

## ARTICLE



# The association between fine particulate matter (PM<sub>2.5</sub>) and chronic kidney disease using electronic health record data in urban Minnesota

Lama Ghazi<sup>1✉</sup>, Paul E. Drawz<sup>2</sup> and Jesse D. Berman<sup>3</sup>

The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

**BACKGROUND:** Recent evidence has shown that fine particulate matter (PM<sub>2.5</sub>) may be an important environmental risk factor for chronic kidney disease (CKD), but few studies have examined this association for individual patients using fine spatial data.

**OBJECTIVE:** To investigate the association between PM<sub>2.5</sub> and CKD (estimated glomerular filtration rate [eGFR]<45 ml/min/1.73 m<sup>2</sup>) in the Twin-Cities area in Minnesota using a large electronic health care database (2012–2019).

**METHODS:** We estimated the previous 1-year average PM<sub>2.5</sub> from the first eGFR (measured with the CKD Epidemiology Collaboration equation using the first available creatinine measure during the baseline period [2012–2014]) using Environmental Protection Agency downscaler modeling data at the census tract level. We evaluated the spatial relative risk and clustering of CKD prevalence using a K-function test statistic. We assessed the prevalence ratio of the PM<sub>2.5</sub> association with CKD incidence using a mixed effect Cox model, respectively.

**RESULTS:** Patients ( $n = 20,289$ ) in the fourth (PM<sub>2.5</sub> > 10.4), third (10.3 < PM<sub>2.5</sub> < 10.8) and second quartile (9.9 < PM<sub>2.5</sub> < 10.3) vs. the first quartile (<9.9 μg/m<sup>3</sup>) had a 2.52[2.21, 2.87], 2.18[1.95, 2.45], and 1.72[1.52, 1.97] hazard rate of developing CKD in the fully adjusted models, respectively. We identified spatial heterogeneities and evidence of CKD clustering across our study region, but this spatial variation was accounted for by air pollution and individual covariates.

**SIGNIFICANCE:** Exposure to higher PM<sub>2.5</sub> is associated with a greater risk for incident CKD. Improvements in air quality, specifically at hotspots, may reduce CKD.

**Keywords:** Air pollution; Health studies; Epidemiology

*Journal of Exposure Science & Environmental Epidemiology* (2022) 32:583–589; <https://doi.org/10.1038/s41370-021-00351-3>

## INTRODUCTION

Chronic kidney disease (CKD) is a significant public health problem in the US affecting 14.8% of the US adult population [1]. Common risk factors for CKD include: history of diabetes or hypertension, older age, male sex, black race, and having lower socioeconomic status [2]. In addition, CKD is associated with increased risk of cardiovascular disease, frailty, poor quality of life, premature death, and is costly in the private and public sectors [3]. The risk of adverse events increases with CKD progression [4]. Patients with CKD are also at increased risk of severe illness from SARS-CoV-2 infection [5, 6]. The burden of CKD risk is not equitably distributed with greater prevalence among low-income countries and socio-economically disadvantaged and minority communities in the US [7]. However, even among these groups, it has been observed that CKD risk shows a substantial spatial variation that cannot be explained by individual risk factors, such as incident hypertension or diabetes [8, 9]. This indicates that potentially unmeasured environmental exposures may be attributing to unexplained CKD incidence.

A growing literature has documented that air pollution exposure, notably fine particulate matter (PM<sub>2.5</sub>), maybe a novel risk factor for CKD [10–12]. PM<sub>2.5</sub> can induce systemic-wide effects through mechanisms involving inflammation, oxidation, and direct toxicity from particle constituents [13, 14]. In animal models, inhalation of diesel exhaust has been shown to induce renal oxidative stress, inflammation, and DNA damage, while a 6-year study of Boston hospital stroke patients found that individuals living closer to roadways showed reduced renal function [15, 16]. On a global scale, Bowe et al. estimated that PM<sub>2.5</sub> can be attributed to 17–20% of CKD burden with an estimated 6.95 million annual cases caused by PM<sub>2.5</sub> exposure [8]. In the US, the estimated attributable burden is closer to 4% [8]. Previous studies in the US have looked at the association between PM<sub>2.5</sub> and CKD in specific populations (veterans and established cohorts), but few have examined the general population and the spatial variation in risk of CKD [17, 18].

