5}$, and especially to the primarily anthropogenic constituents, SO_4^{2-}, NO_3^-, and NH_4^+, was associated with substantial increases in mortality. Multi-constituent analyses also demonstrated consistent harmful impacts of these pollutants on baseline lung function and the rate of lung function decline. This work highlights the regional variability in $PM_{2.5}$-associated harms in patients with fILD, underscoring the need to target reductions in anthropogenic sources of $PM_{2.5}$ emissions that contribute most significantly to fILD morbidity and mortality.

Aim 3 demonstrates the first evidence linking $PM_{2.5}$ and constituent exposures with DNAm and telomere length alterations in patients with fILD. We demonstrate that higher exposures to pollutants are associated with increased global DNAm in patients with IPF. Subsequently, epigenome-wide DNAm analysis identified multiple differentially-methylated probes, with the identification of cg25354716 as the top CpG that may mediate a portion of the $PM_{2.5}$-mortality relationship. Lastly, we demonstrate that higher $PM_{2.5}$ exposure is associated with shortened leukocyte telomere length in patients with fILD, demonstrating how pollution may contribute to disease development and progression through this important pathway.

5.2 Strengths and Limitations

As mentioned, this work reflects the largest and most geographically-diverse evaluation of neighborhood and environmental impacts on clinical outcomes in patients with fILD to date. It is also the first research to explore the mechanistic underpinnings of the relationship between pollution and fILD progression, thus laying the groundwork for establishing a causal pathway linking these diseases with adverse environmental exposures.

As mentioned in previous chapters, each aspect of this work has limitations. Aim 1 was limited by including only a single U.S. cohort, which had limited racial diversity to explore the intersectionalities of neighborhood-level disadvantage and systemic racism in patients with fILD. Additionally, we were unable to directly compare access to care and medication use, which will be critical in future studies evaluating the role of healthcare system structure in influencing clinical outcomes in patients with fILD. Aim 2 was limited by our inability to capture changes in patient address over time, leading to some risk of exposure misclassification, as well as an inability to directly apportion the sources of different $PM_{2.5}$ constituent exposures. Aim 3 was limited by our lack of ability to compare DNAm in blood and lung tissue, as well as inherent assumptions underlying mediation analyses. Further work on this aim will alleviate some aspects of these limitations.

5.3 Future Directions

Extensions for Aim 1 and Aim 2 are ongoing. We are in the process of evaluating the impact of neighborhood disadvantage and $PM_{2.5}$ on healthcare utilization outcomes in patients with

fILD. We are also applying the methods from Aims 1& 2 to other clinical cohorts, including patients with systemic sclerosis, rheumatoid arthritis, pediatric asthma, and obstructive sleep apnea. Future work will also explore the interaction effects between neighborhood disadvantage and $PM_{2.5}$ exposure in patients with fILD.

My postdoctoral training will largely extend upon Aim 3 of my research. We will begin by evaluating DNAm patterns and the influence of $PM_{2.5}$ exposures on these patterns in lung tissue from patients with fILD, comparing our findings to those in blood. Using single cell imaging techniques, we will then consider how cellular distributions are altered in response to air pollution exposures in these patients. Further work will leverage whole genome sequencing (WGS) data to perform gene-environment interaction studies and to identify methylation quantitative trait loci (meQTLs) that may influence methylation at our top CpGs noted in Aim 3. This research will lay the groundwork for our understanding of the environmental pathophysiology of fILDs. This work has central public health relevance as it provides critical mechanistic evidence that has the potential to influence environmental health policies which impact everyone.

Appendix A – Data Supplement for Neighborhood Disadvantage Impacts in fILD

Neighborhood-level disadvantage impacts on patients with fibrotic interstitial lung disease: an international multicohort study – Online Data Supplement.

Gillian C. Goobie, MD[1,2,3]; Christopher J. Ryerson, MD, MAS[4,5]; Kerri A. Johannson, MD, MPH[6]; Erin Schikowski, MD[7]; Richard H. Zou, MD[8]; Nasreen Khalil, MD[4]; Veronica Marcoux, MD[9]; Deborah Assayag, MD, MAS[10]; Hélène Manganas, MD[11]; Jolene H. Fisher, MD[12]; Martin Kolb, MD, PhD[13]; Kevin F. Gibson, MD[2,8]; Daniel J. Kass, MD[2,8]; Yingze Zhang, PhD[8]; Kathleen O. Lindell, RN, PhD[2,15*]; S. Mehdi Nouraie, MD, PhD[8*]

1. Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, U.S.A. Email: goobiegc@upmc.edu.
2. Simmons Center for Interstitial Lung Disease, Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A.
3. Clinician Investigator Program, Department of Medicine, University of British Columbia, Vancouver, B.C., Canada.
4. Division of Respiratory Medicine, Department of Medicine, University of British Columbia, Vancouver, B.C., Canada. Email: chris.ryerson@hli.ubc.ca, nkhalil@mail.ubc.ca.
5. Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, B.C., Canada.
6. Division of Respiratory Medicine, Department of Medicine, University of Calgary, Calgary, AB, Canada. Email: kerri.johannson@albertahealthservices.ca.
7. Division of General Internal Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A. Email: schikowskiem@upmc.edu.
8. Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A. Emails: zour@upmc.edu, gibsonkf@upmc.edu, kassd2@upmc.edu, zhanyx@upmc.edu, nouraies@upmc.edu.
9. Division of Respirology, Critical Care and Sleep Medicine, College of Medicine, University of Saskatchewan, Saskatoon, SK, Canada. Email: vsm940@mail.usask.ca.
10. Division of Respiratory Medicine, Department of Medicine, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada. Email: deborah.assayag@mcgill.ca.
11. Département de Médecine, Centre Hospitalier de l'Université de Montréal, Montreal, QC, Canada. Email: helene.manganas.med@ssss.gouv.qc.ca.
12. Division of Respirology, Department of Medicine, University of Toronto, Toronto, ON, Canada. Email: jolene.fisher@uhn.ca.
13. Department of Medicine, Firestone Institute for Respiratory Health, The Research Institute of St. Joe's Hamilton, St. Joseph's Healthcare, McMaster University, Hamilton, ON, Canada. Email: kolbm@mcmaster.ca.
14. College of Nursing, Medical University of South Carolina, Charleston, SC, U.S.A. Email: lindellk@musc.edu.

*Denotes co-senior authorship

Appendix Table 1 – Baseline characteristics in U.S. and Canadian cohorts broken down by ADI or CIMD quartile. Abbreviations: ADI, area deprivation index; CIMD, Canadian index of multiple deprivation; CTD-ILD, connective tissue disease-interstitial lung disease; D_LCO, diffusion capacity for carbon monoxide; fHP, fibrotic hypersensitivity pneumonitis; FVC, forced vital capacity; IIP, idiopathic interstitial pneumonia; ILD, interstitial lung disease; IPF, idiopathic pulmonary fibrosis; IQR, interquartile range; NYD, not yet diagnosed; Q1-Q4, Quartiles 1-4.

Patient Characteristics	**U.S. Cohort**				**Canadian Cohort**			
	ADI Q1 (1 to 44)	ADI Q2 (45 to 61)	ADI Q3 (62 to 78)	ADI Q4 (79 to 100)	CIMD Q1 (-1.30 to -0.35)	CIMD Q2 (-0.34 to -0.03)	CIMD Q3 (-0.02 to 0.38)	CIMD Q4 (0.39 to 2.57)
	N = 348	N = 339	N = 348	N = 337	N = 840	N = 839	N = 839	N = 839
Age at diagnosis, median (IQR), years	67 (59, 74)	66 (59, 73)	66 (58, 73)	63 (56, 70)	64 (56, 72)	66 (58, 72)	67 (58, 73)	66 (56, 74)
Male sex, n (%)	208 (60)	183 (54)	203 (58)	167 (50)	461 (55)	412 (49)	409 (49)	382 (46)
Self-reported Race (n, %)								
White	312 (90)	307 (91)	313 (90)	277 (82)	731 (87)	708 (84)	645 (77)	574 (69)
Black	6 (2)	5 (1.5)	10 (3)	35 (10)	11 (1)	4 (1)	18 (2)	20 (2)
Asian	4 (1)	1 (0.3)	0 (0)	0 (0)	63 (8)	70 (8)	108 (13)	141 (17)
Indigenous*	1 (0.3)	1 (0.3)	0 (0)	0 (0)	8 (1)	9 (1)	25 (3)	43 (5)
Pacific Islander	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.1)	7 (1)	5 (1)	12 (1)
Unknown	25 (7)	25 (7)	25 (7)	25 (7)	26 (3)	41 (5)	38 (4)	49 (6)
Self-reported Ethnicity (n, %)								
Not Hispanic	293 (84)	282 (83)	293 (84)	280 (83)	705 (84)	696 (83)	697 (83)	674 (80)
Hispanic	1 (0.3)	1 (0.3)	0 (0)	0 (0)	18 (2)	13 (2)	14 (2)	23 (3)
Unknown	54 (16)	56 (17)	55 (16)	57 (17)	117 (14)	130 (15)	128 (15)	142 (17)
Smoker (n, %)								
Never	106 (30)	104 (31)	90 (26)	103 (30)	331 (39)	298 (36)	323 (39)	317 (38)
Former	161 (46)	160 (47)	174 (50)	144 (43)	465 (55)	507 (60)	479 (57)	455 (54)
Current	6 (2)	5 (1)	10 (3)	16 (5)	38 (5)	33 (4)	36 (4)	65 (8)
Unknown	75 (22)	70 (21)	74 (21)	74 (22)	6 (1)	1 (0.1)	1 (0.1)	2 (0.2)
ILD Diagnostic Group (n, %)								
IPF	180 (52)	168 (50)	176 (51)	164 (49)	226 (27)	247 (29)	243 (29)	202 (24)
CTD-ILD	76 (22)	84 (25)	65 (19)	65 (19)	332 (40)	303 (36)	311 (37)	337 (40)
fHP	10 (3)	13 (4)	19 (5)	12 (4)	73 (9)	75 (9)	53 (6)	56 (7)
Pneumoconiosis	3 (1)	8 (2)	8 (2)	5 (1)	13 (2)	4 (1)	7 (1)	4 (1)
Non-IPF IIP	12 (3)	18 (5)	12 (3)	23 (7)	19 (2)	31 (4)	33 (4)	25 (3)
Other ILD	10 (3)	12 (3)	14 (4)	14 (4)	38 (4)	26 (3)	32 (4)	24 (3)
Unclassifiable or NYD	57 (16)	36 (11)	54 (16)	54 (16)	139 (16)	153 (18)	160 (19)	191 (22)
ADI or CIMD, median score (IQR)	32 (23, 38)	54 (50, 59)	71 (67, 75)	87 (82, 93)	-0.54 (-0.68, -0.43)	-0.21 (-0.28, -0.12)	0.16 (0.06, 0.28)	0.69 (0.52, 0.91)
Baseline % Predicted FVC, median (IQR)	69 (55, 82)	64 (53, 78)	66 (54, 80)	66 (53, 81)	76 (62, 90)	75 (63, 89)	75 (61, 88)	74 (60, 91)

Baseline % Predicted D_LCO, Median (IQR)	53 (41, 66)	48 (37, 62)	48 (35, 62)	45 (33, 61)	60 (47, 76)	58 (45, 71)	54 (43, 68)	56 (42, 69)
Follow-up Duration, median (IQR), years	2.5 (1.2, 4.7)	2.4 (0.9, 4.7)	2.5 (1.1, 4.9)	2.1 (0.8, 4.6)	2.7 (1.5, 4.4)	2.2 (1.3, 3.8)	2.2 (1.2, 4.0)	2.7 (1.3, 4.4)
Cause of censoring, n (%)								
Death	168 (48)	177 (49)	164 (50)	174 (52)	185 (22)	175 (21)	197 (23)	204 (24)
Lung Transplantation	45 (13)	58 (16)	52 (16)	41 (12)	45 (5)	59 (7)	33 (4)	37 (5)
Lost to follow-up or censored by data extraction	135 (39)	126 (35)	110 (34)	122 (36)	610 (73)	605 (72)	609 (73)	598 (71)

*Includes Native American, American Indian, Alaskan First Nations, & other Indigenous persons in the U.S.; First Nations, Métis, Inuit, and other Indigenous persons in Canada.

Appendix Table 2 – State and province breakdown of patients in U.S. and Canadian cohorts.

U.S. Cohort (n=1372)		Canadian Cohort (N=3357)	
State, n (%)		**Province, n (%)**	
Arkansas	1 (0.07%)	Alberta	346 (10.31%)
California	1 (0.07%)	British Columbia	1438 (42.84%)
Colorado	1 (0.07%)	Newfoundland and Labrador	1 (0.03%)
Connecticut	1 (0.07%)	Ontario	797 (23.74%)
District of Columbia	1 (0.07%)	Quebec	687 (20.46%)
Delaware	1 (0.07%)	Saskatchewan	88 (2.62%)
Florida	12 (0.87%)		
Georgia	1 (0.07%)		
Illinois	2 (0.15%)		
Indiana	4 (0.29%)		
Kentucky	3 (0.22%)		
Massachusetts	4 (0.29%)		
Maryland	11 (0.80%)		
Maine	1 (0.07%)		
Michigan	7 (0.51%)		
Missouri	1 (0.07%)		
North Carolina	6 (0.44%)		
New Jersey	5 (0.36%)		
New York	34 (2.48%)		
Ohio	61 (4.45%)		
Pennsylvania	1096 (79.88%)		
Rhode Island	1 (0.07%)		
South Carolina	2 (0.15%)		
Tennessee	1 (0.07%)		
Texas	1 (0.07%)		
Virginia	8 (0.58%)		
West Virginia	105 (7.65%)		

Appendix Table 3 – Unadjusted, partially, and fully adjusted Cox proportional hazards models evaluating the impact of ADI or CIMD on the composite outcome of death or lung transplant in full (all patients with fILD) and IPF-only U.S. and Canadian cohorts. Abbreviations: ADI, area deprivation index; CIMD, Canadian index of multiple deprivation; D_LCO, diffusion capacity for carbon monoxide; fILD, fibrotic interstitial lung disease; FVC, forced vital capacity; HR, hazard ratio; IPF, idiopathic pulmonary fibrosis.

Full U.S. Cohort			**Full Canadian Cohort**		
	HR	**95% CI**		**HR**	**95% CI**
Continuous ADI Unadjusted Model (N = 1372)			**Continuous CIMD Unadjusted Model (N = 3333)**		
Continuous ADI	1.001	0.998, 1.004	Continuous CIMD	1.04	0.93, 1.17
Continuous ADI Partially Adjusted Model (N = 1372)			**Continuous CIMD Partially Adjusted Model (N = 3332)**		
Continuous ADI	1.004	1.002, 1.007	Continuous CIMD	1.03	0.91, 1.16
Female sex	0.56	0.49, 0.64	Female sex	0.61	0.53, 0.70
Age at diagnosis	1.04	1.03, 1.04	Age at diagnosis	1.04	1.03, 1.05
Continuous ADI Fully Adjusted Model (N = 797)			**Continuous CIMD Fully Adjusted Model (N = 2362)**		
Continuous ADI	1.006*	1.002, 1.010	Continuous CIMD	0.97	0.83, 1.14
Female sex	0.44	0.36, 0.54	Female sex	0.56	0.47, 0.67
Age at diagnosis	1.04	1.03, 1.05	Age at diagnosis	1.04	1.03, 1.05
Former smoking	1.40	1.14, 1.71	Former smoking	1.40	1.16, 1.69
Always smoking	0.92	0.52, 1.63	Always smoking	0.99	0.61, 1.62
Non-White race	0.75	0.54, 1.04	Non-White race	0.69	0.55, 0.87
Baseline FVC	0.99	0.98, 0.99	Baseline FVC	0.98	0.98, 0.99
Baseline D_LCO	0.96	0.96, 0.97	Baseline D_LCO	0.96	0.96, 0.97
Quartiled ADI Unadjusted Model (N = 1372)			**Quartiled CIMD Unadjusted Model (N = 3333)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	1.10	0.91, 1.32	Quartile 2	1.09	0.90, 1.30
Quartile 3	1.18	0.98, 1.42	Quartile 3	1.02	0.85, 1.23
Quartile 4	1.10	0.91, 1.33	Quartile 4	1.04	0.86, 1.24
Quartiled ADI Partially Adjusted Model (N = 1372)			**Quartiled CIMD Partially Adjusted Model (N = 3332)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	1.13	0.94, 1.36	Quartile 2	1.09	0.91, 1.31
Quartile 3	1.17	0.97, 1.42	Quartile 3	1.01	0.84, 1.22
Quartile 4	1.31	1.08, 1.59	Quartile 4	1.02	0.85, 1.22
Female sex	0.56	0.49, 0.64	Female sex	0.61	0.53, 0.69
Age at diagnosis	1.04	1.03, 1.04	Age at diagnosis	1.04	1.03, 1.05
Quartiled ADI Fully Adjusted Model (N = 797)			**Quartiled CIMD Fully Adjusted Model (N = 2362)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	1.08	0.84, 1.39	Quartile 2	1.09	0.86, 1.37
Quartile 3	1.13	0.89, 1.44	Quartile 3	0.98	0.78, 1.24
Quartile 4	1.51†	1.17, 1.95	Quartile 4	0.94	0.74, 1.19
Female sex	0.44	0.36, 0.54	Female sex	0.56	0.47, 0.67
Age at diagnosis	1.04	1.03, 1.05	Age at diagnosis	1.04	1.03, 1.05
Former smoking	1.41	1.15, 1.72	Former smoking	1.39	1.15, 1.68
Always smoking	0.90	0.51, 1.60	Always smoking	1.00	0.61, 1.64
Non-White race	0.72	0.52, 1.01	Non-White race	0.70	0.55, 0.88
Baseline FVC	0.99	0.98, 0.99	Baseline FVC	0.98	0.98, 0.99
Baseline D_LCO	0.96	0.96, 0.97	Baseline D_LCO	0.96	0.96, 0.97
IPF-Only U.S. Cohort			**IPF-Only Canadian Cohort**		
	HR	**95% CI**		**HR**	**95% CI**
Continuous ADI Unadjusted Model (N = 688)			**Continuous CIMD Unadjusted Model (N = 912)**		
Continuous ADI	1.003	0.9997, 1.007	Continuous CIMD	1.03	0.84, 1.27

Continuous ADI Partially Adjusted Model (N = 688)			**Continuous CIMD Partially Adjusted Model (N = 912)**		
Continuous ADI	1.004	1.0002, 1.008	Continuous CIMD	1.03	0.84, 1.27
Female sex	0.68	0.57, 0.81	Female sex	0.67	0.52, 0.86
Age at diagnosis	1.008	1.00-1.02	Age at diagnosis	1.02	1.00, 1.03
Continuous ADI Fully Adjusted Model (N = 444)			**Continuous CIMD Fully Adjusted Model (N = 591)**		
Continuous ADI	1.003	0.998, 1.008	Continuous CIMD	1.00	0.76, 1.32
Female sex	0.44	0.34, 0.56	Female sex	0.65	0.46, 0.93
Age at diagnosis	1.02	1.01, 1.03	Age at diagnosis	1.02	1.00, 1.04
Former smoking	1.11	0.87, 1.41	Former smoking	0.98	0.70, 1.37
Always smoking	0.51	0.21, 1.23	Always smoking	0.78	0.33, 1.73
Non-White race	1.38	0.93, 2.04	Non-White race	0.86	0.56, 1.31
Baseline FVC	0.99	0.98, 0.99	Baseline FVC	0.98	0.97, 0.99
Baseline D_LCO	0.96	0.95, 0.97	Baseline D_LCO	0.97	0.96, 0.98
Quartiled ADI Unadjusted Model (N = 688)			**Quartiled CIMD Unadjusted Model (N = 912)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	1.17	0.93, 1.48	Quartile 2	1.12	0.84, 1.49
Quartile 3	1.18	0.93, 1.49	Quartile 3	1.00	0.74, 1.34
Quartile 4	1.31	1.03, 1.66	Quartile 4	0.99	0.72, 1.35
Quartiled ADI Partially Adjusted Model (N = 688)			**Quartiled CIMD Partially Adjusted Model (N = 912)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	1.17	0.92, 1.48	Quartile 2	1.19	0.89, 1.59
Quartile 3	1.14	0.90, 1.44	Quartile 3	1.05	0.78, 1.42
Quartile 4	1.35	1.06, 1.71	Quartile 4	1.02	0.74, 1.39
Female sex	0.68	0.57, 0.81	Female sex	0.66	0.51, 0.86
Age at diagnosis	1.01	1.00, 1.02	Age at diagnosis	1.02	1.01, 1.03
Quartiled ADI Fully Adjusted Model (N = 444)			**Quartiled CIMD Fully Adjusted Model (N = 591)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	1.06	0.79, 1.42	Quartile 2	1.16	0.79, 1.71
Quartile 3	0.98	0.73, 1.30	Quartile 3	1.04	0.71, 1.54
Quartile 4	1.35	1.01, 1.82	Quartile 4	0.90	0.59, 1.38
Female sex	0.43	0.34, 0.56	Female sex	0.65	0.46, 0.93
Age at diagnosis	1.02	1.01, 1.03	Age at diagnosis	1.02	1.00, 1.04
Former smoking	1.12	0.88, 1.44	Former smoking	0.96	0.68, 1.34
Always smoking	0.47	0.19, 1.14	Always smoking	0.78	0.34, 1.77
Non-White race	1.36	0.91, 2.01	Non-White race	0.88	0.58, 1.36
Baseline FVC	0.99	0.98, 0.99	Baseline FVC	0.98	0.97, 0.99
Baseline D_LCO	0.96	0.95, 0.97	Baseline D_LCO	0.97	0.96, 0.98

*Interpretation: For each 1-point increase in ADI (ranging from 1-100), the risk of mortality/transplant increases by 0.6%.

†Interpretation: Compared to quartile 1, quartile 4 in U.S. cohort experiences 51% higher risk of mortality/transplant.

Appendix Table 4 – ADI and CIMD at risk table for competing hazards survival analyses. Abbreviations: ADI, area deprivation index; CIMD, Canadian index of multiple deprivation.

U.S. Cohort						
Status	**0 years**	**2 years**	**4 years**	**6 years**	**8 years**	**10 years**
Alive (total)	**1372**	**851**	**559**	**365**	**223**	**149**
ADI Quartile 1	348	221	148	99	57	36
ADI Quartile 2	339	218	138	93	64	39
ADI Quartile 3	348	210	140	81	47	31
ADI Quartile 4	337	202	133	92	55	43
Dead (total)	**0**	**275**	**165**	**105**	**63**	**29**
ADI Quartile 1	0	65	41	30	16	9
ADI Quartile 2	0	68	43	24	11	8
ADI Quartile 3	0	67	35	32	20	8
ADI Quartile 4	0	75	46	19	16	4
Transplanted (total)	**0**	**113**	**49**	**18**	**11**	**2**
ADI Quartile 1	0	20	15	6	3	1
ADI Quartile 2	0	30	14	5	2	1
ADI Quartile 3	0	34	15	5	2	0
ADI Quartile 4	0	29	5	2	4	0
Censored (total)	**0**	**133**	**78**	**71**	**68**	**43**
ADI Quartile 1	0	42	17	13	23	11
ADI Quartile 2	0	23	23	16	16	16
ADI Quartile 3	0	37	20	22	12	8
ADI Quartile 4	0	31	18	20	17	8
Canadian Cohort						
Status	**0 years**	**2 years**	**4 years**	**6 years**	**8 years**	**10 years**
At risk (total)	**3357**	**2362**	**1099**	**626**	**291**	**149**
CIMD Quartile 1	840	617	322	144	59	35
CIMD Quartile 2	839	564	298	151	70	34
CIMD Quartile 3	839	584	315	160	74	41
CIMD Quartile 4	839	597	328	171	88	39
Dead (total)	**0**	**296**	**259**	**110**	**48**	**29**
CIMD Quartile 1	0	59	69	39	10	4
CIMD Quartile 2	0	75	59	20	10	6
CIMD Quartile 3	0	84	61	24	13	8
CIMD Quartile 4	0	78	70	27	15	11
Transplanted (total)	**0**	**76**	**59**	**31**	**5**	**1**
CIMD Quartile 1	0	15	20	9	0	0
CIMD Quartile 2	0	31	18	8	2	0
CIMD Quartile 3	0	15	11	6	1	0
CIMD Quartile 4	0	15	10	8	2	1
Censored (total)	**0**	**623**	**781**	**496**	**282**	**112**
CIMD Quartile 1	0	149	206	130	75	20
CIMD Quartile 2	0	169	189	119	69	30
CIMD Quartile 3	0	156	197	125	72	25
CIMD Quartile 4	0	149	189	122	66	37

Appendix Table 5 – Competing hazards models evaluating the impact of ADI or CIMD on survival in full U.S. and Canadian cohorts with lung transplant as competing risk for death. Abbreviations: ADI, area deprivation index; CIMD, Canadian index of multiple deprivation; D_LCO, diffusion capacity for carbon monoxide; FVC, forced vital capacity; HR, hazard ratio.

U.S. Cohort			**Canadian Cohort**		
	HR	**95% CI**		**HR**	**95% CI**
Continuous ADI Unadjusted Model – Mortality Outcome (N = 1372)			**Continuous CIMD Unadjusted Model – Mortality Outcome (N = 3333)**		
Continuous ADI	1.00	0.998, 1.001	Continuous CIMD	1.12	0.98, 1.27
Continuous ADI Unadjusted Model – Transplant Outcome (N = 1372)			**Continuous CIMD Unadjusted Model – Transplant Outcome (N = 3333)**		
Continuous ADI	0.998	0.993, 1.004	Continuous CIMD	0.80	0.62, 1.04
Continuous ADI Adjusted Model – Mortality Outcome (N = 797)			**Continuous CIMD Adjusted Model – Mortality Outcome (N = 2362)**		
Continuous ADI	1.007*	1.003, 1.011	Continuous CIMD	0.90	0.76, 1.08
Female sex	0.79	0.63, 0.99	Female sex	0.68	0.56, 0.82
Ever smoking	1.23	0.98, 1.54	Ever smoking	1.43	1.15, 1.78
Age at diagnosis	1.07	1.06, 1.08	Age at diagnosis	1.07	1.05, 1.08
Non-White race	1.39	0.98, 1.97	Non-White race	0.93	0.72, 1.22
Baseline FVC	0.99	0.98, 0.995	Baseline FVC	0.99	0.98, 0.996
Baseline D_LCO	0.99	0.98, 0.992	Baseline D_LCO	0.97	0.96, 0.97
Continuous ADI Adjusted Model – Transplant Outcome (N = 797)			**Continuous CIMD Adjusted Model – Transplant Outcome (N = 2362)**		
Continuous ADI	0.997	0.99, 1.01	Continuous CIMD	1.08	0.79, 1.47
Female sex	0.48	0.32, 0.71	Female sex	0.46	0.31, 0.68
Ever smoking	1.47	0.96, 2.26	Ever smoking	1.26	0.83, 1.92
Age at diagnosis	0.97	0.96, 0.98	Age at diagnosis	0.98	0.96, 0.99
Non-White race	0.29†	0.11, 0.72	Non-White race	0.37	0.20, 0.67
Baseline FVC	0.99	0.98, 1.00	Baseline FVC	0.96	0.95, 0.97
Baseline D_LCO	0.96	0.95, 0.98	Baseline D_LCO	0.98	0.97, 0.99
Quartiled ADI Unadjusted Model – Mortality Outcome (N = 1372)			**Quartiled CIMD Unadjusted Model – Mortality Outcome (N = 3333)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	1.01	0.82, 1.25	Quartile 2	0.98	0.78, 1.20
Quartile 3	1.05	0.85, .29	Quartile 3	1.09	0.89, 1.33
Quartile 4	1.11	0.90, 1.38	Quartile 4	1.09	0.90, 1.33
Quartiled ADI Unadjusted Model – Transplant Outcome (N = 1372)			**Quartiled CIMD Unadjusted Model – Transplant Outcome (N = 3333)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	1.20	0.81, 1.78	Quartile 2	1.42	0.96, 2.09
Quartile 3	1.31	0.89, 1.94	Quartile 3	0.76	0.48, 1.19
Quartile 4	0.94	0.62, 1.43	Quartile 4	0.83	0.54, 1.28
Quartiled ADI Adjusted Model – Mortality Outcome (N = 797)			**Quartiled CIMD Adjusted Model – Mortality Outcome (N = 2362)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	1.06	0.80, 1.39	Quartile 2	0.88	0.68, 1.15
Quartile 3	0.99	0.75, 1.30	Quartile 3	0.89	0.69, 1.15
Quartile 4	1.57‡	1.17, 2.11	Quartile 4	0.78	0.60, 1.02

Female sex	0.79	0.63, 0.99	Female sex	0.68	0.56, 0.82
Ever smoking	1.27	1.01, 1.60	Ever smoking	1.43	1.15, 1.78
Age at diagnosis	1.08	1.06, 1.09	Age at diagnosis	1.07	1.05, 1.08
Non-White race	1.34	0.94, 1.91	Non-White race	0.95	0.73, 1.24
Baseline FVC	0.99	0.98, 0.995	Baseline FVC	0.99	0.98, 0.996
Baseline D_LCO	0.99	0.98, 0.99	Baseline D_LCO	0.97	0.96, 0.97
Quartiled ADI Adjusted Model – Transplant Outcome (N = 797)			**Quartiled CIMD Adjusted Model – Transplant Outcome (N = 2362)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	0.89	0.53, 1.50	Quartile 2	1.98	1.19, 3.31
Quartile 3	1.04	0.64, 1.70	Quartile 3	1.37	0.79, 2.40
Quartile 4	0.93	0.55, 1.57	Quartile 4	1.52	0.87, 2.66
Female sex	0.48	0.32, 0.71	Female sex	0.45	0.30, 0.66
Ever smoking	1.45	0.95, 2.22	Ever smoking	1.23	0.81, 1.86
Age at diagnosis	0.97	0.96, 0.98	Age at diagnosis	0.97	0.96, 0.99
Non-White race	0.28	0.11, 0.71	Non-White race	0.36	0.81, 1.86
Baseline FVC	0.99	0.98, 1.003	Baseline FVC	0.96	0.95, 0.97
Baseline D_LCO	0.97	0.95, 0.98	Baseline D_LCO	0.98	0.97, 0.99

*Interpretation: For each 1-point increase in ADI (ranging from 1-100) in the U.S. cohort, the hazard ratio for mortality increases by 0.6%. †Interpretation: As compared to patients of White race, patients of non-White race have 0.29 the likelihood of experiencing the "hazard" of lung transplant. I.e. they are approximately 70% less likely than White individuals to receive transplant. ‡Interpretation: As compared to quartile 1, patients living in quartile 4 of neighborhood disadvantage in the U.S. cohort have 57% higher risk of mortality.

Appendix Table 6 – Unadjusted, partially adjusted, and fully adjusted generalized linear models evaluating the impact of ADI or CIMD on odds of receiving lung transplant in full cohort (all patients with fILD) and IPF-only cohort. Abbreviations: ADI, area deprivation index; CIMD, Canadian index of multiple deprivation; D_LCO, diffusion capacity for carbon monoxide; fILD, fibrotic interstitial lung disease; FVC, forced vital capacity; IPF, idiopathic pulmonary fibrosis; OR, odds ratio.

Full U.S. Cohort			**Full Canadian Cohort**		
	OR	**95% CI**		**OR**	**95% CI**
Continuous ADI Unadjusted Model (N = 1372)			**Continuous CIMD Unadjusted Model (N = 3357)**		
Continuous ADI	0.998	0.992, 1.004	Continuous CIMD	0.80	0.59, 1.07
Continuous ADI Partially Adjusted Model (N = 1372)			**Continuous CIMD Partially Adjusted Model (N = 3332)**		
Continuous ADI	0.997	0.990, 1.004	Continuous CIMD	0.88	0.65, 1.18
Female sex	0.42	0.30, 0.58	Female sex	0.44	0.31, 0.61
Age at diagnosis	0.96	0.95, 0.97	Age at diagnosis	0.96	0.95, 0.97
Continuous ADI Fully Adjusted Model (N = 797)			**Continuous CIMD Fully Adjusted Model (N = 2362)**		
Continuous ADI	0.995	0.985, 1.004	Continuous CIMD	1.01	0.68, 1.46
Female sex	0.45	0.28, 0.71	Female sex	0.45	0.30, 0.68
Age at diagnosis	0.95	0.94, 0.97	Age at diagnosis	0.96	0.95, 0.98
Ever smoking	1.56	0.98, 2.55	Ever smoking	1.33	0.87, 2.06
Non-White race	0.23	0.08, 0.56	Non-White race	0.35	0.18, 0.63
Baseline FVC	0.99	0.98, 1.00	Baseline FVC	0.96	0.94, 0.97
Baseline D_LCO	0.96	0.95, 0.97	Baseline D_LCO	0.98	0.96, 0.99
Quartiled ADI Unadjusted Model (N = 1372)			**Quartiled CIMD Unadjusted Model (N = 3357)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	1.29	0.81, 1.92	Quartile 2	1.34	0.89, 2.00
Quartile 3	1.32	0.87, 2.02	Quartile 3	0.72	0.45, 1.14
Quartile 4	0.93	0.59, 1.47	Quartile 4	0.82	0.52, 1.27
Quartiled ADI Partially Adjusted Model (N = 1372)			**Quartiled CIMD Partially Adjusted Model (N = 3332)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	1.24	0.80, 1.94	Quartile 2	1.52	1.01, 2.30
Quartile 3	1.31	0.85, 2.02	Quartile 3	0.82	0.51, 1.30
Quartile 4	0.87	0.55, 1.38	Quartile 4	0.94	0.60, 1.48
Female sex	0.42	0.30, 0.59	Female sex	0.44	0.31, 0.60
Age at diagnosis	0.96	0.95, 0.97	Age at diagnosis	0.96	0.95, 0.97
Quartiled ADI Fully Adjusted Model (N = 797)			**Quartiled CIMD Fully Adjusted Model (N = 2362)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	0.95	0.53, 1.70	Quartile 2	1.79*	1.06, 3.10
Quartile 3	1.05	0.60, 1.86	Quartile 3	1.17	0.66, 2.08
Quartile 4	0.85	0.47, 1.55	Quartile 4	1.32	0.73, 2.38
Female sex	0.46	0.28, 0.72	Female sex	0.45	0.29, 0.67
Age at diagnosis	0.95	0.94, 0.97	Age at diagnosis	0.96	0.95, 0.98
Ever smoking	1.53	0.96, 2.50	Ever smoking	1.32	0.86, 2.04
Non-White race	0.23	0.08, 0.56	Non-White race	0.35	0.18, 0.62
Baseline FVC	0.99	0.97, 1.002	Baseline FVC	0.96	0.94, 0.97
Baseline D_LCO	0.96	0.95, 0.98	Baseline D_LCO	0.98	0.96, 0.99
IPF-Only U.S. Cohort			**IPF-Only Canadian Cohort**		
	OR	**95% CI**		**OR**	**95% CI**
Continuous ADI Unadjusted Model (N = 688)			**Continuous CIMD Unadjusted Model (N = 918)**		
Continuous ADI	0.996	0.988, 1.004	Continuous CIMD	0.66	0.39, 1.10
Continuous ADI Partially Adjusted Model (N = 688)			**Continuous CIMD Partially Adjusted Model (N = 912)**		
Continuous ADI	0.991	0.982, 0.9999	Continuous CIMD	0.74	0.42, 1.27

Female sex	0.69	0.44, 1.07	Female sex	0.72	0.38, 1.29
Age at diagnosis	0.89	0.87, 0.91	Age at diagnosis	0.89	0.86, 0.92
Continuous ADI Fully Adjusted Model (N = 444)			**Continuous CIMD Fully Adjusted Model (N = 591)**		
Continuous ADI	0.986*	0.975, 0.996	Continuous CIMD	1.1	0.50, 1.98
Female sex	0.69	0.38, 1.21	Female sex	0.71	0.30, 1.54
Age at diagnosis	0.89	0.87, 0.92	Age at diagnosis	0.90	0.86, 0.93
Ever smoking	1.43	0.79, 2.63	Ever smoking	0.59	0.29, 1.24
Non-White race	0.35	0.098, 1.004	Non-White race	0.80	0.24, 2.22
Baseline FVC	0.99	0.98, 1.01	Baseline FVC	0.95	0.93, 0.97
Baseline D_LCO	0.97	0.95, 0.99	Baseline D_LCO	1.00	0.98, 1.02
Quartiled ADI Unadjusted Model (N = 688)			**Quartiled CIMD Unadjusted Model (N = 918)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	1.21	0.72, 2.02	Quartile 2	1.57	0.84, 3.01
Quartile 3	1.18	0.71, 1.96	Quartile 3	0.81	0.39, 1.66
Quartile 4	0.79	0.45, 1.36	Quartile 4	0.64	0.28, 1.41
Quartiled ADI Partially Adjusted Model (N = 688)			**Quartiled CIMD Partially Adjusted Model (N = 912)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	1.11	0.63, 1.95	Quartile 2	1.81	0.93, 3.63
Quartile 3	1.06	0.61, 1.84	Quartile 3	0.96	0.44, 2.07
Quartile 4	0.59	0.32, 1.07	Quartile 4	0.75	0.31, 1.72
Female sex	0.71	0.45, 1.10	Female sex	0.72	0.38, 1.28
Age at diagnosis	0.89	0.87, 0.91	Age at diagnosis	0.89	0.86, 0.91
Quartiled ADI Fully Adjusted Model (N = 444)			**Quartiled CIMD Fully Adjusted Model (N = 591)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	1.18	0.59, 2.35	Quartile 2	2.31	0.96, 5.96
Quartile 3	0.83	0.42, 1.64	Quartile 3	1.49	0.57, 4.05
Quartile 4	0.46†	0.22, 0.95	Quartile 4	1.48	0.49, 4.40
Female sex	0.70	0.38, 1.25	Female sex	0.75	0.32, 1.63
Age at diagnosis	0.89	0.87, 0.92	Age at diagnosis	0.89	0.86, 0.93
Ever smoking	1.44	0.79, 2.68	Ever smoking	0.61	0.30, 1.30
Non-White race	0.36	0.10, 1.05	Non-White race	0.76	0.22, 2.13
Baseline FVC	0.99	0.98, 1.01	Baseline FVC	0.95	0.93, 0.97
Baseline D_LCO	0.97	0.95, 0.99	Baseline D_LCO	1.00	0.98, 1.02

*Interpretation: Patients with fILD living in quartile 2 of CIMD in the Canadian cohort have 1.79 higher odds of receiving lung transplant compared to patients living in quartile 1.

Appendix Table 7 – Linear regression models evaluating the impact of ADI or CIMD on baseline FVC.
Abbreviations: ADI, area deprivation index; CIMD, Canadian index of multiple deprivation; FVC, forced vital capacity.

Full U.S. Cohort			**Full Canadian Cohort**		
	β-Value	**95% CI**		**β-Value**	**95% CI**
Continuous ADI Unadjusted Model (N = 1078)			**Continuous CIMD Unadjusted Model (N = 2941)**		
Continuous ADI	-0.005*	-0.05, 0.05	Continuous CIMD	-0.97†	-2.31, 0.36
Continuous ADI Partially Adjusted Model (N = 1076)			**Continuous CIMD Partially Adjusted Model (N = 2938)**		
Continuous ADI	-0.006	-0.06, 0.04	Continuous CIMD	-1.31	-2.63, 0.02
Female sex	5.34	3.03, 7.65	Female sex	2.45	0.99, 3.90
Age at diagnosis	0.12	0.02, 0.22	Age at diagnosis	0.24	0.18, 0.30
Continuous ADI Fully Adjusted Model (N = 848)			**Continuous CIMD Fully Adjusted Model (N = 2929)**		
Continuous ADI	0.03	-0.03, 0.09	Continuous CIMD	-1.47	-2.79, -0.16
Female sex	7.65	4.93, 10.37	Female sex	3.08	1.62, 4.54
Age at diagnosis	0.14	0.02, 0.25	Age at diagnosis	0.23	0.17, 0.29
Former smoking	1.24	-1.57, 4.05	Former smoking	4.15	2.62, 5.68
Always smoking	11.48	4.56, 18.41	Always smoking	11.71	8.39, 15.03
Quartiled ADI Unadjusted Model (N = 1076)			**Quartiled CIMD Unadjusted Model (N = 2939)**		
Quartile 1	*Reference*		Quartile 1	*reference*	
Quartile 2	-3.12	-6.35, 0.11	Quartile 2	-0.30	-2.33, 1.73
Quartile 3	-1.43	-4.66, 1.81	Quartile 3	-1.96	-3.99, 0.07
Quartile 4	-1.29	-4.55, 1.98	Quartile 4	-1.40	-3.43, 0.63
Quartiled ADI Partially Adjusted Model (N = 1074)			**Quartiled CIMD Partially Adjusted Model (N = 2936)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	-3.42	-6.62, -0.22	Quartile 2	-0.76	-2.77, 1.25
Quartile 3	-1.39	-4.59, 1.81	Quartile 3	-2.57	-4.59, -0.56
Quartile 4	-1.35	-4.60, 1.91	Quartile 4	-1.90	-3.92, 0.11
Female sex	5.45	3.12, 7.75	Female sex	2.46	1.00, 3.91
Age at diagnosis	0.12	0.02, 0.22	Age at diagnosis	0.24	0.18, 0.30
Quartiled ADI Fully Adjusted Model (N = 846)			**Quartiled CIMD Fully Adjusted Model (N = 2927)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	-2.38	-6.00, 1.25	Quartile 2	-0.82	-2.82, 1.17
Quartile 3	0.83	-2.79, 4.45	Quartile 3	-2.63	-4.63, -0.63
Quartile 4	0.80	-2.89, 4.48	Quartile 4	-2.10	-4.10, -0.10
Female sex	7.78	5.06, 10.50	Female sex	3.08	1.63, 4.54
Age at diagnosis	0.13	0.02, 0.25	Age at diagnosis	0.23	0.17, 0.29
Former smoking	1.22	-1.58, 4.03	Former smoking	4.16	2.64, 5.69
Always smoking	11.24	4.32, 18.17	Always smoking	11.62	8.30, 14.93
IPF-Only U.S. Cohort			**IPF-Only Canadian Cohort**		
	β-Value	**95% CI**		**β-Value**	**95% CI**
Continuous ADI Unadjusted Model (N = 529)			**Continuous CIMD Unadjusted Model (N = 807)**		
Continuous ADI	0.04	-0.03, 0.12	Continuous CIMD	2.00	-0.56, 4.56
Continuous ADI Partially Adjusted Model (N = 527)			**Continuous CIMD Partially Adjusted Model (N = 805)**		
Continuous ADI	0.05	-0.02, 0.12	Continuous CIMD	1.23	-1.33, 3.79
Female sex	1.96	-1.43, 5.35	Female sex	4.63	1.71, 7.55
Age at diagnosis	0.31	0.14, 0.49	Age at diagnosis	0.22	0.07, 0.37
Continuous ADI Fully Adjusted Model (N = 479)			**Continuous CIMD Fully Adjusted Model (N = 802)**		
Continuous ADI	0.07	-0.001, 0.14	Continuous CIMD	1.24	-1.31, 3.79
Female sex	4.59	0.91, 8.27	Female sex	4.76	1.86, 7.65
Age at diagnosis	0.37	0.18, 0.55	Age at diagnosis	0.26	0.11, 0.41

Former smoking	4.45	0.72, 8.18	Former smoking	5.01	1.98, 8.04
Always smoking	20.37	9.27, 31.47	Always smoking	10.30	4.17, 16.43
Quartiled ADI Unadjusted Model (N = 527)			**Quartiled CIMD Unadjusted Model (N = 805)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	-1.33	-5.91, 3.25	Quartile 2	-1.12	-4.70, 2.46
Quartile 3	0.88	-3.61, 5.36	Quartile 3	-1.63	-5.19, 1.93
Quartile 4	2.83	-1.74, 7.41	Quartile 4	3.29	-0.51, 7.08
Quartiled ADI Partially Adjusted Model (N = 525)			**Quartiled CIMD Partially Adjusted Model (N = 803)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	-1.59	-6.13, 2.94	Quartile 2	-1.51	-5.07, 2.04
Quartile 3	0.89	-3.54, 5.33	Quartile 3	-2.21	-5.74, 1.33
Quartile 4	3.37	-1.17, 7.91	Quartile 4	2.26	-1.53, 6.06
Female sex	1.94	-1.46, 5.34	Female sex	4.60	1.69, 7.52
Age at diagnosis	0.32	0.15, 0.50	Age at diagnosis	0.23	0.07, 0.38
Quartiled ADI Fully Adjusted Model (N = 477)			**Quartiled CIMD Fully Adjusted Model (N = 800)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	-1.70	-6.41, 3.02	Quartile 2	-1.27	-4.79, 2.26
Quartile 3	1.71	-2.91, 6.33	Quartile 3	-2.49	-5.99, 1.02
Quartile 4	4.40	-0.32, 9.12	Quartile 4	2.38	-1.40, 6.16
Female sex	4.61	0.92, 8.29	Female sex	4.73	1.84, 7.62
Age at diagnosis	0.37	0.19, 0.56	Age at diagnosis	0.26	0.11, 0.41
Former smoking	4.51	0.78, 8.24	Former smoking	5.24	2.21, 8.28
Always smoking	19.42	8.27, 30.57	Always smoking	10.00	3.86, 16.14

*Interpretation: For each 1-point increase in ADI (ranging from 1-100) in U.S. cohort, the baseline percent predicted FVC is decreased by 0.005 units. This effect is non-significant.

