OXFORD

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

American Journal of Epidemiology, 2025, **194**, 730–737
https://doi.org/10.1093/aje/kwae271
Advance access publication date August 9, 2024
Original Contribution

The association between long-term $PM_{2.5}$ exposure and risk for pancreatic cancer: an application of social informatics

Nrupen A. Bhavsar*[,1,2], Kay Jowers[3], Lexie Z. Yang[2], Sharmistha Guha[4], Xuan Lin[5], Sarah Peskoe[2], Hannah McManus[6], Lisa McElroy[1], Mercedes Bravo[7], Jerome P. Reiter[8], Eric Whitsel[9], Christopher Timmins[5]

[1]Department of Surgery, School of Medicine, Duke University, Durham, NC, United States
[2]Department of Biostatistics and Bioinformatics, School of Medicine, Duke University, Durham, NC, United States
[3]Nicholas Institute for Energy, Environment & Sustainability, Duke University, Durham, NC, United States
[4]Department of Statistics, Texas A&M University, College Station, TX, United States
[5]Department of Economics, Duke University, Durham, NC, United States
[6]Department of Medicine, School of Medicine, Duke University, Durham, NC, United States
[7]Global Health Institute, Duke University, Durham, NC, United States
[8]Department of Statistical Science, Duke University, Durham, NC, United States
[9]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

*Corresponding author: Nrupen Bhavsar, 710 W. Main Street, Durham, NC (nrupen.bhavsar@duke.edu)

Abstract

There is a profound need to identify modifiable risk factors to screen and prevent pancreatic cancer. Air pollution, including fine particulate matter ($PM_{2.5}$), is increasingly recognized as a risk factor for cancer. We conducted a case-control study using data from the electronic health record (EHR) of Duke University Health System, 15-year residential history, NASA satellite fine particulate matter ($PM_{2.5}$), and neighborhood socioeconomic data. Using deterministic and probabilistic linkage algorithms, we linked residential history and EHR data to quantify long-term $PM_{2.5}$ exposure. Logistic regression models quantified the association between a 1 interquartile range (IQR) increase in $PM_{2.5}$ concentration and pancreatic cancer risk. The study included 203 cases and 5027 controls (median age of 59 years, 62% female, 26% Black). Individuals with pancreatic cancer had higher average annual exposure (9.4 $\mu g/m^3$) as compared to an IQR increase in average annual $PM_{2.5}$, which was associated with greater odds of pancreatic cancer (odds ratio = 1.20; 95% CI, 1.00-1.44). These findings highlight the link between elevated $PM_{2.5}$ exposure and increased pancreatic cancer risk. They may inform screening strategies for high-risk populations and guide air pollution policies to mitigate exposure.

This article is part of a Special Collection on Environmental Epidemiology.

Key words: social informatics; electronic health records; fine particulate matter; $PM_{2.5}$; pancreatic cancer; data linkage; social determinants of health.

Introduction

Pancreatic cancer is the seventh leading cause of cancer-related death worldwide and the third leading cause of cancer-related mortality in the United States.[1] By 2040, it is projected to be the second leading cause of cancer mortality.[2] The 5-year survival rate in the United States is 12.5%, among the lowest of all cancers. Known risk factors for pancreatic cancer include smoking, type 2 diabetes, body mass index, and genetics.[3] There is a need to identify novel modifiable risk factors to inform pancreatic cancer screening. Air pollution exposure is a modifiable environmental exposure that may increase the risk for pancreatic cancer and inform screening practices.

Environmental exposures to air pollution, specifically to fine particulate matter ($PM_{2.5}$), have been linked with several health outcomes, including cardiovascular disease, inflammation, oxidative stress, and various forms of cancer, including bladder, breast, lung, kidney, and, increasingly, pancreatic cancer.[4-11] Despite several studies demonstrating a potential link between $PM_{2.5}$ and pancreatic cancer, the mechanism of this relationship is not well understood.[4,7,8] Potential hypotheses for the role of $PM_{2.5}$ in pancreatic cancer include PM-mediated inflammation, oxidative stress, and subsequent DNA damage, as has been better characterized in lung cancer.[12] Additionally, several pancreatic cancer–specific mechanisms have been proposed, including the direct effect on the intestinal epithelium and local immune response from PM ingestion as well as heavy metal accumulation in the pancreas, which has been found in patients with pancreatic cancer. However, these studies have been relatively small and limited to largely racially and ethnically homogeneous populations.[13,14]

To address some of these limitations, we use electronic health record (EHR) data linked to residential history, air pollution, and socioeconomic data and multidisciplinary approaches at the intersection of informatics and social epidemiology—which we

Received: July 7, 2023. **Accepted:** August 6, 2024

term *social informatics*—to quantify the association between long-term $PM_{2.5}$ exposure and risk for pancreatic cancer. Results from this study may inform future studies' screening strategies for populations at highest risk for pancreatic cancer and air pollution policy that can mitigate exposure.

Methods

Overview

We conducted an unmatched case-control study among individuals with and without pancreatic cancer who had encounters within the Duke University Health System.

Data sources

Electronic health records

Data on demographic and clinical characteristics were obtained from the EHR of the Duke University Health System (DUHS) and the Duke Cancer Registry from 2014 to 2019. DUHS is the predominant health care provider for insured Durham County residents as 85% of Durham County residents have a primary or specialty encounter in the DUHS.[15] It includes an academic medical center and 2 community hospitals linked with a network of outpatient clinics. The Duke Cancer Center is a National Cancer Institute–designated comprehensive cancer center.