†Interpretation: For each 1-point increase in CIMD (ranging from -1.3 to 2.57) in the Canadian cohort, the baseline percent predicted FVC is decreased by 0.97 units. This effect is non-significant.

Appendix Table 8 – Unadjusted and fully adjusted linear regression models evaluating the impact of ADI or CIMD on baseline D_LCO. Abbreviations: ADI, area deprivation index; CIMD, Canadian index of multiple deprivation; D_LCO, diffusion capacity for carbon monoxide.

U.S. Cohort			**Canadian Cohort**		
	β-Value	**95% CI**		**β-Value**	**95% CI**
Continuous ADI Unadjusted Model (N = 1007)			**Continuous CIMD Unadjusted Model (N = 2366)**		
Continuous ADI	-0.07	-0.12, -0.02	Continuous CIMD	-3.32	-4.78, -1.86
Continuous ADI Partially Adjusted Model (N = 1005)			**Continuous CIMD Partially Adjusted Model (N = 2364)**		
Continuous ADI	-0.07	-0.12, -0.02	Continuous CIMD	-3.24	-4.69, -1.78
Female sex	-3.57	-5.94, -1.20	Female sex	1.60	-0.01, 3.21
Age at diagnosis	-0.03	-0.14, 0.07	Age at diagnosis	-0.13	-0.19, -0.06
Continuous ADI Fully Adjusted Model (N = 795)			**Continuous CIMD Fully Adjusted Model (N = 2357)**		
Continuous ADI	-0.05	-0.11, 0.01	Continuous CIMD	-3.23†	-4.70, -1.77
Female sex	-4.86	-7.64, -2.08	Female sex	1.30	-0.32, 2.92
Age at diagnosis	-0.009	-0.13, 0.11	Age at diagnosis	-0.11	-0.18, -0.04
Former smoking	-6.68	-9.54, -3.82	Former smoking	-2.54	-4.27, -0.82
Always smoking	-1.79	-8.74, 5.16	Always smoking	-2.51	-6.08, 1.07
Quartiled ADI Unadjusted Model (N = 1005)			**Quartiled CIMD Unadjusted Model (N = 2364)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	-5.05	-8.34, -1.76	Quartile 2	-2.89	-5.12, -0.66
Quartile 3	-3.89	-7.16, -0.62	Quartile 3	-5.07	-7.29, -2.84
Quartile 4	-5.51	-8.83, -2.19	Quartile 4	-4.69	-6.92, -2.46
Quartiled ADI Partially Adjusted Model (N = 1003)			**Quartiled CIMD Partially Adjusted Model (N = 2362)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	-4.94	-8.22, -1.67	Quartile 2	-2.82	-5.04, -0.60
Quartile 3	-4.06	-7.31, -0.80	Quartile 3	-4.84	-7.06, -2.62
Quartile 4	-5.46	-8.79, -2.13	Quartile 4	-4.58	-6.81, -2.35
Female sex	-3.54	-5.91, -1.17	Female sex	1.57	-0.03, 3.18
Age at diagnosis	-0.03	-0.13, 0.07	Age at diagnosis	-0.13	-0.19, -0.06
Quartiled ADI Fully Adjusted Model (N = 793)			**Quartiled CIMD Fully Adjusted Model (N = 2355)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	-3.57	-7.28, 0.14	Quartile 2	-2.78	-5.00, -0.56
Quartile 3	-2.11	-5.79, 1.57	Quartile 3	-4.82	-7.04, -2.60
Quartile 4	-4.32*	-8.08, -0.55	Quartile 4	-4.57	-6.80, -2.34
Female sex	-4.84	-7.62, -2.06	Female sex	1.28	-0.35, 2.90
Age at diagnosis	-0.01	-0.13, 0.10	Age at diagnosis	-0.11	-0.18, -0.04
Former smoking	-6.77	-9.63, -3.91	Former smoking	-2.48	-4.21, -0.76
Always smoking	-1.90	-8.85, 5.05	Always smoking	-2.73	-6.30, 0.84
IPF-Only U.S. Cohort			**IPF-Only Canadian Cohort**		
	β-Value	**95% CI**		**β-Value**	**95% CI**
Continuous ADI Unadjusted Model (N = 487)			**Continuous CIMD Unadjusted Model (N = 591)**		
Continuous ADI	-0.03	-0.10, 0.05	Continuous CIMD	-3.09	-6.01, -0.17
Continuous ADI Partially Adjusted Model (N = 485)			**Continuous CIMD Partially Adjusted Model (N = 589)**		
Continuous ADI	-0.02	-0.09, 0.06	Continuous CIMD	-2.42	-5.38, 0.54
Female sex	-5.53	-9.04, -2.2	Female sex	0.36	-3.07, 3.78
Age at diagnosis	0.33	0.15, 0.51	Age at diagnosis	-0.23	-0.41, -0.06
Continuous ADI Fully Adjusted Model (N = 441)			**Continuous CIMD Fully Adjusted Model (N = 586)**		
Continuous ADI	-0.003	-0.08, 0.07	Continuous CIMD	-2.33	-5.29, 0.64
Female sex	-6.62	-10.46, -2.79	Female sex	0.26	-3.14, 3.67
Age at diagnosis	0.34	0.15, 0.53	Age at diagnosis	-0.25	-0.42, -0.07
Former smoking	-5.35	-9.20, -1.50	Former smoking	-5.87	-9.50, -2.24

Always smoking	-0.90	-12.37, 10.57	Always smoking	-8.33	-15.24, -1.43
Quartiled ADI Unadjusted Model (N = 485)			**Quartiled CIMD Unadjusted Model (N = 589)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	-3.23	-8.01, 1.55	Quartile 2	-1.17	-5.38, 3.03
Quartile 3	-2.29	-6.90, 2.31	Quartile 3	-3.74	-7.92, 0.45
Quartile 4	-2.69	-7.45, 2.07	Quartile 4	-2.84	-7.27, 1.58
Quartiled ADI Partially Adjusted Model (N = 483)			**Quartiled CIMD Partially Adjusted Model (N = 587)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	-3.47	-8.14, 1.20	Quartile 2	-1.10	-5.29, 3.09
Quartile 3	-2.98	-7.49, 1.54	Quartile 3	-3.20	-7.38, 0.99
Quartile 4	-1.80	-6.48, 2.87	Quartile 4	-2.04	-6.50, 2.42
Female sex	-5.63	-9.15, -2.11	Female sex	0.22	-3.21, 3.66
Age at diagnosis	0.34	0.16, 0.52	Age at diagnosis	-0.24	-0.41, -0.07
Quartiled ADI Fully Adjusted Model (N = 439)			**Quartiled CIMD Fully Adjusted Model (N = 584)**		
Quartile 1	*reference*		Quartile 1	*reference*	
Quartile 2	-3.09	-8.00, 1.83	Quartile 2	-1.39	-5.56, 2.77
Quartile 3	-1.37	-6.11, 3.38	Quartile 3	-2.84	-7.00, 1.32
Quartile 4	-1.16	-6.05, 3.74	Quartile 4	-1.95	-6.40, 2.51
Female sex	-6.70	-10.54, -2.85	Female sex	0.15	-3.27, 3.57
Age at diagnosis	0.34	0.15, 0.53	Age at diagnosis	-0.26	-0.43, -0.08
Former smoking	-5.42	-9.27, -1.56	Former smoking	-5.70	-9.36, -2.05
Always smoking	-1.37	-12.93, 10.19	Always smoking	-8.50	-15.45, -1.55

*Interpretation: Compared to quartile 1, living in quartile 4 of ADI in the U.S. cohort is associated with 4.32 unit lower baseline percent predicted D_LCO.

†Interpretation: For each 1-point increase in CIMD (ranging from -1.3 to 2.57) in the Canadian cohort, the baseline percent predicted D_LCO is decreased by 3.23 units.

Appendix Table 9 – Linear mixed models with random intercept and slope evaluating the impact of ADI or CIMD on rate of change in FVC. Abbreviations: ADI, area deprivation index; CIMD, Canadian index of multiple deprivation; FVC, forced vital capacity.

Full U.S. Cohort			**Full Canadian Cohort**		
	β-Value	**95% CI**		**β-Value**	**95% CI**
Continuous ADI Unadjusted Model (N = 1083)			**Continuous CIMD Unadjusted Model (N = 2942)**		
Continuous ADI*time	-0.003*	-0.02, 0.01	Continuous CIMD*time	-0.004	-0.34, 0.33
Continuous ADI Partially Adjusted Model (N = 1083)			**Continuous CIMD Partially Adjusted Model (N = 2941)**		
Continuous ADI*time	-0.006	-0.02, 0.008	Continuous CIMD*time	-0.02	-0.35, 0.31
Female sex*time	1.19	0.56, 1.82	Female sex*time	0.54	0.16, 0.91
Age at diagnosis*time	-0.07	-0.09, -0.04	Age at diagnosis*time	-0.02	-0.04, -0.008
Continuous ADI Fully Adjusted Model (N = 857)			**Continuous CIMD Fully Adjusted Model (N = 2934)**		
Continuous ADI*time	-0.009	-0.03, 0.008	Continuous CIMD*time	-0.04	-0.37, 0.30
Female sex*time	1.17	0.41, 1.93	Female sex*time	0.49	0.11, 0.87
Age at diagnosis*time	-0.05	-0.08, -0.02	Age at diagnosis*time	-0.02	-0.04, -0.005
Former smoking*time	0.14	-0.65, 0.93	Former smoking*time	-0.34	-0.73, 0.06
Always smoking*time	1.24	-0.57, 3.05	Always smoking*time	0.10	-0.77, 0.96
Quartiled ADI Unadjusted Model (N = 1083)			**Quartiled CIMD Unadjusted Model (N = 2942)**		
Quartile 1*time	*reference*		Quartile 1*time	*reference*	
Quartile 2*time	0.33	-0.56, 1.22	Quartile 2*time	0.40	-0.11, 0.92
Quartile 3*time	-0.45	-1.37, 0.46	Quartile 3*time	0.20	-0.32, 0.71
Quartile 4*time	0.02	-0.91, 0.95	Quartile 4*time	0.14	-0.36, 0.65
Quartiled ADI Partially Adjusted Model (N = 1083)			**Quartiled CIMD Partially Adjusted Model (N = 2941)**		
Quartile 1*time	*reference*		Quartile 1*time	*reference*	
Quartile 2*time	0.33	-0.52, 1.19	Quartile 2*time	0.39	-0.13, 0.90
Quartile 3*time	-0.24	-1.12, 0.64	Quartile 3*time	0.19	-0.33, 0.70
Quartile 4*time	-0.29	-1.19, 0.60	Quartile 4*time	0.12	-0.38, 0.62
Female sex*time	1.18	0.55, 1.82	Female sex*time	0.52	0.15, 0.90
Age at diagnosis*time	-0.07	-0.09, -0.04	Age at diagnosis*time	-0.02	-0.04, -0.008
Quartiled ADI Fully Adjusted Model (N = 857)			**Quartiled CIMD Fully Adjusted Model (N = 2934)**		
Quartile 1*time	*reference*		Quartile 1*time	*reference*	
Quartile 2*time	-0.04	-1.03, 0.95	Quartile 2*time	0.41	-0.11, 0.93
Quartile 3*time	-0.25	-1.26, 0.77	Quartile 3*time	0.20	-0.32, 0.72
Quartile 4*time	-0.54	-1.60, 0.51	Quartile 4*time	0.11	-0.40, 0.61
Female sex*time	1.18	0.41, 1.94	Female sex*time	0.47	0.09, 0.85
Age at diagnosis*time	-0.05	-0.08, -0.02	Age at diagnosis*time	-0.02	-0.04, -0.006
Former smoking*time	0.13	-0.66, 0.92	Former smoking*time	-0.35	-0.74, 0.05
Always smoking*time	1.23	-0.58, 3.04	Always smoking*time	0.10	-0.76, 0.96
IPF-Only U.S. Cohort			**IPF-Only Canadian Cohort**		
	β-Value	**95% CI**		**β-Value**	**95% CI**
Continuous ADI Unadjusted Model (N = 531)			**Continuous CIMD Unadjusted Model (N = 808)**		
Continuous ADI*time	-0.007	-0.03, 0.02	Continuous CIMD*time	0.12	-0.61, 0.85
Continuous ADI Partially Adjusted Model (N = 531)			**Continuous CIMD Partially Adjusted Model (N = 808)**		
Continuous ADI*time	-0.07	-0.03, 0.02	Continuous CIMD*time	0.02	-0.72, 0.76
Female sex*time	0.54	-0.51, 1.59	Female sex*time	0.09	-0.73, 0.92
Age at diagnosis*time	0.02	-0.04, 0.07	Age at diagnosis*time	0.03	-0.01, 0.08
Continuous ADI Fully Adjusted Model (N = 485)			**Continuous CIMD Fully Adjusted Model (N = 807)**		
Continuous ADI*time	-0.007	-0.03, 0.02	Continuous CIMD*time	-0.04	-0.79, 0.71
Female sex*time	0.82	-0.31, 1.95	Female sex*time	0.09	-0.74, 0.92
Age at diagnosis*time	0.03	-0.03, 0.08	Age at diagnosis*time	0.04	-0.005, 0.09
Former smoking*time	0.93	-0.24, 2.09	Former smoking*time	-0.07	-0.95, 0.81
Always smoking*time	3.11	-0.34, 6.57	Always smoking*time	1.35	-0.56, 3.26

Quartiled ADI Unadjusted Model (N = 531)			**Quartiled CIMD Unadjusted Model (N = 808)**		
Quartile 1*time	*reference*		Quartile 1*time	*reference*	
Quartile 2*time	-1.03	-2.42, 0.37	Quartile 2*time	-0.17	-1.18, 0.84
Quartile 3*time	-0.63	-1.99, 0.73	Quartile 3*time	0.15	-0.87, 1.17
Quartile 4*time	-0.59	-2.07, 0.90	Quartile 4*time	-0.34	-1.40, 0.73
Quartiled ADI Partially Adjusted Model (N = 531)			**Quartiled CIMD Partially Adjusted Model (N = 808)**		
Quartile 1*time	*reference*		Quartile 1*time	*reference*	
Quartile 2*time	-1.04	-2.45, 0.36	Quartile 2*time	-0.18	-1.20, 0.83
Quartile 3*time	-0.62	-1.99, 0.74	Quartile 3*time	0.12	-0.90, 1.14
Quartile 4*time	-0.53	-2.02, 0.95	Quartile 4*time	-0.47	-1.55, 0.61
Female sex*time	0.51	-0.54, 1.56	Female sex*time	0.13	-0.70, 0.96
Age at diagnosis*time	0.02	-0.03, 0.08	Age at diagnosis*time	0.04	-0.007, 0.08
Quartiled ADI Fully Adjusted Model (N = 485)			**Quartiled CIMD Fully Adjusted Model (N = 807)**		
Quartile 1*time	*reference*		Quartile 1*time	*reference*	
Quartile 2*time	-1.04	-2.50, 0.42	Quartile 2*time	-0.17	-1.19, 0.85
Quartile 3*time	-0.69	-2.09, 0.71	Quartile 3*time	0.11	-0.91, 1.14
Quartile 4*time	-0.63	-2.17, 0.90	Quartile 4*time	-0.57	-1.66, 0.52
Female sex*time	0.77	-0.37, 1.91	Female sex*time	0.13	-0.70, 0.96
Age at diagnosis*time	0.03	-0.03, 0.09	Age at diagnosis*time	0.04	-0.002, 009
Former smoking*time	0.89	-0.28, 2.06	Former smoking*time	-0.10	-0.98, 0.78
Always smoking*time	2.94	-0.53, 6.41	Always smoking*time	1.40	-0.52, 3.32

*Interpretation: for every 1-point increase in ADI (ranging from 1-100), the percent predicted FVC declines by 0.003 units more per year, although this effect is not significant.

†Interpretation: compared to males, the percent predicted FVC declines by 1.12 units less per year, i.e. female patients have slower decline in FVC.

‡Interpretation: for every 1-year increase in age at diagnosis, the percent predicted FVC declines by 0.05 units more per year, i.e. older patients experience more rapid decline in FVC.

Appendix Table 10 – Unadjusted and fully adjusted linear mixed models with random intercept and slope evaluating the impact of ADI or CIMD on rate of change in D_LCO. Abbreviations: ADI, area deprivation index; CIMD, Canadian index of multiple deprivation; D_LCO, diffusion capacity for carbon monoxide.

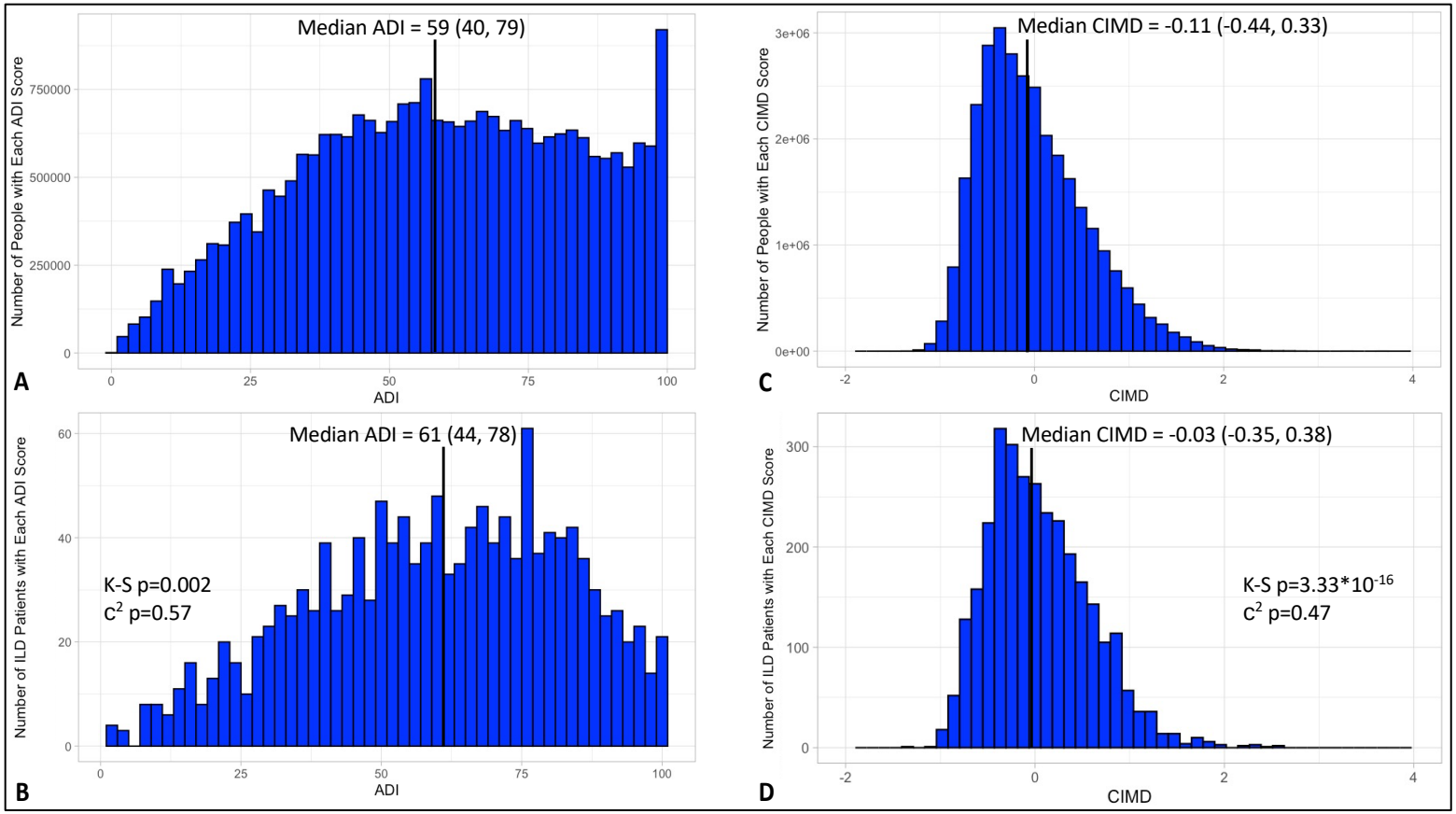
U.S. Cohort			**Canadian Cohort**		
	β-Value	**95% CI**		**β-Value**	**95% CI**
Continuous ADI Unadjusted Model (N = 1036)			**Continuous CIMD Unadjusted Model (N = 2758)**		
Continuous ADI*time	0.002*	-0.01, 0.02	Continuous CIMD*time	0.33	-0.04, 0.70
Continuous ADI Partially Adjusted Model (N = 1036)			**Continuous CIMD Partially Adjusted Model (N = 2758)**		
Continuous ADI*time	-0.003	-0.02, 0.01	Continuous CIMD*time	0.28	-0.08, 0.64
Female sex*time	1.64	1.01, 2.27	Female sex*time	1.62	1.22, 2.02
Age at diagnosis*time	-0.08	-0.10, -0.05	Age at diagnosis*time	-0.05	-0.06, -0.03
Continuous ADI Fully Adjusted Model (N = 822)			**Continuous CIMD Fully Adjusted Model (N = 2751)**		
Continuous ADI*time	-0.001	-0.02, 0.01	Continuous CIMD*time	0.26	-0.10, 0.62
Female sex*time	1.65	0.92, 2.38	Female sex*time	1.56	1.15, 1.96
Age at diagnosis*time	-0.06	-0.09, -0.03	Age at diagnosis*time	-0.05	-0.06, -0.03
Former smoking*time	-0.34	-1.10, 0.42	Former smoking*time	-0.49	-0.91, -0.07
Always smoking*time	0.99	-0.69, 2.66	Always smoking*time	-0.18	-1.09, 0.73
Quartiled ADI Unadjusted Model (N = 1036)			**Quartiled CIMD Unadjusted Model (N = 2758)**		
Quartile 1*time	*reference*		Quartile 1*time	*reference*	
Quartile 2*time	0.76	-0.13, 1.65	Quartile 2*time	0.59	0.02, 1.17
Quartile 3*time	-0.08	-1.00, 0.85	Quartile 3*time	0.69	0.11, 1.26
Quartile 4*time	0.49	-0.46, 1.44	Quartile 4*time	0.56	0.0004, 1.13
Quartiled ADI Partially Adjusted Model (N = 1036)			**Quartiled CIMD Partially Adjusted Model (N = 2758)**		
Quartile 1*time	*reference*		Quartile 1*time	*reference*	
Quartile 2*time	0.78	-0.07, 1.62	Quartile 2*time	0.51	-0.04, 1.06
Quartile 3*time	0.20	-0.68, 1.07	Quartile 3*time	0.63	0.08, 1.18
Quartile 4*time	0.05	-0.85, 0.95	Quartile 4*time	0.49	-0.05, 1.03
Female sex*time	1.64	1.01, 2.28	Female sex*time	1.60	1.20, 2.00
Age at diagnosis*time	-0.08	-0.11, -0.05	Age at diagnosis*time	-0.05	-0.07, -0.03
Quartiled ADI Fully Adjusted Model (N = 822)			**Quartiled CIMD Fully Adjusted Model (N = 2751)**		
Quartile 1*time	*reference*		Quartile 1*time	*reference*	
Quartile 2*time	0.61	-0.33, 1.55	Quartile 2*time	0.52	-0.03, 1.08
Quartile 3*time	0.34	-0.63, 1.31	Quartile 3*time	0.64	0.09, 1.19
Quartile 4*time	0.06	-0.96, 1.08	Quartile 4*time	0.46	-0.08, 1.01
Female sex*time	1.68	0.95, 2.42	Female sex*time	1.54	1.14, 1.94
Age at diagnosis*time	-0.07	-0.10, -0.04	Age at diagnosis*time	-0.05	-0.06, -0.03
Former smoking*time	-0.34	1.10, 0.42	Former smoking*time	-0.50	-0.93, -0.08
Always smoking*time	0.99	-0.69, 2.67	Always smoking*time	-0.15	-1.05, 0.76
IPF-Only U.S. Cohort			**IPF-Only Canadian Cohort**		
	β-Value	**95% CI**		**β-Value**	**95% CI**
Continuous ADI Unadjusted Model (N = 499)			**Continuous CIMD Unadjusted Model (N = 745)**		
Continuous ADI*time	0.006	-0.02, 0.03	Continuous CIMD*time	0.60	-0.16, 1.37
Continuous ADI Partially Adjusted Model (N = 499)			**Continuous CIMD Partially Adjusted Model (N = 745)**		
Continuous ADI*time	0.005	-0.02, 0.03	Continuous CIMD*time	0.49	-0.29, 1.26
Female sex*time	1.07	0.09, 2.05	Female sex*time	1.26	0.42, 2.10
Age at diagnosis*time	-0.02	-0.07, 0.03	Age at diagnosis*time	0.002	-0.04, 0.05
Continuous ADI Fully Adjusted Model (N = 457)			**Continuous CIMD Fully Adjusted Model (N = 744)**		
Continuous ADI*time	0.007	-0.02, 0.03	Continuous CIMD*time	0.43	-0.34, 1.21
Female sex*time	1.27	0.20, 2.34	Female sex*time	1.27	0.43, 2.11
Age at diagnosis*time	-0.01	-0.07, 0.04	Age at diagnosis*time	0.008	-0.04, 0.05
Former smoking*time	0.45	-0.65, 1.56	Former smoking*time	0.08	-0.82, 0.98
Always smoking*time	3.17	0.05, 6.28	Always smoking*time	1.60	-0.34, 3.55
Quartiled ADI Unadjusted Model (N = 499)			**Quartiled CIMD Unadjusted Model (N = 745)**		

Quartile 1*time	*reference*		Quartile 1*time	*reference*	
Quartile 2*time	0.10	-1.21, 1.41	Quartile 2*time	0.50	-0.55, 1.56
Quartile 3*time	0.41	-0.86, 1.68	Quartile 3*time	0.68	-0.38, 1.74
Quartile 4*time	0.30	-1.14, 1.73	Quartile 4*time	0.71	-0.39, 1.82
Quartiled ADI Partially Adjusted Model (N = 499)			**Quartiled CIMD Partially Adjusted Model (N = 745)**		
Quartile 1*time	*reference*		Quartile 1*time	*reference*	
Quartile 2*time	0.20	-1.12, 1.52	Quartile 2*time	0.32	-0.74, 1.37
Quartile 3*time	0.58	-0.69, 1.86	Quartile 3*time	0.53	-0.53, 1.59
Quartile 4*time	0.20	-1.24, 1.63	Quartile 4*time	0.52	-0.60, 1.63
Female sex*time	1.10	0.11, 2.10	Female sex*time	1.26	0.41, 2.11
Age at diagnosis*time	-0.03	-0.08, 0.03	Age at diagnosis*time	0.004	-0.04, 0.05
Quartiled ADI Fully Adjusted Model (N = 457)			**Quartiled CIMD Fully Adjusted Model (N = 744)**		
Quartile 1*time	*reference*		Quartile 1*time	*reference*	
Quartile 2*time	0.21	-1.16, 1.59	Quartile 2*time	0.34	-0.71, 1.40
Quartile 3*time	0.58	-0.73, 1.89	Quartile 3*time	0.50	-0.56, 1.55
Quartile 4*time	0.19	-1.29, 1.68	Quartile 4*time	0.45	-0.66, 1.57
Female sex*time	1.31	0.23, 2.39	Female sex*time	1.26	0.41, 2.11
Age at diagnosis*time	-0.02	-0.07, 0.04	Age at diagnosis*time	0.01	-0.04, 0.06
Former smoking*time	0.47	-0.65, 1.59	Former smoking*time	0.06	-0.84, 0.97
Always smoking*time	3.22	0.06, 6.37	Always smoking*time	1.62	-0.33, 3.57

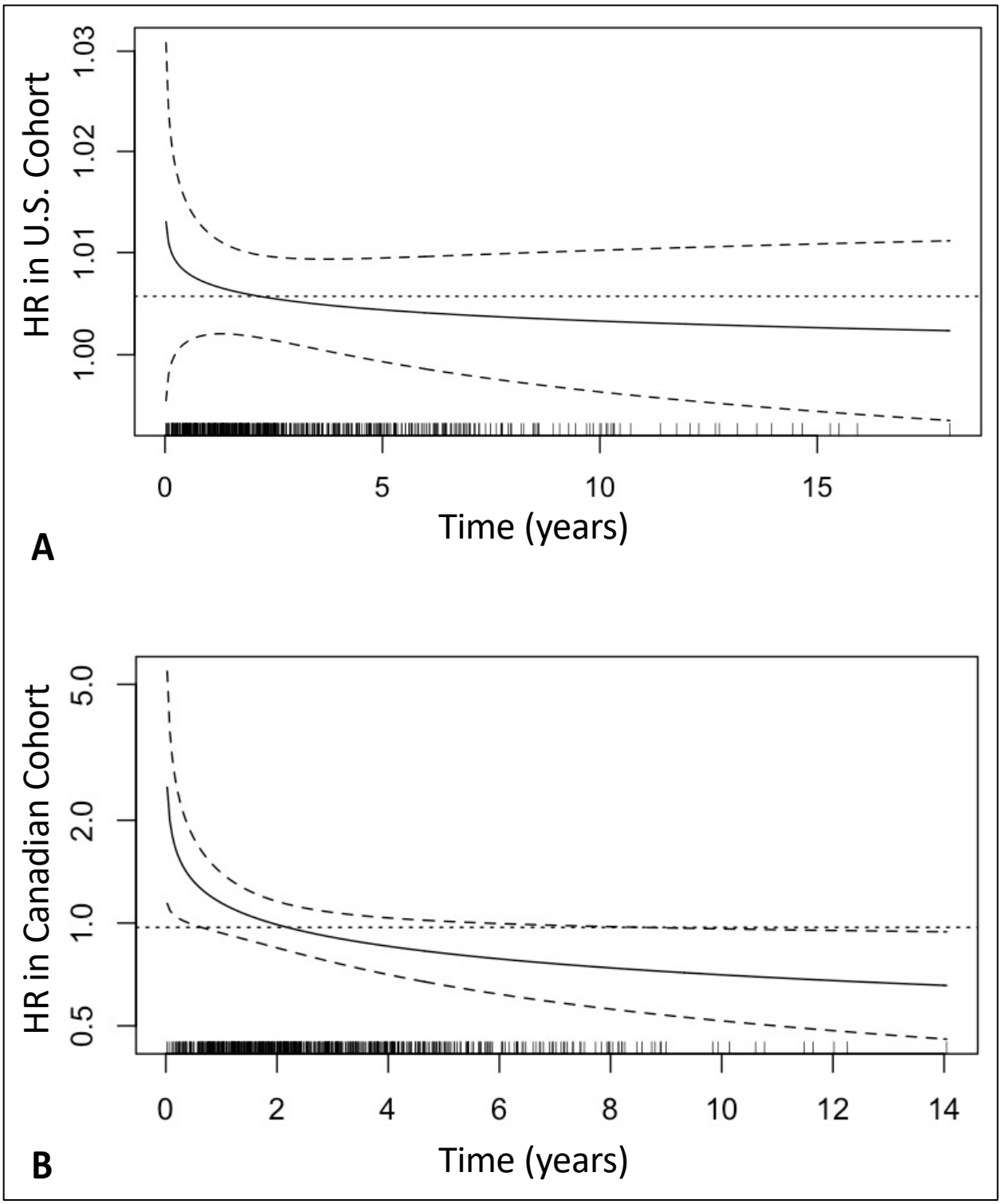
*Interpretation: for every 1-point increase in ADI (ranging from 1-100), the percent predicted D_LCO declines by 0.002 units less per year, although this effect is not significant.

†Interpretation: compared to males, the percent predicted D_LCO declines by 1.61 units less per year, i.e. female patients have slower decline in D_LCO.

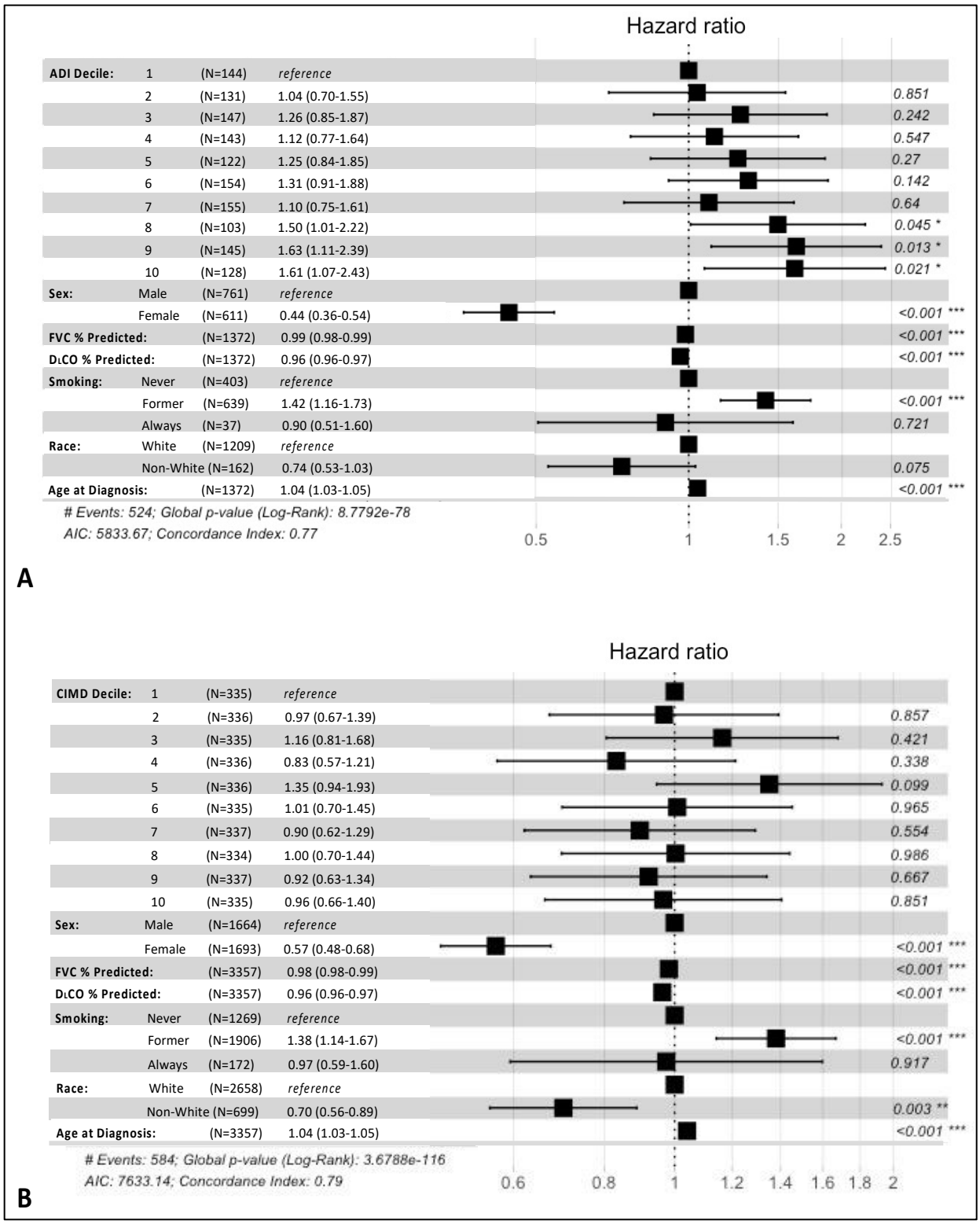
‡Interpretation: for every 1-year increase in age at diagnosis, the percent predicted D_LCO declines by 0.06 units more per year, i.e. older patients experience more rapid decline in D_LCO.



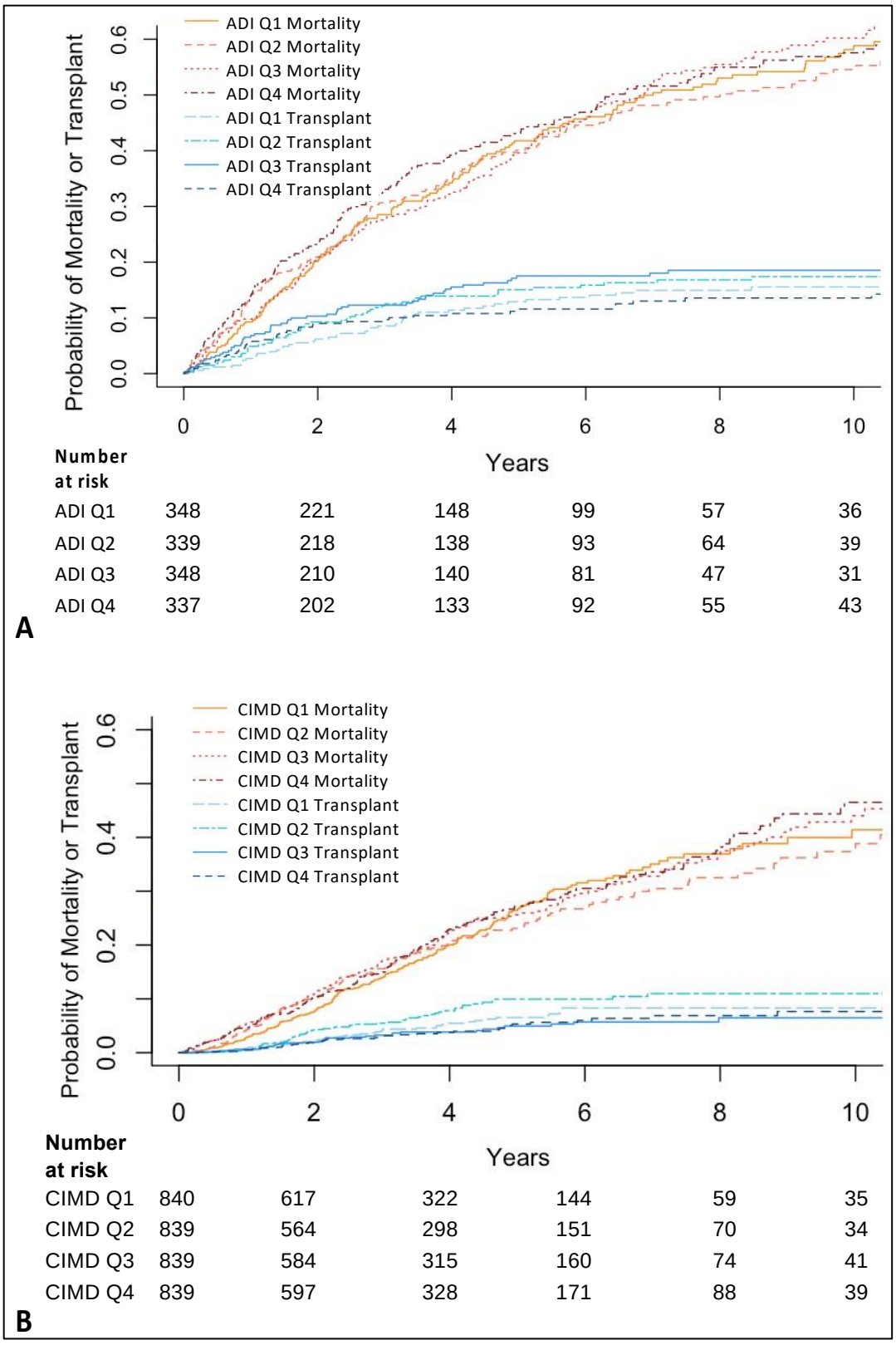
Appendix Figure 1 – Distribution of ADI or CIMD scores across referral areas and U.S. and Canadian cohorts. To determine if cohorts matched distributions of neighborhood disadvantage across referral regions, we performed Kolmorogov-Smirnov (K-S) tests on the continuous distribution and Chi-squared (χ^2) tests for goodness of fit after breaking the distributions of patients into deciles. For the U.S. cohort, we compared the distribution of ADI in the cohort of patients with fILD evaluated at the Simmons Center with the distribution of ADI across the states of Pennsylvania, Ohio, and West Virginia where over 90% of the referrals for this center reside. For the Canadian cohort, we compared the distribution of CIMD in the cohort of patients with fILD evaluated at CARE-PF sites with the distribution of CIMD across the provinces of British Columbia, Alberta, Saskatchewan, Ontario, and Quebec where over 99% of the referrals in this registry with available CIMD scores reside. **A)** Distribution of ADI across the entire population of Pennsylvania, Ohio, and West Virginia. **B)** Distribution of ADI across the Simmons Center cohort of patients with fibrotic ILD. **C)** Distribution of CIMD across the entire population of British Columbia, Alberta, Saskatchewan, Ontario, and Quebec. **D)** Distribution of CIMD across the CARE-PF cohort of patients with fibrotic ILD. The black vertical lines reflect the median ADI or CIMD, with interquartile range in brackets. The p-values in panels B and D reflect the statistical significance of the difference between distributions A and B or C and D by K-S test and χ^2 test. ADI, area deprivation index; CIMD, Canadian index of multiple deprivation; fILD, fibrotic interstitial lung disease.



Appendix Figure 2 – Hazard ratio (HR) over time for A) continuous area deprivation index (ADI) score in U.S. cohort and B) continuous Canadian index of multiple deprivation (CIMD) in Canadian cohort.



Appendix Figure 3 – Forest plots for Cox proportional hazards models evaluating the impact of neighborhood-level disadvantage decile on mortality (composite outcome of time to death or lung transplant) controlling for relevant covariates in A) U.S. and B) Canadian cohort. ADI, area deprivation index; AIC, Akaike information criterion; CIMD, Canadian index of multiple deprivation, FVC, forced vital capacity; D_LCO, diffusion capacity for carbon monoxide.



Appendix Figure 4 – Cumulative incidence curves for competing hazards baseline models with red lines for mortality and blue lines for lung transplant in A) U.S. and B) Canadian cohorts. Breakdown of full mortality, lung transplant and censoring outcomes reported in Table E4.

Appendix B – Data Supplement for $PM_{2.5}$ Impacts on Lung Function and Mortality in fILD

The following supplemental appendix has been accepted for publication as an Open Access manuscript in the *Journal of the American Medical Association Internal Medicine*, which allows for it to be reproduced in full in this dissertation.

Cite: Goobie, GC, *et al*. Association of particulate matter exposure with lung function and mortality in fibrotic interstitial lung disease: A multinational cohort study. *JAMA Intern Med* 2022.

Online-Only Material - Association of particulate matter exposure with lung function and mortality in fibrotic interstitial lung disease: A multinational cohort study

Gillian C. Goobie, MD[1,2,3]; Christopher Carlsten, MD MPH[4,5]; Kerri A. Johannson, MD, MPH[6]; Nasreen Khalil, MD[4]; Veronica Marcoux, MD[7]; Deborah Assayag, MD[8]; Hélène Manganas, MD[9]; Jolene H. Fisher, MD[10]; Martin R.J. Kolb, MD, PhD[11]; Kathleen O. Lindell, RN, PhD[2,12] James P. Fabisiak, PhD[13]; Xiaoping Chen, MS[2,14]; Kevin F. Gibson, MD[2,14]; Yingze Zhang, PhD[14]; Daniel J. Kass, MD[2,14]; Christopher J. Ryerson, MD, MAS*[4,15]; S. Mehdi Nouraie, MD, PhD*[14]

1. Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, U.S.A. Email: goobiegc@upmc.edu.
2. Simmons Center for Interstitial Lung Disease, Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A.
3. Clinician Investigator Program, Department of Medicine, University of British Columbia, Vancouver, B.C., Canada.
4. Division of Respiratory Medicine, Department of Medicine, University of British Columbia, Vancouver, B.C., Canada. Email: carlsten@mail.ubc.ca, chris.ryerson@hli.ubc.ca, nkhalil@mail.ubc.ca.
5. Air Pollution Exposure Laboratory, Vancouver Coastal Health Research Institute, Vancouver, B.C., Canada.
6. Division of Respiratory Medicine, Department of Medicine, University of Calgary, Calgary, AB, Canada. Email: kerri.johannson@ahs.ca.
7. Division of Respirology, Critical Care and Sleep Medicine, College of Medicine, University of Saskatchewan, Saskatoon, SK, Canada. Email: vsm940@mail.usask.ca.
8. Division of Respiratory Medicine, Department of Medicine, McGill University, Montreal, QC, Canada. Email: deborah.assayag@mcgill.ca.
9. Département de Médecine, Centre Hospitalier de l'Université de Montréal, Montreal, QC, Canada. Email: helene.manganas.med@ssss.gouv.qc.ca.
10. Division of Respirology, Department of Medicine, University of Toronto, Toronto, ON, Canada. Email: jolene.fisher@uhn.ca.
11. Department of Medicine, Firestone Institute for Respiratory Health, The Research Institute of St. Joe's Hamilton, St. Joseph's Healthcare, McMaster University, Hamilton, ON, Canada. Email: kolbm@mcmaster.ca.
12. College of Nursing, Medical University of South Carolina, Charleston, SC, U.S.A. Email: lindellk@musc.edu.
13. Department of Environmental and Occupational Health, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, U.S.A. Email: fabs@pitt.edu.
14. Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A. Emails: gibsonkf@upmc.edu, kassd2@upmc.edu, zhanyx@upmc.edu, nouraies@upmc.edu.
15. Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, B.C., Canada.

*Denotes co-senior authorship.

Correspondence to:
Gillian C. Goobie, MD FRCPC

Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh
130 DeSoto Street, Pittsburgh, PA, 15261
Email: goobiegc@upmc.edu and gcgoobie@alumni.ubc.ca; Phone: 412-624-3018; Fax: 412-624-3020

Appendix Table 11 – $PM_{2.5}$ and constituent component breakdown by study site. Breakdown of cohorts, enrollment sites within cohorts, number of patients recruited from each site, and median (interquartile range/IQR) of $PM_{2.5}$ and constituent exposures of patients at each site. BC, black carbon; IQR, interquartile range; NH_4^+, ammonium; NO_3^-, nitrate; OM, organic matter; $PM_{2.5}$, particulate matter with a diameter ≤2.5μm; SO_4^{2-}, sulfate; SS, sea salt.

Cohort	State or Province	City	Site Name	Patients Enrolled n (%)	$PM_{2.5}$ and Constituent Exposures – median (IQR) in μg/m³							
					$PM_{2.5}$	SO_4^{2-}	NO_3^-	NH_4^+	BC	OM	SS	Soil
Simmons Center for Interstitial Lung Disease	Pennsylvania	Pittsburgh	Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease at the University of Pittsburgh	1424 (100%)	9.4 (7.8-11.4)	2.1 (1.6-3.7)	0.9 (0.8-1.1)	0.8 (0.5-1.5)	0.8 (0.7-1.0)	3.2 (2.8-3.7)	0.2 (0.2-0.3)	0.5 (0.4-0.5)
Pulmonary Fibrosis Foundation (PFF)	Alabama	Birmingham	University of Alabama at Birmingham	94 (5.0%)	8.6 (8.2-9.3)	1.5 (1.4-1.6)	0.4 (0.3-0.5)	0.3 (0.2-0.3)	0.8 (0.7-0.8)	4.1 (3.8-4.5)	0.2 (0.2-0.3)	0.8 (0.7-0.8)
	Arizona	Phoenix	St. Joseph's Hospital	41 (2.2%)	8.4 (7.7-8.9)	0.7 (0.6-0.8)	0.6 (0.5-0.7)	0.2 (0.2-0.2)	0.5 (0.4-0.7)	2.8 (2.2-3.5)	0.3 (0.1-0.4)	2.0 (1.8-2.2)
		Tucson	University of Arizona – Banner Health	32 (1.7%)	4.8 (4.3-5.9)	0.5 (0.5-0.6)	0.2 (0.1-0.2)	0.1 (0.1-0.1)	0.2 (0.2-0.4)	1.3 (1.0-1.7)	0.5 (0.4-0.6)	1.4 (1.2-1.6)
	California	Emeryville	Stanford Health Center	60 (3.2%)	8.8 (6.9-10.5)	0.7 (0.6-0.8)	1.1 (0.8-1.6)	0.3 (0.2-0.4)	0.6 (0.4-0.7)	3.3 (2.2-4.1)	1.6 (1.1-2.0)	0.5 (0.3-0.6)
		Los Angeles	University of California at Los Angeles	40 (2.1%)	12.7 (11.7-13.4)	1.2 (1.1-1.2)	2.5 (2.2-2.6)	0.8 (0.7-0.9)	1.1 (0.8-1.3)	5.0 (4.4-5.4)	0.7 (0.6-0.8)	1.0 (1.0-1.1)
		San Francisco	University of California at San Francisco	54 (2.9%)	7.9 (6.6-9.8)	0.7 (0.5-0.7)	0.9 (0.6-1.3)	0.2 (0.2-0.3)	0.5 (0.3-0.6)	2.9 (2.2-3.7)	1.5 (0.8-2.1)	0.4 (0.3-0.5)
	Colorado	Denver	National Jewish Health	45 (2.4%)	6.1 (3.4-8.1)	0.5 (0.4-0.6)	0.5 (0.2-1.0)	0.1 (0.1-0.2)	0.5 (0.2-0.8)	2.3 (1.1-3.8)	0.1 (0.01-0.2)	0.9 (0.5-1.4)
	Connecticut	New Haven	Yale School of Medicine	37 (2.0%)	6.8 (6.2-7.8)	0.9 (0.9-1.2)	0.6 (0.6-0.9)	0.2 (0.2-0.3)	0.5 (0.4-0.7)	2.2 (1.9-3.4)	0.4 (0.3-0.4)	0.2 (0.2-0.4)
	Florida	Miami	University of Miami	28 (1.5%)	7.1 (6.8-7.3)	0.8 (0.8-0.9)	0.3 (0.3-0.3)	0.1 (0.1-0.1)	0.3 (0.3-0.3)	1.6 (1.5-2.0)	1.3 (1.1-1.4)	0.9 (0.8-1.1)
	Georgia	Atlanta	Piedmont Healthcare	58 (3.1%)	9.4 (8.9-9.7)	1.4 (1.4-1.6)	0.5 (0.4-0.6)	0.2 (0.2-0.3)	0.8 (0.8-0.9)	4.2 (3.9-4.5)	0.2 (0.2-0.2)	0.7 (0.6-0.7)
	Illinois	Chicago	Northwestern Memorial Hospital	52 (2.8%)	9.6 (8.8-10.0)	1.2 (1.2-1.3)	1.4 (1.3-1.6)	0.4 (0.4-0.5)	0.7 (0.7-0.8)	2.8 (2.6-3.0)	0.3 (0.2-0.4)	0.6 (0.5-0.7)

		Chicago	University of Chicago	37 (2.0%)	9.0 (8.6-9.3)	1.2 (1.2-1.3)	1.4 (1.4-1.5)	0.4 (0.4-0.5)	0.7 (0.6-0.7)	2.6 (2.5-2.7)	0.2 (0.2-0.3)	0.6 (0.5-0.6)
	Kansas	Kansas City	The University of Kansas Hospital	28 (1.5%)	7.0 (6.7-7.2)	1.2 (1.1-1.3)	0.8 (0.7-0.9)	0.3 (0.3-0.4)	0.5 (0.5-0.5)	2.6 (2.5-2.8)	0.1 (0.1-0.1)	0.6 (0.5-0.6)
	Kentucky	Louisville	University of Louisville School of Medicine	34 (1.8%)	9.1 (8.4-9.7)	1.6 (1.4-1.8)	1.0 (0.9-1.1)	0.5 (0.4-0.6)	0.7 (0.6-0.8)	3.0 (2.7-3.9)	0.1 (0.1-0.2)	0.4 (0.4-0.6)
	Louisiana	New Orleans	Tulane University School of Medicine	48 (2.6%)	8.4 (8.1-8.6)	1.3 (1.3-1.4)	0.4 (0.3-0.5)	0.2 (0.2-0.3)	0.6 (0.6-0.6)	3.3 (3.2-3.6)	0.1 (0.1-0.2)	0.8 (0.8-0.9)
	Maryland	Baltimore	Johns Hopkins University	12 (0.6%)	8.0 (7.4-8.6)	1.4 (1.3-1.6)	0.9 (0.8-1.0)	0.5 (0.4-0.5)	0.6 (0.6-0.7)	2.7 (2.6-2.8)	0.3 (0.2-0.4)	0.3 (0.3-0.3)
		Baltimore	University of Maryland Medical Center	35 (1.9%)	8.6 (8.3-9.1)	1.6 (1.5-1.6)	1.0 (1.0-1.1)	0.5 (0.5-0.6)	0.7 (0.7-0.7)	2.8 (2.7-3.0)	0.3 (0.2-0.3)	0.3 (0.3-0.4)
	Massachusetts	Boston	Massachusetts General Hospital	42 (2.3%)	6.0 (5.4-6.9)	0.9 (0.7-0.9)	0.5 (0.4-0.6)	0.2 (0.1-0.2)	0.4 (0.4-0.5)	2.2 (1.8-2.7)	0.4 (0.3-0.5)	0.1 (0.1-0.3)
	Michigan	Ann Arbor	University of Michigan Health System	68 (3.6%)	8.0 (7.6-8.5)	1.2 (1.1-1.4)	1.3 (1.2-1.4)	0.5 (0.4-0.6)	0.6 (0.5-0.6)	2.8 (2.6-3.2)	0.3 (0.2-0.3)	0.5 (0.5-0.7)
	Minnesota	Minneapolis	University of Minnesota Medical Center	64 (3.4%)	7.0 (6.7-7.2)	1.0 (0.9-1.0)	1.2 (1.1-1.3)	0.4 (0.3-0.4)	0.4 (0.4-0.5)	2.1 (2.0-2.3)	0.1 (0.1-0.2)	0.4 (0.4-0.4)
		Rochester	Mayo Clinic	56 (3.0%)	6.7 (6.2-7.2)	0.9 (0.8-1.1)	1.0 (0.8-1.2)	0.3 (0.2-0.4)	0.4 (0.4-0.5)	2.1 (1.8-2.3)	0.2 (0.2-0.3)	0.4 (0.3-0.5)
	Missouri	St. Louis	Washington University School of Medicine	33 (1.8%)	9.0 (8.0-9.4)	1.6 (1.4-1.7)	1.1 (0.9-1.2)	0.5 (0.4-0.6)	0.6 (0.6-0.7)	3.0 (2.8-3.1)	0.4 (0.3-0.5)	0.6 (0.6-0.7)
	New York	New York	Columbia University Medical Center	42 (2.3%)	7.8 (7.1-8.1)	1.2 (1.1-1.3)	1.0 (0.7-1.2)	0.4 (0.3-0.4)	0.6 (0.5-0.8)	2.6 (2.2-3.2)	0.5 (0.3-0.6)	0.4 (0.3-0.5)
		Setauket	Stony Brook University Hospital	16 (0.9%)	7.2 (6.8-7.5)	1.1 (1.0-1.2)	0.7 (0.7-0.7)	0.2 (0.2-0.3)	0.5 (0.5-0.6)	2.3 (2.1-2.5)	0.6 (0.6-0.8)	0.3 (0.3-0.3)
		Rochester	University of Rochester Medical Center	64 (3.4%)	6.9 (6.4-7.2)	1.1 (1.0-1.3)	0.7 (0.7-0.8)	0.3 (0.3-0.5)	0.5 (0.5-0.5)	2.3 (2.2-2.5)	0.3 (0.2-0.3)	0.3 (0.3-0.3)

	New York	Weill-Cornell Medical Center	57 (3.1%)	7.6 (7.2-8.3)	1.0 (0.9-1.1)	1.0 (0.8-1.2)	0.3 (0.2-0.4)	0.8 (0.6-0.8)	2.8 (2.4-3.2)	0.4 (0.3-0.4)	0.5 (0.3-0.6)
North Carolina	Durham	Duke University Medical Center	50 (2.7%)	7.5 (6.9-8.0)	1.3 (1.2-1.4)	0.4 (0.4-0.5)	0.2 (0.2-0.3)	0.6 (0.6-0.7)	3.3 (3.0-3.5)	0.3 (0.2-0.3)	0.4 (0.4-0.4)
Ohio	Columbus	The Ohio State University Wexner Medical Center	54 (2.9%)	8.0 (7.5-8.4)	1.3 (1.3-1.6)	1.2 (1.1-1.3)	0.2 (0.1-0.2)	0.7 (0.7-0.7)	3.1 (3.0-3.3)	0.2 (0.2-0.3)	0.5 (0.5-0.6)
	Cincinnati	University of Cincinnati Medical Center	40 (2.1%)	8.9 (8.4-9.6)	1.7 (1.5-1.9)	1.1 (1.0-1.2)	0.5 (0.1-0.6)	0.7 (0.7-0.8)	3.0 (2.9-3.5)	01 (0.1-0.2)	0.4 (0.4-0.5)
Pennsylvania	Hershey	Penn State Milton S. Hershey Medical Center	49 (2.6%)	8.5 (7.9-9.0)	1.4 (1.2-1.6)	1.2 (1.1-1.3)	0.5 (0.4-0.6)	0.7 (0.6-0.7)	3.0 (2.7-3.2)	0.3 (0.2-0.3)	0.4 (0.3-0.4)
	Philadelphia	Temple Health	66 (3.5%)	8.5 (8.0-9.1)	1.3 (1.1-1.5)	1.1 (1.0-1.2)	0.4 (0.4-0.6)	0.6 (0.6-0.8)	2.5 (2.3-3.0)	0.4 (0.3-0.5)	0.3 (0.3-0.5)
	Philadelphia	University of Pennsylvania	10 (0.5%)	8.0 (7.5-8.9)	1.3 (1.2-1.4)	1.0 (0.8-1.2)	0.4 (0.4-0.5)	0.6 (0.6-0.7)	2.5 (2.3-3.0)	0.2 (0.2-0.3)	0.4 (0.2-0.5)
	Pittsburgh	University of Pittsburgh	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded	Excluded
South Carolina	Charleston	Medical University of South Carolina	36 (1.9%)	7.8 (7.6-8.0)	1.3 (1.2-1.3)	0.3 (0.3-0.4)	0.2 (0.1-0.2)	0.6 (0.6-0.7)	3.3 (3.2-3.7)	0.4 (0.3-0.5)	0.5 (0.5-0.5)
Tennessee	Nashville	Vanderbilt University Medical Center	51 (2.7%)	8.3 (7.9-8.9)	1.6 (1.5-1.8)	0.6 (0.5-0.8)	0.4 (0.4-0.5)	0.6 (0.6-0.7)	3.1 (2.9-3.6)	0.1 (0.1-0.2)	0.5 (0.5-0.6)
Texas	Dallas	University of Texas Southwestern Medical Center	64 (3.4%)	8.3 (8.0-8.8)	1.5 (1.5-1.6)	0.6 (0.6-0.7)	0.4 (0.3-0.4)	0.6 (0.5-0.6)	3.0 (2.8-3.2)	0.3 (0.3-0.3)	1.1 (1.1-1.2)
	Houston	University of Texas Health Science Center	45 (2.4%)	8.2 (6.9-8.6)	1.6 (1.3-1.8)	0.5 (0.4-0.5)	0.4 (0.3-0.4)	0.5 (0.4-0.5)	2.6 (2.0-2.8)	0.7 (0.6-0.8)	1.1 (1.0-1.2)
	San Antonio	University of Texas Health Science Center	54 (2.9%)	7.6 (7.1-7.9)	1.6 (1.5-1.7)	0.5 (0.4-0.6)	0.4 (0.4-0.4)	0.5 (0.5-0.5)	2.4 (2.2-2.7)	0.5 (0.4-0.6)	1.2 (1.0-1.3)
Utah	Salt Lake City	University of Utah Health Care	26 (1.4%)	5.6 (4.8-8.4)	0.5 (0.4-0.6)	0.5 (0.4-1.4)	0.2 (0.1-0.3)	0.4 (0.3-0.7)	2.5 (1.4-3.5)	0.1 (0.0-0.2)	0.6 (0.6-1.1)

	Virginia	Fairfax	Inova Fairfax Medical Campus	85 (4.6%)	7.6 (7.3-8.1)	1.3 (1.2-1.4)	0.8 (0.7-0.9)	0.3 (0.3-0.4)	0.6 (0.6-0.7)	2.8 (2.7-3.0)	0.1 (0.1-0.1)	0.3 (0.3-0.4)
		Charlottesville	University of Virginia Health Systems	40 (2.1%)	5.9 (5.4-7.0)	1.2 (1.0-1.3)	0.4 (0.4-0.5)	0.2 (0.2-0.3)	0.5 (0.5-0.6)	2.5 (2.3-2.7)	0.1 (0.1-0.2)	0.3 (0.2-0.3)
	Washington	Seattle	University of Washington Medical Center	23 (1.2%)	6.3 (5.3-6.8)	0.4 (0.3-0.4)	0.4 (0.3-0.4)	0.1 (0.1-0.1)	0.4 (0.3-0.6)	2.6 (1.9-3.7)	0.2 (0.1-0.2)	0.2 (0.1-0.3)
Canadian Registry for Pulmonary Fibrosis (CARE-PF)	Alberta	Calgary	University of Calgary	356 (10.5%)	5.6 (4.9-6.7)	0.4 (0.4-0.5)	0.5 (0.3-0.8)	0.1 (0.1-0.2)	0.3 (0.3-0.6)	2.3 (1.9-3.5)	0.1 (0.0-0.1)	0.3 (0.3-0.4)
	British Columbia	Vancouver	St. Paul's Hospital, University of British Columbia	1159 (34.2%)	5.6 (4.7-6.3)	0.4 (0.3-0.5)	0.4 (0.2-0.5)	0.1 (0.1-0.1)	0.4 (0.3-0.7)	2.5 (1.8-4.2)	0.3 (0.2-0.3)	0.2 (0.1-0.3)
		Vancouver	Vancouver General Hospital, University of British Columbia	298 (8.8%)	5.8 (5.2-6.4)	0.4 (0.3-0.4)	0.4 (0.3-0.5)	0.1 (0.1-0.1)	0.4 (0.3-0.7)	2.8 (2.2-4.3)	0.3 (0.2-0.3)	0.2 (0.2-0.3)
	Ontario	Hamilton	Firestone Institute for Respiratory Health, McMaster University	449 (13.3%)	7.9 (7.4-8.7)	1.2 (1.1-1.4)	1.0 (1.0-1.1)	0.4 (0.4-0.5)	0.6 (0.5-0.7)	2.7 (2.4-3.4)	0.3 (0.3-0.3)	0.5 (0.4-0.7)
		Toronto	University of Toronto	353 (10.4%)	7.1 (6.1-7.9)	1.0 (0.9-1.2)	0.9 (0.7-1.0)	0.4 (0.3-0.4)	0.5 (0.4-0.6)	2.2 (2.0-3.1)	0.3 (0.2-0.3)	0.4 (0.3-0.6)
	Quebec	Montreal	McGill University	85 (2.5%)	6.8 (5.9-7.5)	0.9 (0.8-0.9)	0.7 (0.5-0.8)	0.3 (0.2-0.3)	0.5 (0.4-0.5)	3.1 (2.5-3.5)	0.2 (0.2-0.2)	0.3 (0.2-0.4)
		Montreal	Centre Hospitalier de l'Université de Montréal	604 (17.8%)	6.9 (6.0-7.7)	0.9 (0.8-1.0)	0.7 (0.5-0.8)	0.3 (0.2-0.3)	0.5 (0.4-0.5)	3.0 (2.5-3.6)	0.2 (0.2-0.2)	0.3 (0.2-0.4)
	Saskatchewan	Saskatoon	University of Saskatchewan	85 (2.5%)	5.5 (5.1-5.9)	0.5 (0.5-0.6)	0.6 (0.5-0.8)	0.2 (0.1-0.2)	0.3 (0.3-0.4)	2.1 (1.9-2.4)	0.1 (0.0-0.1)	0.4 (0.3-0.4)

Appendix Table 12 – Demographic characteristics for cohorts split by low versus high $PM_{2.5}$ exposures (<8 or ≥8μg/m³) in the 5-years pre-censoring.

Cohort characteristic	Simmons Cohort		PFF Cohort		CARE-PF Cohort	
	Low (<8μg/m³)	High (≥8μg/m³)	Low (<8μg/m³)	High (≥8μg/m³)	Low (<8μg/m³)	High (≥8μg/m³)
	N=394	N=1030	N=958	N=900	N=2960	N=406
Age at diagnosis, median (IQR), years	64 (55,71)	67 (59,73)	67 (61,73)	68 (60,73)	66 (57, 72)	68 (59, 74)
Male sex, n (%)	184 (47%)	611 (59%)	605 (63%)	574 (64%)	1433 (48%)	227 (56%)
Self-reported race, n (%)						
White	361 (92%)	897 (87%)	879 (92%)	778 (86%)	2307 (78%)	358 (88%)
Black	5 (1%)	51 (5%)	31 (3%)	65 (7%)	44 (1%)	8 (2%)
Asian	1 (0.3%)	4 (0.4%)	24 (2%)	24 (3%)	356 (12%)	25 (6%)
Indigenous[a]	0 (0%)	2 (0.2%)	2 (0.2%)	1 (0.1%)	81 (3%)	5 (1%)
Pacific Islander	0 (0%)	0 (0%)	2 (0.2%)	1 (0.1%)	19 (1%)	7 (2%)
Unknown	27 (7%)	76 (7%)	20 (2%)	31 (4%)	153 (5%)	3 (1%)
Smoking history, n (%)						
Never	133 (34%)	280 (27%)	417 (44%)	359 (40%)	1149 (39%)	120 (30%)
Former	139 (35%)	525 (51%)	"Ever" 541 (56%)	"Ever" 541 (60%)	1651 (56%)	261 (64%)
Current	14 (4%)	24 (2%)			152 (5%)	23 (6%)
Unknown	108 (27%)	201 (20%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Residential location, n (%)						
Metropolitan (>50,000 people)	227 (58%)	851 (83%)	769 (80%)	836 (93%)	1993 (67%)	320 (79%)
Micropolitan (10,000-50,000)	99 (25%)	124 (12%)	93 (10%)	41 (5%)	535 (18%)	62 (15%)
Rural (<10,000 people)	68 (17%)	54 (5%)	95 (10%)	23 (2%)	432 (15%)	24 (6%)
ILD diagnostic group, n(%)						
IPF	119 (30%)	597 (58%)	625 (65%)	571 (63%)	759 (26%)	160 (39%)
CTD-ILD	102 (26%)	198 (19%)	145 (15%)	163 (18%)	1174 (40%)	116 (29%)
fHP	24 (6%)	31 (3%)	91 (10%)	59 (7%)	243 (8%)	15 (4%)
Pneumoconiosis	11 (3%)	15 (2%)	0 (0%)	0 (0%)	22 (1%)	6 (1%)
Non-IPF IIP	25 (6%)	43 (4%)	70 (7%)	73 (8%)	87 (3%)	21 (5%)
Other ILD[b]	15 (4%)	35 (3%)	0 (0%)	0 (0%)	105 (3%)	15 (4%)
Unclassifiable/NYD	98 (25%)	111 (11%)	27 (3%)	34 (4%)	570 (19%)	73 (18%)
ADI or CIMD, median (IQR)	68 (50,81)	60 (41,77)	N/A	N/A	-0.04 (-0.36, 0.36)	0.04 (-0.29, 0.51)
Percent of 5-digit zip below poverty, median (IQR)	N/A	N/A	8% (5-13%)	9% (6-14%)	N/A	N/A
Baseline FVC % Predicted, median (IQR)[c]	74 (60,86)	64 (51,78)	68 (56, 81)	66 (54, 79)	76 (62, 90)	71 (58, 85)
Baseline D_LCO % Predicted, median (IQR)[d]	57 (43,67)	47 (34,61)	41 (33, 52)	39 (30, 50)	58 (44, 72)	55 (42, 70)

Follow-up Duration, median (IQR), years	4.35 (1.78, 7.87)	2.61 (1.06, 5.46)	2.57 (1.59, 3.51)	2.34 (1.21, 3.44)	3.17 (1.76, 5.24)	3.22 (1.83, 4.69)
Cause of censoring, n (%)						
Death	65 (16%)	642 (62%)	203 (21%)	223 (25%)	642 (22%)	257 (64%)
Lung Transplantation	22 (6%)	179 (17%)	129 (13%)	126 (14%)	150 (5%)	123 (30%)
Lost to follow-up or censored by dataextraction	307 (78%)	209 (20%)	626 (66%)	551 (61%)	2168 (73%)	26 (6%)

[a] – Includes Native American, American Indian, Alaskan First Nations, & other Indigenous persons in the U.S.; First Nations, Métis, Inuit, and other Indigenous persons in Canada.
[b] – Includes drug-, radiation-, aspiration-, or acute lung injury-induced fILD.
[c] – FVC was available for 76% of patients enrolled in Simmons, 91% of PFF, and 88% of CARE-PF.
[d] – D_LCO was available for 70% of patients enrolled in Simmons, 85% of PFF, and 70% of CARE-PF
Abbreviations: ADI, area deprivation index; CIMD, Canadian Index of Multiple Deprivation; CARE-PF, Canadian Registry for Pulmonary Fibrosis; D_LCO, diffusion capacity of lung for carbon monoxide; fILD, fibrotic interstitial lung disease; FVC, forced vital capacity; IPF, idiopathic pulmonary fibrosis; IQR, interquartile range; PFF, Pulmonary Fibrosis Foundation.

Appendix Table 13 – Full cohort mortality models. Unadjusted and adjusted Cox models for continuous and dichotomized $PM_{2.5}$ and constituent components (sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic matter (OM), sea salt (SS), soil) in the 5-years pre-censoring (defined as time of death, lung transplant, or cessation of follow-up). Adjustments made for age at enrollment, sex, smoking history, race, a socioeconomic variable, and site (for PFF and CARE-PF). Significant associations are bolded.

Cohort	Pollutant	Method	Model	HR	2.5 CI	97.5 CI	p	n
Simmons	$PM_{2.5}$	continuous	unadjusted	**1.34**	**1.30**	**1.37**	**<0.001**	1416
Simmons	$PM_{2.5}$	continuous	adjusted	**1.33**	**1.29**	**1.36**	**<0.001**	1372
PFF	$PM_{2.5}$	continuous	unadjusted	**1.06**	**1.00**	**1.12**	**0.04**	1858
PFF	$PM_{2.5}$	continuous	adjusted	**1.20**	**1.10**	**1.31**	**<0.001**	1832
CARE-PF	$PM_{2.5}$	continuous	unadjusted	0.99	0.95	1.03	0.58	3364
CARE-PF	$PM_{2.5}$	continuous	adjusted	1.00	0.96	1.05	0.89	3353
Simmons	$PM_{2.5}$	dichotomized	unadjusted	**4.80**	**3.85**	**5.99**	**<0.001**	1416
Simmons	$PM_{2.5}$	dichotomized	adjusted	**4.40**	**3.51**	**5.51**	**<0.001**	1372
PFF	$PM_{2.5}$	dichotomized	unadjusted	**1.24**	**1.03**	**1.50**	**0.02**	1858
PFF	$PM_{2.5}$	dichotomized	adjusted	**1.71**	**1.32**	**2.21**	**<0.001**	1832
CARE-PF	$PM_{2.5}$	dichotomized	unadjusted	**1.43**	**1.20**	**1.71**	**<0.001**	3364
CARE-PF	$PM_{2.5}$	dichotomized	adjusted	**1.45**	**1.18**	**1.79**	**<0.001**	3353
Simmons	SO_4^{2-}	continuous	unadjusted	**1.81**	**1.73**	**1.89**	**<0.001**	1416
Simmons	SO_4^{2-}	continuous	adjusted	**1.79**	**1.70**	**1.88**	**<0.001**	1372
PFF	SO_4^{2-}	continuous	unadjusted	**3.23**	**2.42**	**4.31**	**<0.001**	1858
PFF	SO_4^{2-}	continuous	adjusted	**132.19**	**78.12**	**223.70**	**<0.001**	1832
CARE-PF	SO_4^{2-}	continuous	unadjusted	**1.81**	**1.58**	**2.07**	**<0.001**	3364
CARE-PF	SO_4^{2-}	continuous	adjusted	**2.26**	**2.05**	**2.48**	**<0.001**	3353
Simmons	SO_4^{2-}	dichotomized	unadjusted	**18.94**	**4.73**	**75.87**	**<0.001**	1416
Simmons	SO_4^{2-}	dichotomized	adjusted	**15.34**	**3.83**	**61.50**	**<0.001**	1372
PFF	SO_4^{2-}	dichotomized	unadjusted	**1.41**	**1.15**	**1.74**	**0.001**	1858
PFF	SO_4^{2-}	dichotomized	adjusted	**5.46**	**3.79**	**7.86**	**<0.001**	1832
CARE-PF	SO_4^{2-}	dichotomized	unadjusted	**1.75**	**1.52**	**2.01**	**<0.001**	3364
CARE-PF	SO_4^{2-}	dichotomized	adjusted	**4.97**	**3.83**	**6.45**	**<0.001**	3353
Simmons	NO_3^-	continuous	unadjusted	**3.50**	**3.07**	**3.99**	**<0.001**	1416
Simmons	NO_3^-	continuous	adjusted	**3.59**	**3.10**	**4.16**	**<0.001**	1372
PFF	NO_3^-	continuous	unadjusted	**1.34**	**1.11**	**1.62**	**0.002**	1858
PFF	NO_3^-	continuous	adjusted	**2.48**	**1.74**	**3.53**	**<0.001**	1832
CARE-PF	NO_3^-	continuous	unadjusted	**1.52**	**1.22**	**1.90**	**<0.001**	3364
CARE-PF	NO_3^-	continuous	adjusted	**6.26**	**4.16**	**9.42**	**<0.001**	3353
Simmons	NO_3^-	dichotomized	unadjusted	**2.94**	**2.34**	**3.70**	**<0.001**	1416
Simmons	NO_3^-	dichotomized	adjusted	**2.68**	**2.12**	**3.40**	**<0.001**	1372
PFF	NO_3^-	dichotomized	unadjusted	1.14	0.94	1.37	0.18	1858
PFF	NO_3^-	dichotomized	adjusted	**1.63**	**1.15**	**2.30**	**0.006**	1832
CARE-PF	NO_3^-	dichotomized	unadjusted	**1.19**	**1.04**	**1.36**	**0.01**	3364
CARE-PF	NO_3^-	dichotomized	adjusted	**1.69**	**1.34**	**2.13**	**<0.001**	3353

Simmons	NH_4^+	continuous	unadjusted	**4.39**	**3.93**	**4.91**	**<0.001**	1416
Simmons	NH_4^+	continuous	adjusted	**4.31**	**3.83**	**4.86**	**<0.001**	1372
PFF	NH_4^+	continuous	unadjusted	**15.36**	**9.47**	**24.90**	**<0.001**	1858
PFF	NH_4^+	continuous	adjusted	**903.17**	**408.4**	**1998.0**	**<0.001**	1832
CARE-PF	NH_4^+	continuous	unadjusted	**7.67**	**5.35**	**11.01**	**<0.001**	3364
CARE-PF	NH_4^+	continuous	adjusted	**36.22**	**27.32**	**48.03**	**<0.001**	3353
Simmons	NH_4^+	dichotomized	unadjusted	**7.94**	**3.77**	**16.70**	**<0.001**	1416
Simmons	NH_4^+	dichotomized	adjusted	**6.50**	**3.08**	**13.72**	**<0.001**	1372
PFF	NH_4^+	dichotomized	unadjusted	**1.98**	**1.62**	**2.41**	**<0.001**	1858
PFF	NH_4^+	dichotomized	adjusted	**5.05**	**3.72**	**6.84**	**<0.001**	1832
CARE-PF	NH_4^+	dichotomized	unadjusted	**1.61**	**1.40**	**1.84**	**<0.001**	3364
CARE-PF	NH_4^+	dichotomized	adjusted	**5.11**	**3.92**	**6.66**	**<0.001**	3353
Simmons	BC	continuous	unadjusted	0.86	0.62	1.21	0.39	1416
Simmons	BC	continuous	adjusted	0.88	0.62	1.25	0.48	1372
PFF	BC	continuous	unadjusted	0.86	0.51	1.46	0.58	1858
PFF	BC	continuous	adjusted	1.14	0.54	2.41	0.74	1832
CARE-PF	BC	continuous	unadjusted	0.73	0.50	1.06	0.1	3364
CARE-PF	BC	continuous	adjusted	0.73	0.49	1.08	0.12	3353
Simmons	BC	dichotomized	unadjusted	**1.41**	**1.05**	**1.90**	**0.02**	1416
Simmons	BC	dichotomized	adjusted	1.26	0.93	1.70	0.14	1372
PFF	BC	dichotomized	unadjusted	0.92	0.77	1.11	0.4	1858
PFF	BC	dichotomized	adjusted	0.98	0.74	1.30	0.89	1832
CARE-PF	BC	dichotomized	unadjusted	1.02	0.88	1.17	0.83	3364
CARE-PF	BC	dichotomized	adjusted	0.95	0.81	1.11	0.51	3353
Simmons	OM	continuous	unadjusted	1.02	0.92	1.13	0.67	1416
Simmons	OM	continuous	adjusted	1.03	0.93	1.14	0.63	1372
PFF	OM	continuous	unadjusted	**0.89**	**0.80**	**0.99**	**0.04**	1858
PFF	OM	continuous	adjusted	**0.86**	**0.74**	**1.00**	**0.05**	1832
CARE-PF	OM	continuous	unadjusted	**1.06**	**1.00**	**1.13**	**0.05**	3364
CARE-PF	OM	continuous	adjusted	**1.08**	**1.02**	**1.15**	**0.01**	3353
Simmons	OM	dichotomized	unadjusted	0.99	0.86	1.14	0.93	1416
Simmons	OM	dichotomized	adjusted	1.04	0.90	1.20	0.62	1372
PFF	OM	dichotomized	unadjusted	0.84	0.70	1.01	0.07	1858
PFF	OM	dichotomized	adjusted	**0.77**	**0.60**	**0.99**	**0.04**	1832
CARE-PF	OM	dichotomized	unadjusted	1.05	0.92	1.19	0.48	3364
CARE-PF	OM	dichotomized	adjusted	1.08	0.95	1.24	0.25	3353
Simmons	SS	continuous	unadjusted	**0.00**	**0.00**	**0.01**	**<0.001**	1416
Simmons	SS	continuous	adjusted	**0.01**	**0.00**	**0.02**	**<0.001**	1372
PFF	SS	continuous	unadjusted	0.88	0.67	1.15	0.36	1858
PFF	SS	continuous	adjusted	**0.62**	**0.39**	**1.00**	**0.05**	1832
CARE-PF	SS	continuous	unadjusted	**4.16**	**2.53**	**6.83**	**<0.001**	3364

CARE-PF	SS	continuous	adjusted	**4.24**	**2.16**	**8.34**	**<0.001**	3353
Simmons	SS	dichotomized	unadjusted	**0.39**	**0.34**	**0.45**	**<0.001**	1416
Simmons	SS	dichotomized	adjusted	**0.43**	**0.37**	**0.50**	**<0.001**	1372
PFF	SS	dichotomized	unadjusted	0.84	0.70	1.01	0.07	1858
PFF	SS	dichotomized	adjusted	**0.71**	**0.55**	**0.93**	**0.01**	1832
CARE-PF	SS	dichotomized	unadjusted	**1.49**	**1.28**	**1.73**	**<0.001**	3364
CARE-PF	SS	dichotomized	adjusted	**1.41**	**1.18**	**1.69**	**<0.001**	3353
Simmons	Soil	continuous	unadjusted	**3.79**	**2.44**	**5.91**	**<0.001**	1416
Simmons	Soil	continuous	adjusted	**3.09**	**1.95**	**4.87**	**<0.001**	1372
PFF	Soil	continuous	unadjusted	0.97	0.76	1.24	0.81	1858
PFF	Soil	continuous	adjusted	1.31	0.76	2.25	0.34	1832
CARE-PF	Soil	continuous	unadjusted	0.77	0.53	1.13	0.18	3364
CARE-PF	Soil	continuous	adjusted	0.84	0.51	1.38	0.49	3353
Simmons	Soil	dichotomized	unadjusted	**1.54**	**1.32**	**1.81**	**<0.001**	1416
Simmons	Soil	dichotomized	adjusted	**1.50**	**1.28**	**1.76**	**<0.001**	1372
PFF	Soil	dichotomized	unadjusted	0.91	0.75	1.12	0.38	1858
PFF	Soil	dichotomized	adjusted	0.91	0.69	1.21	0.52	1832
CARE-PF	Soil	dichotomized	unadjusted	0.94	0.82	1.08	0.38	3364
CARE-PF	Soil	dichotomized	adjusted	0.93	0.78	1.12	0.45	3353
Simmons	Multi-Constituent	continuous	unadjusted	**2.13**	**1.89**	**2.40**	**<0.001**	1416
Simmons	Multi-Constituent	continuous	adjusted	**2.19**	**1.93**	**2.48**	**<0.001**	1372
PFF	Multi-Constituent	continuous	unadjusted	**1.19**	**1.01**	**1.41**	**0.03**	1858
PFF	Multi-Constituent	continuous	adjusted	**2.76**	**2.15**	**3.54**	**<0.001**	1832
CARE-PF	Multi-Constituent	continuous	unadjusted	**1.47**	**1.34**	**1.62**	**<0.001**	3364
CARE-PF	Multi-Constituent	continuous	adjusted	**2.30**	**2.02**	**2.61**	**<0.001**	3353

Appendix Table 14 – Meta-analysis results. Random effects meta-analysis results where hazard ratio (HR) or beta value (β) reported represents the result from the meta-analysis of the pooled effect estimates from each of the three cohorts (Simmons, PFF, CARE-PF). Associations of $PM_{2.5}$, constituents (sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic matter (OM), sea salt (SS), soil), and multi-constituent model containing all constituent exposures in the 5-years pre-censoring with mortality are reported. Associations of $PM_{2.5}$, constituents, and multi-constituent exposures in the 5-years pre-enrollment with baseline forced vital capacity (FVC) and diffusion capacity of the lung for carbon monoxide (D_LCO) are reported. Associations of $PM_{2.5}$ and constituents with rate of decline in FVC and D_LCO are reported. All models are adjusted for age at enrollment, sex, smoking history, race, a socioeconomic variable, and site (in PFF and CARE-PF). I^2 is reported for all models, which is a measure of heterogeneity with its own 95% confidence interval (CI) reported. Significant effect estimates are **bolded**.