Duke cancer registry

Clinical data from the EHR were supplemented with data from the Duke Cancer Registry. Comprehensive cancer centers are required to report all newly diagnosed cases to the National Cancer Database. The cancer registry updates the vital status of patients treated at the cancer center with routine queries of the state death registry.

Study population

The source population for this study included adults with at least 1 primary care encounter from 2014 to 2019 (Figure 1). A patient was considered a case if they were diagnosed with pancreatic cancer after a primary care encounter. They must have been ≥ 19 years of age at their first primary care encounter and included in the Duke Cancer Registry. Controls included patients who did not have any cancer diagnosis before or after a comparable primary care encounter in the EHR who were ≥ 19 years old at the earliest encounter. The age of controls was restricted to the age range of cases (ie, 37-90 years).

Exposure assessment

Residential mobility

The InfoUSA Historical Residential File data set was used to quantify residential mobility. InfoUSA is a marketing and sales company that maintains a consumer database tracking 120 million households and 292 million individuals primarily for use by private-sector companies to connect with customers. These data—available from 2006 to 2021—are built and maintained using 29 billion records from hundreds of sources, such as census statistics, billing statements, telephone directory listings, and magazine subscribers. These data are used to quantify residential mobility (ie, when people move from one residence to another residence), regardless of where in the United States an individual lived before their encounter in the Duke University Health System. We created residential history timelines by defining start and end dates for residence at a specific address. The start date of one address equals the end date of another address.

Residential data linkage

We used deterministic and probabilistic linkage approaches to match patient ID from EHR data with family ID in the InfoUSA data. Deterministic linkage matched records in EHR data with InfoUSA data by shared keys, including first name, last name, and address. Probabilistic linkage was implemented using R package "fastLink," which calculates the string similarities of linkage fields based on the Fellegi-Sunter methods.[16,17] Manual review was used to identify correct matches that were missed through deterministic linkage due to alternative spellings, spelling mistakes, and spacing errors. The matching algorithm lists all potential pairs and sorts them by similarity scores. A manual check was conducted for pairs above a certain threshold. After InfoUSA family ID was linked to name and address from the EHR, residential history was quantified by tracking family IDs across prior years of InfoUSA data. Past addresses from the EHR were used to address discrepancies if they were recorded but were not found in InfoUSA.

$PM_{2.5}$

Ground-level monitors can provide accurate measures of $PM_{2.5}$, but their locations and distribution are limited.[18] Satellite-based $PM_{2.5}$ estimates were therefore derived from aerosol optical depth (AOD) measures. The AOD measures the amount of direct sunlight that is prevented from reaching the ground by particles like particulate matter with a diameter < 2.5 μm ($PM_{2.5}$), which absorbs, scatters, and thereby blocks sunlight. Exposure to $PM_{2.5}$ was quantified using satellite AOD measures from the Moderate Resolution Imaging Spectroradiometer and Multiangle Imaging Spectroradiometer satellite instruments onboard the National Aeronautics and Space Administration (NASA) Terra satellite. Geographically weighted regression was then used to estimate $PM_{2.5}$ concentrations from the AOD measures. These estimates are correlated with measures from ground-level monitors. The developers of the $PM_{2.5}$ model have published a validation paper that reports $PM_{2.5}$ concentrations from the model have a correlation of 0.82 with in situ concentrations (ie, Canadian National Air Pollution Surveillance Network and US Environmental Protection Agency's Air Quality System).[19,20] The data are publicly available,[21] and the models have been used in numerous studies.[22-28]

Air pollution exposures are assigned at census geographies (eg, block group), grid cell, or latitude/longitude. The $PM_{2.5}$ estimates we use are available at a resolution of 0.01° × 0.01°; these estimates are then assigned to block group shape files. AOD-based, block group–level $PM_{2.5}$ mass concentrations from 2001 to 2018 were then linked to individual patient residences and used to calculate annual mean $PM_{2.5}$ concentrations. As $PM_{2.5}$ data for 2019 were not released at the time of analysis, we assigned 2018 concentrations to 2019 as there was minimal change in $PM_{2.5}$ concentrations from one year to the next in recent years.

Outcome assessment

Pancreatic cancer cases were obtained from the Duke Cancer Registry, which contains the date of diagnosis, cancer stage, and histology.

Statistical analysis

Continuous variables were summarized with means, medians, standard deviations, Quartile 1 (Q1), Quartile 3 (Q3), and ranges. Categorical variables were analyzed with frequency counts and percentages. We used logistic regression to quantify the association between a 1 interquartile range (IQR) increase in $PM_{2.5}$