Pollutant	Model	HR/β	2.5 CI	97.5 CI	p	n	I^2 (%)	I^2 2.5 CI (%)	I^2 97.5 CI (%)
Mortality Models – Meta-Analysis of Pooled Effect Estimates from Simmons, PFF, and CARE-PF									
$PM_{2.5}$	**Adjusted**	**1.18**	**1.02**	**1.37**	**0.03**		98	96	99
SO_4^{2-}	**Adjusted**	8.02	0.52	122.63	0.13		99	99	99.5
NO_3^-	**Adjusted**	**3.78**	**2.30**	**6.20**	**<0.001**		82	46	94
NH_4^+	**Adjusted**	**50.99**	**2.46**	**1056.64**	**0.01**		99	99	99.6
BC	**Adjusted**	0.84	0.66	1.08	0.17	6557	0	0	90
OM	**Adjusted**	1.00	0.88	1.13	0.995		73	11	92
SS	**Adjusted**	0.27	0.01	10.77	0.49		99	98	99
Soil	**Adjusted**	1.51	0.71	3.23	0.29		87	61	95
Multi-Constituent	**Adjusted**	**2.30**	**2.11**	**2.50**	**<0.001**		25	0	92
Baseline FVC Models – Meta-Analysis of Pooled Effect Estimates from Simmons, PFF, and CARE-PF									
$PM_{2.5}$	**Adjusted**	-0.30	-1.00	0.41	0.41		82	44	94
SO_4^{2-}	**Adjusted**	-0.90	-2.52	0.73	0.28		65	0	90
NO_3^-	**Adjusted**	-0.29	-3.78	3.20	0.87		73	10	92
NH_4^+	**Adjusted**	-0.85	-6.28	4.58	0.76		87	62	95
BC	**Adjusted**	-1.42	-5.14	2.30	0.46		43	0	83
OM	**Adjusted**	-0.53	-2.53	1.48	0.61	5678	86	58	95
SS	**Adjusted**	-8.92	-27.74	9.91	0.35		91	76	97
Soil	**Adjusted**	1.16	-5.83	8.16	0.74		68	0	91
Multi-Constituent	**Adjusted**	**-3.37**	**-4.88**	**-1.87**	**<0.001**		56	0	87
Baseline D_LCO Models – Meta-Analysis of Pooled Effect Estimates from Simmons, PFF, and CARE-PF									
$PM_{2.5}$	**Adjusted**	-0.32	-0.84	0.19	0.22		64	0	90
SO_4^{2-}	**Adjusted**	-1.43	-3.87	1.02	0.25		88	68	96
NO_3^-	**Adjusted**	-0.14	-1.76	1.47	0.86		0	0	90
NH_4^+	**Adjusted**	-1.54	-6.88	3.80	0.57		88	67	96
BC	**Adjusted**	-2.23	-5.68	1.23	0.21		39	0	81
OM	**Adjusted**	-0.35	-0.82	0.11	0.14	4908	0	0	90
SS	**Adjusted**	-4.77	-16.34	6.80	0.42		78	29	93
Soil	**Adjusted**	-0.74	-4.37	2.88	0.69		0	0	90
Multi-Constituent	**Adjusted**	**-3.64**	**-4.61**	**-2.66**	**<0.001**		9	0	91
FVC Decline Models – Meta-Analysis of Pooled Effect Estimates from Simmons, PFF, and CARE-PF									
$PM_{2.5}$	**Adjusted**	-0.15	-0.42	0.12	0.29		90	74	96
SO_4^{2-}	**Adjusted**	**-2.53**	**-4.45**	**-0.62**	**0.01**		92	81	97
NO_3^-	**Adjusted**	**-1.72**	**-2.86**	**-0.58**	**0.003**	5167	67	0	90
NH_4^+	**Adjusted**	**-5.93**	**-10.18**	**-1.69**	**0.006**		95	90	98

BC	**Adjusted**	0.70	-0.21	1.60	0.13		0	0	90
OM	**Adjusted**	0.10	-0.07	0.27	0.23		0	0	90
SS	**Adjusted**	2.13	-2.49	6.74	0.37		89	70	96
Soil	**Adjusted**	0.46	-0.94	1.85	0.52		48	0	85
D_LCO Decline Models – Meta-Analysis of Pooled Effect Estimates from Simmons, PFF, and CARE-PF									
$PM_{2.5}$	**Adjusted**	-0.05	-0.31	0.21	0.70	4878	88	65	96
SO_4^{2-}	**Adjusted**	**-2.12**	**-3.93**	**-0.30**	**0.02**		88	68	96
NO_3^-	**Adjusted**	-1.21	-2.66	0.24	0.10		75	15	92
NH_4^+	**Adjusted**	**-4.66**	**-8.77**	**-0.54**	**0.03**		93	84	97
BC	**Adjusted**	**1.12**	**0.16**	**2.08**	**0.02**		0	0	90
OM	**Adjusted**	0.14	-0.04	0.32	0.13		0	0	90
SS	**Adjusted**	2.34	-2.97	7.65	0.39		91	76	97
Soil	**Adjusted**	0.55	-0.52	1.61	0.31		19	0	92

Appendix Table 15 – Multi-constituent model results. The hazard ratio (HR) or beta-value (β), 95% confidence interval (CI), and p-value are reported for a one quantile increase in the overall mixture of exposure to all $PM_{2.5}$ constituents (sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic matter (OM), sea salt (SS), soil) in the 5-years pre-censoring for mortality models and the 5-years pre-enrollment for baseline forced vital capacity (FVC) and diffusion capacity of the lung for carbon monoxide (D_LCO) models. Effect estimates are provided for each constituent, where the sum of all negative effects adds to -1 and the sum of all positive effects adds to +1, so positive and negative effects cannot be directly compared. Positive effects are harmful (i.e. increased hazard) in mortality models, but beneficial (i.e. higher baseline FVC or D_LCO) in baseline lung function models. Adjusted models include covariates of age at enrollment, sex, smoking history, race, a socioeconomic variable, and site (in PFF and CARE-PF). Significant overall mixture effects are **bolded**.

Cohort	Model	Mixture HR/β	2.5 CI	97.5 CI	p	n	SO_4^{2-}	NO_3^-	NH_4^+	BC	OM	SS	Soil
Mortality Analyses – Full Cohort													
Simmons	**Unadjusted**	**2.13**	**1.89**	**2.40**	**<0.001**	1416	0.42	0.11	0.40	-0.85	0.07	-0.09	-0.06
Simmons	**Adjusted**	**2.19**	**1.93**	**2.48**	**<0.001**	1372	0.39	0.10	0.42	-0.91	0.07	-0.09	0.02
PFF	**Unadjusted**	**1.19**	**1.01**	**1.41**	**0.03**	1858	0.26	-0.37	0.74	-0.16	-0.23	-0.11	-0.12
PFF	**Adjusted**	**2.76**	**2.15**	**3.54**	**<0.001**	1832	0.49	-0.29	0.46	-0.23	-0.31	-0.16	0.04
CARE-PF	**Unadjusted**	**1.47**	**1.34**	**1.62**	**<0.001**	3364	0.25	-0.37	0.29	-0.25	0.27	0.19	-0.38
CARE-PF	**Adjusted**	**2.30**	**2.02**	**2.61**	**<0.001**	3353	0.49	-0.22	0.24	-0.27	0.18	0.09	-0.52
Baseline FVC Analyses – Full Cohort													
Simmons	**Unadjusted**	**-4.94**	**-6.67**	**-3.22**	**<0.001**	1073	-0.29	-0.01	-0.42	0.78	-0.28	0.08	0.14
Simmons	**Adjusted**	**-4.44**	**-6.20**	**-2.69**	**<0.001**	1048	-0.32	-0.04	-0.38	0.71	-0.26	0.16	0.13
PFF	**Unadjusted**	**-2.78**	**-4.21**	**-1.35**	**<0.001**	1696	-0.66	-0.04	-0.15	0.36	0.64	-0.11	-0.04
PFF	**Adjusted**	-1.61	-3.66	0.44	0.12	1672	-0.61	-0.11	-0.29	0.36	0.14	0.26	0.24
CARE-PF	**Unadjusted**	**-3.38**	**-4.34**	**-2.42**	**<0.001**	2967	-0.43	0.14	-0.09	-0.27	0.06	-0.21	0.80
CARE-PF	**Adjusted**	**-3.75**	**-5.13**	**-2.37**	**<0.001**	2958	-0.43	0.20	-0.22	-0.06	-0.14	-0.15	0.80
Baseline D_LCO Analyses – Full Cohort													
Simmons	**Unadjusted**	**-3.82**	**-5.61**	**-2.03**	**<0.001**	998	-0.48	-0.01	-0.29	0.94	-0.18	0.06	-0.03
Simmons	**Adjusted**	**-4.14**	**-5.96**	**-2.33**	**<0.001**	978	-0.56	-0.05	-0.26	0.99	-0.09	0.01	-0.03
PFF	**Unadjusted**	**-1.37**	**-2.72**	**-0.03**	**0.045**	1570	-0.76	0.34	-0.18	0.02	0.05	-0.06	0.59
PFF	**Adjusted**	**-2.40**	**-4.31**	**-0.48**	**0.01**	1547	-0.65	0.20	-0.35	0.17	0.15	0.18	0.29
CARE-PF	**Unadjusted**	**-2.03**	**-3.10**	**-0.96**	**<0.001**	2389	-0.14	0.15	-0.03	0.35	-0.50	-0.33	0.50
CARE-PF	**Adjusted**	**-4.02**	**-5.47**	**-2.57**	**<0.001**	2383	-0.51	0.01	-0.14	0.29	-0.30	-0.06	0.71

Appendix Table 16 – Cohort attributable risk fractions for $PM_{2.5}$, SO_4^{2-}, NO_3^-, and NH_4^+. Cohort attributable risk fraction for mortality for high exposures compared to low exposures to $PM_{2.5}$ and the constituents of sulfate (SO_4^{2-}), nitrate (NO_3^-), and ammonium (NH_4^+) in the 5-years pre-censoring. Hazard ratios used for attributable risk fraction calculations are from adjusted models (covariates including age at enrollment, sex, smoking history, race, a socioeconomic variable, and site (for PFF and CARE-PF)).

	Simmons	**PFF**	**CARE-PF**
	N=1416	**N=1858**	**N=3364**
High $PM_{2.5}$ (>8μg/m^3)	0.71[a]	0.26	0.05[a]
High SO_4^{2-} (>1.04μg/m^3)	0.93	0.75	0.44
High NO_3^- (>0.71μg/m^3)	0.57	0.25	0.17
High NH_4^+ (>0.31μg/m^3)	0.84	0.70	0.52

[a] – Interpretation: if high exposures to $PM_{2.5}$ were removed (i.e. all patients had 5-year pre-censoring $PM_{2.5}$ average <8μg/m^3), then 71% of the mortality in the Simmons cohort could be avoided, as compared to only 5% of the mortality in the CARE-PF cohort. These estimates must be interpreted with caution as they do not include consideration of interactions with other environmental, clinical, and molecular prognostic factors.

Appendix Table 17 – IPF subgroup analyses for primary outcomes. Results of adjusted models for associations of continuous $PM_{2.5}$ or constituents (sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic matter (OM), sea salt (SS), soil) in the 5-years pre-censoring with mortality and rate of decline in forced vital capacity (FVC) or diffusion capacity for carbon monoxide (D_LCO) and in the 5-years pre-enrollment with baseline FVC or D_LCO. Adjustments made for age at enrollment, sex, smoking history, race, a socioeconomic variable, and site (for PFF and CARE-PF). Significant associations are **bolded.**

Cohort	Pollutant	Method	Model	HR/β	2.5 CI	97.5 CI	p	n
Mortality Models								
Simmons	$PM_{2.5}$	continuous	adjusted	**1.25**	**1.20**	**1.29**	**<0.001**	688
PFF	$PM_{2.5}$	continuous	adjusted	**1.30**	**1.17**	**1.44**	**<0.001**	1181
CARE-PF	$PM_{2.5}$	continuous	adjusted	1.01	0.95	1.07	0.80	916
Simmons	SO_4^{2-}	continuous	adjusted	**1.59**	**1.49**	**1.69**	**<0.001**	688
PFF	SO_4^{2-}	continuous	adjusted	**405.60**	**206.40**	**797.30**	**<0.001**	1181
CARE-PF	SO_4^{2-}	continuous	adjusted	**55.41**	**29.48**	**104.13**	**<0.001**	916
Simmons	NO_3^-	continuous	adjusted	**2.41**	**1.99**	**2.92**	**<0.001**	688
PFF	NO_3^-	continuous	adjusted	**3.35**	**2.22**	**5.05**	**<0.001**	1181
CARE-PF	NO_3^-	continuous	adjusted	**6.16**	**3.30**	**11.48**	**<0.001**	916
Simmons	NH_4^+	continuous	adjusted	**3.15**	**2.70**	**3.68**	**<0.001**	688
PFF	NH_4^+	continuous	adjusted	**$2.20*10^3$**	**794.9**	**$6.08*10^3$**	**<0.001**	1181
CARE-PF	NH_4^+	continuous	adjusted	**$9.00*10^3$**	**$2.45*10^3$**	**$3.30*10^4$**	**<0.001**	916
Simmons	BC	continuous	adjusted	1.17	0.75	1.82	0.49	688
PFF	BC	continuous	adjusted	1.60	0.67	3.81	0.29	1181
CARE-PF	BC	continuous	adjusted	0.60	0.31	1.17	0.13	916
Simmons	OM	continuous	adjusted	1.11	0.98	1.25	0.09	688
PFF	OM	continuous	adjusted	0.90	0.75	1.07	0.22	1181
CARE-PF	OM	continuous	adjusted	1.04	0.94	1.16	0.46	916
Simmons	SS	continuous	adjusted	**0.09**	**0.04**	**0.24**	**<0.001**	688
PFF	SS	continuous	adjusted	0.70	0.39	1.25	0.23	1181
CARE-PF	SS	continuous	adjusted	0.99	0.23	4.23	0.99	916
Simmons	Soil	continuous	adjusted	**2.07**	**1.17**	**3.67**	**0.01**	688
PFF	Soil	continuous	adjusted	1.35	0.70	2.61	0.38	1181
CARE-PF	Soil	continuous	adjusted	0.68	0.32	1.46	0.32	916
Baseline FVC Models								
Simmons	$PM_{2.5}$	continuous	adjusted	-0.54	-1.23	0.16	0.13	517
PFF	$PM_{2.5}$	continuous	adjusted	-0.35	-1.08	0.38	0.35	1068
CARE-PF	$PM_{2.5}$	continuous	adjusted	-0.24	-1.29	0.81	0.66	810
Simmons	SO_4^{2-}	continuous	adjusted	-1.19	-2.50	0.12	0.08	517
PFF	SO_4^{2-}	continuous	adjusted	**-2.34**	**-4.26**	**-0.42**	**0.02**	1068
CARE-PF	SO_4^{2-}	continuous	adjusted	**5.50**	**0.69**	**10.30**	**0.02**	810
Simmons	NO_3^-	continuous	adjusted	-2.17	-6.83	2.48	0.36	517
PFF	NO_3^-	continuous	adjusted	0.17	-2.73	3.06	0.91	1068
CARE-PF	NO_3^-	continuous	adjusted	3.05	-3.20	9.30	0.34	810
Simmons	NH_4^+	continuous	adjusted	-3.10	-6.60	0.40	0.08	517

PFF	NH_4^+	continuous	adjusted	**-3.60**	**-7.21**	**-0.001**	**0.05**	1068
CARE-PF	NH_4^+	continuous	adjusted	**11.02**	**2.47**	**19.57**	**0.01**	810
Simmons	BC	continuous	adjusted	-1.48	-10.48	7.51	0.75	517
PFF	BC	continuous	adjusted	-1.59	-6.69	3.52	0.54	1068
CARE-PF	BC	continuous	adjusted	-2.09	-10.39	6.20	0.62	810
Simmons	OM	continuous	adjusted	-2.10	-4.38	0.17	0.07	517
PFF	OM	continuous	adjusted	0.30	-0.95	1.55	0.64	1068
CARE-PF	OM	continuous	adjusted	-0.59	-1.62	0.45	0.26	810
Simmons	SS	continuous	adjusted	**-19.63**	**-34.50**	**-4.75**	**0.009**	517
PFF	SS	continuous	adjusted	-0.09	-4.81	4.62	0.97	1068
CARE-PF	SS	continuous	adjusted	-4.06	-21.13	13.01	0.64	810
Simmons	Soil	continuous	adjusted	2.49	-10.25	15.22	0.70	517
PFF	Soil	continuous	adjusted	-0.34	-7.34	6.65	0.92	1068
CARE-PF	Soil	continuous	adjusted	-1.12	-11.58	9.34	0.83	810
Baseline D_LCO Models								
Simmons	$PM_{2.5}$	continuous	adjusted	-0.01	-0.73	0.71	0.97	476
PFF	$PM_{2.5}$	continuous	adjusted	**-1.09**	**-1.78**	**-0.39**	**0.002**	978
CARE-PF	$PM_{2.5}$	continuous	adjusted	-0.67	-1.76	0.42	0.23	594
Simmons	SO_4^{2-}	continuous	adjusted	-0.20	-1.55	1.16	0.78	476
PFF	SO_4^{2-}	continuous	adjusted	**-4.49**	**-6.33**	**-2.65**	**<0.001**	978
CARE-PF	SO_4^{2-}	continuous	adjusted	-0.56	-5.72	4.59	0.83	594
Simmons	NO_3^-	continuous	adjusted	1.19	-3.64	6.02	0.63	476
PFF	NO_3^-	continuous	adjusted	-0.15	-2.94	2.64	0.92	978
CARE-PF	NO_3^-	continuous	adjusted	-0.51	-7.10	6.08	0.88	594
Simmons	NH_4^+	continuous	adjusted	0.11	-3.50	3.71	0.95	476
PFF	NH_4^+	continuous	adjusted	**-7.35**	**-10.80**	**-3.90**	**<0.001**	978
CARE-PF	NH_4^+	continuous	adjusted	4.81	-4.33	13.94	0.30	594
Simmons	BC	continuous	adjusted	5.52	-3.95	14.99	0.25	476
PFF	BC	continuous	adjusted	**-6.59**	**-11.44**	**-1.75**	**0.008**	978
CARE-PF	BC	continuous	adjusted	-4.77	-13.82	4.27	0.30	594
Simmons	OM	continuous	adjusted	-0.34	-2.70	2.02	0.78	476
PFF	OM	continuous	adjusted	-0.71	-1.92	0.50	0.25	978
CARE-PF	OM	continuous	adjusted	-0.61	-1.74	0.52	0.29	594
Simmons	SS	continuous	adjusted	-10.64	-25.90	4.62	0.17	476
PFF	SS	continuous	adjusted	1.58	-3.12	6.29	0.51	978
CARE-PF	SS	continuous	adjusted	-14.25	-32.40	3.90	0.12	594
Simmons	Soil	continuous	adjusted	5.43	-7.58	18.43	0.41	476
PFF	Soil	continuous	adjusted	-3.86	-10.53	2.81	0.26	978
CARE-PF	Soil	continuous	adjusted	0.61	-11.06	12.29	0.92	594
FVC Decline Models								
Simmons	$PM_{2.5}$	continuous	adjusted	**-0.32**	**-0.52**	**-0.12**	**0.002**	519

PFF	$PM_{2.5}$	continuous	adjusted	0.01	-0.36	0.38	0.96	812
CARE-PF	$PM_{2.5}$	continuous	adjusted	-0.03	-0.34	0.28	0.85	811
Simmons	SO_4^{2-}	continuous	adjusted	**-0.62**	**-0.99**	**-0.24**	**0.001**	519
PFF	SO_4^{2-}	continuous	adjusted	**-4.97**	**-7.43**	**-2.52**	**<0.001**	812
CARE-PF	SO_4^{2-}	continuous	adjusted	**-3.98**	**-6.24**	**-1.72**	**<0.001**	811
Simmons	NO_3^-	continuous	adjusted	**-2.06**	**-3.58**	**-0.54**	**0.008**	519
PFF	NO_3^-	continuous	adjusted	-1.21	-2.91	0.50	0.17	812
CARE-PF	NO_3^-	continuous	adjusted	-1.80	-3.94	0.35	0.10	811
Simmons	NH_4^+	continuous	adjusted	**-1.56**	**-2.45**	**-0.67**	**<0.001**	519
PFF	NH_4^+	continuous	adjusted	**-8.90**	**-13.09**	**-4.71**	**<0.001**	812
CARE-PF	NH_4^+	continuous	adjusted	**-8.38**	**-12.37**	**-4.39**	**<0.001**	811
Simmons	BC	continuous	adjusted	-0.39	-3.18	2.39	0.78	519
PFF	BC	continuous	adjusted	0.25	-3.19	3.69	0.89	812
CARE-PF	BC	continuous	adjusted	0.97	-1.57	3.51	0.45	811
Simmons	OM	continuous	adjusted	-0.005	-0.79	0.78	0.99	519
PFF	OM	continuous	adjusted	0.61	-0.09	1.30	0.09	812
CARE-PF	OM	continuous	adjusted	0.03	-0.38	0.45	0.88	811
Simmons	SS	continuous	adjusted	3.85	-0.93	8.63	0.11	519
PFF	SS	continuous	adjusted	0.08	-1.99	2.14	0.94	812
CARE-PF	SS	continuous	adjusted	5.77	-0.15	11.69	0.06	811
Simmons	Soil	continuous	adjusted	-0.26	-4.28	3.77	0.90	519
PFF	Soil	continuous	adjusted	2.28	-0.24	4.81	0.08	812
CARE-PF	Soil	continuous	adjusted	1.00	-1.46	3.46	0.43	811
D_LCO Decline Models								
Simmons	$PM_{2.5}$	continuous	adjusted	**-0.20**	**-0.40**	**-0.00005**	**0.05**	491
PFF	$PM_{2.5}$	continuous	adjusted	0.08	-0.35	0.51	0.71	762
CARE-PF	$PM_{2.5}$	continuous	adjusted	-0.14	-0.44	0.17	0.38	748
Simmons	SO_4^{2-}	continuous	adjusted	**-0.43**	**-0.79**	**-0.06**	**0.02**	491
PFF	SO_4^{2-}	continuous	adjusted	**-4.29**	**-7.16**	**-1.42**	**0.003**	762
CARE-PF	SO_4^{2-}	continuous	adjusted	**-4.11**	**-6.45**	**-1.77**	**<0.001**	748
Simmons	NO_3^-	continuous	adjusted	**-2.34**	**-3.94**	**-0.74**	**0.004**	491
PFF	NO_3^-	continuous	adjusted	-0.78	-2.78	1.21	0.44	762
CARE-PF	NO_3^-	continuous	adjusted	-1.14	-3.26	0.99	0.30	748
Simmons	NH_4^+	continuous	adjusted	**-1.22**	**-2.08**	**-0.37**	**0.005**	491
PFF	NH_4^+	continuous	adjusted	-5.79	-10.85	-0.72	**0.03**	762
CARE-PF	NH_4^+	continuous	adjusted	**-7.96**	**-12.10**	**-3.82**	**<0.001**	748
Simmons	BC	continuous	adjusted	0.46	-2.19	3.12	0.73	491
PFF	BC	continuous	adjusted	-0.53	-4.54	3.48	0.80	762
CARE-PF	BC	continuous	adjusted	0.37	-2.18	2.92	0.78	748
Simmons	OM	continuous	adjusted	0.21	-0.55	0.96	0.59	491
PFF	OM	continuous	adjusted	0.50	-0.30	1.31	0.22	762

CARE-PF	OM	continuous	adjusted	-0.12	-0.54	0.30	0.58	748
Simmons	SS	continuous	adjusted	**7.21**	**2.75**	**11.66**	**0.002**	491
PFF	SS	continuous	adjusted	0.05	-2.31	2.41	0.97	762
CARE-PF	SS	continuous	adjusted	5.59	-0.38	11.55	0.07	748
Simmons	Soil	continuous	adjusted	-0.50	-4.40	3.40	0.80	491
PFF	Soil	continuous	adjusted	2.15	-0.71	5.00	0.14	762
CARE-PF	Soil	continuous	adjusted	0.12	-2.37	2.62	0.92	748

Appendix Table 18 – Pre- and Post-2015 year of enrollment subgroup analyses for mortality outcome. Results of adjusted models for associations of continuous $PM_{2.5}$ or constituents (sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic matter (OM), sea salt (SS), soil) in the 5-years pre-censoring with mortality. Adjustments made for age at enrollment, sex, smoking history, race, a socioeconomic variable, and site (for PFF and CARE-PF). Significant associations are **bolded.**

Cohort	Pollutant	Enrollment Year	Method	Model	HR	2.5 CI	97.5 CI	p	n
Simmons	$PM_{2.5}$	2015 & Pre-2015	continuous	adjusted	**1.36**	**1.32**	**1.40**	**<0.001**	1127
PFF	$PM_{2.5}$	2015 & Pre-2015	continuous	adjusted	**1.39**	**1.23**	**1.57**	**<0.001**	1059
CARE-PF	$PM_{2.5}$	2015 & Pre-2015	continuous	adjusted	0.99	0.94	1.05	0.84	1385
Simmons	$PM_{2.5}$	Post-2015	continuous	adjusted	**1.31**	**1.13**	**1.52**	**<0.001**	245
PFF	$PM_{2.5}$	Post-2015	continuous	adjusted	1.04	0.91	1.18	0.57	773
CARE-PF	$PM_{2.5}$	Post-2015	continuous	adjusted	1.01	0.96	1.07	0.68	1968
Simmons	SO_4^{2-}	2015 & Pre-2015	continuous	adjusted	**1.90**	**1.80**	**2.00**	**<0.001**	1127
PFF	SO_4^{2-}	2015 & Pre-2015	continuous	adjusted	**485.40**	**228.20**	**1.03*10^3**	**<0.001**	1059
CARE-PF	SO_4^{2-}	2015 & Pre-2015	continuous	adjusted	**2.27**	**2.01**	**2.56**	**<0.001**	1385
Simmons	SO_4^{2-}	Post-2015	continuous	adjusted	**18.35**	**8.21**	**41.03**	**<0.001**	245
PFF	SO_4^{2-}	Post-2015	continuous	adjusted	**45.75**	**17.68**	**118.35**	**<0.001**	773
CARE-PF	SO_4^{2-}	Post-2015	continuous	adjusted	**217.77**	**115.87**	**409.31**	**<0.001**	1968
Simmons	NO_3^-	2015 & Pre-2015	continuous	adjusted	**3.52**	**3.02**	**4.10**	**<0.001**	1127
PFF	NO_3^-	2015 & Pre-2015	continuous	adjusted	**5.82**	**3.29**	**10.30**	**<0.001**	1059
CARE-PF	NO_3^-	2015 & Pre-2015	continuous	adjusted	**7.73**	**4.41**	**13.56**	**<0.001**	1385
Simmons	NO_3^-	Post-2015	continuous	adjusted	**63.20**	**17.51**	**228.04**	**<0.001**	245
PFF	NO_3^-	Post-2015	continuous	adjusted	1.15	0.67	1.96	0.61	773
CARE-PF	NO_3^-	Post-2015	continuous	adjusted	**6.77**	**3.71**	**12.36**	**<0.001**	1968
Simmons	NH_4^+	2015 & Pre-2015	continuous	adjusted	**5.00**	**4.38**	**5.70**	**<0.001**	1127
PFF	NH_4^+	2015 & Pre-2015	continuous	adjusted	**2.73*10^4**	**761.10**	**9.76*10^4**	**<0.001**	1059
CARE-PF	NH_4^+	2015 & Pre-2015	continuous	adjusted	**39.32**	**27.04**	**57.16**	**<0.001**	1385
Simmons	NH_4^+	Post-2015	continuous	adjusted	**753.40**	**164.92**	**3.44*10^3**	**<0.001**	245
PFF	NH_4^+	Post-2015	continuous	adjusted	**43.22**	**11.80**	**158.32**	**<0.001**	773
CARE-PF	NH_4^+	Post-2015	continuous	adjusted	**1.08*10^5**	**2.97*10^4**	**3.95*10^5**	**<0.001**	1968
Simmons	BC	2015 & Pre-2015	continuous	adjusted	0.98	0.68	1.42	0.92	1127
PFF	BC	2015 & Pre-2015	continuous	adjusted	1.51	0.60	3.80	0.38	1059
CARE-PF	BC	2015 & Pre-2015	continuous	adjusted	0.26	0.43	1.22	0.23	1385
Simmons	BC	Post-2015	continuous	adjusted	**0.61**	**0.09**	**0.94**	**0.04**	245

PFF	BC	Post-2015	continuous	adjusted	0.81	0.24	2.77	0.74	773
CARE-PF	BC	Post-2015	continuous	adjusted	0.67	0.36	1.25	0.21	1968
Simmons	OM	2015 & Pre-2015	continuous	adjusted	1.02	0.92	1.14	0.71	1127
PFF	OM	2015 & Pre-2015	continuous	adjusted	0.93	0.77	1.13	0.46	1059
CARE-PF	OM	2015 & Pre-2015	continuous	adjusted	1.05	0.97	1.14	0.21	1385
Simmons	OM	Post-2015	continuous	adjusted	1.11	0.75	1.63	0.60	245
PFF	OM	Post-2015	continuous	adjusted	0.82	0.64	1.05	0.11	773
CARE-PF	OM	Post-2015	continuous	adjusted	**1.14**	**1.03**	**1.26**	**0.01**	1968
Simmons	SS	2015 & Pre-2015	continuous	adjusted	**0.007**	**0.003**	**0.02**	**<0.001**	1127
PFF	SS	2015 & Pre-2015	continuous	adjusted	0.57	0.25	1.30	0.18	1059
CARE-PF	SS	2015 & Pre-2015	continuous	adjusted	**7.45**	**3.56**	**15.57**	**<0.001**	1385
Simmons	SS	Post-2015	continuous	adjusted	**0.001**	**0.00003**	**0.03**	**<0.001**	245
PFF	SS	Post-2015	continuous	adjusted	0.69	0.37	1.28	0.24	773
CARE-PF	SS	Post-2015	continuous	adjusted	1.06	0.27	4.19	0.93	1968
Simmons	Soil	2015 & Pre-2015	continuous	adjusted	**2.58**	**1.59**	**4.18**	**<0.001**	1127
PFF	Soil	2015 & Pre-2015	continuous	adjusted	**2.36**	**1.14**	**4.86**	**0.02**	1059
CARE-PF	Soil	2015 & Pre-2015	continuous	adjusted	0.99	0.49	1.98	0.97	1385
Simmons	Soil	Post-2015	continuous	adjusted	**15.40**	**2.96**	**80.22**	**0.001**	245
PFF	Soil	Post-2015	continuous	adjusted	0.70	0.29	1.73	0.44	773
CARE-PF	Soil	Post-2015	continuous	adjusted	0.65	0.32	1.34	0.25	1968

Appendix Table 19 – Warm-months vs cold-months sensitivity analysis for mortality. Results of adjusted models for associations of continuous $PM_{2.5}$ exposure averaged over the warm months (April-September) in the 5-years pre-censoring with mortality versus $PM_{2.5}$ exposure averaged over the cold months (October-March) in the 5-years pre-censoring with mortality. Adjustments made for age at enrollment, sex, smoking history, race, a socioeconomic variable, and site (for PFF and CARE-PF). Significant associations are **bolded.**

Cohort	Pollutant	Exposure Period	Method	Model	HR	2.5 CI	97.5 CI	p	n
Simmons	$PM_{2.5}$	Warm Months in 5yrs Pre-Censoring	continuous	adjusted	**1.25**	**1.22**	**1.27**	**<0.001**	1372
PFF	$PM_{2.5}$	Warm Months in 5yrs Pre-Censoring	continuous	adjusted	**1.35**	**1.22**	**1.50**	**<0.001**	1832
CARE-PF	$PM_{2.5}$	Warm Months in 5yrs Pre-Censoring	continuous	adjusted	**0.95**	**0.91**	**0.99**	**0.009**	3353
Simmons	$PM_{2.5}$	Cold Months in 5yrs Pre-Censoring	continuous	adjusted	**1.41**	**1.36**	**1.46**	**<0.001**	1372
PFF	$PM_{2.5}$	Cold Months in 5yrs Pre-Censoring	continuous	adjusted	**1.09**	**1.02**	**1.16**	**0.008**	1832
CARE-PF	$PM_{2.5}$	Cold Months in 5yrs Pre-Censoring	continuous	adjusted	**1.13**	**1.08**	**1.18**	**<0.001**	3353

Appendix Table 20 – Full cohort baseline FVC models. Associations of 5-year pre-enrollment exposures to $PM_{2.5}$ or constituent components (sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic matter (OM), sea salt (SS), soil) with baseline forced vital capacity (FVC). Adjustments made for age at enrollment, sex, smoking history, race, a socioeconomic variable, and site (for PFF and CARE-PF). Significant associations are **bolded.**

Cohort	Pollutant	Method	Model	β	2.5 CI	97.5 CI	p	n
Simmons	$PM_{2.5}$	continuous	unadjusted	**-1.03**	**-1.51**	**-0.56**	**<0.001**	1073
Simmons	$PM_{2.5}$	continuous	adjusted	**-0.98**	**-1.45**	**-0.50**	**<0.001**	1048
PFF	$PM_{2.5}$	continuous	unadjusted	**-0.74**	**-1.17**	**-0.30**	**<0.001**	1696
PFF	$PM_{2.5}$	continuous	adjusted	0.20	-0.40	0.79	0.52	1672
CARE-PF	$PM_{2.5}$	continuous	unadjusted	**-0.75**	**-1.13**	**-0.37**	**<0.001**	2967
CARE-PF	$PM_{2.5}$	continuous	adjusted	-0.07	-0.59	0.46	0.80	2958
Simmons	$PM_{2.5}$	dichotomized	unadjusted	-0.78	-5.18	3.63	0.73	1073
Simmons	$PM_{2.5}$	dichotomized	adjusted	-1.12	-5.60	3.35	0.62	1048
PFF	$PM_{2.5}$	dichotomized	unadjusted	**-3.89**	**-5.86**	**-1.93**	**<0.001**	1696
PFF	$PM_{2.5}$	dichotomized	adjusted	-0.41	-2.83	2.02	0.74	1672
CARE-PF	$PM_{2.5}$	dichotomized	unadjusted	**-1.78**	**-3.40**	**-0.16**	**0.03**	2967
CARE-PF	$PM_{2.5}$	dichotomized	adjusted	0.99	-1.13	3.11	0.36	2958
Simmons	SO_4^{2-}	continuous	unadjusted	**-1.92**	**-2.80**	**-1.05**	**<0.001**	1073
Simmons	SO_4^{2-}	continuous	adjusted	**-1.85**	**-2.73**	**-0.98**	**<0.001**	1048
PFF	SO_4^{2-}	continuous	unadjusted	**-2.77**	**-3.80**	**-1.74**	**<0.001**	1696
PFF	SO_4^{2-}	continuous	adjusted	-1.17	-2.72	0.38	0.14	1672
CARE-PF	SO_4^{2-}	continuous	unadjusted	**-3.21**	**-4.49**	**-1.93**	**<0.001**	2967
CARE-PF	SO_4^{2-}	continuous	adjusted	1.38	-1.17	3.93	0.29	2958
Simmons	SO_4^{2-}	dichotomized	unadjusted	-2.47	-9.60	4.67	0.50	1073
Simmons	SO_4^{2-}	dichotomized	adjusted	-3.59	-10.77	3.59	0.32	1048
PFF	SO_4^{2-}	dichotomized	unadjusted	**-4.51**	**-6.38**	**-2.63**	**<0.001**	1696
PFF	SO_4^{2-}	dichotomized	adjusted	-0.75	-3.74	2.23	0.62	1672
CARE-PF	SO_4^{2-}	dichotomized	unadjusted	**-4.49**	**-6.31**	**-2.67**	**<0.001**	2967
CARE-PF	SO_4^{2-}	dichotomized	adjusted	-0.37	-2.78	2.04	0.76	2958
Simmons	NO_3^-	continuous	unadjusted	**-4.85**	**-8.32**	**-1.38**	**0.006**	1073
Simmons	NO_3^-	continuous	adjusted	**-4.13**	**-7.61**	**-0.65**	**0.02**	1048
PFF	NO_3^-	continuous	unadjusted	-0.19	-1.55	1.18	0.79	1696
PFF	NO_3^-	continuous	adjusted	1.11	-1.21	3.42	0.35	1672
CARE-PF	NO_3^-	continuous	unadjusted	**-2.74**	**-4.71**	**-0.77**	**0.006**	2967
CARE-PF	NO_3^-	continuous	adjusted	1.77	-1.47	5.01	0.28	2958
Simmons	NO_3^-	dichotomized	unadjusted	-0.40	-3.35	2.54	0.79	1073
Simmons	NO_3^-	dichotomized	adjusted	-0.38	-3.37	2.60	0.80	1048
PFF	NO_3^-	dichotomized	unadjusted	-0.18	-1.93	1.56	0.84	1696
PFF	NO_3^-	dichotomized	adjusted	-0.04	-2.90	2.83	0.98	1672
CARE-PF	NO_3^-	dichotomized	unadjusted	**-1.90**	**-3.41**	**-0.40**	**0.01**	2967
CARE-PF	NO_3^-	dichotomized	adjusted	1.73	-0.42	3.88	0.12	2958

Simmons	NH_4^+	continuous	unadjusted	**-5.01**	**-7.28**	**-2.74**	**<0.001**	1073
Simmons	NH_4^+	continuous	adjusted	**-4.80**	**-7.09**	**-2.52**	**<0.001**	1048
PFF	NH_4^+	continuous	unadjusted	**-4.35**	**-6.52**	**-2.19**	**<0.001**	1696
PFF	NH_4^+	continuous	adjusted	-1.83	-4.74	1.08	0.22	1672
CARE-PF	NH_4^+	continuous	unadjusted	**-3.93**	**-6.53**	**-1.34**	**0.003**	2967
CARE-PF	NH_4^+	continuous	adjusted	**4.87**	**0.49**	**9.25**	**0.03**	2958
Simmons	NH_4^+	dichotomized	unadjusted	1.01	-4.64	6.66	0.73	1073
Simmons	NH_4^+	dichotomized	adjusted	0.57	-5.16	6.30	0.85	1048
PFF	NH_4^+	dichotomized	unadjusted	**-2.07**	**-3.85**	**-0.28**	**0.02**	1696
PFF	NH_4^+	dichotomized	adjusted	0.23	-2.11	2.57	0.85	1672
CARE-PF	NH_4^+	dichotomized	unadjusted	**-4.11**	**-5.75**	**-2.47**	**<0.001**	2967
CARE-PF	NH_4^+	dichotomized	adjusted	-0.18	-2.53	2.18	0.88	2958
Simmons	BC	continuous	unadjusted	**-7.46**	**-13.96**	**-0.95**	**0.03**	1073
Simmons	BC	continuous	adjusted	-5.53	-12.04	0.98	0.10	1048
PFF	BC	continuous	unadjusted	-1.99	-5.15	1.17	0.22	1696
PFF	BC	continuous	adjusted	1.63	-2.65	5.92	0.45	1672
CARE-PF	BC	continuous	unadjusted	**-9.24**	**-13.18**	**-5.31**	**<0.001**	2967
CARE-PF	BC	continuous	adjusted	-2.01	-6.33	2.32	0.36	2958
Simmons	BC	dichotomized	unadjusted	-3.00	-7.07	1.07	0.15	1073
Simmons	BC	dichotomized	adjusted	-2.13	-6.19	1.93	0.30	1048
PFF	BC	dichotomized	unadjusted	-0.61	-2.38	1.16	0.50	1696
PFF	BC	dichotomized	adjusted	1.15	-1.00	3.29	0.29	1672
CARE-PF	BC	dichotomized	unadjusted	**-2.22**	**-3.85**	**-0.58**	**0.008**	2967
CARE-PF	BC	dichotomized	adjusted	-0.26	-1.93	1.41	0.76	2958
Simmons	OM	continuous	unadjusted	**-3.48**	**-5.09**	**-1.86**	**<0.001**	1073
Simmons	OM	continuous	adjusted	**-2.68**	**-4.33**	**-1.02**	**0.002**	1048
PFF	OM	continuous	unadjusted	0.15	-0.69	0.99	0.73	1696
PFF	OM	continuous	adjusted	1.02	-0.01	2.06	0.05	1672
CARE-PF	OM	continuous	unadjusted	**-0.58**	**-1.09**	**-0.08**	**0.02**	2967
CARE-PF	OM	continuous	adjusted	-0.21	-0.74	0.32	0.43	2958
Simmons	OM	dichotomized	unadjusted	-2.32	-4.75	0.10	0.06	1073
Simmons	OM	dichotomized	adjusted	-1.30	-3.76	1.16	0.30	1048
PFF	OM	dichotomized	unadjusted	-0.01	-1.76	1.75	0.99	1696
PFF	OM	dichotomized	adjusted	**2.20**	**0.11**	**4.28**	**0.04**	1672
CARE-PF	OM	dichotomized	unadjusted	-0.83	-2.27	0.60	0.26	2967
CARE-PF	OM	dichotomized	adjusted	-0.12	-1.60	1.35	0.87	2958
Simmons	SS	continuous	unadjusted	**-29.31**	**-40.56**	**-18.06**	**<0.001**	1073
Simmons	SS	continuous	adjusted	**-29.41**	**-41.35**	**-17.47**	**<0.001**	1048
PFF	SS	continuous	unadjusted	-1.14	-3.44	1.17	0.33	1696
PFF	SS	continuous	adjusted	0.02	-3.98	4.01	0.99	1672
CARE-PF	SS	continuous	unadjusted	**-9.45**	**-15.29**	**-3.62**	**0.002**	2967

CARE-PF	SS	continuous	adjusted	0.77	-7.04	8.58	0.85	2958
Simmons	SS	dichotomized	unadjusted	**-6.05**	**-8.56**	**-3.53**	**<0.001**	1073
Simmons	SS	dichotomized	adjusted	**-5.23**	**-7.84**	**-2.61**	**<0.001**	1048
PFF	SS	dichotomized	unadjusted	-0.36	-2.10	1.38	0.68	1696
PFF	SS	dichotomized	adjusted	**4.22**	**1.64**	**6.80**	**0.001**	1672
CARE-PF	SS	dichotomized	unadjusted	3.68	-2.07	9.42	0.21	2967
CARE-PF	SS	dichotomized	adjusted	5.12	-0.50	10.74	0.07	2958
Simmons	Soil	continuous	unadjusted	-8.46	-17.59	0.68	0.07	1073
Simmons	Soil	continuous	adjusted	-6.77	-15.97	2.42	0.15	1048
PFF	Soil	continuous	unadjusted	0.15	-2.00	2.29	0.89	1696
PFF	Soil	continuous	adjusted	**6.51**	**1.16**	**11.86**	**0.02**	1672
CARE-PF	Soil	continuous	unadjusted	-4.34	-9.17	0.48	0.08	2967
CARE-PF	Soil	continuous	adjusted	1.38	-4.43	7.19	0.64	2958
Simmons	Soil	dichotomized	unadjusted	0.59	-2.24	3.41	0.68	1073
Simmons	Soil	dichotomized	adjusted	0.85	-1.99	3.68	0.56	1048
PFF	Soil	dichotomized	unadjusted	-1.11	-2.93	0.70	0.23	1696
PFF	Soil	dichotomized	adjusted	2.37	-0.41	5.16	0.09	1672
CARE-PF	Soil	dichotomized	unadjusted	-0.34	-1.91	1.24	0.68	2967
CARE-PF	Soil	dichotomized	adjusted	0.59	-1.10	2.29	0.49	2958
Simmons	Multi-Constituent	continuous	unadjusted	**-4.94**	**-6.67**	**-3.22**	**<0.001**	1073
Simmons	Multi-Constituent	continuous	adjusted	**-4.44**	**-6.20**	**-2.69**	**<0.001**	1048
PFF	Multi-Constituent	continuous	unadjusted	**-2.78**	**-4.21**	**-1.35**	**<0.001**	1696
PFF	Multi-Constituent	continuous	adjusted	-1.61	-3.66	0.44	0.12	1672
CARE-PF	Multi-Constituent	continuous	unadjusted	**-3.38**	**-4.34**	**-2.42**	**<0.001**	2967
CARE-PF	Multi-Constituent	continuous	adjusted	**-3.75**	**-5.13**	**-2.37**	**<0.001**	2958

Appendix Table 21 – Full cohort baseline D_LCO models. Associations of 5-year pre-enrollment exposures to $PM_{2.5}$ or constituent components (sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic matter (OM), sea salt (SS), soil) with baseline diffusion capacity of the lung for carbon monoxide (D_LCO). Adjustments made for age at enrollment, sex, smoking history, race, a socioeconomic variable, and site (for PFF and CARE-PF). Significant associations are **bolded.**

Cohort	Pollutant	Method	Model	β	2.5 CI	97.5 CI	p	n
Simmons	$PM_{2.5}$	continuous	unadjusted	-0.20	-0.69	0.29	0.41	998
Simmons	$PM_{2.5}$	continuous	adjusted	-0.13	-0.63	0.36	0.60	978
PFF	$PM_{2.5}$	continuous	unadjusted	**-0.80**	**-1.21**	**-0.39**	**<0.001**	1570
PFF	$PM_{2.5}$	continuous	adjusted	**-0.86**	**-1.42**	**-0.31**	**0.002**	1547
CARE-PF	$PM_{2.5}$	continuous	unadjusted	0.28	-0.14	0.70	0.19	2389
CARE-PF	$PM_{2.5}$	continuous	adjusted	0.01	-0.54	0.56	0.98	2383
Simmons	$PM_{2.5}$	dichotomized	unadjusted	2.62	-1.85	7.08	0.25	998
Simmons	$PM_{2.5}$	dichotomized	adjusted	3.66	-0.90	0.82	0.12	978
PFF	$PM_{2.5}$	dichotomized	unadjusted	**-3.71**	**-5.57**	**-1.85**	**<0.001**	1570
PFF	$PM_{2.5}$	dichotomized	adjusted	**-2.61**	**-4.88**	**-0.33**	**0.02**	1547
CARE-PF	$PM_{2.5}$	dichotomized	unadjusted	1.05	-0.80	2.90	0.27	2389
CARE-PF	$PM_{2.5}$	dichotomized	adjusted	0.09	-2.18	2.36	0.94	2383
Simmons	SO_4^{2-}	continuous	unadjusted	-0.31	-1.22	0.59	0.50	998
Simmons	SO_4^{2-}	continuous	adjusted	-0.23	-1.14	0.69	0.62	978
PFF	SO_4^{2-}	continuous	unadjusted	**-3.08**	**-4.05**	**-2.11**	**<0.001**	1570
PFF	SO_4^{2-}	continuous	adjusted	**-3.76**	**-5.19**	**-2.32**	**<0.001**	1547
CARE-PF	SO_4^{2-}	continuous	unadjusted	**2.66**	**1.14**	**4.18**	**<0.001**	2389
CARE-PF	SO_4^{2-}	continuous	adjusted	-0.03	-2.83	2.78	0.98	2383
Simmons	SO_4^{2-}	dichotomized	unadjusted	2.83	-4.52	10.18	0.45	998
Simmons	SO_4^{2-}	dichotomized	adjusted	5.57	-1.88	13.02	0.14	978
PFF	SO_4^{2-}	dichotomized	unadjusted	**-4.36**	**-6.13**	**-2.58**	**<0.001**	1570
PFF	SO_4^{2-}	dichotomized	adjusted	**-2.81**	**-5.60**	**-0.02**	**0.048**	1547
CARE-PF	SO_4^{2-}	dichotomized	unadjusted	**2.47**	**0.23**	**4.70**	**0.03**	2389
CARE-PF	SO_4^{2-}	dichotomized	adjusted	1.25	-1.36	3.86	0.35	2383
Simmons	NO_3^-	continuous	unadjusted	-0.31	-3.88	3.26	0.86	998
Simmons	NO_3^-	continuous	adjusted	0.49	-3.12	4.10	0.79	978
PFF	NO_3^-	continuous	unadjusted	0.87	-0.41	2.14	0.18	1570
PFF	NO_3^-	continuous	adjusted	-0.51	-2.66	1.63	0.64	1547
CARE-PF	NO_3^-	continuous	unadjusted	**6.26**	**3.98**	**8.55**	**<0.001**	2389
CARE-PF	NO_3^-	continuous	adjusted	0.23	-3.16	3.62	0.89	2383
Simmons	NO_3^-	dichotomized	unadjusted	0.11	-2.92	3.14	0.94	998
Simmons	NO_3^-	dichotomized	adjusted	-0.10	-3.17	2.98	0.95	978
PFF	NO_3^-	dichotomized	unadjusted	0.55	-1.10	2.02	0.51	1570
PFF	NO_3^-	dichotomized	adjusted	0.57	-2.11	3.26	0.67	1547
CARE-PF	NO_3^-	dichotomized	unadjusted	**4.66**	**2.96**	**6.35**	**<0.001**	2389
CARE-PF	NO_3^-	dichotomized	adjusted	1.32	-0.86	3.50	0.24	2383

Simmons	NH_4^+	continuous	unadjusted	-0.64	-2.98	1.71	0.59	998
Simmons	NH_4^+	continuous	adjusted	-0.25	-2.63	2.13	0.83	978
PFF	NH_4^+	continuous	unadjusted	**-4.39**	**-6.43**	**-2.36**	**<0.001**	1570
PFF	NH_4^+	continuous	adjusted	**-6.54**	**-9.23**	**-3.85**	**<0.001**	1547
CARE-PF	NH_4^+	continuous	unadjusted	**8.45**	**5.38**	**11.51**	**<0.001**	2389
CARE-PF	NH_4^+	continuous	adjusted	2.84	-1.89	7.57	0.24	2383
Simmons	NH_4^+	dichotomized	unadjusted	1.62	-4.13	7.36	0.58	998
Simmons	NH_4^+	dichotomized	adjusted	4.46	-1.41	10.33	0.14	978
PFF	NH_4^+	dichotomized	unadjusted	**-2.94**	**-4.63**	**-1.25**	**<0.001**	1570
PFF	NH_4^+	dichotomized	adjusted	**-3.05**	**-5.21**	**-0.88**	**0.006**	1547
CARE-PF	NH_4^+	dichotomized	unadjusted	**2.35**	**0.42**	**4.28**	**0.02**	2389
CARE-PF	NH_4^+	dichotomized	adjusted	-0.47	-2.91	1.96	0.70	2383
Simmons	BC	continuous	unadjusted	0.42	-6.31	7.15	0.90	998
Simmons	BC	continuous	adjusted	2.90	-3.91	9.71	0.40	978
PFF	BC	continuous	unadjusted	**-4.33**	**-7.28**	**-1.39**	**0.004**	1570
PFF	BC	continuous	adjusted	**-4.34**	**-8.28**	**-0.40**	**0.03**	1547
CARE-PF	BC	continuous	unadjusted	-1.73	-6.16	2.71	0.45	2389
CARE-PF	BC	continuous	adjusted	-2.59	-7.26	2.08	0.28	2383
Simmons	BC	dichotomized	unadjusted	1.82	-2.26	5.91	0.381	998
Simmons	BC	dichotomized	adjusted	2.23	-1.85	6.31	0.28	978
PFF	BC	dichotomized	unadjusted	**-1.77**	**-3.45**	**-0.10**	**0.04**	1570
PFF	BC	dichotomized	adjusted	-1.01	-3.01	0.99	0.32	1547
CARE-PF	BC	dichotomized	unadjusted	-0.95	-2.81	0.92	0.32	2389
CARE-PF	BC	dichotomized	adjusted	0.53	-1.33	2.38	0.58	2383
Simmons	OM	continuous	unadjusted	-1.59	-3.24	0.06	0.06	998
Simmons	OM	continuous	adjusted	-0.97	-2.68	0.74	0.27	978
PFF	OM	continuous	unadjusted	-0.38	-1.17	0.41	0.35	1570
PFF	OM	continuous	adjusted	-0.19	-1.15	0.78	0.70	1547
CARE-PF	OM	continuous	unadjusted	-0.48	-1.04	0.07	0.08	2389
CARE-PF	OM	continuous	adjusted	-0.34	-0.91	0.22	0.23	2383
Simmons	OM	dichotomized	unadjusted	0.24	-2.25	2.73	0.85	998
Simmons	OM	dichotomized	adjusted	0.24	-2.31	2.78	0.86	978
PFF	OM	dichotomized	unadjusted	-0.88	-2.54	0.77	0.30	1570
PFF	OM	dichotomized	adjusted	0.01	-1.94	1.96	0.99	1547
CARE-PF	OM	dichotomized	unadjusted	-0.27	-1.85	1.32	0.74	2389
CARE-PF	OM	dichotomized	adjusted	-0.30	-1.91	1.32	0.72	2383
Simmons	SS	continuous	unadjusted	**-13.95**	**-25.56**	**-2.34**	**0.02**	998
Simmons	SS	continuous	adjusted	**-18.86**	**-31.21**	**-6.52**	**0.003**	978
PFF	SS	continuous	unadjusted	**2.94**	**0.68**	**5.20**	**0.01**	1570
PFF	SS	continuous	adjusted	0.95	-2.83	4.73	0.62	1547
CARE-PF	SS	continuous	unadjusted	**-23.24**	**-29.39**	**-17.10**	**<0.001**	2389

CARE-PF	SS	continuous	adjusted	0.27	-7.98	8.51	0.95	2383
Simmons	SS	dichotomized	unadjusted	**-3.80**	**-6.41**	**-1.18**	**0.004**	998
Simmons	SS	dichotomized	adjusted	**-4.42**	**-7.13**	**-1.72**	**0.001**	978
PFF	SS	dichotomized	unadjusted	0.16	-1.48	1.81	0.85	1570
PFF	SS	dichotomized	adjusted	2.28	-0.10	4.65	0.06	1547
CARE-PF	SS	dichotomized	unadjusted	-5.22	-11.39	0.95	0.10	2389
CARE-PF	SS	dichotomized	adjusted	0.38	-5.58	6.34	0.90	2383
Simmons	Soil	continuous	unadjusted	-3.40	-12.69	5.90	0.47	998
Simmons	Soil	continuous	adjusted	-0.96	-10.40	8.48	0.84	978
PFF	Soil	continuous	unadjusted	**2.05**	**0.02**	**4.09**	**0.048**	1570
PFF	Soil	continuous	adjusted	-0.41	-5.38	4.56	0.87	1547
CARE-PF	Soil	continuous	unadjusted	**8.54**	**2.92**	**14.16**	**0.003**	2389
CARE-PF	Soil	continuous	adjusted	-1.20	-7.63	5.23	0.71	2383
Simmons	Soil	dichotomized	unadjusted	2.05	-0.85	4.95	0.17	998
Simmons	Soil	dichotomized	adjusted	2.23	-0.71	5.16	0.14	978
PFF	Soil	dichotomized	unadjusted	1.33	-0.38	3.05	0.13	1570
PFF	Soil	dichotomized	adjusted	0.10	-2.46	2.67	0.94	1547
CARE-PF	Soil	dichotomized	unadjusted	1.59	-0.18	3.36	0.08	2389
CARE-PF	Soil	dichotomized	adjusted	-0.66	-2.50	1.19	0.48	2383
Simmons	Multi-Constituent	continuous	unadjusted	**-3.82**	**-5.61**	**-2.03**	**<0.001**	998
Simmons	Multi-Constituent	continuous	adjusted	**-4.14**	**-5.96**	**-2.33**	**<0.001**	978
PFF	Multi-Constituent	continuous	unadjusted	**-1.37**	**-2.72**	**-0.03**	**0.046**	1570
PFF	Multi-Constituent	continuous	adjusted	**-2.40**	**-4.31**	**-0.48**	**0.01**	1547
CARE-PF	Multi-Constituent	continuous	unadjusted	**-2.03**	**-3.10**	**-0.96**	**<0.001**	2389
CARE-PF	Multi-Constituent	continuous	adjusted	**-4.02**	**-5.47**	**-2.57**	**<0.001**	2383

Appendix Table 22 – Full cohort FVC decline models. Associations of 5-years pre-censoring and 5-years pre-enrollment exposures to $PM_{2.5}$ or constituent components (sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic matter (OM), sea salt (SS), soil) with rate of change in forced vital capacity (FVC). Adjusted models include age at enrollment, sex, smoking history, race, a socioeconomic variable, and site (for PFF and CARE-PF). Significant associations are **bolded.**

Cohort	Pollutant	Method	Model	β	2.5 CI	97.5 CI	p	n
FVC Decline – 5-years pre-censoring models								
Simmons	$PM_{2.5}$	continuous	unadjusted	**-0.39**	**-0.52**	**-0.27**	**<0.001**	1080
Simmons	$PM_{2.5}$	continuous	adjusted	**-0.40**	**-0.53**	**-0.27**	**<0.001**	1055
PFF	$PM_{2.5}$	continuous	unadjusted	0.12	-0.08	0.33	0.23	1173
PFF	$PM_{2.5}$	continuous	adjusted	0.01	-0.28	0.30	0.96	1153
CARE-PF	$PM_{2.5}$	continuous	unadjusted	0.08	-0.02	0.19	0.12	2968
CARE-PF	$PM_{2.5}$	continuous	adjusted	-0.01	-0.13	0.11	0.86	2959
Simmons	$PM_{2.5}$	dichotomized	unadjusted	**-1.45**	**-2.12**	**-0.77**	**<0.001**	1080
Simmons	$PM_{2.5}$	dichotomized	adjusted	**-1.31**	**-1.98**	**-0.63**	**<0.001**	1055
PFF	$PM_{2.5}$	dichotomized	unadjusted	0.53	-0.10	1.15	0.10	1173
PFF	$PM_{2.5}$	dichotomized	adjusted	0.02	-0.81	0.85	0.97	1153
CARE-PF	$PM_{2.5}$	dichotomized	unadjusted	-0.04	-0.57	0.49	0.87	2968
CARE-PF	$PM_{2.5}$	dichotomized	adjusted	-0.39	-0.96	0.18	0.18	2959
Simmons	SO_4^{2-}	continuous	unadjusted	**-0.88**	**-1.13**	**-0.64**	**<0.001**	1080
Simmons	SO_4^{2-}	continuous	adjusted	**-0.88**	**-1.13**	**-0.64**	**<0.001**	1055
PFF	SO_4^{2-}	continuous	unadjusted	-0.08	-1.05	0.89	0.87	1173
PFF	SO_4^{2-}	continuous	adjusted	**-3.39**	**-5.37**	**-1.40**	**<0.001**	1153
CARE-PF	SO_4^{2-}	continuous	unadjusted	0.50	-0.03	1.03	0.07	2968
CARE-PF	SO_4^{2-}	continuous	adjusted	**-3.73**	**-4.95**	**-2.52**	**<0.001**	2959
Simmons	SO_4^{2-}	dichotomized	unadjusted	-1.67	-3.44	0.11	0.07	1080
Simmons	SO_4^{2-}	dichotomized	adjusted	-1.74	-3.49	0.02	0.05	1055
PFF	SO_4^{2-}	dichotomized	unadjusted	-0.13	-0.80	0.54	0.7	1173
PFF	SO_4^{2-}	dichotomized	adjusted	**-1.30**	**-2.52**	**-0.07**	**0.04**	1153
CARE-PF	SO_4^{2-}	dichotomized	unadjusted	-0.06	-0.49	0.38	0.8	2968
CARE-PF	SO_4^{2-}	dichotomized	adjusted	**-1.31**	**-1.97**	**-0.65**	**<0.001**	2959
Simmons	NO_3^-	continuous	unadjusted	**-2.86**	**-3.95**	**-1.76**	**<0.001**	1080
Simmons	NO_3^-	continuous	adjusted	**-2.82**	**-3.90**	**-1.74**	**<0.001**	1055
PFF	NO_3^-	continuous	unadjusted	-0.30	-1.03	0.43	0.42	1173
PFF	NO_3^-	continuous	adjusted	-0.89	-2.21	0.42	0.18	1153
CARE-PF	NO_3^-	continuous	unadjusted	0.34	-0.30	0.98	0.30	2968
CARE-PF	NO_3^-	continuous	adjusted	**-1.35**	**-2.40**	**-0.29**	**0.01**	2959
Simmons	NO_3^-	dichotomized	unadjusted	**-1.24**	**-2.02**	**-0.45**	**0.002**	1080
Simmons	NO_3^-	dichotomized	adjusted	**-1.03**	**-1.81**	**-0.25**	**0.01**	1055
PFF	NO_3^-	dichotomized	unadjusted	-0.15	-0.77	0.48	0.65	1173
PFF	NO_3^-	dichotomized	adjusted	-0.58	-1.71	0.55	0.31	1153
CARE-PF	NO_3^-	dichotomized	unadjusted	0.07	-0.31	0.46	0.70	2968

CARE-PF	NO_3^-	dichotomized	adjusted	-0.53	-1.10	0.04	0.07	2959
Simmons	NH_4^+	continuous	unadjusted	**-2.17**	**-2.75**	**-1.59**	**<0.001**	1080
Simmons	NH_4^+	continuous	adjusted	**-2.16**	**-2.73**	**-1.58**	**<0.001**	1055
PFF	NH_4^+	continuous	unadjusted	**-2.51**	**-4.63**	**-0.38**	**0.02**	1173
PFF	NH_4^+	continuous	adjusted	**-7.04**	**-10.41**	**-3.68**	**<0.001**	1153
CARE-PF	NH_4^+	continuous	unadjusted	-0.58	-1.77	0.60	0.34	2968
CARE-PF	NH_4^+	continuous	adjusted	**-9.05**	**-11.19**	**-6.91**	**<0.001**	2959
Simmons	NH_4^+	dichotomized	unadjusted	**-1.93**	**-3.39**	**-0.47**	**0.01**	1080
Simmons	NH_4^+	dichotomized	adjusted	**-1.86**	**-3.30**	**-0.43**	**0.01**	1055
PFF	NH_4^+	dichotomized	unadjusted	-0.55	-1.17	0.07	0.08	1173
PFF	NH_4^+	dichotomized	adjusted	**-1.33**	**-2.25**	**-0.42**	**0.004**	1153
CARE-PF	NH_4^+	dichotomized	unadjusted	-0.15	-0.55	0.25	0.47	2968
CARE-PF	NH_4^+	dichotomized	adjusted	**-1.76**	**-2.39**	**-1.13**	**<0.001**	2959
Simmons	BC	continuous	unadjusted	1.01	-0.65	2.68	0.23	1080
Simmons	BC	continuous	adjusted	0.62	-1.05	2.28	0.47	1055
PFF	BC	continuous	unadjusted	0.83	-1.02	2.68	0.38	1173
PFF	BC	continuous	adjusted	-0.45	-3.13	2.22	0.74	1153
CARE-PF	BC	continuous	unadjusted	1.09	-0.03	2.20	0.06	2968
CARE-PF	BC	continuous	adjusted	0.96	-0.22	2.13	0.11	2959
Simmons	BC	dichotomized	unadjusted	-0.07	-1.39	1.24	0.92	1080
Simmons	BC	dichotomized	adjusted	-0.02	-1.31	1.28	0.98	1055
PFF	BC	dichotomized	unadjusted	0.50	-0.13	1.13	0.12	1173
PFF	BC	dichotomized	adjusted	0.26	-0.62	1.15	0.56	1153
CARE-PF	BC	dichotomized	unadjusted	0.26	-0.16	0.69	0.22	2968
CARE-PF	BC	dichotomized	adjusted	**0.55**	**0.10**	**1.00**	**0.02**	2959
Simmons	OM	continuous	unadjusted	0.26	-0.24	0.77	0.31	1080
Simmons	OM	continuous	adjusted	0.14	-0.36	0.65	0.58	1055
PFF	OM	continuous	unadjusted	0.27	-0.09	0.62	0.14	1173
PFF	OM	continuous	adjusted	0.32	-0.21	0.86	0.23	1153
CARE-PF	OM	continuous	unadjusted	0.06	-0.12	0.25	0.50	2968
CARE-PF	OM	continuous	adjusted	0.07	-0.12	0.26	0.48	2959
Simmons	OM	dichotomized	unadjusted	0.34	-0.37	1.04	0.35	1080
Simmons	OM	dichotomized	adjusted	0.10	-0.59	0.80	0.77	1055
PFF	OM	dichotomized	unadjusted	0.27	-0.35	0.90	0.39	1173
PFF	OM	dichotomized	adjusted	0.13	-0.70	0.96	0.76	1153
CARE-PF	OM	dichotomized	unadjusted	0.18	-0.19	0.55	0.33	2968
CARE-PF	OM	dichotomized	adjusted	0.23	-0.15	0.60	0.23	2959
Simmons	SS	continuous	unadjusted	**6.43**	**3.60**	**9.26**	**<0.001**	1080
Simmons	SS	continuous	adjusted	**7.09**	**4.03**	**10.16**	**<0.001**	1055
PFF	SS	continuous	unadjusted	0.02	-0.92	0.96	0.97	1173
PFF	SS	continuous	adjusted	-0.34	-2.08	1.40	0.7	1153

CARE-PF	SS	continuous	unadjusted	**2.22**	**0.55**	**3.90**	**0.009**	2968
CARE-PF	SS	continuous	adjusted	0.03	-2.20	2.26	0.98	2959
Simmons	SS	dichotomized	unadjusted	**1.54**	**0.92**	**2.16**	**<0.001**	1080
Simmons	SS	dichotomized	adjusted	**1.34**	**0.73**	**1.95**	**<0.001**	1055
PFF	SS	dichotomized	unadjusted	0.28	-0.35	0.90	0.39	1173
PFF	SS	dichotomized	adjusted	0.18	-0.65	1.01	0.67	1153
CARE-PF	SS	dichotomized	unadjusted	0.09	-0.44	0.62	0.73	2968
CARE-PF	SS	dichotomized	adjusted	0.36	-0.24	0.97	0.24	2959
Simmons	Soil	continuous	unadjusted	-1.79	-4.17	0.58	0.14	1080
Simmons	Soil	continuous	adjusted	-1.50	-3.86	0.86	0.21	1055
PFF	Soil	continuous	unadjusted	0.44	-0.35	1.22	0.27	1173
PFF	Soil	continuous	adjusted	1.35	-0.48	3.18	0.15	1153
CARE-PF	Soil	continuous	unadjusted	**1.07**	**0.04**	**2.11**	**0.04**	2968
CARE-PF	Soil	continuous	adjusted	0.85	-0.48	2.19	0.21	2959
Simmons	Soil	dichotomized	unadjusted	-0.20	-0.92	0.51	0.58	1080
Simmons	Soil	dichotomized	adjusted	0.01	-0.70	0.73	0.97	1055
PFF	Soil	dichotomized	unadjusted	0.33	-0.35	1.00	0.34	1173
PFF	Soil	dichotomized	adjusted	-0.02	-0.98	0.94	0.96	1153
CARE-PF	Soil	dichotomized	unadjusted	**0.37**	**0.00**	**0.74**	**0.05**	2968
CARE-PF	Soil	dichotomized	adjusted	0.14	-0.31	0.58	0.54	2959
FVC Decline – 5-years pre-enrollment models								
Simmons	$PM_{2.5}$	continuous	unadjusted	**0.15**	**0.01**	**0.29**	**0.03**	1076
Simmons	$PM_{2.5}$	continuous	adjusted	0.06	-0.09	0.20	0.44	1051
PFF	$PM_{2.5}$	continuous	unadjusted	**0.18**	**0.02**	**0.35**	**0.03**	1168
PFF	$PM_{2.5}$	continuous	adjusted	0.03	-0.20	0.26	0.78	1148
CARE-PF	$PM_{2.5}$	continuous	unadjusted	**0.21**	**0.12**	**0.30**	**<0.001**	2967
CARE-PF	$PM_{2.5}$	continuous	adjusted	0.02	-0.12	0.16	0.81	2958
Simmons	$PM_{2.5}$	dichotomized	unadjusted	**2.06**	**0.35**	**3.77**	**0.02**	1076
Simmons	$PM_{2.5}$	dichotomized	adjusted	**1.86**	**0.13**	**3.59**	**0.04**	1051
PFF	$PM_{2.5}$	dichotomized	unadjusted	0.31	-0.42	1.04	0.41	1168
PFF	$PM_{2.5}$	dichotomized	adjusted	-0.36	-1.28	0.56	0.44	1148
CARE-PF	$PM_{2.5}$	dichotomized	unadjusted	**0.75**	**0.37**	**1.13**	**<0.001**	2967
CARE-PF	$PM_{2.5}$	dichotomized	adjusted	0.13	-0.40	0.66	0.63	2958
Simmons	SO_4^{2-}	continuous	unadjusted	**0.35**	**0.09**	**0.60**	**0.008**	1076
Simmons	SO_4^{2-}	continuous	adjusted	0.15	-0.12	0.41	0.27	1051
PFF	SO_4^{2-}	continuous	unadjusted	0.36	-0.02	0.74	0.07	1168
PFF	SO_4^{2-}	continuous	adjusted	0.04	-0.53	0.61	0.88	1148
CARE-PF	SO_4^{2-}	continuous	unadjusted	**0.65**	**0.36**	**0.94**	**<0.001**	2967
CARE-PF	SO_4^{2-}	continuous	adjusted	-0.15	-0.72	0.41	0.60	2958
Simmons	SO_4^{2-}	dichotomized	unadjusted	2.91	-1.01	6.83	0.15	1076

Simmons	SO_4^{2-}	dichotomized	adjusted	3.11	-1.36	7.58	0.17	1051
PFF	SO_4^{2-}	dichotomized	unadjusted	0.18	-0.55	0.90	0.63	1168
PFF	SO_4^{2-}	dichotomized	adjusted	-0.48	-1.60	0.65	0.41	1148
CARE-PF	SO_4^{2-}	dichotomized	unadjusted	**0.54**	**0.14**	**0.95**	**0.008**	2967
CARE-PF	SO_4^{2-}	dichotomized	adjusted	-0.19	-0.74	0.35	0.49	2958
Simmons	NO_3^-	continuous	unadjusted	0.59	-0.45	1.62	0.27	1076
Simmons	NO_3^-	continuous	adjusted	0.08	-0.94	1.11	0.87	1051
PFF	NO_3^-	continuous	unadjusted	0.09	-0.40	0.59	0.71	1168
PFF	NO_3^-	continuous	adjusted	0.11	-0.82	1.03	0.82	1148
CARE-PF	NO_3^-	continuous	unadjusted	**0.53**	**0.04**	**1.02**	**0.03**	2967
CARE-PF	NO_3^-	continuous	adjusted	-0.14	-0.92	0.64	0.72	2958
Simmons	NO_3^-	dichotomized	unadjusted	0.09	-0.81	1.00	0.84	1076
Simmons	NO_3^-	dichotomized	adjusted	0.01	-0.89	0.90	0.99	1051
PFF	NO_3^-	dichotomized	unadjusted	-0.07	-0.69	0.56	0.84	1168
PFF	NO_3^-	dichotomized	adjusted	-0.13	-1.35	1.09	0.83	1148
CARE-PF	NO_3^-	dichotomized	unadjusted	0.33	-0.04	0.69	0.08	2967
CARE-PF	NO_3^-	dichotomized	adjusted	-0.04	-0.56	0.47	0.87	2958
Simmons	NH_4^+	continuous	unadjusted	**0.76**	**0.07**	**1.45**	**0.03**	1076
Simmons	NH_4^+	continuous	adjusted	0.23	-0.47	0.94	0.51	1051
PFF	NH_4^+	continuous	unadjusted	0.53	-0.26	1.31	0.19	1168
PFF	NH_4^+	continuous	adjusted	0.08	-1.03	1.18	0.89	1148
CARE-PF	NH_4^+	continuous	unadjusted	**0.78**	**0.18**	**1.38**	**0.01**	2967
CARE-PF	NH_4^+	continuous	adjusted	-0.80	-1.77	0.17	0.11	2958
Simmons	NH_4^+	dichotomized	unadjusted	**4.16**	**1.06**	**7.27**	**0.009**	1076
Simmons	NH_4^+	dichotomized	adjusted	**1.41**	**0.68**	**7.49**	**0.02**	1051
PFF	NH_4^+	dichotomized	unadjusted	0.39	-0.27	1.04	0.25	1168
PFF	NH_4^+	dichotomized	adjusted	0.51	-0.37	1.40	0.26	1148
CARE-PF	NH_4^+	dichotomized	unadjusted	**0.61**	**0.23**	**0.99**	**0.002**	2967
CARE-PF	NH_4^+	dichotomized	adjusted	-0.07	-0.64	0.50	0.81	2958
Simmons	BC	continuous	unadjusted	-0.01	-1.88	1.86	0.99	1076
Simmons	BC	continuous	adjusted	-0.47	-2.32	1.39	0.62	1051
PFF	BC	continuous	unadjusted	0.45	-0.79	1.70	0.48	1168
PFF	BC	continuous	adjusted	-0.43	-2.17	1.30	0.62	1148
CARE-PF	BC	continuous	unadjusted	**1.46**	**0.47**	**2.44**	**0.004**	2967
CARE-PF	BC	continuous	adjusted	0.49	-0.62	1.59	0.39	2958
Simmons	BC	dichotomized	unadjusted	0.13	-1.16	1.42	0.85	1076
Simmons	BC	dichotomized	adjusted	-0.05	-1.32	1.22	0.94	1051
PFF	BC	dichotomized	unadjusted	0.16	-0.48	0.79	0.63	1168
PFF	BC	dichotomized	adjusted	-0.38	-1.18	0.43	0.36	1148
CARE-PF	BC	dichotomized	unadjusted	0.18	-0.22	0.57	0.38	2967

CARE-PF	BC	dichotomized	adjusted	0.12	-0.28	0.52	0.56	2958
Simmons	OM	continuous	unadjusted	0.29	-0.17	0.75	0.22	1076
Simmons	OM	continuous	adjusted	0.04	-0.42	0.50	0.87	1051
PFF	OM	continuous	unadjusted	0.10	-0.23	0.42	0.55	1168
PFF	OM	continuous	adjusted	-0.12	-0.54	0.31	0.59	1148
CARE-PF	OM	continuous	unadjusted	**0.14**	**0.02**	**0.27**	**0.02**	2967
CARE-PF	OM	continuous	adjusted	0.08	-0.05	0.21	0.22	2958
Simmons	OM	dichotomized	unadjusted	0.09	-0.59	0.77	0.80	1076
Simmons	OM	dichotomized	adjusted	-0.20	-0.88	0.48	0.56	1051
PFF	OM	dichotomized	unadjusted	0.37	-0.26	0.99	0.25	1168
PFF	OM	dichotomized	adjusted	-0.15	-0.94	0.64	0.70	1148
CARE-PF	OM	dichotomized	unadjusted	**0.36**	**0.00**	**0.73**	**0.05**	2967
CARE-PF	OM	dichotomized	adjusted	0.26	-0.12	0.64	0.18	2958
Simmons	SS	continuous	unadjusted	2.52	-0.67	5.71	0.12	1076
Simmons	SS	continuous	adjusted	1.00	-2.39	4.39	0.56	1051
PFF	SS	continuous	unadjusted	0.04	-0.96	1.04	0.94	1168
PFF	SS	continuous	adjusted	-0.74	-2.54	1.06	0.42	1148
CARE-PF	SS	continuous	unadjusted	1.29	-0.11	2.68	0.07	2967
CARE-PF	SS	continuous	adjusted	0.16	-1.76	2.08	0.87	2958
Simmons	SS	dichotomized	unadjusted	0.56	-0.15	1.27	0.12	1076
Simmons	SS	dichotomized	adjusted	0.09	-0.63	0.81	0.81	1051
PFF	SS	dichotomized	unadjusted	0.32	-0.31	0.95	0.33	1168
PFF	SS	dichotomized	adjusted	-0.06	-0.99	0.88	0.91	1148
CARE-PF	SS	dichotomized	unadjusted	-0.28	-1.63	1.07	0.68	2967
CARE-PF	SS	dichotomized	adjusted	-0.24	-1.57	1.09	0.73	2958
Simmons	Soil	continuous	unadjusted	0.17	-2.53	2.87	0.90	1076
Simmons	Soil	continuous	adjusted	-0.58	-3.31	2.15	0.68	1051
PFF	Soil	continuous	unadjusted	0.25	-0.48	0.98	0.50	1168
PFF	Soil	continuous	adjusted	0.65	-1.53	2.83	0.56	1148
CARE-PF	Soil	continuous	unadjusted	1.08	-0.21	2.36	0.10	2967
CARE-PF	Soil	continuous	adjusted	1.11	-0.36	2.59	0.14	2958
Simmons	Soil	dichotomized	unadjusted	-0.27	-1.13	0.58	0.53	1076
Simmons	Soil	dichotomized	adjusted	-0.47	-1.30	0.37	0.27	1051
PFF	Soil	dichotomized	unadjusted	**0.89**	**0.24**	**1.54**	**0.007**	1168
PFF	Soil	dichotomized	adjusted	**1.06**	**0.04**	**2.08**	**0.04**	1148
CARE-PF	Soil	dichotomized	unadjusted	-0.08	-0.41	0.31	0.69	2967
CARE-PF	Soil	dichotomized	adjusted	-0.02	-0.44	0.39	0.91	2958

Appendix Table 23 – Full cohort D_LCO decline models. Associations of 5-years pre-censoring and 5-years pre-enrollment exposures to $PM_{2.5}$ or constituent components (sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic matter (OM), sea salt (SS), soil) with rate of change in diffusion capacity of the lung for carbon monoxide (D_LCO). Adjusted models include age at enrollment, sex, smoking history, race, a socioeconomic variable, and site (for PFF and CARE-PF). Significant associations are **bolded.**

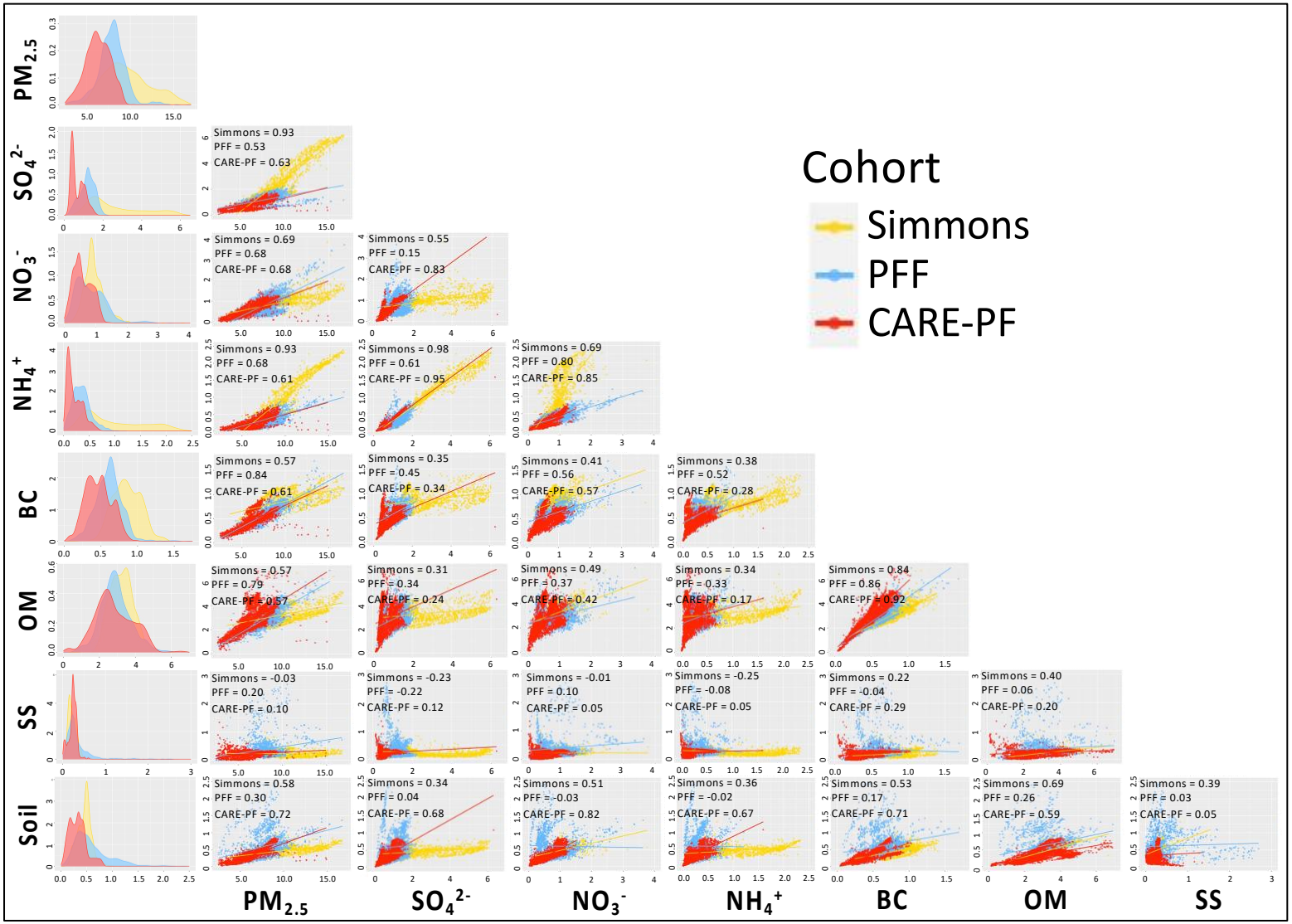
Cohort	Pollutant	Method	Model	β	2.5 CI	97.5 CI	p	n
D_LCO Decline – 5-years pre-censoring models								
Simmons	$PM_{2.5}$	continuous	unadjusted	**-0.31**	**-0.45**	**-0.17**	**<0.001**	1036
Simmons	$PM_{2.5}$	continuous	adjusted	**-0.28**	**-0.42**	**-0.15**	**<0.001**	1013
PFF	$PM_{2.5}$	continuous	unadjusted	**0.27**	**0.03**	**0.52**	**0.03**	1109
PFF	$PM_{2.5}$	continuous	adjusted	0.09	-0.24	0.43	0.58	1090
CARE-PF	$PM_{2.5}$	continuous	unadjusted	0.09	-0.03	0.21	0.14	2783
CARE-PF	$PM_{2.5}$	continuous	adjusted	0.08	-0.04	0.21	0.19	2775
Simmons	$PM_{2.5}$	dichotomized	unadjusted	**-1.06**	**-1.74**	**-0.37**	**0.002**	1036
Simmons	$PM_{2.5}$	dichotomized	adjusted	**-0.81**	**-1.49**	**-0.14**	**0.02**	1013
PFF	$PM_{2.5}$	dichotomized	unadjusted	**1.07**	**0.31**	**1.82**	**0.006**	1109
PFF	$PM_{2.5}$	dichotomized	adjusted	0.41	-0.55	1.38	0.40	1090
CARE-PF	$PM_{2.5}$	dichotomized	unadjusted	-0.42	-1.04	0.20	0.19	2783
CARE-PF	$PM_{2.5}$	dichotomized	adjusted	-0.35	-0.99	0.28	0.27	2775
Simmons	SO_4^{2-}	continuous	unadjusted	**-0.75**	**-1.01**	**-0.49**	**<0.001**	1036
Simmons	SO_4^{2-}	continuous	adjusted	**-0.67**	**-0.93**	**-0.41**	**<0.001**	1013
PFF	SO_4^{2-}	continuous	unadjusted	0.46	-0.74	1.65	0.46	1109
PFF	SO_4^{2-}	continuous	adjusted	**-2.93**	**-5.28**	**-0.58**	**0.02**	1090
CARE-PF	SO_4^{2-}	continuous	unadjusted	-0.19	-0.81	0.43	0.54	2783
CARE-PF	SO_4^{2-}	continuous	adjusted	**-3.29**	**-4.64**	**-1.95**	**<0.001**	2775
Simmons	SO_4^{2-}	dichotomized	unadjusted	-1.15	-2.86	0.56	0.19	1036
Simmons	SO_4^{2-}	dichotomized	adjusted	-0.84	-2.50	0.82	0.32	1013
PFF	SO_4^{2-}	dichotomized	unadjusted	0.30	-0.50	1.10	0.46	1109
PFF	SO_4^{2-}	dichotomized	adjusted	**-1.46**	**-2.85**	**-0.07**	**0.04**	1090
CARE-PF	SO_4^{2-}	dichotomized	unadjusted	**-0.80**	**-1.31**	**-0.29**	**0.002**	2783
CARE-PF	SO_4^{2-}	dichotomized	adjusted	**-1.67**	**-2.38**	**-0.96**	**<0.001**	2775
Simmons	NO_3^-	continuous	unadjusted	**-2.94**	**-4.14**	**-1.75**	**<0.001**	1036
Simmons	NO_3^-	continuous	adjusted	**-2.61**	**-3.80**	**-1.43**	**<0.001**	1013
PFF	NO_3^-	continuous	unadjusted	-0.19	-1.07	0.69	0.67	1109
PFF	NO_3^-	continuous	adjusted	-0.22	-1.77	1.34	0.79	1090
CARE-PF	NO_3^-	continuous	unadjusted	-0.47	-1.21	0.26	0.21	2783
CARE-PF	NO_3^-	continuous	adjusted	-0.66	-1.79	0.46	0.25	2775
Simmons	NO_3^-	dichotomized	unadjusted	**-1.30**	**-2.10**	**-0.51**	**0.001**	1036
Simmons	NO_3^-	dichotomized	adjusted	**-0.92**	**-1.69**	**-0.14**	**0.02**	1013
PFF	NO_3^-	dichotomized	unadjusted	-0.08	-0.83	0.68	0.84	1109
PFF	NO_3^-	dichotomized	adjusted	-0.94	-2.26	0.38	0.16	1090
CARE-PF	NO_3^-	dichotomized	unadjusted	**-0.67**	**-1.10**	**-0.24**	**0.002**	2783