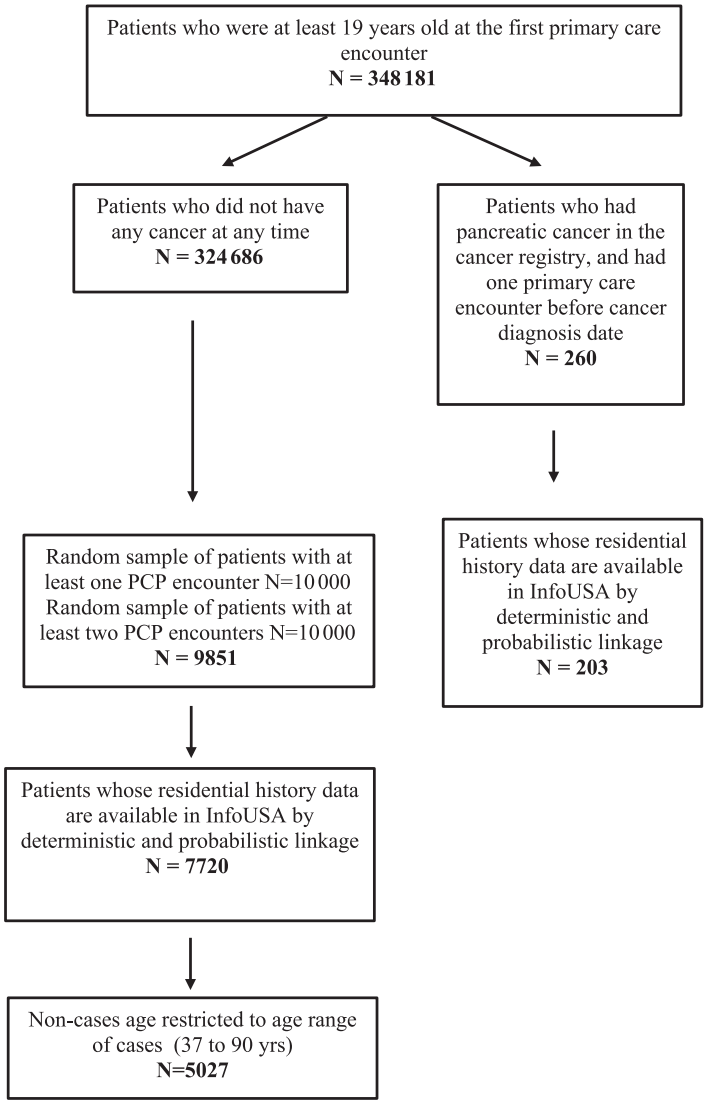

Figure 1. Consort diagram of cases and controls included in the study.

concentration and risk for pancreatic cancer. For both cases and noncases, the date of primary care encounter was used as the index date to quantify $PM_{2.5}$ concentration. The annual average of $PM_{2.5}$ was calculated by taking the average of $PM_{2.5}$ measured in all years before the initial primary care encounter at Duke. Years where we did not have residential history data and therefore unable to link $PM_{2.5}$ were dropped from this calculation. Models were adjusted for demographic (ie, age, race, ethnicity, sex), clinical (ie, diabetes mellitus, body mass index [BMI]), and social characteristics (ie, smoking status, neighborhood socioeconomic status [SES], number of primary care encounters during 2014-2019). Statistical significance was assessed at α level of .05. Analyses were conducted using SAS 9.4 (SAS Institute) and R 4.1.0 (R Core Team).

Sensitivity analysis

We tested the sensitivity of our results to missing residential history data by imputing $PM_{2.5}$ values for these years with (1) average of values in years adjacent to missing years and (2) the smaller concentration of adjacent values. We also examined the association by (1) only using $PM_{2.5}$ values from the index year and (2) assuming static addresses for individuals after the index year (ie, no residential mobility).

Results

This analysis included 5230 patients with pancreatic cancer and comparable controls who received care within the Duke University Health System from 2014 to 2019. The study population was 62% female, 26% Black, and 3% Hispanic, with a median age of 59 years (IQR, 48-69 years) (Table 1). The study population had a median BMI of 29 (IQR, 25-34), 40% were current or former smokers, and 30% had a diagnosis of type 2 diabetes.

Baseline characteristics

Within the cohort, 203 individuals were diagnosed with pancreatic cancer, and 5027 individuals did not have a diagnosis for pancreatic cancer. As compared to individuals without pancreatic cancer, individuals with pancreatic cancer were older (69 years vs 58 years), less likely to be female (50% vs 62%), and more likely to self-identify as Black (32% vs 26%) and non-Hispanic (98% vs 94%). As compared to individuals without pancreatic cancer, individuals

Table 1. Baseline characteristics of the study population.

Characteristic	No pancreatic cancer ($n = 5027$)	Pancreatic cancer ($n = 203$)	Total ($n = 5230$)
Age, y			
Median	58.1	68.8	58.6
Q1, Q3	47.6, 68.8	61.1, 77.0	47.8, 69.1
Female	3130 (62.3)	101 (49.8)	3231 (61.8)
Race			
Asian	125 (2.5)	3 (1.5)	128 (2.4)
Black	1308 (26.0)	64 (31.5)	1372 (26.2)
White	3330 (66.2)	130 (64.0)	3460 (66.2)
Other	264 (5.3)	6 (3.0)	270 (5.2)
Hispanic			
No	4731 (94.1)	198 (97.5)	4929 (94.2)
Yes	131 (2.6)	0 (0.0)	131 (2.5)
Unknown/missing	165 (3.3)	5 (2.5)	170 (3.3)
BMI			
Median	29.1	28.7	29.1
Q1, Q3	25.3, 34.0	24.6, 33.4	25.3, 34.0
Missing	95	1	96
BMI group			
Underweight (< 18.5)	51 (1.0)	7 (3.4)	58 (1.1)
Normal (18.5-24.9)	1086 (21.6)	47 (23.2)	1133 (21.7)
Overweight (25-29.9)	1606 (31.9)	59 (29.1)	1665 (31.8)
Obese (≥ 30)	2189 (43.5)	89 (43.8)	2278 (43.6)
Unknown/missing	95 (1.9)	1 (0.5)	96 (1.8)
Smoking status			
Current/former	1959 (39.0)	113 (55.7)	2072 (39.6)
Never	2932 (58.3)	85 (41.9)	3017 (57.7)
Unknown/missing	136 (2.7)	5 (2.5)	141 (2.7)
Diabetes			
Yes	1505 (29.9)	72 (35.5)	1577 (30.2)
Years of exposure history			
Median	12.0	12.0	12.0
Q1, Q3	7.0, 14.0	8.0, 14.0	7.0, 14.0
Annual average $PM_{2.5}$ ($\mu g/m^3$) (Method 1: drop missing years)			
Median	9.0	9.4	9.0
Q1, Q3	8.3, 9.6	8.6, 9.8	8.3, 9.6
Annual average $PM_{2.5}$ ($\mu g/m^3$) (Method 2: impute with adjacent values)			
Median	9.1	9.3	9.1
Q1, Q3	8.5, 9.6	8.7, 9.7	8.5, 9.6
Annual average $PM_{2.5}$ ($\mu g/m^3$) (Method 3: impute with smaller values)			
Median	9.0	9.3	9.0
Q1, Q3	8.4, 9.5	8.6, 9.7	8.4, 9.5
AHRQ SES index			
Median	55.2	55.3	55.2
Q1, Q3	50.9, 59.4	51.1, 59.7	50.9, 59.4
Missing	238	13	251
Individual-level SES before index			
Median	69.0	72.0	69.0
Q1, Q3	37.0, 114.0	35.0, 113.0	37.0, 114.0
Missing	1116	41	1157
Cancer group—cases only			
Intraductal papillary	-	2 (1.0)	2 (0.0)
Pancreatic adenocarcinoma	-	158 (77.8)	158 (3.0)
Pancreatic neuroendocrine carcinoma (aggressive grade 3)	-	21 (10.3)	21 (0.4)
Pancreatic neuroendocrine tumor (grades 1-2)	-	18 (8.9)	18 (0.3)
Sarcomatous carcinoma	-	1 (0.5)	1 (0.0)
Vague	-	3 (1.5)	3 (0.1)
NA	5027 (100)	0 (0.0)	5027 (96.1)
Number of PCP encounters in 2014-2019			
1	2532 (50.4)	57 (28.1)	2589 (49.5)
≥ 2	2495 (49.6)	146 (71.9)	2641 (50.5)
Cancer stage—cases only			
1-3	-	79 (38.9)	79 (1.5)
4	-	74 (36.5)	74 (1.4)
Unknown/missing	-	50 (24.6)	50 (1.0)
Number of encounters before diagnosis (cases only)			
Median		29.0	29.0
Q1, Q3		13.0, 72.0	13.0, 72.0

Abbreviations: AHRQ, Agency for Healthcare Research and Quality; BMI, body mass index; NA, Not applicable; PCP, Primary care provider; SES, socioeconomic status.
Values are presented as *n* (%) unless otherwise indicated.

Table 2. Air pollution concentration by pancreatic cancer type.

Annual average $PM_{2.5}$ (μg/m3)	Total	Intraductal papillary ($n = 2$)	Pancreatic adenocarcinoma ($n = 158$)	Pancreatic neuroendocrine tumor (grades 1-2) ($n = 18$)	Pancreatic neuroendocrine carcinoma (aggressive grade 3) ($n = 21$)	Sarcomatous carcinoma ($n = 1$)	Vague ($n = 3$)
Method 1: drop missing years							
Mean (SD)	8.9 (1.1)	9.0 (1.0)	9.2 (1.1)	9.4 (1.0)	8.8 (1.3)	6.9 (.)	10.3 (0.9)
Median	9.0	9.0	9.3	9.5	8.6	6.9	10.0
Q1, Q3	8.3, 9.6	8.4, 9.7	8.6, 9.8	9.0, 9.9	8.0, 9.6	6.9, 6.9	9.7, 11.4
Range	(3.2-13.9)	(8.4-9.7)	(6.1-12.1)	(6.2-10.9)	(6.4-11.1)	(6.9-6.9)	(9.7-11.4)
Method 2: impute with adjacent values							
Mean (SD)	9.0 (0.9)	9.0 (0.9)	9.2 (0.8)	9.3 (0.9)	8.7 (1.2)	6.9 (.)	10.0 (0.6)
Median	9.1	9.0	9.3	9.5	8.7	6.9	9.7
Q1, Q3	8.5, 9.6	8.4, 9.6	8.8, 9.7	9.1, 9.9	8.2, 9.6	6.9, 6.9	9.7, 10.6
Range	(3.2-13.8)	(8.4-9.6)	(6.4-12.1)	(6.2-10.5)	(6.2-10.3)	(6.9-6.9)	(9.7-10.6)
Method 3: impute with smaller values							
Mean (SD)	8.9 (0.9)	8.9 (0.8)	9.1 (0.9)	9.3 (0.9)	8.7 (1.2)	6.9 (.)	10.0 (0.6)
Median	9.0	8.9	9.2	9.5	8.7	6.9	9.7
Q1, Q3	8.4, 9.5	8.4, 9.5	8.6, 9.7	9.0, 9.9	8.0, 9.6	6.9, 6.9	9.7, 10.6
Range	(3.2-13.8)	(8.4-9.5)	(6.4-12.1)	(6.2-10.5)	(6.2-10.3)	(6.9-6.9)	(9.7-10.6)

with pancreatic cancer were more likely to be current or former smokers (56% vs 39%) and more likely to have diabetes (36% vs 30%). The median BMI was similar between individuals with and without pancreatic cancer (29 kg/m^2).