CARE-PF	NO_3^-	dichotomized	adjusted	**-0.74**	**-1.33**	**-0.15**	**0.01**	2775
Simmons	NH_4^+	continuous	unadjusted	**-1.93**	**-2.55**	**-1.32**	**<0.001**	1036
Simmons	NH_4^+	continuous	adjusted	**-1.74**	**-2.35**	**-1.12**	**<0.001**	1013
PFF	NH_4^+	continuous	unadjusted	-1.80	-4.44	0.84	0.18	1109
PFF	NH_4^+	continuous	adjusted	-4.04	-8.15	0.07	0.05	1090
CARE-PF	NH_4^+	continuous	unadjusted	**-2.29**	**-3.67**	**-0.90**	**0.001**	2783
CARE-PF	NH_4^+	continuous	adjusted	**-8.42**	**-10.78**	**-6.06**	**<0.001**	2775
Simmons	NH_4^+	dichotomized	unadjusted	**-1.56**	**-2.98**	**-0.13**	**0.03**	1036
Simmons	NH_4^+	dichotomized	adjusted	-1.14	-2.53	0.24	0.11	1013
PFF	NH_4^+	dichotomized	unadjusted	-0.47	-1.23	0.28	0.22	1109
PFF	NH_4^+	dichotomized	adjusted	-0.83	-1.93	0.27	0.14	1090
CARE-PF	NH_4^+	dichotomized	unadjusted	**-0.82**	**-1.27**	**-0.36**	**<0.001**	2783
CARE-PF	NH_4^+	dichotomized	adjusted	**-1.88**	**-2.54**	**-1.23**	**<0.001**	2775
Simmons	BC	continuous	unadjusted	1.17	-0.51	2.85	1.361	1036
Simmons	BC	continuous	adjusted	0.88	-0.79	2.55	0.30	1013
PFF	BC	continuous	unadjusted	**3.54**	**1.31**	**5.78**	**0.002**	1109
PFF	BC	continuous	adjusted	0.43	-2.71	3.58	0.79	1090
CARE-PF	BC	continuous	unadjusted	**1.28**	**0.03**	**2.53**	**0.045**	2783
CARE-PF	BC	continuous	adjusted	**1.36**	**0.11**	**2.62**	**0.03**	2775
Simmons	BC	dichotomized	unadjusted	-0.36	-1.66	0.94	0.58	1036
Simmons	BC	dichotomized	adjusted	-0.16	-1.43	1.11	0.80	1013
PFF	BC	dichotomized	unadjusted	**0.98**	**0.22**	**1.74**	**0.01**	1109
PFF	BC	dichotomized	adjusted	0.18	-0.82	1.19	0.72	1090
CARE-PF	BC	dichotomized	unadjusted	0.04	-0.12	0.85	0.14	2783
CARE-PF	BC	dichotomized	adjusted	**0.63**	**0.14**	**1.12**	**0.01**	2775
Simmons	OM	continuous	unadjusted	0.40	-0.12	0.92	0.13	1036
Simmons	OM	continuous	adjusted	0.28	-0.23	0.80	0.29	1013
PFF	OM	continuous	unadjusted	**0.64**	**0.22**	**1.07**	**0.003**	1109
PFF	OM	continuous	adjusted	0.41	-0.21	1.04	0.19	1090
CARE-PF	OM	continuous	unadjusted	0.13	-0.08	0.34	0.21	2783
CARE-PF	OM	continuous	adjusted	0.09	-0.11	0.29	0.38	2775
Simmons	OM	dichotomized	unadjusted	0.11	-0.60	0.83	0.76	1036
Simmons	OM	dichotomized	adjusted	-0.14	-0.85	0.56	0.69	1013
PFF	OM	dichotomized	unadjusted	0.47	-0.29	1.23	0.22	1109
PFF	OM	dichotomized	adjusted	0.03	-0.95	1.01	0.96	1090
CARE-PF	OM	dichotomized	unadjusted	0.32	-0.10	0.73	0.13	2783
CARE-PF	OM	dichotomized	adjusted	0.27	-0.13	0.68	0.19	2775
Simmons	SS	continuous	unadjusted	**8.02**	**5.17**	**10.87**	**<0.001**	1036
Simmons	SS	continuous	adjusted	**7.97**	**4.86**	**11.08**	**<0.001**	1013
PFF	SS	continuous	unadjusted	-0.82	-2.00	0.36	0.17	1109
PFF	SS	continuous	adjusted	-0.32	-2.35	1.72	0.76	1090

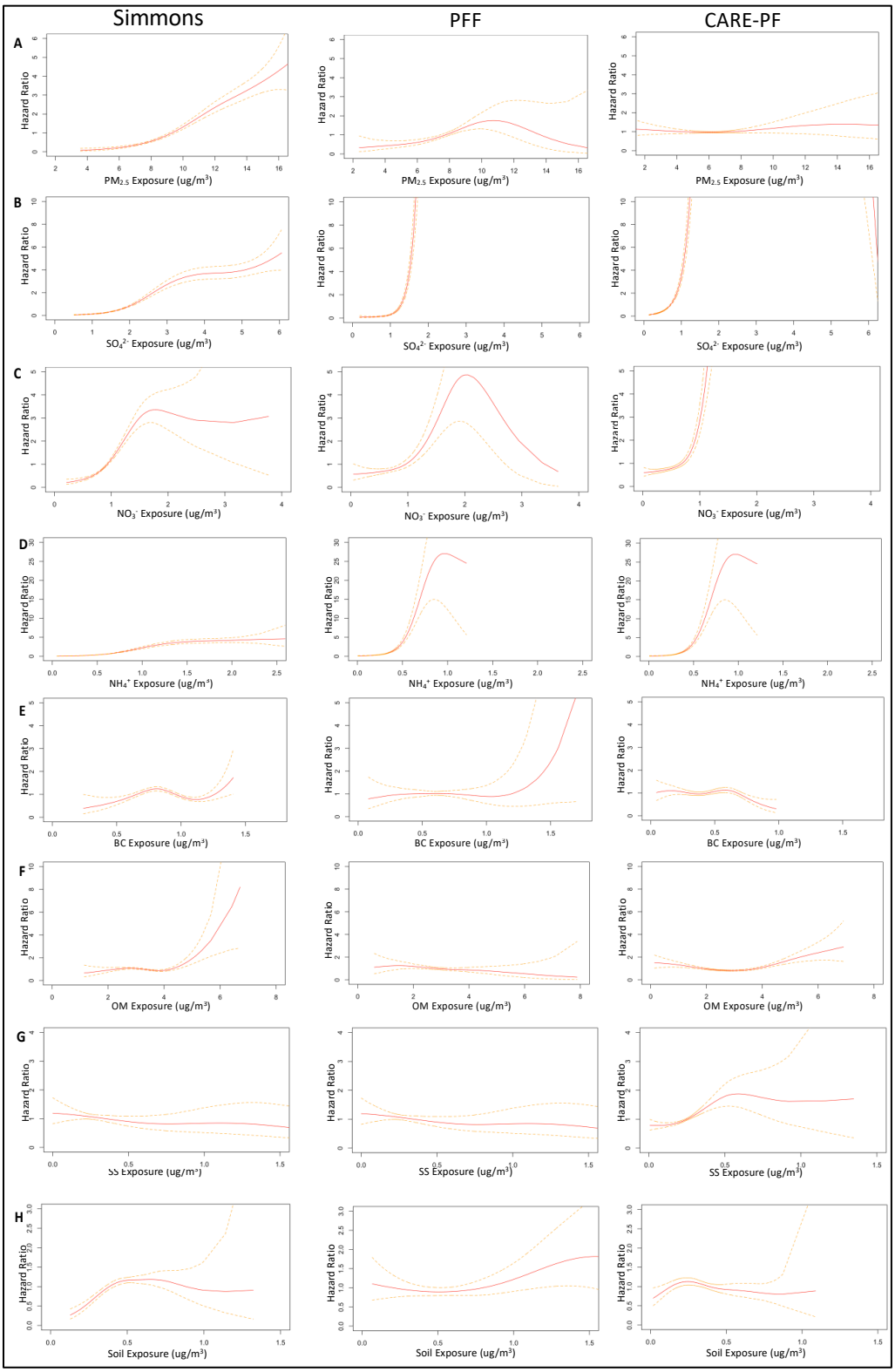
CARE-PF	SS	continuous	unadjusted	**3.24**	**1.40**	**5.08**	**<0.001**	2783
CARE-PF	SS	continuous	adjusted	-0.33	-2.64	1.99	0.78	2775
Simmons	SS	dichotomized	unadjusted	**1.64**	**1.02**	**2.27**	**<0.001**	1036
Simmons	SS	dichotomized	adjusted	**1.40**	**0.79**	**2.02**	**<0.001**	1013
PFF	SS	dichotomized	unadjusted	0.28	-0.48	1.03	0.48	1109
PFF	SS	dichotomized	adjusted	0.35	-0.60	1.30	0.47	1090
CARE-PF	SS	dichotomized	unadjusted	**0.61**	**0.01**	**1.21**	**0.047**	2783
CARE-PF	SS	dichotomized	adjusted	0.35	-0.31	1.02	0.30	2775
Simmons	Soil	continuous	unadjusted	-1.14	-3.56	1.28	0.36	1036
Simmons	Soil	continuous	adjusted	-1.00	-3.39	1.40	0.42	1013
PFF	Soil	continuous	unadjusted	0.08	-0.85	1.01	0.87	1109
PFF	Soil	continuous	adjusted	1.53	-0.55	3.62	0.15	1090
CARE-PF	Soil	continuous	unadjusted	-0.40	-1.59	0.79	0.51	2783
CARE-PF	Soil	continuous	adjusted	0.63	-0.81	2.07	0.39	2775
Simmons	Soil	dichotomized	unadjusted	-0.43	-1.15	0.29	0.24	1036
Simmons	Soil	dichotomized	adjusted	-0.16	-0.88	0.56	0.67	1013
PFF	Soil	dichotomized	unadjusted	0.25	-0.58	1.07	0.56	1109
PFF	Soil	dichotomized	adjusted	0.45	-0.67	1.57	0.43	1090
CARE-PF	Soil	dichotomized	unadjusted	-0.20	-0.62	0.21	0.34	2783
CARE-PF	Soil	dichotomized	adjusted	-0.06	-0.53	0.41	0.81	2775
D_LCO Decline – 5-years pre-enrollment models								
Simmons	$PM_{2.5}$	continuous	unadjusted	0.12	-0.02	0.26	0.10	1032
Simmons	$PM_{2.5}$	continuous	adjusted	0.02	-0.13	0.17	0.79	1009
PFF	$PM_{2.5}$	continuous	unadjusted	**0.34**	**0.15**	**0.53**	**<0.001**	1104
PFF	$PM_{2.5}$	continuous	adjusted	0.09	-0.17	0.35	0.51	1085
CARE-PF	$PM_{2.5}$	continuous	unadjusted	**0.14**	**0.04**	**0.24**	**0.008**	2782
CARE-PF	$PM_{2.5}$	continuous	adjusted	0.00	-0.14	0.15	0.95	2774
Simmons	$PM_{2.5}$	dichotomized	unadjusted	1.58	-0.36	3.51	0.11	1032
Simmons	$PM_{2.5}$	dichotomized	adjusted	1.78	-0.18	3.75	0.08	1009
PFF	$PM_{2.5}$	dichotomized	unadjusted	**1.07**	**0.19**	**1.95**	**0.02**	1104
PFF	$PM_{2.5}$	dichotomized	adjusted	0.49	-0.56	1.55	0.36	1085
CARE-PF	$PM_{2.5}$	dichotomized	unadjusted	0.32	-0.11	0.75	0.15	2782
CARE-PF	$PM_{2.5}$	dichotomized	adjusted	-0.16	-0.74	0.41	0.58	2774
Simmons	SO_4^{2-}	continuous	unadjusted	0.25	-0.02	0.52	0.07	1032
Simmons	SO_4^{2-}	continuous	adjusted	0.05	-0.23	0.32	0.74	1009
PFF	SO_4^{2-}	continuous	unadjusted	**0.73**	**0.27**	**1.19**	**0.002**	1104
PFF	SO_4^{2-}	continuous	adjusted	0.03	-0.62	0.68	0.92	1085
CARE-PF	SO_4^{2-}	continuous	unadjusted	0.27	-0.06	0.59	0.11	2782
CARE-PF	SO_4^{2-}	continuous	adjusted	-0.47	-1.06	0.12	0.12	2774
Simmons	SO_4^{2-}	dichotomized	unadjusted	2.05	-2.55	6.65	0.38	1032

Simmons	SO_4^{2-}	dichotomized	adjusted	3.91	-1.65	9.47	0.17	1009
PFF	SO_4^{2-}	dichotomized	unadjusted	0.84	-0.03	1.71	0.06	1104
PFF	SO_4^{2-}	dichotomized	adjusted	-0.26	-1.53	1.02	0.69	1085
CARE-PF	SO_4^{2-}	dichotomized	unadjusted	-0.12	-0.58	0.34	0.61	2782
CARE-PF	SO_4^{2-}	dichotomized	adjusted	**-0.95**	**-1.52**	**-0.37**	**0.001**	2774
Simmons	NO_3^-	continuous	unadjusted	0.41	-0.65	1.47	0.45	1032
Simmons	NO_3^-	continuous	adjusted	-0.11	-1.16	0.94	0.84	1009
PFF	NO_3^-	continuous	unadjusted	0.16	-0.43	0.74	0.60	1104
PFF	NO_3^-	continuous	adjusted	0.43	-0.63	1.50	0.43	1085
CARE-PF	NO_3^-	continuous	unadjusted	-0.13	-0.68	0.42	0.64	2782
CARE-PF	NO_3^-	continuous	adjusted	-0.11	-0.92	0.70	0.79	2774
Simmons	NO_3^-	dichotomized	unadjusted	-0.32	-1.26	0.62	0.5	1032
Simmons	NO_3^-	dichotomized	adjusted	-0.26	-1.18	0.66	0.58	1009
PFF	NO_3^-	dichotomized	unadjusted	0.22	-0.53	0.97	0.57	1104
PFF	NO_3^-	dichotomized	adjusted	0.31	-1.10	1.72	0.67	1085
CARE-PF	NO_3^-	dichotomized	unadjusted	-0.25	-0.66	0.16	0.23	2782
CARE-PF	NO_3^-	dichotomized	adjusted	-0.28	-0.82	0.26	0.32	2774
Simmons	NH_4^+	continuous	unadjusted	0.52	-0.19	1.24	0.15	1032
Simmons	NH_4^+	continuous	adjusted	-0.01	-0.75	0.72	0.97	1009
PFF	NH_4^+	continuous	unadjusted	**1.11**	**0.18**	**2.04**	**0.02**	1104
PFF	NH_4^+	continuous	adjusted	0.19	-1.07	1.46	0.76	1085
CARE-PF	NH_4^+	continuous	unadjusted	-0.06	-0.73	0.60	0.85	2782
CARE-PF	NH_4^+	continuous	adjusted	**-1.34**	**-2.35**	**-0.34**	**0.009**	2774
Simmons	NH_4^+	dichotomized	unadjusted	1.68	-2.11	5.46	0.38	1032
Simmons	NH_4^+	dichotomized	adjusted	2.36	-1.98	6.71	0.29	1009
PFF	NH_4^+	dichotomized	unadjusted	**1.07**	**0.28**	**1.86**	**0.008**	1104
PFF	NH_4^+	dichotomized	adjusted	0.68	-0.35	1.71	0.19	1085
CARE-PF	NH_4^+	dichotomized	unadjusted	0.02	-0.41	0.45	0.93	2782
CARE-PF	NH_4^+	dichotomized	adjusted	-0.57	-1.17	0.03	0.06	2774
Simmons	BC	continuous	unadjusted	-0.11	-2.02	1.79	0.91	1032
Simmons	BC	continuous	adjusted	-0.60	-2.48	1.28	0.53	1009
PFF	BC	continuous	unadjusted	**2.31**	**0.83**	**3.79**	**0.002**	1104
PFF	BC	continuous	adjusted	0.42	-1.60	2.44	0.68	1085
CARE-PF	BC	continuous	unadjusted	0.98	-0.12	2.08	0.08	2782
CARE-PF	BC	continuous	adjusted	0.03	-1.15	1.21	0.96	2774
Simmons	BC	dichotomized	unadjusted	1.00	-0.34	2.35	0.14	1032
Simmons	BC	dichotomized	adjusted	1.08	-0.24	2.40	0.11	1009
PFF	BC	dichotomized	unadjusted	**0.89**	**0.13**	**1.65**	**0.02**	1104
PFF	BC	dichotomized	adjusted	0.23	-0.69	1.15	0.63	1085
CARE-PF	BC	dichotomized	unadjusted	0.21	-0.23	0.65	0.35	2782

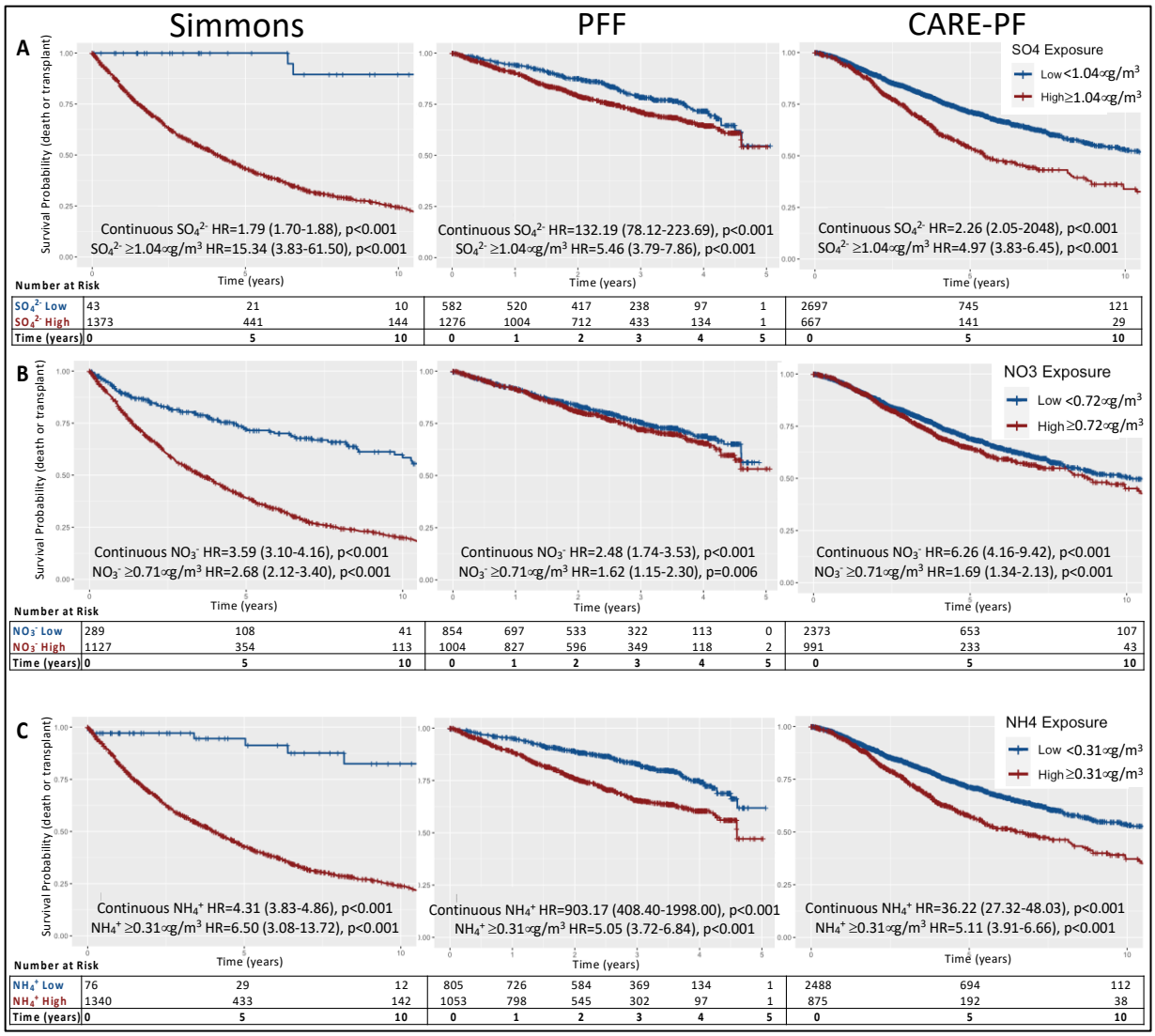
CARE-PF	BC	dichotomized	adjusted	0.06	-0.37	0.49	0.79	2774
Simmons	OM	continuous	unadjusted	0.33	-0.13	0.80	0.16	1032
Simmons	OM	continuous	adjusted	0.02	-0.45	0.49	0.93	1009
PFF	OM	continuous	unadjusted	**0.45**	**0.06**	**0.83**	**0.03**	1104
PFF	OM	continuous	adjusted	0.17	-0.32	0.67	0.49	1085
CARE-PF	OM	continuous	unadjusted	0.13	-0.01	0.26	0.07	2782
CARE-PF	OM	continuous	adjusted	0.02	-0.12	0.15	0.82	2774
Simmons	OM	dichotomized	unadjusted	-0.07	-0.76	0.63	0.85	1032
Simmons	OM	dichotomized	adjusted	-0.41	-1.11	0.28	0.25	1009
PFF	OM	dichotomized	unadjusted	**0.96**	**0.21**	**1.71**	**0.01**	1104
PFF	OM	dichotomized	adjusted	0.43	-0.49	1.34	0.36	1085
CARE-PF	OM	dichotomized	unadjusted	0.38	-0.03	0.79	0.07	2782
CARE-PF	OM	dichotomized	adjusted	0.19	-0.22	0.60	0.37	2774
Simmons	SS	continuous	unadjusted	**4.42**	**1.22**	**7.62**	**0.007**	1032
Simmons	SS	continuous	adjusted	2.19	-1.18	5.56	0.20	1009
PFF	SS	continuous	unadjusted	-0.85	-2.09	0.39	0.18	1104
PFF	SS	continuous	adjusted	-0.24	-2.32	1.84	0.82	1085
CARE-PF	SS	continuous	unadjusted	**2.75**	**1.22**	**4.27**	**<0.001**	2782
CARE-PF	SS	continuous	adjusted	-0.26	-2.27	1.75	0.80	2774
Simmons	SS	dichotomized	unadjusted	**0.96**	**0.25**	**1.68**	**0.008**	1032
Simmons	SS	dichotomized	adjusted	0.46	-0.26	1.18	0.21	1009
PFF	SS	dichotomized	unadjusted	0.00	-0.76	0.76	0.995	1104
PFF	SS	dichotomized	adjusted	-0.02	-1.09	1.05	0.97	1085
CARE-PF	SS	dichotomized	unadjusted	-0.14	-1.65	1.36	0.85	2782
CARE-PF	SS	dichotomized	adjusted	-0.94	-2.37	0.49	0.20	2774
Simmons	Soil	continuous	unadjusted	1.24	-1.54	4.02	0.38	1032
Simmons	Soil	continuous	adjusted	0.18	-2.63	2.99	0.90	1009
PFF	Soil	continuous	unadjusted	-0.07	-0.92	0.79	0.88	1104
PFF	Soil	continuous	adjusted	1.31	-1.18	3.81	0.30	1085
CARE-PF	Soil	continuous	unadjusted	-1.18	-2.67	0.32	0.12	2782
CARE-PF	Soil	continuous	adjusted	-0.17	-1.79	1.45	0.84	2774
Simmons	Soil	dichotomized	unadjusted	-0.26	-1.13	0.62	0.57	1032
Simmons	Soil	dichotomized	adjusted	-0.38	-1.24	0.48	0.39	1009
PFF	Soil	dichotomized	unadjusted	0.71	-0.07	1.49	0.08	1104
PFF	Soil	dichotomized	adjusted	1.05	-0.12	2.22	0.08	1085
CARE-PF	Soil	dichotomized	unadjusted	-0.66	-1.10	-0.22	0.004	2782
CARE-PF	Soil	dichotomized	adjusted	-0.29	-0.74	0.15	0.20	2774



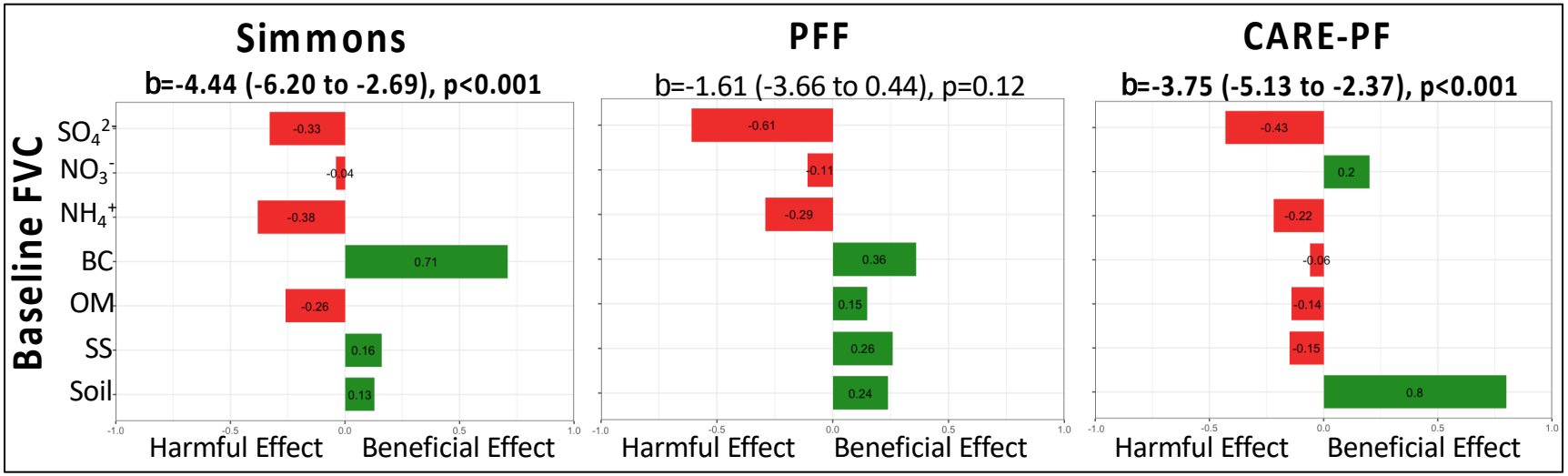
Appendix Figure 5 – Scatterplot matrix of 5-year pre-censoring exposures to $PM_{2.5}$ and constituent components. Correlation coefficients between pollutants are shown on each plot for each cohort. Points in yellow reflect values for patients in the Simmons cohort, blue for the Pulmonary Fibrosis Foundation (PFF) cohort, and red for the Canadian Registry for Pulmonary Fibrosis (CARE-PF) cohort. Density plots are shown on the left-most column to reflect the distribution of $PM_{2.5}$ and constituent components.



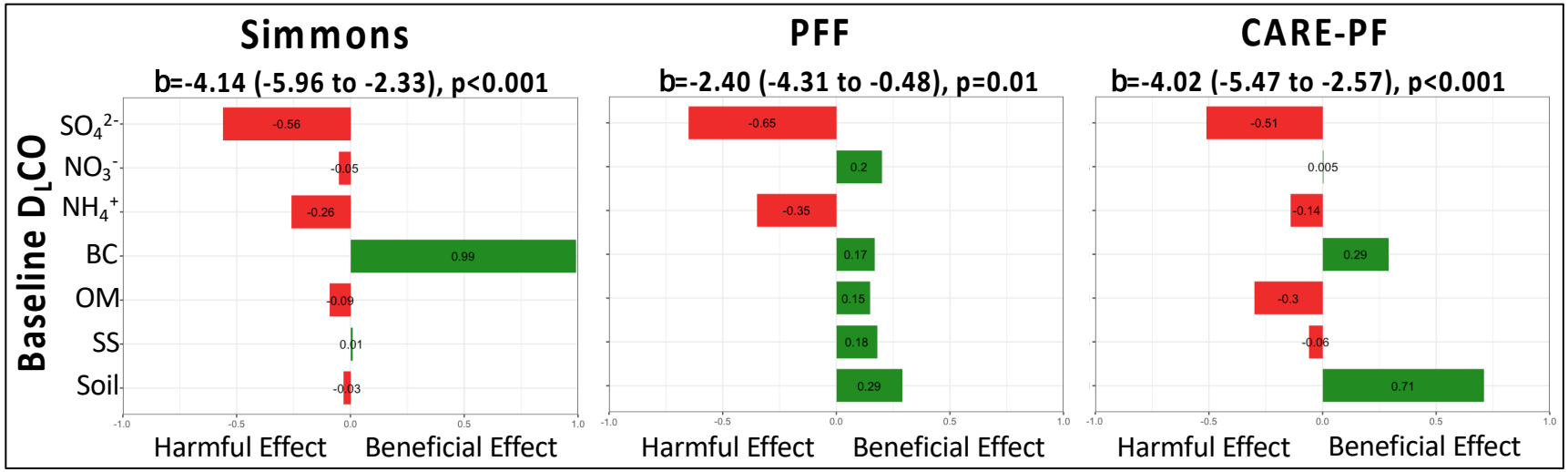
Appendix Figure 6 – $PM_{2.5}$ and constituent spline models. Adjusted spline models of continuous hazard ratio for increasing 5-year pre-censoring exposures to A) $PM_{2.5}$ and constituents B) sulfate (SO_4^{2-}), C) nitrate (NO_3^-), D) ammonium (NH_4^+), E) black carbon (BC), F) organic matter (OM), G) sea salt (SS), and H) soil all in $\mu g/m^3$ across the three cohorts (Simmons, CARE-PF, PFF). Adjustments made for age at diagnosis, sex, smoking history, race, a socioeconomic variable, and site (for PFF and CARE-PF only).



Appendix Figure 7 – Survival by low versus high anthropogenic constituent exposures in 5-years pre-censoring. Kaplan-Meier survival curves for associations of exposures to A) sulfate (SO_4^{2-}), B) nitrate (NO_3^-), and C) ammonium (NH_4^+) in the 5-years pre-censoring, where death and transplant are considered composite outcomes. Hazard ratios (HR) reported for dichotomized and continuous models are adjusted for age at enrollment, sex, race, smoking history, a socioeconomic variable, and site (PFF and CARE-PF only).



Appendix Figure 8 – Tornado plots of $PM_{2.5}$ constituent impacts on baseline forced vital capacity (FVC) in multi-pollutant models. Results are reported from adjusted quantile-based g-computation linear regression models where 5-year pre-censoring estimates for SO_4^{2-}, NO_3^-, NH_4^+, BC, OM, SS, and soil were included. All models were adjusted for age at enrollment, sex, race, smoking history, a socioeconomic variable, and site (for PFF and CARE-PF). The weight of effect in a direction is displayed over the bars of each plot with bars representing harmful effects displayed in red and bars representing protective effects in green. The sum of all positive weights equals 1 and all negative weights equals -1 (i.e. cannot directly compare effect size between positive and negative weights). The β-value (95% CI) and p-value for a 1-quantile increase in the overall mixture is reported above each plot.



Appendix Figure 9 – Tornado plots of $PM_{2.5}$ constituent impacts on baseline diffusion capacity of the lung for carbon monoxide (D_LCO) in multi-pollutant models. Results are reported from adjusted quantile-based g-computation linear regression models where 5-year pre-censoring estimates for SO_4^{2-}, NO_3^-, NH_4^+, BC, OM, SS, and soil were included. All models were adjusted for age at enrollment, sex, race, smoking history, a socioeconomic variable, and site (for PFF and CARE-PF). The weight of effect in a direction is displayed over the bars of each plot with bars representing harmful effects displayed in red and bars representing protective effects in green. The sum of all positive weights equals 1 and all negative weights equals -1 (i.e. cannot directly compare effect size between positive and negative weights). The β-value (95% CI) and p-value for a 1-quantile increase in the overall mixture is reported above each plot.

Appendix C – Data Supplement for $PM_{2.5}$ Impacts on Global DNA Methylation

Online Data Supplement: $PM_{2.5}$ and constituent component impacts on global DNA methylation in patients with idiopathic pulmonary fibrosis

Gillian C. Goobie, MD[1,2,3]; Xiaoyun Li, MD[4]; Christopher J. Ryerson, MD, MAS[5,6]; Christopher Carlsten, MD MPH[5,7]; Kerri A. Johannson, MD, MPH[8]; James P. Fabisiak, PhD[9]; Kathleen O. Lindell, RN, PhD[2,10]; Xiaoping Chen, MS[2,4]; Kevin F. Gibson, MD[2,4]; Daniel J. Kass, MD[2,4]; S. Mehdi Nouraie, MD, PhD[4]; Yingze Zhang, PhD[1,2,4]

1. Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, U.S.A. Email: goobiegc@upmc.edu.
2. Simmons Center for Interstitial Lung Disease, Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A.
3. Clinician Investigator Program, Department of Medicine, University of British Columbia, Vancouver, B.C., Canada.
4. Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, U.S.A. Emails: lix11@upmc.edu; chenx9@upmc.edu; gibsonkf@upmc.edu, kassd2@upmc.edu, zhanyx@upmc.edu, nouraies@upmc.edu.
5. Division of Respiratory Medicine, Department of Medicine, University of British Columbia, Vancouver, B.C., Canada. Email: carlsten@mail.ubc.ca, chris.ryerson@hli.ubc.ca.
6. Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, B.C., Canada.
7. Air Pollution Exposure Laboratory, Vancouver Coastal Health Research Institute, Vancouver, B.C., Canada.
8. Division of Respiratory Medicine, Department of Medicine, University of Calgary, Calgary, AB, Canada. Email: kerri.johnnson@albertahealthservices.ca.
9. Department of Environmental and Occupational Health, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, U.S.A. Email: fabs@pitt.edu.
10. College of Nursing, Medical University of South Carolina, Charleston, SC, U.S.A. Email: lindellk@musc.edu.

ORCID IDs: 0000-0001-7982-5635 (G.C.G.); 0000-0003-1205-5511 (K.A.J.); 0000-0003-0283-8985(J.P.F.); 0000-0001-6597-7830 (D.J.K.); 0000-0001-7465-0581 (S.M.N.); 0000-0001-6947-2901 (Y.Z.).

Correspondence to:
Gillian C. Goobie, MD FRCPC
Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh
130 DeSoto Street, Pittsburgh, PA, 15261
Email: goobiegc@upmc.edu; Phone: 412-624-3018; Fax: 412-624-3020

Appendix Table 24 – Breakdown of cohorts, enrollment sites within cohorts, and number of patients recruited from each site.

Cohort	State	City	Site Name	Patients Enrolled n (%)
Simmons Center for Interstitial Lung Disease	Pennsylvania	Pittsburgh	Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease at the University of Pittsburgh	313 (100%)
Pulmonary Fibrosis Foundation (PFF)	Alabama	Birmingham	University of Alabama at Birmingham	68 (9.1%)
	Arizona	Phoenix	St. Joseph's Hospital	13 (1.7%)
		Tucson	University of Arizona – Banner Health	18 (2.4%)
	California	Emeryville	Stanford Health Center	28 (3.8%)
		Los Angeles	University of California at Los Angeles (UCLA)	11 (1.5%)
		San Francisco	University of California at San Francisco (UCSF)	22 (3.0%)
	Colorado	Denver	National Jewish Health	8 (1.1%)
	Connecticut	New Haven	Yale School of Medicine	21 (2.8%)
	Florida	Miami	University of Miami	10 (1.3%)
	Georgia	Atlanta	Piedmont Healthcare	34 (4.6%)
	Illinois	Chicago	Northwestern Memorial Hospital	11 (1.5%)
		Chicago	University of Chicago	7 (0.9%)
	Kansas	Kansas City	The University of Kansas Hospital	18 (2.4%)
	Kentucky	Louisville	University of Louisville School of Medicine	11 (1.5%)
	Louisiana	New Orleans	Tulane University School of Medicine	22 (3.0%)
	Maryland	Baltimore	University of Maryland Medical Center	11 (1.5%)
	Massachusetts	Boston	Massachusetts General Hospital	16 (2.1%)
	Michigan	Ann Arbor	University of Michigan Health System	53 (7.1%)
	Minnesota	Minneapolis	University of Minnesota Medical Center	18 (2.4%)
		Rochester	Mayo Clinic	22 (3.0%)
	Missouri	St. Louis	Washington University School of Medicine	5 (0.7%)
	New York	New York	Columbia University Medical Center	19 (2.6%)
		Setauket	Stony Brook University Hospital	5 (0.7%)
		Rochester	University of Rochester Medical Center	37 (5.0%)
		New York	Weill-Cornell Medical Center	13 (1.7%)
	North Carolina	Durham	Duke University Medical Center	26 (3.5%)
	Ohio	Columbus	The Ohio State University Wexner Medical Center	20 (2.7%)
		Cincinnati	University of Cincinnati Medical Center	5 (0.7%)
	Pennsylvania	Hershey	Penn State Milton S. Hershey Medical Center	19 (2.6%)
		Philadelphia	Temple Health	28 (3.8%)
		Philadelphia	University of Pennsylvania	3 (0.4%)
		Pittsburgh	University of Pittsburgh	Excluded
	South Carolina	Charleston	Medical University of South Carolina (MUSC)	16 (2.1%)
	Tennessee	Nashville	Vanderbilt University Medical Center	24 (3.2%)
	Texas	Dallas	University of Texas Southwestern Medical Center	22 (3.0%)
		Houston	University of Texas Health Science Center	7 (0.9%)
		San Antonio	University of Texas Health Science Center	21 (2.8%)
	Utah	Salt Lake City	University of Utah Health Care	8 (1.1%)
	Virginia	Fairfax	Inova Fairfax Medical Campus	38 (5.1%)

		Charlottesville	University of Virginia Health Systems	8 (1.1%)

Appendix Table 25 – $PM_{2.5}$ and constituent component exposures (in ug/m^3) over multiple time periods pre-sampling (i.e. prior to blood draw used for DNA methylation analysis).

$PM_{2.5}$ or Constituent	Simmons Cohort Exposures Pre-Sampling Median (IQR)					PFF Cohort Exposures Pre-Sampling Median (IQR)				
	5yrs	1yr	6mo	3mo	1mo	5yrs	1yr	6mo	3mo	1mo
$PM_{2.5}$	12.0 (10.0-13.9)	11.2 (9.5-13.4)	10.9 (9.1-12.6)	10.4 (8.8-12.7)	10.5 (8.7-12.9)	8.3 (7.4-9.1)	7.6 (6.7-8.3)	7.6 (6.6-8.5)	7.4 (6.4-8.5)	7.4 (6.3-8.7)
SO_4^{2-}	4.2 (2.9-5.2)	3.6 (2.5-5.0)	3.2 (2.3-4.5)	3.0 (2.2-4.1)	3.1 (2.2-4.2)	1.5 (1.2-1.7)	1.1 (0.9-1.3)	1.1 (0.9-1.3)	1.1 (0.8-1.3)	1.1 (0.8-1.3)
NO_3^-	1.1 (0.9-1.3)	1.0 (0.9-1.2)	1.1 (0.7-1.5)	1.0 (0.6-1.6)	0.9 (0.5-1.6)	0.8 (0.5-1.3)	0.7 (0.4-1.0)	0.6 (0.4-1.0)	0.5 (0.3-0.9)	0.5 (0.3-0.9)
NH_4^+	1.6 (1.1-1.9)	1.4 (0.9-1.8)	1.4 (0.9-1.7)	1.3 (0.9-1.8)	1.3 (0.9-1.8)	0.5 (0.3-0.6)	0.3 (0.2-0.4)	0.3 (0.2-0.4)	0.2 (0.1-0.3)	0.2 (0.2-0.4)
BC	0.8 (0.8-1.0)	0.9 (0.7-1.0)	0.9 (0.7-1.0)	0.8 (0.7-1.0)	0.9 (0.7-1.1)	0.6 (0.5-0.7)	0.6 (0.5-0.7)	0.6 (0.5-0.8)	0.6 (0.4-0.7)	0.6 (0.4-0.8)
OM	3.3 (2.9-3.6)	3.3 (2.8-3.7)	3.2 (2.7-3.8)	3.1 (2.7-3.8)	3.2 (2.6-4.0)	2.7 (2.3-3.4)	3.0 (2.4-3.6)	3.0 (2.4-3.8)	3.0 (2.2-3.7)	2.9 (2.1-3.8)
SS	0.2 (0.1-0.3)	0.2 (0.1-0.3)	0.2 (0.1-0.3)	0.2 (0.1-0.3)	0.2 (0.1-0.3)	0.2 (0.2-0.4)	0.2 (0.2-0.4)	0.2 (0.1-0.4)	0.2 (0.1-0.4)	0.2 (0.1-0.5)
Soil	0.5 (0.4-0.6)	0.5 (0.4-0.6)	0.4 (0.4-0.6)	0.4 (0.3-0.6)	0.5 (0.4-0.6)	0.5 (0.4-0.7)	0.5 (0.3-0.7)	0.5 (0.3-0.7)	0.5 (0.3-0.7)	0.5 (0.3-0.7)

Appendix Table 26 – $PM_{2.5}$ and constituent component association with global DNA methylation (DNAm) proportion. β-values reflect the difference in DNAm proportion per 1ug/m^3 increase in a pollutant. Unadjusted and adjusted models (controlling for age at diagnosis, sex, smoking history, and site in PFF cohort only) are reported. Significant associations are bolded and suggestive associations are *italicized*.