Air pollution exposure

The median annual $PM_{2.5}$ exposure was 9.0 $\mu g/m^3$. Individuals with pancreatic cancer had higher average annual exposure (9.4 $\mu g/m^3$) as compared to individuals without pancreatic cancer (9.0 $\mu g/m^3$). Average annual exposure to $PM_{2.5}$ varied by pancreatic cancer type (6.9-10.3 $\mu g/m^3$) (Table 2). Among cancer types with more than 3 cases, individuals with grade 1 to 2 adenocarcinomas and neuroendocrine tumors had higher average annual $PM_{2.5}$ exposure (9.2 and 9.4 $\mu g/m^3$, respectively) as compared to individuals with grade 3 neuroendocrine carcinoma aggressive (8.8 $\mu g/m^3$).

Association between $PM_{2.5}$ concentration and pancreatic cancer

In fully adjusted models, each IQR increase in average annual $PM_{2.5}$ $\mu g/m^3$ (IQR, 1.3 $\mu g/m^3$) was associated with a 20% greater odds of pancreatic cancer (odds ratio [OR] = 1.20; 95% CI, 1.00-1.44) (Table 3). Imputing missing $PM_{2.5}$ data, with the concentration from the year before and after the missing year averaged, resulted in a similar association (OR = 1.18; 95% CI, 0.98-1.43). Imputing missing $PM_{2.5}$ data with the concentration from the year with a lower concentration (ie, more conservative value) resulted in a slightly attenuated association (OR = 1.15; 95% CI, 0.96-1.39).

Short-term vs long-term PM exposure

Limiting $PM_{2.5}$ concentration to the index year resulted in a lower median concentration of $PM_{2.5}$ (8.0 $\mu g/m^3$) than values calculated using residential history (9.0 $\mu g/m^3$) but a similar association with pancreatic cancer (OR = 1.24; 95% CI, 1.00-1.55) (Table 3). If a static address is assumed for individuals from the index date onward, the concentration of $PM_{2.5}$ was lower (8.3 $\mu g/m^3$) than values calculated using residential history (9.0 $\mu g/m^3$). The association between 1 IQR increase in $PM_{2.5}$ and pancreatic cancer using static addresses was slightly attenuated (OR = 1.13; 95% CI, 0.93-1.38) (Table 3).

Individuals without residential history data

Individuals for whom we were unable to link to their residential history were younger, more likely to be female, and less likely to self-identify as White (Table S1). Individuals not linked were less likely to be obese, have a diagnosis of diabetes, or be a current or former smoker. The neighborhood SES did not differ between individuals linked (neighborhood SES = 55) and not linked (neighborhood SES = 55) to their residential history.

Discussion

In this study of individuals who were diagnosed with pancreatic cancer at the Duke University Health System and a comparable set of controls, we found that individuals with greater annual exposure to $PM_{2.5}$ were at higher risk for pancreatic cancer. The strength of the association did not change appreciably when different imputation strategies were implemented to address missing $PM_{2.5}$ concentration.

Studies that quantify the association between $PM_{2.5}$ and pancreatic cancer have reported mixed results. Within the United States, results from the Cancer Prevention Study II suggested a null association per 4.4-$\mu g/m^3$ increase in $PM_{2.5}$ (hazard ratio [HR] = 0.98; 95% CI, 0.92-1.03).[4] The study population was recruited from across the United States and was predominately White (95%) and female (55%). Mean exposure at baseline among these individuals was 12.6 $\mu g/m^3$. Similarly, results from the National Health Interview Survey suggested a null association between a 10-$\mu g/m^3$ increase in $PM_{2.5}$ and risk for pancreatic cancer mortality (HR = 1.09; 95% CI, 0.83-1.44).[29] The study population was majority White (68%) and female (55%), with a mean estimated $PM_{2.5}$ exposure of 10.7 $\mu g/m^3$. In contrast, using data from the Multiethnic Cohort Study, which includes a population largely residing in the Los Angeles area, researchers found an increased risk for incident pancreatic cancer per 10 $\mu g/m^3$ (HR = 1.61; 95% CI, 1.09-2.37).[8] The study population was largely Latino (42%), African American (32%), and female (57%), with a median baseline

Table 3. Association between 1 IQR increase in average annual $PM_{2.5}$ and odds for pancreatic cancer.