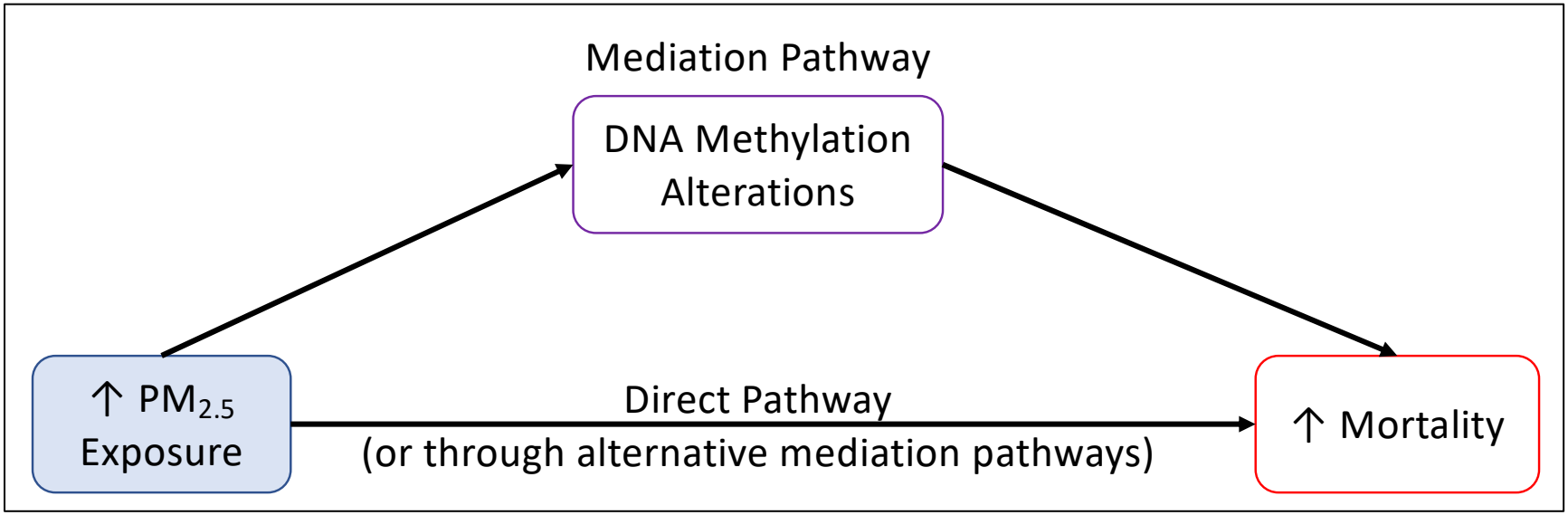
Pollutant	Simmons Cohort						PFF Cohort					
	Unadjusted Models			Adjusted Models			Unadjusted Models			Adjusted Models		
	β (95% CI)	p	n	β (95% CI)	p	n	β (95% CI)	p	n	β (95% CI)	p	n
$PM_{2.5}$ associations with %5mC												
1-yr pre-sampling	**0.03 (0.001 to 0.06)**	**0.045**	313	*0.03 (-0.004 to 0.06)*	*0.09*	313	0.01 (-0.01 to 0.04)	0.34	745	0.03 (-0.007 to 0.06)	0.12	745
3-mo pre-sampling	**0.03 (0.004 to 0.05)**	**0.02**	307	**0.02 (0.0003 to 0.05)**	**0.047**	307	0.005 (-0.02 to 0.03)	0.61	734	0.008 (-0.02 to 0.03)	0.54	734
SO_4^{2-} associations with %5mC												
1-yr pre-sampling	**0.08 (0.03 to 0.13)**	**0.004**	310	**0.07 (0.02 to 0.13)**	**0.008**	310	0.09 (-0.04 to 0.22)	0.19	741	**0.37 (0.08 to 0.65)**	**0.01**	741
3-mo pre-sampling	**0.06 (0.02 to 0.10)**	**0.002**	297	**0.06 (0.02 to 0.09)**	**0.004**	297	0.03 (-0.10 to 0.16)	0.65	621	0.08 (-0.14 to 0.30)	0.48	621
NO_3^- associations with %5mC												
1-yr pre-sampling	0.02 (-0.21 to 0.24)	0.88	310	0.006 (-0.22 to 0.23)	0.96	310	**0.10 (0.01 to 0.19)**	**0.03**	741	**0.18 (0.02 to 0.33)**	**0.02**	741
3-mo pre-sampling	0.01 (-0.10 to 0.12)	0.81	297	0.01 (-0.10 to 0.13)	0.80	297	0.02 (-0.04 to 0.08)	0.55	621	0.02 (-0.05 to 0.09)	0.61	621
NH_4^+ associations with %5mC												
1-yr pre-sampling	**0.20 (0.06 to 0.34)**	**0.004**	310	**0.19 (0.05 to 0.32)**	**0.009**	310	*0.25 (-0.02 to 0.53)*	*0.07*	741	**0.54 (0.07 to 1.00)**	**0.02**	741
3-mo pre-sampling	**0.21 (0.08 to 0.33)**	**0.001**	297	**0.19 (0.07 to 0.32)**	**0.002**	297	0.06 (-0.15 to 0.27)	0.60	621	0.08 (-0.18 to 0.33)	0.55	621
BC associations with %5mC												
1-yr pre-sampling	**0.41 (0.09 to 0.74)**	**0.01**	310	**0.36 (0.03 to 0.69)**	**0.03**	310	0.005 (-0.23 to 0.24)	0.97	741	0.05 (-0.26 to 0.35)	0.76	741
3-mo pre-sampling	**0.32 (0.06 to 0.58)**	**0.02**	297	**0.29 (0.03 to 0.55)**	**0.03**	297	0.07 (-0.12 to 0.27)	0.47	621	0.12 (-0.12 to 0.36)	0.31	621
OM associations with %5mC												
1-yr pre-sampling	0.04 (-0.05 to 0.13)	0.36	310	0.03 (-0.06 to 0.12)	0.52	310	-0.01 (-0.05 to 0.03)	0.61	741	0.02 (-0.04 to 0.08)	0.51	741
3-mo pre-sampling	0.05 (-0.03 to 0.14)	0.21	297	0.05 (-0.04 to 0.13)	0.28	297	0.001 (-0.04 to 0.04)	0.96	621	0.01 (-0.03 to 0.06)	0.54	621
SS associations with %5mC												
1-yr pre-sampling	0.36 (-0.16 to 0.87)	0.17	310	0.31 (-0.21 to 0.82)	0.24	310	-0.009 (-0.10 to 0.08)	0.85	741	0.06 (-0.09 to 0.20)	0.43	741

3-mo pre-sampling	0.003 (-0.44 to 0.45)	0.99	297	-0.03 (-0.47 to 0.42)	0.91	297	0.02 (-0.07 to 0.10)	0.72	621	0.09 (-0.03 to 0.22)	0.13	621
Soil associations with %5mC												
1-yr pre-sampling	-0.11 (-0.63 to 0.41)	0.69	310	-0.15 (-0.67 to 0.37)	0.58	310	-0.008 (-0.13 to 0.11)	0.90	741	0.10 (-0.16 to 0.35)	0.46	741
3-mo pre-sampling	-0.02 (-0.34 to 0.31)	0.92	297	-0.03 (-0.36 to 0.30)	0.86	297	0.01 (-0.08 to 0.11)	0.77	621	0.05 (-0.08 to 0.18)	0.44	621

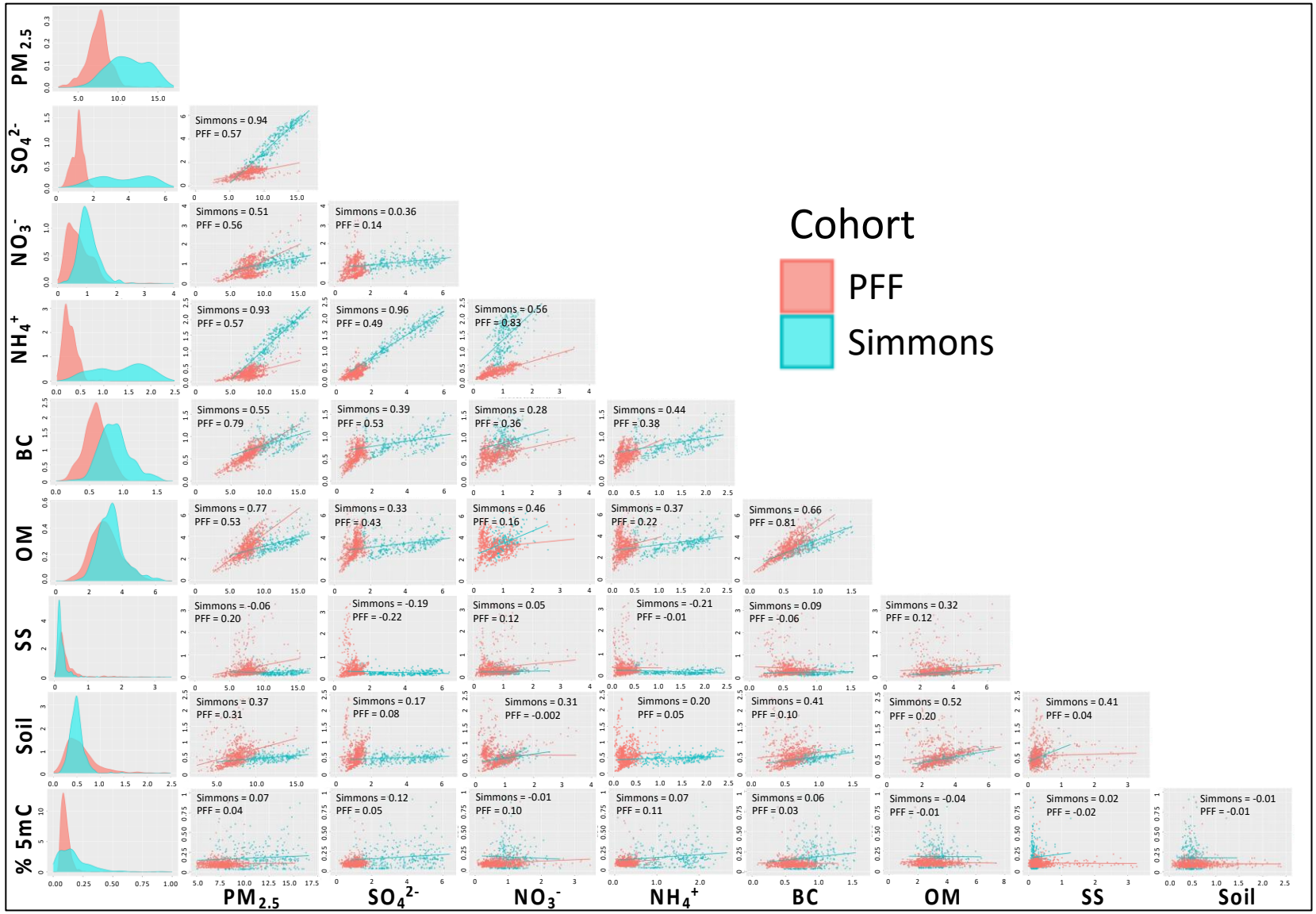
Appendix Table 27 – Associations of $PM_{2.5}$ or constituent components with mortality. Partially-adjusted models control for age at diagnosis, sex, and smoking history. Fully-adjusted models additionally control for race, baseline forced vital capacity (FVC), baseline diffusion capacity of the lung for carbon monoxide (D_LCO), and site (in PFF only). Significant associations are bolded and suggestive associations *italicized*.

Model	Simmons Cohort				PFF Cohort			
	1 Year Pre-Sampling Model							
	HR	95% CI	p	n	HR	95% CI	p	n
$PM_{2.5}$ associations with mortality								
Unadjusted	**1.08**	**1.03-1.13**	**0.002**	313	1.02	0.93-1.11	0.73	745
Partially-adjusted	**1.08**	**1.03-1.13**	**0.003**	313	1.01	0.93-1.10	0.76	745
Fully-adjusted	0.99	0.94-1.05	0.78	286	**1.16**	**1.003-1.33**	**0.045**	636
SO_4^{2-} associations with mortality								
Unadjusted	**1.16**	**1.07-1.27**	**<0.001**	310	0.95	0.62-1.45	0.81	741
Partially-adjusted	**1.16**	**1.07-1.27**	**<0.001**	310	0.94	0.62-1.44	0.78	741
Fully-adjusted	1.01	0.92-1.11	0.8	283	**3.82**	**1.22-12.02**	**0.02**	633
NO_3^- associations with mortality								
Unadjusted	**1.46**	**1.03-2.08**	**0.04**	310	0.91	0.67-1.24	0.56	741
Partially-adjusted	**1.45**	**1.01-2.09**	**0.046**	310	0.93	0.68-1.26	0.64	741
Fully-adjusted	1.25	0.85-1.86	0.26	283	1.47	0.80-2.69	0.22	633
NH_4^+ associations with mortality								
Unadjusted	**1.45**	**1.15-1.81**	**0.001**	310	0.59	0.23-1.50	0.27	741
Partially-adjusted	**1.44**	**1.15-1.80**	**0.002**	310	0.63	0.25-1.61	0.34	741
Fully-adjusted	1.04	0.82-1.31	0.77	283	2.90	0.44-19.11	0.27	633
BC associations with mortality								
Unadjusted	1.47	0.86-2.51	0.16	310	1.00	0.46-2.15	0.995	741
Partially-adjusted	1.26	0.72-2.18	0.42	310	1.00	0.46-2.16	0.99	741
Fully-adjusted	0.76	0.43-1.34	0.34	283	**3.23**	**1.02-10.30**	**0.047**	633
OM associations with mortality								
Unadjusted	1.10	0.96-1.26	0.18	310	1.01	0.88-1.18	0.85	741
Partially-adjusted	1.11	0.96-1.29	0.18	310	1.01	0.87-1.16	0.94	741
Fully-adjusted	0.90	0.77-1.06	0.22	283	**1.26**	**1.008-1.58**	**0.04**	633
SS associations with mortality								
Unadjusted	1.09	0.44-2.67	0.85	310	1.01	0.73-1.40	0.94	741
Partially-adjusted	0.87	0.35-2.15	0.76	310	0.96	0.69-1.33	0.79	741
Fully-adjusted	0.59	0.21-1.68	0.33	283	1.27	0.70-2.29	0.44	633
Soil associations with mortality								
Unadjusted	1.51	0.67-3.41	0.33	310	*1.34*	*0.95-1.89*	*0.095*	741
Partially-adjusted	1.35	0.59-3.10	0.48	310	*1.34*	*0.95-1.90*	*0.099*	741
Fully-adjusted	0.64	0.23-1.73	0.38	283	1.31	0.52-3.32	0.57	633
Model	**3 Month Pre-Sampling Models**							
	HR	**95% CI**	**p**	**n**	**HR**	**95% CI**	**p**	**n**
$PM_{2.5}$ associations with mortality								
Unadjusted	**1.06**	**1.02-1.10**	**0.002**	307	1.05	0.98-1.12	0.18	734
Partially-adjusted	**1.05**	**1.02-1.09**	**0.004**	307	1.05	0.98-1.12	0.17	734
Fully-adjusted	1.01	0.97-1.05	0.64	280	**1.13**	**1.02-1.24**	**0.01**	628
SO_4^{2-} associations with mortality								
Unadjusted	**1.09**	**1.03-1.16**	**0.003**	297	1.17	0.79-1.76	0.43	621
Partially-adjusted	**1.08**	**1.02-1.15**	**0.01**	297	1.12	0.76-1.67	0.56	621
Fully-adjusted	1.01	0.95-1.08	0.71	270	**3.16**	**1.41-7.09**	**0.005**	531
NO_3^- associations with mortality								
Unadjusted	1.01	0.84-1.20	0.94	297	**1.19**	**1.002-1.42**	**0.047**	621
Partially-adjusted	1.01	0.84-1.21	0.94	297	*1.18*	*0.998-1.40*	*0.053*	621
Fully-adjusted	0.91	0.75-1.11	0.36	270	**1.36**	**1.10-1.70**	**0.005**	531
NH_4^+ associations with mortality								

Unadjusted	**1.37**	**1.12-1.68**	**0.003**	297	1.66	0.89-3.09	0.11	621
Partially-adjusted	**1.33**	**1.08-1.63**	**0.007**	297	1.60	0.88-2.91	0.12	621
Fully-adjusted	0.98	0.79-1.23	0.88	270	**2.57**	**1.23-5.36**	**0.01**	531
BC associations with mortality								
Unadjusted	1.11	0.73-1.69	0.61	297	1.45	0.78-2.72	0.24	621
Partially-adjusted	0.98	0.64-1.50	0.92	297	1.44	0.77-2.69	0.25	621
Fully-adjusted	0.72	0.46-1.14	0.17	270	**2.68**	**1.17-6.14**	**0.02**	531
OM associations with mortality								
Unadjusted	1.05	0.92-1.20	0.44	297	*1.11*	*0.996-1.25*	*0.06*	621
Partially-adjusted	1.04	0.91-1.19	0.57	297	*1.11*	*0.99-1.24*	*0.07*	621
Fully-adjusted	*0.87*	*0.75-1.01*	*0.07*	270	**1.23**	**1.05-1.44**	**0.008**	531
SS associations with mortality								
Unadjusted	1.34	0.64-2.79	0.43	297	0.88	0.62-1.26	0.49	621
Partially-adjusted	1.16	0.56-2.41	0.69	297	0.85	0.60-1.22	0.38	621
Fully-adjusted	0.88	0.39-1.97	0.76	270	0.83	0.49-1.41	0.49	531
Soil associations with mortality								
Unadjusted	0.73	0.41-1.31	0.3	297	1.09	0.85-1.41	0.49	621
Partially-adjusted	0.66	0.37-1.20	0.17	297	1.08	0.84-1.38	0.57	621
Fully-adjusted	**0.43**	**0.22-0.84**	**0.01**	270	0.99	0.67-1.45	0.95	531



Appendix Figure 10 – Mediation framework. The purpose of mediation analysis in this study is to determine the proportion of the association between $PM_{2.5}$ and mortality that is mediated through alterations in global DNA methylation. Whatever proportion of the association between $PM_{2.5}$ and mortality remains unaccounted for is mediated through alternative pathways or direct pathways.



Appendix Figure 11 – Scatterplot matrix of percent 5mC (%5mC), $PM_{2.5}$, and constituent components with each other. Correlations are shown on each plot for both cohorts. Points in blue reflects values for patients in the Simmons cohort and red reflects values for patients in the Pulmonary Fibrosis Foundation (PFF) cohort. Density plots are shown on the left-most column to reflect the distribution of 5mC (%5mC), $PM_{2.5}$, and constituent components.

Appendix D – List of Manuscripts During PhD

During the course of my PhD, the following manuscripts have been published, are in-press, have been submitted, or are in progress.

Published:

*Mentees underlined.

1. Ong S, Koo JN, Johannson KA, Ryerson CJ, **Goobie GC**. Twitter content analysis of idiopathic pulmonary fibrosis-related posts. *ATS Scholar* 2022 Nov 16. DOI: 10.34197/ats-scholar.2022-0054OC.
2. **Goobie GC**, Ryerson CJ, Johannson KA, Schikowski E, Khalil N, Marcoux V, Assayag D, Manganas H, Fisher JH, Kolb MRJ, Gibson KF, Kass DJ, Zhang Y, Lindell KO, Nouraie SM. Neighborhood-level disadvantage impacts on lung function in patients with sarcoidosis. *ERJ Open Res* 2022 Oct 24; 8(4): 00357-2022. DOI: 10.1183/23120541.00357-2022. PMID: 36299359.
3. **Goobie GC**, Carlsten C, Johannson KA, Khalil N, Marcoux V, Assayag D, Manganas H, Fisher JH, Kolb MRJ, Lindell KO, Fabisiak JP, Chen X, Gibson KF, Zhang Y, Kass DJ, Ryerson CJ, Nouraie SM. Association of particulate matter exposure with lung function and mortality in patients with fibrotic interstitial lung disease: a multinational cohort study. *JAMA Intern Med* 2022 Oct 17: e224696. DOI: 10.1001/jamainternmed.2022.4696. PMID: 36251286.
4. **Goobie GC**, Kass DJ. Genomic classifiers in diagnosing interstitial lung disease – finding the right place at the right time. *Ann Am Thorac Soc* 2022; 19(6): 895-897. DOI: 10.1513/AnnalsATS.2021112-1353ED. PMID: 35648084.
5. Wilkens FM, Ganter C, Kreigsmann K, Wilkens H, Kahn N, **Goobie GC**, Ryerson CJ, Kreuter M. YouTube videos for patient education in lymphangioleiomyomatosis? *Respiratory Research* 2022; 23(1): 103. DOI: 10.1186/s12931-022-02022-9. PMID: 35477513.
6. **Goobie GC**, Ryerson, CJ, Johannson KA, Schikowski E, Zou RH, Khalil N, Marcoux V, Assayag D, Manganas H, Fisher JH, Kolb M, Gibson KF, Kass DJ, Zhang Y, Lindell KO, Nouraie SM. Neighborhood-level disadvantage impacts on patients with fibrotic interstitial lung disease. *Am J Respir Crit Care Med* 2022; 205(4): 459-467. DOI: 10.1164/rccm.202109-2065OC. PMID: 34818133.
7. Chen TY, Li X, **Goobie GC**, Hung CH, Hamilton K, Bahudhanapati H, Tan J, Kass DJ, Zhang Y. Identification of a distal RXFP1 gene enhancer with differential activity in fibrotic lung fibroblasts involving AP-1. *PLoS One* 2021; 16(12): e0254466. DOI: 10.1371/journal.pone.0254466. PMID: 34972106.

8. Kochan A, Ong S, Guler S, Johannson KA, Ryerson CJ, **Goobie GC**. Social Media Content of Idiopathic Pulmonary Fibrosis Groups and Pages on Facebook: A Cross Sectional Analysis. *JMIR Public Health Surveill* 2021; 7(5): e24199. DOI: 10.2196/24199. PMID: 34057425.
9. **Goobie GC**, Rice MB, Carlsten C. The EPA's "Strengthening Transparency in Pivotal Science" rule: Don't let history repeat itself. 2021. *Annals Am Thor Soc* 2021; 18(10): 1614-1617. DOI: 10.1513/AnnalsATS.202103-259VP. PMID: 33752570.
10. Li X*, **Goobie GC***, Gregory A, Kass DJ, Zhang Y. TOLL-interacting protein in pulmonary diseases: abiding by the Goldilocks principle. *Am J Respir Cell Mol Biol* 2021; 64(5): 536-546. DOI: 10.116/rcmb.2020-0470TR. PMID: 33233920.
    *Denotes co-first authorship.
11. Li X*, **Goobie GC***, Zhang Y. Toll-interacting protein impacts on inflammation, autophagy, and vacuole trafficking in human disease. *J Mol Med* 2021; 99(1): 21-31. DOI: 10.1007/s00109-020-01999-4. PMID: 33128579.
    *Denotes co-first authorship.
12. Fell CD, **Goobie GC**, Ford-Sahibzada CA, Johannson KA. Clinical Characterization of Patients with Interstitial Lung Disease: Report from a Single Canadian Centre. *Can J Respir Crit Care Sleep Med* 2020; DOI: 10.1080/24745332.2020.1811803.
13. **Goobie GC**, Nouraie SM, Zhang Y, Kass DJ, Ryerson CJ, Carlsten C, Johannson KA. Air pollution and interstitial lung diseases: defining epigenomic effects. *Am J Respir Crit Care Med* 2020; 202(9): 1217-1224. doi: 10.1164/rccm.202003-0836PP. PMID: 32569479.
14. **Goobie GC.** Social media impacts on the dissemination of health-related information and patient-physician relationships. *UBC Med J* 2020;11(2):6-8. https://med-fom-ubcmj.sites.olt.ubc.ca/files/2020/04/UBCMJ-Volume-11-Issue-2_feature_2.pdf.

Submitted:

1. **Goobie GC**, Li X, Ryerson CJ, Carlsten C, Johannson KA, Fabisiak JP, Lindell KO, Chen X, Gibson KF, Kass DJ, Nouraie SM, Zhang Y. $PM_{2.5}$ and constituent component impacts on global DNA methylation in patients with idiopathic pulmonary fibrosis. *Submitted to Environmental Pollution, August 26, 2022.*
2. Da Silva, Lokhandwala A, Al Kaabi N, Semenchuk J, **Goobie GC**, Camacho E, Reid WD, Fisher JH, Ryerson CJ, Rozenberg D. Description and reliability of internet resources on pulmonary rehabilitation for individuals with chronic lung disease. *Submitted to Chronic Respiratory Disease, July 6, 2022*.

In-Progress:

1. **Goobie GC**, Ryerson CJ, Johannson K, Carlsten C, Kass DJ, Lindell KO, Fabisiak J, Nouraie MS, Zhang Y. Epigenome-wide DNA methylation responses to $PM_{2.5}$ exposure in patients with fibrotic interstitial lung disease. 2022.
    a. Status: Currently analyzing DNA methylation data from MethylationEPIC Illumina 850K BeadChip array.

2. **Goobie GC**, Saha P, Fabisiak J, Robinson A, Presto A, Nouraie SM. Impacts of ultrafine particulate matter on patients with fibrotic interstitial lung disease. 2022.
    a. Status: Currently in manuscript preparation phase.
3. Yang X, Li X, Hamilton K, Kim S, Feng J, **Goobie GC**, Zhang Y. Toll interacting protein (TOLLIP) contributes to epithelial cell-mediated fibroblast differentiation by increasing TGF-β secretion. 2022.
    a. Status: Currently in manuscript preparation phase.
4. Keil S, **Goobie GC**, Swabi G, Gibson KF, Kass DJ, Magnani J. Demographic and clinical characteristics of steroid-sparing agent prescriptions in U.S. patients with sarcoidosis: an administrative database study. 2022.
    a. Status: Currently in manuscript preparation phase.
5. **Goobie GC**, Nouraie MS, Jessen H, Sand JMB, Leeming DJ, Karsdal MA, Gibson K, Kass D, Herzog E, Zhang Y. Extracellular matrix biomarkers predict lung function and rate of decline in pulmonary sarcoidosis. 2022.
    a. Status: Currently finalizing analysis from contributing site at Yale.
6. **Goobie GC**, Adegunsoye D, Sutton R, Chen X, Gibson KF, Fabisiak J, Kass DJ, Zhang Y, Alder J, Strek M, Nouraie SM. PM2.5 impacts on telomere length in patients with idiopathic pulmonary fibrosis. 2022.
    a. Status: Currently awaiting data from contributing site at University of Chicago prior to writing manuscript.
7. Semenchuk J, Tania Da Silva, **Goobie GC**, Ryerson CJ, Fisher JH, Rozenberg D. YouTube as a source of information on pulmonary rehabilitation for chronic lung diseases. 2022.
    a. Status: Currently undergoing data collection and analysis.
8. Kim S, Li X, Gibson KF, Kass DJ, Nouraie SM, **Goobie GC***, Zhang Y*. MMP-7 mediates PM2.5-mortality associations in patients with idiopathic pulmonary fibrosis. 2022.
    *Denotes co-senior authorship
    a. Status: Analysis completed for initial abstract. Awaiting data from UBC validation site.

Bibliography

Adegunsoye, A., Oldham, J. M., Bellam, S. K., Chung, J. H., Chung, P. A., Biblowitz, K. M., … Strek, M. E. (2018). African-American race and mortality in interstitial lung disease: a multicentre propensity-matched analysis. *Eur Respir J*, *51*(6), pii: 1800255. https://doi.org/10.1183/13993003.00255-2018

Air pollution. (2022). Retrieved March 2, 2022, from https://www.who.int/data/gho/data/themes/theme-details/GHO/air-pollution

Alder, J. K., J-L Chen, J., Lancaster, L., Danoff, S., Su, S., Cogan, J. D., … Armanios, M. Y. (2008). Short telomeres are a risk factor for idiopathic pulmonary fibrosis. *Proc Natl Acad Sci U.S.A.*, *105*(35), 13051–13056.

Alfano, R., Herceg, Z., Nawrot, T. S., Chadeau-Hyam, M., Ghantous, A., & Plusquin, M. (2018). The Impact of Air Pollution on Our Epigenome: How Far Is the Evidence? (A Systematic Review). *Curr Environ Health Rep*, *5*(4), 544–578. https://doi.org/10.1007/s40572-018-0218-8 [Epub ahead of print]

Algamdi, M., Sadatsafavi, M., Fisher, J. H., Morisset, J., Johannson, K. A., Fell, C. D., … Ryerson, C. J. (2019). Costs of Workplace Productivity Loss in Patients With Fibrotic Interstitial Lung Disease. *CHEST*, *156*(5), 887–895. https://doi.org/10.1016/j.chest.2019.04.016

Allen, R. J., Porte, J., Braybrooke, R., Flores, C., Fingerlin, T. E., Oldham, J. M., … Jenkins, R. G. (2017). Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. *Lancet Respir Med*, *5*(11), 869–880. https://doi.org/10.1016/S2213-2600(17)30387-9

Araki, T., Putman, R. K., Hatabu, H., Gao, W., Jos´, J., Dupuis, J., … Obtained, R. K. P. (2016). Development and Progression of Interstitial Lung Abnormalities in the Framingham Heart Study. *Am J Respir Crit Care Med*, *194*(12), 1514–1522. https://doi.org/10.1164/rccm.201512-2523OC

Atmospheric Composition Analysis Group » Surface PM2.5. (2020). Retrieved February 11, 2019, from http://fizz.phys.dal.ca/~atmos/martin/?page_id=140

Auditor General of Canada. (2015, January 20). Audit at a Glance—Report 4—Access to Health Services for Remote First Nations Communities. Retrieved August 16, 2021, from https://www.oag-bvg.gc.ca/internet/English/att__e_40382.html

Bagnato, G., & Harari, S. (2015). Cellular interactions in the pathogenesis of interstitial lung diseases. *Eur Respir Rev*, *24*(135), 102–114. https://doi.org/10.1183/09059180.00003214

Barbosa, A. M. (n.d.). FDR: False Discovery Rate in fuzzySim: Fuzzy Similarity in Species

Distributions. Retrieved September 21, 2022, from https://rdrr.io/cran/fuzzySim/man/FDR.html

Behr, J., Prasse, A., Kreuter, M., Johow, J., Rabe, K. F., Bonella, F., … Investigators, O. behalf of the R. (2021). Pirfenidone in patients with progressive fibrotic interstitial lung diseases other than idiopathic pulmonary fibrosis (RELIEF): a double-blind, randomised, placebo-controlled, phase 2b trial. *Lancet Respir Med*, *Online ahe*, 1–12. https://doi.org/10.1016/S2213-2600(20)30554-3

Bellizzi, D., D'aquila, P., Montesanto, A., Corsonello, A., Mari, V., Mazzei, B., … Passarino, G. (2012). Global DNA methylation in old subjects is correlated with frailty. *Age (Dodr)*, *34*(1), 169–179. https://doi.org/10.1007/s11357-011-9216-6

Benjamini, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J Royal Stat Soc, Series B (Methodological)*, *57*(1), 289–300. https://rss.onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1995.tb02031.x.

Blanc, P. D., Annesi-Maesano, I., Balmes, J. R., Cummings, K. J., Fishwick, D., Miedinger, D., … Redlich, C. A. (2019). The Occupational Burden of Nonmalignant Respiratory Diseases. An Official American Thoracic Society and European Respiratory Society Statement. *Am J Respir Crit Care Med*, *199*(11), 1312–1334. https://doi.org/10.1164/rccm.201904-0717ST

Borie, R., Cardwell, J., Konigsberg, I. R., Moore, C. M., Zhang, W., Sasse, S. K., … Yang, I. V. (2022). Supplement: Colocalization of Gene Expression and DNA Methylation with Genetic Risk Variants Supports Functional Roles of MUC5B and DSP in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*, *Online Sup*, 1–62. https://doi.org/10.1164/rccm.202110-2308OC

Bowe, B., Xie, Y., Yan, Y., & Al-Aly, Z. (2019). Burden of Cause-Specific Mortality Associated With PM2.5 Air Pollution in the United States. *JAMA Netw Open*, *2*(11), e1915834. https://doi.org/10.1001/jamanetworkopen.2019.15834

Boyd, R. W., Lindo, E. G., Weeks, L. D., & Mclemore, M. R. (2020). On Racism : A New Standard For Publishing On Racial Health Inequities. *Health Affairs Blog*, 1–9. https://www.healthaffairs.org/do/10.1377/forefront.20200630.939347/.

Breton, C. V., Yao, J., Millstein, J., Gao, L., Siegmund, K. D., Mack, W., … Gilliland, F. D. (2016). Prenatal air pollution exposures, DNA methyl transferase genotypes, and associations with newborn line1 and Alu methylation and childhood blood pressure and carotid intima-media thickness in the children's health study. *Environ Health Perspect*, *124*(12), 1905–1912. https://doi.org/10.1289/EHP181

Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Arden, C., … Performed, J. V. S. (2018). Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc Natl Acad Sci U.S.A.*, *115*(38), 9592–9597. https://doi.org/10.1073/pnas.1803222115

Cai, J., Zhao, Y., Liu, P., Zhu, Q., Wang, X., Song, Q., … Zhang, Y. (2017). Exposure to

particulate air pollution during early pregnancy is associated with placental DNA methylation. *Sci Total Environ*, *607–608*, 1103–1108. https://doi.org/10.1016/j.scitotenv.2017.07.029

Carlsten, Chris, Brauer, M., Brinkman, F., Brook, J., Daley, D., McNagny, K., … Denburg, J. (2014). Genes, the environment and personalized medicine: We need to harness both environmental and genetic data to maximize personal and population health. *EMBO Reports*, *15*(7), 736–739. https://doi.org/10.15252/embr.201438480

Carlsten, Christopher. (2018). Synergistic Environmental Exposures and the Airways Capturing Complexity in Humans. An Underappreciated World of Complex Exposures. *CHEST*, 918–924. https://doi.org/10.1016/j.chest.2018.06.004

Carlsten, Christopher, Salvi, S., Wong, G. W. K., & Chung, K. F. (2020). Personal strategies to minimise effects of air pollution on respiratory health: advice for providers, patients and the public. *Eur Respir J*, (pii: 1902056). https://doi.org/10.1183/13993003.02056-2019

Carpenter, D. O., Ma, J., & Lessner, L. (2008). Asthma and infectious respiratory disease in relation to residence near hazardous waste sites. *Ann N.Y. Acad Sci*, *1140*, 201–208. https://doi.org/10.1196/annals.1454.000

Chun-on, P., Hinchie, A. M., Beale, H. C., Silva, A. A. . G., Rush, E., Sander, C., … Alder, J. K. (2022). TPP1 promoter mutations cooperate with TERT promoter mutations to lengthen telomeres in melanoma. *Science*, *378*(6620), 664–668. https://doi.org/10.1126/SCIENCE.ABQ0607

Clay, K., & Muller, N. (2019). Recent Increases in Air Pollution: Evidence and Implications for Mortality. *National Bureau of Economic Research Working Paper No. 26381*. https://doi.org/10.3386/w26381

Clifford, R. L., Jones, M. J., MacIsaac, J. L., McEwen, L. M., Goodman, S. J., Mostafavi, S., … Carlsten, C. (2017). Inhalation of diesel exhaust and allergen alters human bronchial epithelium DNA methylation. *J Allergy Clin Immunol*, *139*(1), 112–121. https://doi.org/10.1016/j.jaci.2016.03.046

Cohen, A. J., Brauer, M., Burnett, R., Ross Anderson, H., Frostad, J., Estep, K., … Forouzanfar, M. H. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet*, *389*(10082), 1907–1918.

Congressional Research Service (CRS) Reports. U.S. Health Care Coverage and Spending. (2021, January 26). Retrieved June 23, 2021, from https://fas.org/sgp/crs/misc/IF10830.pdf

Conti, S., Harari, S., Caminati, A., Zanobetti, A., Schwartz, J. D., Bertazzi, P. A., … Madotto, F. (2018). The association between air pollution and the incidence of idiopathic pulmonary fibrosis in Northern Italy. *Eur Respir J*, *51*(1), 1700397. https://doi.org/10.1183/13993003.00397-2017

Cottin, V., Hirani, N. A., Hotchkin, D. L., Nambiar, A. M., Ogura, T., Otaola, M., … Wells, A. U. (2018). Presentation, diagnosis and clinical course of the spectrum of progressive-fibrosing interstitial lung diseases. *Eur Respir Rev*, *27*(150), 180076. https://doi.org/10.1183/16000617.0076-2018

Cottin, V., Wollin, L., Fischer, A., Quaresma, M., Stowasser, S., & Harari, S. (2019). Fibrosing interstitial lung diseases: knowns and unknowns. *Eur Respir Rev*, *28*, 180100. https://doi.org/10.1183/16000617.0100-2018

Coull, B. A., Bobb, J. F., Wellenius, G. A., Kioumourtzoglou, M.-A., Mittleman, M. A., Koutrakis, P., … Spiegelman, C. (2015). *Development of Statistical Methods for Multipollutant Research. Research Report 183, Parts 1 & 2. Boston, MA: Health Effects Institute*.

Courtwright, A. M., & El-Chemaly, S. (2019). Telomeres in Interstitial Lung Disease: The Short and the Long of It. *Ann Am Thorac Soc*, *16*(2), 175–181. https://doi.org/10.1513/AnnalsATS.201808-508CME

Cromar, K. R., Gladson, L. A., Hicks, E. A., Marsh, B., & Ewart, G. (2021). Excess Morbidity and Mortality Associated with Air Pollution above American Thoracic Society Recommended Standards, 2017-2019. *Annals Am Thorac Soc*, *Epub ahead*, 1–57. https://doi.org/10.1513/AnnalsATS.202107-860OC

Cronkhite, J. T., Xing, C., Raghu, G., Chin, K. M., Torres, F., Rosenblatt, R. L., & Garcia, C. K. (2008). Telomere Shortening in Familial and Sporadic Pulmonary Fibrosis. *Am J Respir Crit Care Med*, *178*(7), 729–737.

CRTAP Gene - GeneCards. (2022, August 30). Retrieved September 21, 2022, from https://www.genecards.org/cgi-bin/carddisp.pl?gene=CRTAP

Dales, R., Blanco-Vidal, C., & Cakmak, S. (2020). The Association Between Air Pollution and Hospitalization of Patients With Idiopathic Pulmonary Fibrosis in Chile A Daily Time Series Analysis. *CHEST*, *158*(2), 630–636. https://doi.org/10.1016/j.chest.2020.02.017

de F.C. Lichtenfels, A. J., Van Der Plaat, D. A., De Jong, K., Van Diemen, C. C., Postma, D. S., Nedeljkovic, I., … Vonk, J. M. (2018). Long-term Air Pollution Exposure, Genome-wide DNA Methylation and Lung Function in the LifeLines Cohort Study. *Environ Health Perspect*, *126*(2), 027004. https://doi.org/10.1289/EHP2045

Dickman, S. L., Himmelstein, D. U., & Woolhandler, S. (2017). America: Equity and Equality in Health 1. Inequality and the health-care system in the USA. *Lancet*, *389*(10077), 1431–1441. Retrieved from www.thelancet.com/

Ding, R., Jin, Y., Liu, X., Ye, H., Zhu, Z., Zhang, Y., … Xu, Y. (2017). Dose-and time-effect responses of DNA methylation and histone H3K9 acetylation changes induced by traffic-related air pollution. *Sci Rep*, *7*, 43737. https://doi.org/10.1038/srep43737

Ding, Z., Mangino, M., Aviv, A., Spector, T., & Durbin, R. (2014). Estimating telomere length from whole genome sequence data. *Nucleic Acids Res*, *42*(9).

https://doi.org/10.1093/NAR/GKU181

Duchemann, B., Annesi-Maesano, I., Jacobe de Naurois, C., Sanyal, S., Brillet, P.-Y., Brauner, M., … Nunes, H. (2017). Prevalence and incidence of interstitial lung diseases in a multi-ethnic county of Greater Paris. *Eur Respir J*, *50*(2), pii: 1602419.

Dutzik, T., Group, F., Barber, Z., Research, P., & Center, P. (2019). *Cutting through the Smoke: Why the Allegheny County Health Department Must Turn the Corner on Decades of Weak Clean Air Enforcement*. Retrieved from www.frontiergroup.org.

Ejike, C. O., Woo, H., Galiatsatos, P., Paulin, L. M., Krishnan, J. A., Cooper, C. B., … co-authors Affiliations, F. (2020). Contribution of Individual and Neighborhood Factors to Racial Disparities in Respiratory Outcomes. *Am J Respir Crit Care Med, Epub ahead of print*, 1–42. https://doi.org/10.1164/rccm.202002-0253OC

Fernandez, I. E., & Eickelberg, O. (2012). The impact of TGF-β on lung fibrosis: From targeting to biomarkers. *Proc Am Thorac Soc*, *9*(3), 111–116. https://doi.org/10.1513/PATS.201203-023AW

Fisher, J. H., Shapera, S., To, T., Marras, T. K., Gershon, A., & Dell, S. (2019). Procedure volume and mortality after surgical lung biopsy in interstitial lung disease. *Eur Respir J*, *53*(2), 1801164. https://doi.org/10.1183/13993003.01164-2018

Flaherty, K., Wells, A., Cottin, V., Devaraj, A., Walsh, S., Inoue, Y., … INBULD Trial Investigators. (2019). Nintedanib in Progressive Fibrosing Interstitial Lung Diseases. *N Engl J Med*. https://doi.org/doi: 10.1056/NEJMoa1908681

Foltescu, V. L., Selin Lindgren, E., Isakson, J., Oblad, M., Pacyna, J. M., & Benson, S. (1996). Gas-to-particle conversion of sulphur and nitrogen compounds as studied at marine stations in Northern Europe. *Atmos Environ*, *30*(18), 3129.

Forbes, M. S., Raison, R. J., & Skjemstad, J. O. (2006). Formation, transformation and transport of black carbon (charcoal) in terrestrial and aquatic ecosystems. *Sci Tot Environ*, *370*, 190–206. https://doi.org/10.1016/j.scitotenv.2006.06.007

Ford-Sahibzada, C. A., Johannson, K. A., Goobie, G. C., & Fell, C. D. (2017). Clinical characterization of interstitial lung disease patients: report from a single center longitudinal database [abstract]. *Am J Respir Crit Care Med*, *195*, A3473.

Fortin, J.-P., Labbe, A., Lemire, M., Zanke, B. W., Hudson, T. J., Fertig, E. J., … Hansen, K. D. (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biology*, *15*, 503. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4283580/.