Method	Estimate	95% CI	
		LL	UL
Drop missing years[a]	1.20	1.00	1.44
Impute missing values			
Impute with average of adjacent years	1.18	0.98	1.43
Impute with smaller PM value	1.15	0.96	1.39
Use $PM_{2.5}$ concentration from index year	1.24	1.00	1.55
Use $PM_{2.5}$ concentration from a static address (use patient address from index year throughout exposure history)	1.13	0.93	1.38

Abbreviations: LL, lower limit; UL, upper limit.
All models adjusted for demographic (age, race, ethnicity, sex), clinical (diabetes mellitus, body mass index), social factors (smoking status, neighborhood socioeconomic status), and number of encounters.
[a]Years with missing $PM_{2.5}$ data were dropped.

and study period concentration of $PM_{2.5}$ between 16 and 17 $\mu g/m^3$. Our study population was largely White (66%) and female (62%), with a median $PM_{2.5}$ concentration of 9.0 $\mu g/m^3$. There have been a limited number of studies conducted outside of the United States. One study conducted in China using pancreatic cancer mortality data from 1991 to 2009 reported a 16% increased risk for mortality per 10-unit increase in $PM_{2.5}$ (relative risk (RR) = 1.16; 95% CI, 1.13-1.20).[7] Average concentration of $PM_{2.5}$ ranged from 39 to 46 $\mu g/m^3$ between rural and urban areas. The association we report per IQR (ie, 1.3 $\mu g/m^3$) increase in $PM_{2.5}$ (OR = 1.20; 95% CI, 1.00-1.44) and risk for incident pancreatic cancer was stronger than associations reported in these other studies. The baseline levels of $PM_{2.5}$ in our study population were substantially lower than values reported in other studies. For our population, a sustained, modest increase in $PM_{2.5}$ exposure may substantially increase the risk for health outcomes.

The potential carcinogenic mechanism of $PM_{2.5}$ in the development of pancreatic cancer has not been well characterized, but it has been more thoroughly studied in respiratory epithelial cells in relation to lung cancer.[30] Proposed mechanisms primarily have included the role of PM and its components in inducing oxidative stress, generating reactive oxygen species, and resulting in oxidative DNA damage, as well as increased inflammation.[12,31] These compounds may affect cancer risk through increasing oxidative stress and inflammation, as observed in the airways,[12] and formation of DNA adducts.[31] Similarly, in gastrointestinal conditions such as pancreatic cancer, exposure to $PM_{2.5}$ is somewhat direct due to potential ingestion and thus may have a direct impact on the gut epithelial cells, leading to direct injury of epithelial cells, alterations in immune response, and effects on the gut microbiota.[32] In addition to these mechanisms, there may be pancreatic cancer (PC)-specific links related to heavy metal accumulation in the pancreas with exposure to several trace elements, including arsenic, cadmium, and lead, which have been linked to PC and are contained in $PM_{2.5}$.[13] In several studies, increased levels of heavy metals, both systemically and in pancreatic juice, were found in patients with pancreatic cancer.[13,14]

Our study is unique and has multiple strengths. There is increasing interest in the subfield of informatics, termed *social informatics*, which aims to link data from the EHR with extent social and environmental data to understand the impact of contextual factors on health. This is the first study, to our knowledge, to use social informatics approaches to link data from the EHR to residential mobility and air pollution data to quantify the association between long-term $PM_{2.5}$ exposure and odds of pancreatic cancer. We were able to follow individuals up to 15 years before their encounter within the DUHS regardless of whether they had an encounter during that time. This allowed us to quantify long-term $PM_{2.5}$ exposure. We examined multiple air pollution averaging intervals and found that associations between air pollution exposure and PC were similar. Our population was relatively diverse and representative of the broader community within Durham County and North Carolina. We were able to quantify exposure to $PM_{2.5}$ across subtypes of pancreatic cancer. This may suggest that $PM_{2.5}$ may be associated not just with cancer incidence but also with the aggressiveness of pancreatic cancer or responsiveness to systemic treatment. In addition, we were able to test different approaches to impute missing $PM_{2.5}$ data and the impact of short- and long-term $PM_{2.5}$ exposure on pancreatic cancer risk.

Our study has notable limitations. To our knowledge, there are no validation statistics or sources to verify the accuracy of InfoUSA residential history data. We compared the accuracy of InfoUSA with data within the EHR and found that 81% of Federal Information Processing Standards (FIPS) codes matched at the block group level. Other studies have quantified residential history using other third-party vendors, with the most frequently cited vendor being Lexis-Nexis.[33] One study examined the accuracy of Lexis-Nexis data to quantify residential history and found that Lexis-Nexis accounted for 72% of the 3 most recent addresses but only 43% of all addresses. We were not able to link patient addresses and InfoUSA residential history data for all individuals. Individuals for whom we were unable to quantify residential history were younger, more often female, and less likely to self-identify as White. Our results are in line with prior research that has shown that residential mobility linkage rates are lower for younger individuals and those who self-identify as non-White and Hispanic.[34] We did see similar air pollution concentrations (8.0 and 8.1 $\mu g/m^3$, respectively) and proportion with cancer in those linked and not linked (3.1% and 1.8%, respectively). The impact on the association of interest may be minimal. Similarly, access to medical care is somewhat related to residential address, which determines our exposure in this study.[35] Access to health care in the United States varies not only by geography but also by race, ethnicity, and SES; our models adjusted for these variables to minimize potential confounding.[35] Particulate matter air pollution is quantified by its mass and/or composition as a function of size. In this study, we focused on $PM_{2.5}$ concentration, which is measured as the mass of particles < 2.5 μm in aerodynamic diameter per cubic meter of air ($\mu g/m^3$).

We did not have information on air pollution exposure within the home, at work, or if someone lived outside of the United States,

on a military base, or in an institution (eg, prison). Pancreatic cancer diagnoses are not common conditions, and the absolute number of cases limited our ability to conduct subgroup analyses.

In conclusion, exposure to $PM_{2.5}$ may increase the odds of incident pancreatic cancer. The results add to a growing body of evidence that higher air pollution levels increase the risk for cancer outcomes. As increasing research links air pollution exposure to multiple cancers, policy designed to lower air pollution levels may reduce the risk for these cancers. While air pollution alone may not inform screening, including long-term air pollution exposure in conjunction with known nongenetic risk factors, such as type 2 diabetes, adiposity, and smoking, may identify additional high-risk populations who would benefit from screening. A similar evolution has been seen in the approach to lung cancer.[36,37] As noted in the US Preventative Services Task Force updated evidence report, there is a need for more information on whether pancreatic cancer screening impacts morbidity and mortality.[38] The findings from this study and others suggest that environmental factors may need to be considered alongside traditional pancreatic cancer risk factors when designing screening guidelines.

Supplementary material

Supplementary material is available at *American Journal of Epidemiology* online.

Funding

This work was funded by the National Heart, Lung, and Blood Institute (grant K01140146) and NIH UL1TR002553.

Conflict of interest

None declared.

Data availability

The data used in this study are not publicly available due to data use agreements and data privacy laws. Anonymized data may be available upon reasonable request and after approval from the Duke University Health System.

References

1. Siegel RL, Miller KD, Wagle NS, et al. Cancer statistics, 2023. *CA Cancer J Clin*. 2023;73(1):17-48. https://doi.org/10.3322/caac.21763
2. Rahib L, Wehner MR, Matrisian LM, et al. Estimated projection of US cancer incidence and death to 2040. *JAMA Netw Open*. 2021;4(4):e214708. https://doi.org/10.1001/jamanetworkopen.2021.4708
3. Capasso M, Franceschi M, Rodriguez-Castro KI, et al. Epidemiology and risk factors of pancreatic cancer. *Acta Biomed*. 2018;89(9S):141-146. https://doi.org/10.23750/abm.v89i9-S.7923
4. Turner MC, Krewski D, Diver WR, et al. Ambient air pollution and cancer mortality in the Cancer Prevention Study II. *Environ Health Perspect*. 2017;125(8):087013. https://doi.org/10.1289/EHP1249
5. Goldberg MS, Villeneuve PJ, Crouse D, et al. Associations between incident breast cancer and ambient concentrations of nitrogen dioxide from a national land use regression model in the Canadian National Breast Screening Study. *Environ Int*. 2019;133(pt B):105182. https://doi.org/10.1016/j.envint.2019.105182
6. Turner MC, Andersen ZJ, Baccarelli A, et al. Outdoor air pollution and cancer: an overview of the current evidence and public health recommendations. *CA Cancer J Clin*. 2020;70(6):460-479. https://doi.org/10.3322/caac.21632
7. Wang Y, Li M, Wan X, et al. Spatiotemporal analysis of PM(2.5) and pancreatic cancer mortality in China. *Environ Res*. 2018;164:132-139. https://doi.org/10.1016/j.envres.2018.02.026
8. Bogumil D, Wu AH, Stram D, et al. The association between ambient air pollutants and pancreatic cancer in the Multiethnic Cohort Study. *Environ Res*. 2021;202:1-8. https://doi.org/10.1016/j.envres.2021.111608
9. Kaufman JD, Adar SD, Barr RG, et al. Association between air pollution and coronary artery calcification within six metropolitan areas in the USA (the Multi-Ethnic Study of Atherosclerosis and Air Pollution): a longitudinal cohort study. *Lancet*. 2016;388(10045):696-704. https://doi.org/10.1016/S0140-6736(16)00378-0
10. Kaufman JD, Elkind MSV, Bhatnagar A, et al. Guidance to reduce the cardiovascular burden of ambient air pollutants: a policy statement from the American Heart Association. *Circulation*. 2020;142(23):e432-e447. https://doi.org/10.1161/CIR.0000000000000930
11. Dominici F, Peng RD, Bell ML, et al. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA*. 2006;295(10):1127-1134. https://doi.org/10.1001/jama.295.10.1127
12. Zhang L, Li J, Zong L, et al. Reactive oxygen species and targeted therapy for pancreatic cancer. *Oxid Med Cell Longev*. 2016;2016:1616781. https://doi.org/10.1155/2016/1616781
13. Amaral AF, Porta M, Silverman DT, et al. Pancreatic cancer risk and levels of trace elements. *Gut*. 2012;61(11):1583-1588. https://doi.org/10.1136/gutjnl-2011-301086
14. Carrigan PE, Hentz JG, Gordon G, et al. Distinctive heavy metal composition of pancreatic juice in patients with pancreatic carcinoma. *Cancer Epidemiol Biomarkers Prev*. 2007;16(12):2656-2663. https://doi.org/10.1158/1055-9965.EPI-07-0332
15. Lang JE, Tang M, Zhao C, et al. Well-child care attendance and risk of asthma exacerbations. *Pediatrics*. 2020;146(6). https://doi.org/10.1542/peds.2020-1023
16. Fellegi IS, Sunter AB. A theory for record linkage. *J Am Stat Assoc*. 1969;64(328):1183-1210. https://doi.org/10.1080/01621459.1969.10501049
17. Enamorado TFB, Imai K. Using a probabilistic model to assist merging of large-scale administrative records. *Am Polit Sci Rev*. 2019;113(2):353-371. https://doi.org/10.1017/S0003055418000783
18. Johansen K, Phinn S, Taylor M. Mapping woody vegetation clearing in Queensland, Australia from Landsat imagery using the Google Earth Engine. *Remote Sens Appl: Soc Environ*. 2015;1:36-49. https://doi.org/10.1016/j.rsase.2015.06.002
19. van Donkelaar A, Hammer MS, Bindle L, et al. Monthly global estimates of fine particulate matter and their uncertainty. *Environ Sci Technol*. 2021;55(22):15287-15300. https://doi.org/10.1021/acs.est.1c05309
20. van Donkelaar A, Martin RV, Spurr RJ, et al. High-resolution satellite-derived PM2.5 from optimal estimation and geographically weighted regression over North America. *Environ Sci Technol*. 2015;49(17):10482-10491. https://doi.org/10.1021/acs.est.5b02076
21. Group ACA. Surface PM2.5. Washington University in St Louis. https://sites.wustl.edu/acag/datasets/surface-pm2-5/
22. Li C, van Donkelaar A, Hammer MS, et al. Reversal of trends in global fine particulate matter air pollution.