Gaffney, A. W., Himmelstein, D. U., Christiani, D. C., & Woolhandler, S. (2021). Socioeconomic Inequality in Respiratory Health in the US from 1959 to 2018. *JAMA Intern Med*, *181*(7), 968–976. https://doi.org/10.1001/jamainternmed.2021.2441

Gaffney, A. W., & Podolanczuk, A. J. (2022). Inequity and the Interstitium: Pushing Back on Disparities in Fibrosing Lung Disease in the United States and Canada. *Am J Respir Crit Care Med*, *205*(4), 385–387. https://doi.org/10.1164/rccm.202111-2652ED

Gaffney, A. W., Woolhander, S., Himmelstein, D., & Mccormick, D. (2018). Disparities in pulmonary fibrosis care in the United States: an analysis from the Nationwide Inpatient Sample. *BMC Health Serv Res*, *18*(1), 618. https://doi.org/10.1186/s12913-018-3407-0

Galiatsatos, P., Woo, H., Paulin, L. M., Kind, A., Putcha, N., Gassett, A. J., … Hansel, N. N. (2020). The Association Between Neighborhood Socioeconomic Disadvantage and Chronic Obstructive Pulmonary Disease. *Int J Chron Obstruct Pulmon Dis*, *15*, 981–993. https://doi.org/10.2147/COPD.S238933

GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. (2018a). Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*, *392*(10159), 1736–1788. https://doi.org/10.1016/S0140-6736(18)32203-7

GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. (2018b). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*, *392*(10159), 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7

George, P. M., Spagnolo, P., Kreuter, M., Altinisik, G., Bonifazi, M., Martinez, F. J., … Costabel, U. (2020). *Progressive fibrosing interstitial lung disease: clinical uncertainties, consensus recommendations, and research priorities*. *Lancet Respir Med* (Vol. 8). https://doi.org/10.1016/S2213-2600(20)30355-6

Gershon, A. S., Dolmage, T. E., Stephenson, A., & Jackson, B. (2012). Chronic Obstructive Pulmonary Disease and Socioeconomic Status: a Systematic Review. *COPD*, *9*(3), 216–226. https://doi.org/10.3109/15412555.2011.648030

Goobie, G. C., Carlsten, C., Johannson, K. A., Khalil, N., Marcoux, V., Assayag, D., … Nouraie, M. (2022). Association of Particulate Matter Exposure with Lung Function and Mortality in Fibrotic Interstitial Lung Disease: A Multinational Cohort Study. *JAMA Intern Med*, *Oct 17*. https://doi.org/10.1001/jamainternmed.2022.4696

Goobie, G. C., Carlsten, C., Johannson, K. A., Marcoux, V., Assayag, D., Manganas, H., … Nouraie, S. M. (2022). PREPRINT: Association of Particulate Matter Exposure with Lung Function and Mortality in Fibrotic Interstitial Lung Disease: A Multinational Cohort Study. *SSRN Preprint Server*, *April 27*, 1–40. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4094833

Goobie, G. C., Kass, D. J., Zhang, Y., Gibson, K. F., Zou, R. H., Nouraie, S. M., & Lindell, K. O. (2021). A7 - A007: Neighborhood-Level Disadvantage Impacts Mortality and Lung Transplantation in Patients with Idiopathic Pulmonary Fibrosis [ABSTRACT]. *Am J Respir Crit Care Med*, *203*, 1–2.

Goobie, G. C., Nouraie, M., Zhang, Y., Kass, D. J., Ryerson, C. J., Carlsten, C., & Johannson, K. A. (2020). Air Pollution and Interstitial Lung Diseases: Defining Epigenomic Effects. *Am J Respir Crit Care Med*, *202*(9), 1217–1224. https://doi.org/10.1164/rccm.202003-0836PP

Goobie, G. C., Ryerson, C. J., Johannson, K. A., Schikowski, E., Zou, R. H., Khalil, N., … Mehdi Nouraie, S. (2022). Neighborhood-Level Disadvantage Impacts on Patients with Fibrotic Interstitial Lung Disease. *Am J Respir Crit Care Med*, *205*(4), 459–467. https://doi.org/10.1164/rccm.202109-2065OC

Government of Canada. Understand how health care works in Canada. (2021, February 11). Retrieved June 23, 2021, from https://www.canada.ca/en/immigration-refugees-citizenship/services/new-immigrants/new-life-canada/health-care-card.html

Grafe, I., Yang, T., Alexander, S., Homan, E. P., Lietman, C., Jiang, M. M., … Lee, B. (2014). Excessive transforming growth factor-β 2 signaling is a common mechanism in osteogenesis imperfecta. *Nat Med*, *20*(6), 670–675. https://doi.org/10.1038/NM.3544

Gref, A., Merid, S. K., Gruzieva, O., Ballereau, S., Becker, A., Bellander, T., … Melén, E. (2017). Genome-wide interaction analysis of air pollution exposure and childhood asthma with functional follow-up. *Am J Respir Crit Care Med*, *195*(10), 1373–1383. https://doi.org/10.1164/rccm.201605-1026OC

Guiot, J., Struman, I., Chavez, V., Henket, M., Herzog, M., Scoubeau, K., … Louis, R. (2017). Altered epigenetic features in circulating nucleosomes in idiopathic pulmonary fibrosis. *Clin Epigenet*, *9*, 84. https://doi.org/10.1186/s13148-017-0383-x

Gwynn, R. C., Burnett, R. T., & Thurston, G. D. (2000). A time-series analysis of acidic particulate matter and daily mortality and morbidity in the Buffalo, New York, region. *Environ Health Perspect*, *108*(2), 125–133. https://doi.org/10.1289/EHP.00108125

Hajat, A., Hsia, C., & O'Neill, M. S. (2015). Socioeconomic Disparities and Air Pollution Exposure: a Global Review. *Curr Environ Health Rep*, *2*(4), 440–450. https://doi.org/10.1007/s40572-015-0069-5

Hammer, M. S., Van Donkelaar, A., Li, C., Lyapustin, A., Sayer, A. M., Hsu, N. C., … Martin, R. V. (2020). Global Estimates and Long-Term Trends of Fine Particulate Matter Concentrations (1998−2018). *Environ Sci Technol*, *54*(13), 7879–7890. https://doi.org/10.1021/acs.est.0c01764

Hansen, K. D., Aryee, M., Irizarry, R. A., Jaffe, A. E., Maksimovic, J., Houseman, E. A., … Hickey, P. F. (2022). Package "minfi." *R*, 1–75. Retrieved from https://bioconductor.org/packages/release/bioc/manuals/minfi/man/minfi.pdf

Hansen, S., Strøm, M., Olsen, S. F., Dahl, R., Hoffmann, H. J., Granström, C., … Halldorsson, T. I. (2016). Prenatal exposure to persistent organic pollutants and offspring allergic sensitization and lung function at 20 years of age. *Clin Exp Allergy*, *46*(2), 329–336. https://doi.org/10.1111/cea.12631

Hee Cho, K., Mo Nam, C., Jung Lee, E., Choi, Y., Yoo, K.-B., Lee, S.-H., & Park, E.-C. (2016). Effects of individual and neighborhood socioeconomic status on the risk of all-cause mortality in chronic obstructive pulmonary disease: A nationwide population-based cohort study, 2002-2013. *Respir Med*, *114*, 9–17. https://doi.org/10.1016/j.rmed.2016.03.003

Hewitt, C. N. (2002). The atmospheric chemistry of sulphur and nitrogen in power station plumes. *Atmospheric Environ*, *35*, 1155–1170. https://doi.org/10.1016/S1474-8177(02)80012-2

Honkova, K., Rossnerova, A., Chvojkova, I., Milcova, A., Margaryan, H., Pastorkova, A., … Topinka, J. (2022). Genome-Wide DNA Methylation in Policemen Working in Cities Differing by Major Sources of Air Pollution. *Int J Mol Sci*, *23*(3), 1666. https://doi.org/10.3390/ijms23031666

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., … Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, *13*, 86. Retrieved from http://www.biomedcentral.com/1471-2105/13/86

Huang, S. K., Scruggs, A. M., Donaghy, J., Horowitz, J. C., Zaslona, Z., Przybranowski, S., … Peters-Golden, M. (2013). Histone modifications are responsible for decreased Fas expression and apoptosis resistance in fibrotic lung fibroblasts. *Cell Death Dis*, *4*(5), e621. https://doi.org/10.1038/cddis.2013.146

Infinium MethylationEPIC Data Sheet | Illumina. (2019). Retrieved May 3, 2021, from https://filgen.jp/Product/Bioscience/Methyl/humanmethylationepic-data-sheet-1070-2015-008.pdf.

Janjua, S., Powell, P., Atkinson, R., Stovold, E., & Fortescue, R. (2019). Individual-level interventions to reduce personal exposure to outdoor air pollution and their effects on long-term respiratory conditions (Protocol). *Cochrane Database Syst Rev*, (10), CD013441. https://doi.org/10.1002/14651858.CD013441

Jiang, R., Jones, M. J., Sava, F., Kobor, M. S., & Carlsten, C. (2014). Short-term diesel exhaust inhalation in a controlled human crossover study is associated with changes in DNA methylation of circulating mononuclear cells in asthmatics. *Particle Fibre Toxicol*, *11*, 71. https://doi.org/10.1186/s12989-014-0071-3

Johannson, K. A., Vittinghoff, E., Lee, K., Balmes, J. R., Ji, W., Kaplan, G. G., … Kim, D. S. (2014). Acute exacerbation of idiopathic pulmonary fibrosis associated with air pollution exposure. *Eur Respir J*, *43*(4), 1124–1131.

Johannson, K. A., Vittinghoff, E., Morisset, J., Wolters, P. J., Noth, E. M., Balmes, J. R., & Collard, H. R. (2018a). Air Pollution Exposure Is Associated With Lower Lung Function, but Not Changes in Lung Function, in Patients With Idiopathic Pulmonary Fibrosis. *CHEST*, *154*(1), 119–125. https://doi.org/10.1016/j.chest.2018.01.015

Johannson, K. A., Vittinghoff, E., Morisset, J., Wolters, P. J., Noth, E. M., Balmes, J. R., & Collard, H. R. (2018b). Air Pollution Exposure Is Associated With Lower Lung Function, but

Not Changes in Lung Function, in Patients With Idiopathic Pulmonary Fibrosis. *CHEST*, *154*(1), 119–125.

Jones, M. G., & Richeldi, L. (2014). Air pollution and acute exacerbations of idiopathic pulmonary fibrosis: back to miasma? *Eur Respir J*, *43*(4), 956–959. https://doi.org/10.1183/09031936.00204213

Jones, M. J., Goodman, S. J., & Kobor, M. S. (2015). DNA methylation and healthy human aging. *Aging Cell*, *14*(6), 924–932. https://doi.org/10.1111/ACEL.12349

Kawachi, I., Daniels, N., & Robinson, D. E. (2005). Health Disparities By Race And Class: Why Both Matter. *Health Aff (Millwood)*, *24*(2), 343–352. https://doi.org/10.1377/hlthaff.24.2.343

Kazemiparkouhi, F., Honda, T., Eum, K. Do, Wang, B., Manjourides, J., & Suh, H. H. (2022). The impact of Long-Term PM2.5 constituents and their sources on specific causes of death in a US Medicare cohort. *Environ Int*, *159*, 106988. https://doi.org/10.1016/J.ENVINT.2021.106988

Kind, A. J. H., & Buckingham, W. R. (2018). Making Neighborhood-Disadvantage Metrics Accessible - The Neighborhood Atlas. *N Engl J Med*, *378*, 2456–2458. https://doi.org/10.1056/NEJMp1802313

Kind, A. J. H., Jencks, S., Brock, J., Yu, M., Bartels, C., Ehlenbach, W., … Smith, M. (2014). Neighborhood socioeconomic disadvantage and 30-day rehospitalization: A retrospective cohort study. *Ann Intern Med*, *161*(11), 765–774. https://doi.org/10.7326/M13-2946

King, T. E., Bradford, W. Z., Castro-Bernardini, S., Fagan, E. A., Glaspole, I., Glassberg, M. K., … Wzb, B. (2014). A Phase 3 Trial of Pirfenidone in Patients with Idiopathic Pulmonary Fibrosis. *N Engl J Med*, *370*(22), 2083–2092. https://doi.org/10.1056/NEJMoa1402582

Koduri, G., Norton, S., Young, A., Cox, N., Davies, P., Devlin, J., … Williams, P. (2010). Interstitial lung disease has a poor prognosis in rheumatoid arthritis: results from an inception cohort. *Rheumatology (Oxford)*, *49*(8), 1483–1489.

Kolahdooz, F., Nader, F., Yi, K. J., & Sharma, S. (2015). Understanding the social determinants of health among Indigenous Canadians: priorities for health promotion policies and actions. *Glob Health Action*, *8*, 27968. https://doi.org/10.3402/gha.v8.27968

Kotsianidis, I., Nakou, E., Bouchliou, I., Tzouvelekis, A., Spanoudakis, E., Steiropoulos, P., … Bouros, D. (2009). Global Impairment of CD4 1 CD25 1 FOXP3 1 Regulatory T Cells in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*, *179*(12), 1121–1130. https://doi.org/10.1164/rccm.200812-1936OC

Kurdyukov, S., & Bullock, M. (2016). DNA methylation analysis: Choosing the right method. *Biology*, *5*(1), 1–21. https://doi.org/10.3390/biology5010003

Leclercq, B., Platel, A., Antherieu, S., Alleman, L. Y., Hardy, E. M., Perdrix, E., … Garçon, G. (2017). Genetic and epigenetic alterations in normal and sensitive COPD-diseased human

bronchial epithelial cells repeatedly exposed to air pollution-derived PM 2.5. *Environ Pollut, 230*, 163–177. https://doi.org/10.1016/j.envpol.2017.06.028

Lederer, D. J., Arcasoy, S. M., Barr, R. G., Wilt, J. S., Bagiella, E., D'Ovidio, F., … Kawut, S. M. (2006). Racial and ethnic disparities in idiopathic pulmonary fibrosis: A UNOS/OPTN database analysis. *Am J Transplantation*, *6*(10), 2436–2442. https://doi.org/10.1111/j.1600-6143.2006.01480.x

Levine, B. (2007). What does the population attributable fraction mean? *Preventing Chronic Disease*, *4*(1), 1–5.

Ley, B., Collard, H. R., & King, T. E. (2011). Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, *183*(4), 431-440. doi: 10.1164/rccm.201006-0894CI.

Ley, B., Newton, C. A., Arnould, I., Elicker, B. M., Henry, T. S., Vittinghoff, E., … Wolters, P. J. (2017). The MUC5B promoter polymorphism and telomere length in patients with chronic hypersensitivity pneumonitis. *Lancet Respir Med*, *5*(8), 639–647.

Ley, B., Ryerson, C. J., Vittinghoff, E., Ryu, J. H., Tomassetti, S., Lee, J. S., … Collard, H. R. (2012). A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med*, *156*(10), 684–691.

Li, H., Chen, R., Cai, J., Cui, X., Huang, N., & Kan, H. (2018). Short-term exposure to fine particulate air pollution and genome-wide DNA methylation: A randomized, double-blind, crossover trial. *Environ Int*, *120*, 130–136. https://doi.org/10.1016/j.envint.2018.07.041

Li, Z., Li, N., Guo, C., Li, X., Qian, Y., Wu, J., … Wei, Y. (2019). Genomic DNA methylation signatures in different tissues after ambient air particulate matter exposure. *Ecotoxicol Environ Saf*, *179*, 175–181. https://doi.org/10.1016/j.ecoenv.2019.04.049

Liu, C., Xu, J., Chen, Y., Guo, X., Zheng, Y., Wang, Q., … Hou, L. (2015). Characterization of genome-wide H3K27ac profiles reveals a distinct PM 2.5-associated histone modification signature. *Environ Health*, *14*, 65. https://doi.org/10.1186/s12940-015-0052-5

Ma, X., Zhu, L., Kurche, J. S., Xiao, H., Dai, H., & Wang, C. (2022). Global and regional burden of interstitial lung disease and pulmonary sarcoidosis from 1990 to 2019: results from the Global Burden of Disease study 2019. *Thorax*, *77*, 596–605. https://doi.org/10.1136/thoraxjnl-2020-216732

Maksimovic, J., Oshlack, A., & Phipson, B. (2021). Gene set enrichment analysis for genome-wide DNA methylation data. *Genom Biol*, *22*, 173. https://doi.org/10.1186/s13059-021-02388-x

Maksimovic, J., Phipson, B., & Oshlack, A. (2022, April 29). A cross-package Bioconductor workflow for analysing methylation array data. Retrieved September 1, 2022, from https://www.bioconductor.org/packages/release/workflows/vignettes/methylationArrayAnalysis/inst/doc/methylationArrayAnalysis.html#ref-Touleimat2012

Margaritopoulos, G. A., Vasarmidi, E., Jacob, J., Wells, A. U., & Antoniou, K. M. (2015). Smoking and interstitial lung diseases. *Eur Respir Rev*, *24*, 428–435. https://doi.org/10.1183/16000617.0050-2015

Martin, E. M., & Fry, R. C. (2018). Environmental Influences on the Epigenome: Exposure-Associated DNA Methylation in Human Populations. *Annu Rev Public Health*, *39*, 309–333.

McCartney, D. L., Walker, R. M., Morris, S. W., McIntosh, A. M., Porteous, D. J., & Evans, K. L. (2016). Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genom Data*, *9*, 24. https://doi.org/10.1016/J.GDATA.2016.05.012

MethylFlash Global DNA Methylation (5-mC) ELISA Easy Kit (Colorimetric) | EpiGentek. (n.d.). Retrieved June 15, 2022, from https://www.epigentek.com/catalog/methylflash-global-dna-methylation-mc-elisa-easy-kit-colorimetric-p-5370.html

Miri, M., Nazarzadeh, M., Alahabadi, A., Ehrampoush, M. H., Rad, A., Lotfi, M. H., … Dadvand, P. (2019). Air pollution and telomere length in adults: A systematic review and meta-analysis of observational studies. *Environ Pollut*, *244*, 636–647. https://doi.org/10.1016/j.envpol.2018.09.130

Mishra, P. K., Bunkar, N., Singh, R. D., Kumar, R., Gupta, P. K., Tiwari, R., … Chaudhury, K. (2021). Comparative profiling of epigenetic modifications among individuals living in different high and low air pollution zones: A pilot study from India. *Environ Adv*, *4*(July), 100052. https://doi.org/10.1016/j.envadv.2021.100052

Montgomery, M. T., Sajuthi, S. P., Cho, S., Everman, J. L., Rios, C. L., Goldfarbmuren, K. C., … Seibold, M. A. (2020). Genome-wide Analysis Reveals Mucociliary Remodeling of the Nasal Airway Epithelium Induced by Urban PM2.5. *Am J Respir Cell Mol Biol*, 1–62. https://doi.org/10.1165/rcmb.2019-0454OC. [Epub ahead of print]

Mooney, J. J., Raimundo, K., Chang, E., & Broder, M. S. (2017). Hospital cost and length of stay in idiopathic pulmonary fibrosis. *J Med Econ*, *20*(5), 518–524.

Nadeau, K., Mcdonald-Hyman, C., Noth, E. M., Pratt, B., Hammond, S. K., Balmes, J., & Tager, I. (2010). Ambient air pollution impairs regulatory T-cell function in asthma. *J Allergy Clin Immunol*, *126*(4), 845–852. https://doi.org/10.1016/j.jaci.2010.08.008

Newton, C. A., Batra, K., Torrealba, J., Kozlitina, J., Glazer, C. S., Aravena, C., … Garcia, C. K. (2016). Telomere-related lung fibrosis is diagnostically heterogeneous but uniformly progressive. *Eur Respir J*, *48*(6), 1710–1720.

Noth, I., Zhang, Y., Ma, S.-F., Flores, C., Barber, M., Huang, Y., … Seiwert, S. D. (2013). Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir Med*, *1*(4), 309–317.

Oates, G. R., Harris, W. T., Rowe, S. M., Solomon, G. M., Dey, S., Zhu, A., … Gutierrez, H. H. (2019). Area Deprivation as a Risk Factor for Methicillin-resistant Staphylococcus aureus

Infection in Pediatric Cystic Fibrosis. *Pediatr Infect Dis J*, *38*(11), e285–e289. https://doi.org/10.1097/INF.0000000000002419

OECD. (2020, July 1). OECD Health Statistics 2020. Health Expenditure and Financing. Retrieved June 23, 2021, from https://www.oecd.org/els/health-systems/health-data.htm

Oldham, J. M., Witt, L. J., Adegunsoye, A., Chung, J. H., Lee, C., Hsu, S., … Noth, I. (2018). N-acetylcysteine exposure is associated with improved survival in anti-nuclear antibody seropositive patients with usual interstitial pneumonia. *BMC Pulm Med*, *18*(1), 30. https://doi.org/10.1186/s12890-018-0599-3

Park, M., Soo Joo, H., Lee, K., Jang, M., Don Kim, S., Kim, I., … Park, K. (2018). Differential toxicities of fine particulate matters from various sources. *Sci Rep*, *8*(1), 17007. https://doi.org/10.1038/s41598-018-35398-0

Peljto, A. L., Zhang, Y., Fingerlin, T. E., Ma, S.-F., N Garcia, J. G., Richards, T. J., … Schwartz, D. A. (2013). Association Between the MUC5B Promoter Polymorphism and Survival in Patients With Idiopathic Pulmonary Fibrosis. *JAMA*, *309*(21), 2232–2239.

Peng, C., Cardenas, A., Rifas-Shiman, S. L., Hivert, M. F., Gold, D. R., Platts-Mills, T. A., … DeMeo, D. L. (2019). Epigenetic age acceleration is associated with allergy and asthma in children in Project Viva. *Journal of Allergy and Clinical Immunology*, *143*(6), 2263-2270.e14. https://doi.org/10.1016/j.jaci.2019.01.034

Peters, T. J., Buckley, M. J., Statham, A. L., Pidsley, R., Samaras, K., V Lord, R., … Molloy, P. L. (2015). De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin*, *8*(1), 1–16. https://epigeneticsandchromatin.biomedcentral.com/articles/10.1186/1756-8935-8-6.

Philip, S., Martin, R. V., Van Donkelaar, A., Lo, J. W. H., Wang, Y., Chen, D., … Macdonald, D. J. (2014). Global chemical composition of ambient fine particulate matter for exposure assessment. *Environ Sci Technol*, *48*(22), 13060–13068. https://doi.org/10.1021/es502965b

Pierce, D. (2021, December 15). ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data File. Retrieved June 16, 2022, from https://cran.r-project.org/web/packages/ncdf4/index.html

Plautz, J. (2018, September 13). Ammonia, a poorly understood smog ingredient, could be key to limiting deadly pollution. https://doi.org/10.1126/SCIENCE.AAV3862

Plusquin, M., Guida, F., Polidoro, S., Vermeulen, R., Raaschou-Nielsen, O., Campanella, G., … Chadeau-Hyam, M. (2017). DNA methylation and exposure to ambient air pollution in two prospective cohorts. *Environ Int*, *108*, 127–136. https://doi.org/10.1016/j.envint.2017.08.006

Podolanczuk, A. J., Oelsner, E. C., Graham Barr, R., Hoffman, E. A., Armstrong, H. F., Austin, J. H. M., … Al, U. (2016). High attenuation areas on chest computed tomography in community-dwelling adults: the MESA study. *Eur Respir J*, *48*(5), 1442–1452. https://doi.org/10.1183/13993003.00129-2016

Prunicki, M., Stell, L., Dinakarpandian, D., De Planell-Saguer, M., Lucas, R. W., Hammond, S. K., … Nadeau, K. C. (2018). Exposure to NO2, CO, and PM2.5 is linked to regional DNA methylation differences in asthma. *Clin Epigenet*, *10*, 2.

Rabinovich, E. I., Kapetanaki, M. G., Steinfeld, I., Gibson, K. F., Pandit, K. V, Yu, G., … Simmons, R. P. (2012). Global Methylation Patterns in Idiopathic Pulmonary Fibrosis. *PLoS One*, *7*(4), e33770. https://doi.org/10.1371/journal.pone.0033770

Raghu, G., Collard, H. R., Egan, J. J., Martinez, F. J., Behr, J., Brown, K. K., … Schünemann, H. J. (2011). An Official ATS/ERS/JRS/ALAT Statement: Idiopathic pulmonary fibrosis: Evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med*, *183*(6), 788–824.

Raghu, G., Remy-Jardin, M., Myers, J. L., Richeldi, L., Ryerson, C. J., Lederer, D. J., … Wilson, K. C. (2018). Diagnosis of Idiopathic Pulmonary Fibrosis An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Resp Crit Care Med*, *198*(5), e44-e68.

Raghu, G., Remy-Jardin, M., Ryerson, C. J., Myers, J. L., Kreuter, M., Vasakova, M., … Wilson, K. C. (2020). Diagnosis of Hypersensitivity Pneumonitis in Adults An Official ATS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med*, *202*(3), 36–69. https://doi.org/10.1164/rccm.202005-2032ST

Raghu, G., Rochwerg, B., Zhang, Y., Garcia, C. A. C., Azuma, A., Behr, J., … Schünemann, H. J. (2015). An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. *American Journal of Respiratory and Critical Care Medicine*, *192*(2), e3–e19. https://doi.org/10.1164/rccm.201506-1063ST

Raimundo, K., Chang, E., Broder, M. S., Alexander, K., Zazzali, J., & Swigris, J. J. (2016). Clinical and economic burden of idiopathic pulmonary fibrosis: a retrospective cohort study. *BMC Pulm Med*, *16*, 2. https://doi.org/10.1186/s12890-015-0165-1

Reiss, R., Anderson, E. L., Cross, C. E., Hidy, G., Hoel, D., McClellan, R., & Moolgavkar, S. (2007). Evidence of health impacts of sulfate-and nitrate-containing particles in ambient air. *Inhal Toxicol*, *19*(5), 419–449. https://doi.org/10.1080/08958370601174941

Rice, M. B., Li, W., Schwartz, J., Di, Q., Kloog, I., Koutrakis, P., … Mittleman, M. A. (2019). Ambient air pollution exposure and risk and progression of interstitial lung abnormalities: the Framingham Heart Study Environmental exposure. *Thorax*, *74*(11), 1063–1069. https://doi.org/10.1136/thoraxjnl-2018-212877

Richeldi, L., Du Bois, R. M., Raghu, G., Azuma, A., Brown, K. K., Costabel, U., … Collard, H. R. (2014). Efficacy and Safety of Nintedanib in Idiopathic Pulmonary Fibrosis. *N Engl J Med*, *370*(22), 2071–2082. https://doi.org/10.1056/NEJMoa1402584

Rider, C. F., & Carlsten, C. (2019). Air pollution and DNA methylation: effects of exposure in humans. *Clin Epigenet*, *11*, 131. https://doi.org/10.1186/s13148-019-0713-2

Rider, C. F., Yamamoto, M., Gunther, O. P., Hirota, J. A., Singh, A., Tebbutt, S. J., & Carlsten, C. (2016). Controlled diesel exhaust and allergen coexposure modulates microRNA and gene expression in humans: Effects on inflammatory lung markers. *J Allergy Clin Immunol*, *138*(6), 1690–1700. https://doi.org/10.1016/j.jaci.2016.02.038

Rozzi, G. (2021, September 22). zipcodeR: Data & Functions for Working with US ZIP Codes. Retrieved June 16, 2022, from https://cran.r-project.org/web/packages/zipcodeR/index.html

Ryerson, C. J., Tan, B., Fell, C. D., Manganas, H., Shapera, S., Mittoo, S., … Kolb, M. (2016). The Canadian Registry for Pulmonary Fibrosis: Design and Rationale of a National Pulmonary Fibrosis Registry. *Can Respir J*, *2016*, 3562923. https://doi.org/10.1155/2016/3562923

Ryerson, C. J., Vittinghoff, E., Ley, B., Lee, J. S., Mooney, J. J., Jones, K. D., … Collard, H. R. (2014). Predicting survival across chronic interstitial lung disease: The ILD-GAP model. *CHEST*, *145*(4), 723–728. https://doi.org/10.1378/chest.13-1474

Sack, C., Vedal, S., Sheppard, L., Raghu, G., Graham Barr, R., Podolanczuk, A., … Kaufman, J. D. (2017). Air pollution and subclinical interstitial lung disease: the Multi-Ethnic Study of Atherosclerosis (MESA) air-lung study. *Eur Respir J*, *50*(6), 1700559. https://doi.org/10.1183/13993003.00559-2017

Sahni, S., Talwar, A., Khanijo, S., & Talwar, A. (2017). Socioeconomic status and its relationship to chronic respiratory disease. *Adv Respir Med*, *85*(2), 97–108. https://doi.org/10.5603/ARM.2017.0016

Sanders, Y. Y., Ambalavanan, N., Halloran, B., Zhang, X., Liu, H., Crossman, D. K., … Hagood, J. S. (2012). Altered DNA Methylation Profile in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*, *186*(6), 525–535. https://doi.org/10.1164/rccm.201201-0077OC

Sanders, Y. Y., Hagood, J. S., Liu, H., Zhang, W., Ambalavanan, N., Thannickal, V. J., & Sanders, Y. Y. (2014). Histone deacetylase inhibition promotes fibroblast apoptosis and ameliorates pulmonary fibrosis in mice. *Eur Respir J*, *43*(5), 1448–1458. https://doi.org/10.1183/09031936.00095113

Sbihi, H., Koehoorn, M., Tamburic, L., & Brauer, M. (2017). Asthma Trajectories in a Population-based Birth Cohort Impacts of Air Pollution and Greenness. *Am J Respir Crit Care Med*, *195*(5), 607–613. https://doi.org/10.1164/rccm.201601-0164OC

Schraufnagel, D. E., Balmes, J. R., Cowl, C. T., De Matteis, S., Jung, S.-H., Mortimer, K., … Wuebbles, D. J. (2019). Air Pollution and Noncommunicable Diseases A Review by the Forum of International Respiratory Societies' Environmental Committee, Part 1: The Damaging Effects of Air Pollution. *CHEST*, *155*(2), 409–416. https://doi.org/10.1016/j.chest.2018.10.042

Sese, L., Caliez, J., Annesi-Maesano, I., Cottin, V., Pesce, G., Didier, M., … Nunes, H. (2021). Low income and outcome in idiopathic pulmonary fibrosis: an association to uncover. *Respir Med*, *183*(April), 106415. https://doi.org/10.1183/13993003.congress-2020.1797

Sesé, L., Cavalin, C., Bernaudin, J.-F., Maesano, I. A., & Nunes, H. (2020). Patient Registries in Idiopathic Pulmonary Fibrosis: Don't Forget Socioeconomic Status. *Am J Respir Crit Care Med*, *201*(8), 1014–1015.

Sesé, L., Nunes, H., Cottin, V., Sanyal, S., Didier, M., Carton, Z., … Annesi-Maesano, I. (2018). Role of atmospheric pollution on the natural history of idiopathic pulmonary fibrosis. *Thorax*, *73*(2), 145–150. https://doi.org/10.1136/thoraxjnl-2017-209967

Shull, J. G., Planas-Cerezales, L., Lara Compte, C., Perona, R., & Molina-Molina, M. (2022). Harnessing PM2.5 Exposure Data to Predict Progression of Fibrotic Interstitial Lung Diseases Based on Telomere Length. *Front Med (Lausanne)*, *9*, 871898. https://www.frontiersin.org/articles/10.3389/fmed.2022.871898/full.

Singh, S., Collins, B. F., Bairwa, M., Joshi, J. M., Talwar, D., Sing, N., … Raghu, G. (2019). Hypersensitivity pneumonitis and its correlation with ambient air pollution in urban India. *Eur Respir J*, *53*(2), pii: 1801563. https://doi.org/10.1183/13993003.01563-2018

Smyth, G., Hu, Y., Ritchie, M., Silver, J., Wettenhall, J., McCarthy, D., … Choi, D. (2019). Package "limma" Title Linear Models for Microarray Data. *R Topics*, 1–242.

Solomon, O., Huen, K., Yousefi, P., Küpers, L. K., González, J. R., Suderman, M., … Holland, N. (2022). Meta-analysis of epigenome-wide association studies in newborns and children show widespread sex differences in blood DNA methylation. *Mutat Res Rev Mutat Res*, *789*(Jan-Jun), 108415. https://doi.org/10.1016/J.MRREV.2022.108415

Spengler, J. D., Koutrakis, P., Dockery, D. W., Raizenne, M., & Speizer, F. E. (1996). Health effects of acid aerosols on North American children: Air pollution exposures. *Environ Health Perspect*, *104*(5), 492–499. https://doi.org/10.1289/EHP.96104492

Stephenson, A. L., Sykes, J., Stanojevic, S., Quon, B. S., Marshall, B. C., Petren, K., … Goss, C. H. (2017). Survival Comparison of Patients With Cystic Fibrosis in Canada and the United States A Population-Based Cohort Study. *Ann Intern Med*, *166*(8), 537–546. https://doi.org/10.7326/M16-0858

Tashkin, D. P., Elashoff, R., Clements, P. J., Goldin, J., Roth, M. D., Furst, D. E., … Mkc, F. (2006). Cyclophosphamide versus Placebo in Scleroderma Lung Disease. *N Engl J Med*, *354*, 2655–2666. https://www.nejm.org/doi/full/10.1056/nejmoa055120.

Tashkin, D. P., Roth, M. D., Clements, P. J., Furst, D. E., Khanna, D., Kleerup, E. C., … Elashoff, R. M. (2016). Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (SLS II): a randomised controlled, double-blind, parallel group trial. *Lancet Respir Med*, *4*(9), 708–719. https://doi.org/10.1016/S2213-2600(16)30152-7

Tellez-Plaza, M., Tang, W. Y., Shang, Y., Umans, J. G., Francesconi, K. A., Goessler, W., … Navas-Acien, A. (2014). Association of global DNA methylation and global DNA hydroxymethylation with metals and other exposures in human blood DNA samples. *Environ Health Perspect*, *122*(9), 946–954. https://doi.org/10.1289/EHP.1306674

The Canadian Index of Multiple Deprivation: User Guide. (2019). Retrieved from https://www150.statcan.gc.ca/n1/en/pub/45-20-0001/452000012019002-eng.pdf?st=XEILlJq4

Thurston, G. D., Ito, K., Lippmann, M., & Hayes, C. (1989). Reexamination of London, England, mortality in relation to exposure to acidic aerosols during 1963-1972 winters. *Environ Health Perspect*, *79*, 73–82. https://doi.org/10.1289/EHP.897973

Tingley, D., Yamamoto, H. T., Hirose, K., Keele, L., & Princeton, K. I. (2014). mediation: R Package for Causal Mediation Analysis. *J Stat Softw*, *59*, 1–40. Retrieved from http://cran.r-project.org/package=mediation

To, T., Zhu, J., Larsen, K., Simatovic, J., Feldman, L., Ryckman, K., … Carlsten, C. (2016). Progression from asthma to chronic obstructive pulmonary disease is air pollution a risk factor? *Am J Respir Crit Care Med*, *194*(4), 429–438. https://doi.org/10.1164/rccm.201510-1932OC

Toraño, E. G., García, M. G., Luis Fernández-Morera, J., Niño-García, P., & Fernández, A. F. (2016). The Impact of External Factors on the Epigenome: In Utero and over Lifetime. *BioMed Res Int*, *2016*, 2568635. https://doi.org/10.1155/2016/2568635

Touleimat, N., & Tost, J. (2012). Complete pipeline for Infinium® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*, *4*(3), 325–341. https://doi.org/10.2217/epi.12.21

Treon, J., Dutra, F., Cappel, J., Sigmon, H., & Younker, W. (1950). Toxicity of sulfuric acid mist. *Arch Ind Hyg Occup Med*, *2*. https://doi.org/10.1080/15287397909529812

Tzouvelekis, A., & Kaminski, N. (2015). Epigenetics in idiopathic pulmonary fibrosis. *Biochem Cell Biol*, *93*(2), 159–170. https://doi.org/10.1139/bcb-2014-0126

U.S. Census Bureau. (n.d.). S1701: POVERTY STATUS IN THE PAST 12 MONTHS - Census Bureau Table. Retrieved August 8, 2022, from https://data.census.gov/cedsci/table?q=poverty&g=0100000US%248600000&tid=ACSST5Y2020.S1701

Upagupta, C., Shimbori, C., Alsilmi, R., & Kolb, M. (2018). Matrix abnormalities in pulmonary fibrosis. *Eur Respir Rev*, *27*(148), 180033. https://doi.org/10.1183/16000617.0033-2018

Van Donkelaar, A., Martin, R. V, Li, C., & Burnett, R. T. (2019). Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. *Environ Sci Technol*, *53*(5), 2595–2611. https://doi.org/10.1021/acs.est.8b06392

van Iterson, M., van Zwet, E. W., & Heijmans, B. T. (2017). Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol*, *18*(1), 1–13. https://doi.org/10.1186/S13059-016-1131-9

VanderWeele, T. J. (2016). Mediation Analysis: A Practitioner's Guide. *Annu Rev Public Health*, *37*(1), 17–32. https://doi.org/10.1146/annurev-publhealth-032315-021402

Visweswaran, S., Mclay, B., Cappella, N., Morris, M., Milnes, J. T., Reis, S. E., … Becich, M. J. (2022). An atomic approach to the design and implementation of a research data warehouse. *J Am Med Informatics Assoc*, *29*(4), 601–608. https://doi.org/10.1093/jamia/ocab204

Vrijheid, M. (2014). The exposome: a new paradigm to study the impact of environment on health. *Thorax*, *69*(9), 876–878. https://doi.org/10.1136/thoraxjnl-2013-204949

Wang, B. R., Edwards, R., Freiheit, E. A., Ma, Y., Burg, C., De Andrade, J., … Flaherty, K. R. (2020). The Pulmonary Fibrosis Foundation Patient Registry: Rationale, Design, and Methods. *Annals Am Thorac Soc*, *17*(12), 1620–1628. https://doi.org/10.1513/AnnalsATS.202001-035SD

Wang, C., Cardenas, A., Hutchinson, J. N., Just, A., Heiss, J., Hou, L., … Schwartz, J. D. (2022). Short- and intermediate-term exposure to ambient fine particulate elements and leukocyte epigenome-wide DNA methylation in older men: the Normative Aging Study. *Environ Int*, *158*(Jan 2022), 106955. https://doi.org/10.1016/j.envint.2021.106955

Wang, Y., Xiao, S., Zhang, Y., Chang, H., Martin, R. V., Van Donkelaar, A., … Shi, L. (2022). Long-term exposure to PM2.5 major components and mortality in the southeastern United States. *Environ Int*, *158*(Jan 2022), 106969. https://doi.org/10.1016/j.envint.2021.106969

Weber, R. J., Guo, H., Russell, A. G., & Nenes, A. (2016). High aerosol acidity despite declining atmospheric sulfate concentrations over the past 15 years. *Nat Geosci*, *9*, 282–286. https://doi.org/10.1038/NGEO2665

Wielscher, M., Vierlinger, K., Kegler, U., Ziesche, R., Gsur, A., & Weinhäusel, A. (2015). Diagnostic Performance of Plasma DNA Methylation Profiles in Lung Cancer, Pulmonary Fibrosis and COPD. *EBioMedicine*, *2*(8), 2015. https://doi.org/10.1016/j.ebiom.2015.06.025

Wilhelm-Benartzi, C. S., Koestler, D. C., Karagas, M. R., Flanagan, J. M., Christensen, B. C., Kelsey, K. T., … Brown, R. (2013). Review of processing and analysis methods for DNA methylation array data. *Brit J Cancer*, *109*, 1394–1402. https://doi.org/10.1038/bjc.2013.496

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, *26*(17), 2190–2191. https://doi.org/10.1093/bioinformatics/btq340

Winterbottom, C. J., Shah, R. J., Patterson, K. C., Kreider, M. E., Panettieri, R. A., Rivera-Lebron, B., … Christie, J. D. (2018). Exposure to Ambient Particulate Matter Is Associated With Accelerated Functional Decline in Idiopathic Pulmonary Fibrosis. *CHEST*, *153*(5), 1221–1228. https://doi.org/10.1016/j.chest.2017.07.034

Wolters, P. J., Blackwell, T. S., Eickelberg, O., Loyd, J. E., Kaminski, N., Jenkins, G., … Richeldi, L. (2018). Time for a change: is idiopathic pulmonary fibrosis still idiopathic and only fibrotic? *Lancet Respir Med*, *6*(2), 154–160.

Wong, A. W., Ryerson, C. J., & Guler, S. A. (2020). Progression of fibrosing interstitial lung disease. *Respir Res*, *21*(1), 32. https://doi.org/10.1186/s12931-020-1296-3

World Health Organization. (2018). Ambient (outdoor) air pollution. Retrieved January 20, 2019, from https://www.who.int/en/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

Yamamoto, M., Singh, A., Sava, F., Pui, M., Tebbutt, S. J., & Carlsten, C. (2013). MicroRNA Expression in Response to Controlled Exposure to Diesel Exhaust: Attenuation by the Antioxidant N-Acetylcysteine in a Randomized Crossover Study. *Environ Health Perspect*, *121*, 670–675. https://doi.org/10.1289/ehp.1205963

Yang, I. V, Pedersen, B. S., Rabinovich, E., Hennessy, C. E., Davidson, E. J., Murphy, E., … Schwartz, D. A. (2014). Relationship of DNA Methylation and Gene Expression in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*, *190*(11), 1263–1272. https://doi.org/10.1164/rccm.201408-1452OC

Zhang, H., Hu, J., Kleeman, M., & Ying, Q. (2014). Source apportionment of sulfate and nitrate particulate matter in the Eastern United States and effectiveness of emission control programs. *Sci Total Environ*, *490*, 171–181. https://doi.org/10.1016/j.scitotenv.2014.04.064

Zhang, Y. (2017). MUC5B and short telomere length in hypersensitivity pneumonitis. *Lancet Respir Med*, *5*(8), 603–604. https://www.thelancet.com/journals/lanres/article/PIIS2213-2600(17)30210-2/fulltext.

Zhang, Y., Noth, I., Garcia, J. G. N., & Kaminski, N. (2011). A Variant in the Promoter of MUC5B and Idiopathic Pulmonary Fibrosis. *N Engl J Med*, *364*(16), 1576–1577.

Zhao, N., Al-Aly, Z., Zheng, B., van Donkelaar, A., Martin, R. V., Pineau, C. A., & Bernatsky, S. (2022). Fine particles matter components and interstitial lung disease in rheumatoid arthritis. *Eur Respir J*, *60*, 2102149. https://doi.org/10.1183/13993003.02149-2021

Zheng, Y., Sanchez-Guerra, M., Zhang, Z., Joyce, B. T., Zhong, J., Kresovich, J. K., … Hou, L. (2016). Traffic-derived particulate matter exposure and histone H3 modification: A repeated measures study. *Environ Res*, *153*, 112–119. https://doi.org/10.1016/j.envres.2016.11.015

Zhong, J., Karlsson, O., Wang, G., Li, J., Guo, Y., Lin, X., … Baccarelli, A. A. (2017). B vitamins attenuate the epigenetic effects of ambient fine particles in a pilot human intervention trial. *Proc Natl Acad Sci U.S.A.*, *114*(13), 3503–3508. https://doi.org/10.1073/pnas.1618545114