Nat Commun. 2023;14(1):5349. https://doi.org/10.1038/s41467-023-41086-z
23. Bowe B, Gibson AK, Xie Y, et al. Ambient fine particulate matter air pollution and risk of weight gain and obesity in United States veterans: an observational cohort study. *Environ Health Perspect*. 2021;129(4):47003. https://doi.org/10.1289/EHP7944
24. Chen C, Chen H, van Donkelaar A, et al. Using parametric g-computation to estimate the effect of long-term exposure to air pollution on mortality risk and simulate the benefits of hypothetical policies: the Canadian community health survey cohort (2005 to 2015). *Environ Health Perspect*. 2023;131(3):37010. https://doi.org/10.1289/EHP11095
25. Chen H, Burnett RT, Copes R, et al. Ambient fine particulate matter and mortality among survivors of myocardial infarction: population-based cohort study. *Environ Health Perspect*. 2016;124(9):1421-1428. https://doi.org/10.1289/EHP185
26. Johnson M, Shin HH, Roberts E, et al. Critical time windows for air pollution exposure and birth weight in a multicity Canadian pregnancy cohort. *Epidemiology*. 2022;33(1):7-16. https://doi.org/10.1097/EDE.0000000000001428
27. Nunez Y, Boehme AK, Weisskopf MG, et al. Fine particle exposure and clinical aggravation in neurodegenerative diseases in New York state. *Environ Health Perspect*. 2021;129(2):27003. https://doi.org/10.1289/EHP7425
28. Yu X, Mostafijur Rahman M, Carter SA, et al. Prenatal air pollution, maternal immune activation, and autism spectrum disorder. *Environ Int*. 2023;179:108148. https://doi.org/10.1016/j.envint.2023.108148
29. Coleman NC, Burnett RT, Higbee JD, et al. Cancer mortality risk, fine particulate air pollution, and smoking in a large, representative cohort of US adults. *Cancer Causes Control*. 2020;31(8):767-776. https://doi.org/10.1007/s10552-020-01317-w
30. Li R, Zhou R, Zhang J. Function of PM2.5 in the pathogenesis of lung cancer and chronic airway inflammatory diseases. *Oncol Lett*. 2018;15(5):7506-7514. https://doi.org/10.3892/ol.2018.8355
31. Demetriou CA, Vineis P. Carcinogenicity of ambient air pollution: use of biomarkers, lessons learnt and future directions. *J Thorac Dis*. 2015;7(1):67-95. https://doi.org/10.3978/j.issn.2072-1439.2014.12.31
32. Beamish LA, Osornio-Vargas AR, Wine E. Air pollution: an environmental factor contributing to intestinal disease. *J Crohns Colitis*. 2011;5(4):279-286. https://doi.org/10.1016/j.crohns.2011.02.017
33. Jacquez GM, Slotnick MJ, Meliker JR, et al. Accuracy of commercially available residential histories for epidemiologic studies. *Am J Epidemiol*. 2011;173(2):236-243. https://doi.org/10.1093/aje/kwq350
34. Tatalovich Z, Stinchcomb DG, Mariotto A, et al. Assessment of interstate residential mobility of SEER patients: SEER and LexisNexis residential address linkage. *J Registry Manag Winter*. 2022;49(4):109-113. https://doi.org/10.1016/j.socscimed.2020.113657
35. Tsui J, Hirsch JA, Bayer FJ, et al. Patterns in geographic access to health care facilities across neighborhoods in the United States based on data from the National Establishment Time-Series between 2000 and 2014. *JAMA Netw Open*. 2020;3(5):e205105. https://doi.org/10.1001/jamanetworkopen.2020.5105
36. Berg CD, Schiller JH, Boffetta P, et al. Air pollution and lung cancer: a review by International Association for the Study of Lung Cancer early detection and screening committee. *J Thorac Oncol*. 2023;18(10):1277-1289. https://doi.org/10.1016/j.jtho.2023.05.024
37. Nightingale C, Bavor C, Stone E, et al. Lung cancer screening: implementation challenges and health equity considerations for the Western Pacific region. *JCO Glob Oncol*. 2023;9:e2200329. https://doi.org/10.1200/GO.22.00329
38. Force USPST, Owens DK, Davidson KW, et al. Screening for pancreatic cancer: US preventive services task Force reaffirmation recommendation statement. *JAMA*. 2019;322(5):438-444. https://doi.org/10.1001/jama.2019.10232

American Journal of Epidemiology, 2025, **194**, 730–737
https://doi.org/10.1093/aje/kwae271
Original Contribution