To gain a better understanding of the PM<sub>2.5</sub> and CKD association at a local and actionable level, we used electronic

<sup>1</sup>Clinical and Translational Research Accelerator, Department of Medicine, Section of Nephrology, Yale University, New Haven, CT, USA. <sup>2</sup>Division of Renal Diseases and Hypertension, University of Minnesota, Minneapolis, MN, USA. <sup>3</sup>Division of Environmental Health Science, School of Public Health, University of Minnesota, Minneapolis, MN, USA. ✉email: lamaghazi@gmail.com

Received: 28 December 2020 Revised: 28 May 2021 Accepted: 3 June 2021  
Published online: 14 June 2021

health record (EHR) data of a health care system servicing the 7-county Twin Cities area in Minnesota. This study aims to use fine-scale environmental data and individual patient cohort data to characterize the association between  $PM_{2.5}$  and CKD prevalence and incidence, as well as evaluate the spatial clustering of CKD cases. Our goal is to add to the current body of literature and estimate the effect of  $PM_{2.5}$  on kidney outcomes.

## METHODS

### Data

Patients were identified from the EHR database of Fairview Health Services, the primary hospital affiliate of the University of Minnesota, from 1 January 2012 to 1 April 2019. The EHR data incorporates all outpatient visits, labs, and billing data, geocoded addresses (latitude and longitude), and census tract of patients' residences [19]. We manually abstracted clinical and laboratory data for 100 random patient charts and compared it with the EHR to ensure data quality. This study was approved by the Institutional Review Board at the University of Minnesota.

### Study population

The population included adult patients ( $\geq 18$  years of age) with a geocoded address from the 7-county Twin Cities area in Minnesota. Included patients' EHR data was restricted to outpatient clinic visits. Patients had at least one measure of outpatient creatinine between 1/1/2012 and 1/1/2014 (baseline period), and at least one primary care physician visit in the Fairview health system at or before the time of index creatinine. We defined index creatinine as the first available creatinine measure during the baseline period. For longitudinal analysis, we included all outpatient creatinines obtained up until 4/1/2019. Patients who opted out of research or moved between 2012 and 2019 were excluded. To assess CKD incidence, we additionally excluded all patients who had CKD during the baseline period and those who had  $< 3$  creatinine measurements (baseline creatinine and two repeat creatinine measures), to ensure adequate measures and patient clinic visits.

### Air pollution exposure

We acquired publicly available daily  $PM_{2.5}$  air pollution data from the Environmental Protection Agency (EPA) downscaler model at the US census tract level ( $\sim 4000$  people) from 2011 to 2014 [20]. Geocoded home address was determined as the latest available address at or before the time of index creatinine. Subsequently, each patient was assigned a census tract based on their geocoded home location and we estimated the 365-day mean  $PM_{2.5}$  concentration prior to their first estimated glomerular filtration rate (eGFR) during the baseline period (2012–2014). A 7- and 30-day mean prior to the date of their first eGFR visit was also estimated to examine short-term exposure.

### Health outcomes

All creatinine measures were traceable to an isotope dilution mass spectrometry reference measurement. Of note, creatinine is a marker of kidney function and is derived from the metabolism of creatine in skeletal muscle and from dietary meat intake. Creatinine is freely filtered across the glomerulus but is not reabsorbed or metabolized by the kidney. According to guidelines, kidney function should be evaluated using eGFR which is calculated using the CKD epidemiology collaboration equation that includes sex, age, creatinine, and race [21–25]. Lower eGFR levels reflect worse kidney function (normal  $eGFR \geq 60$  ml/min/1.73 m<sup>2</sup>). CKD prevalence was defined as having an  $eGFR < 60$  ml/min/1.73 m<sup>2</sup> during the baseline period (1/1/2012–1/1/2014). In the secondary analysis, we used a more restrictive definition of CKD prevalence as an  $eGFR < 45$  ml/min/1.73 m<sup>2</sup> during the baseline period; cutoff reflects moderate to severe kidney disease associated with worse clinical outcomes [4, 22, 25–28]. CKD incidence was assessed between 1/2/2014 and 4/1/2019. CKD incidence was defined using four different criteria: (a) first  $eGFR < 60$ , (b) two consecutive  $eGFR < 60$ , (c) first  $eGFR < 45$ , and (d) two consecutive  $eGFR < 45$  ml/min/1.73 m<sup>2</sup>. Consecutive measures are taken at least one day apart.

### Individual and neighborhood covariates

We identified sex and race from the EHR. Age, smoking status, body mass index, and insurance status as a measure of individual socioeconomic

status (private vs. other [Medicaid or Medicare with no supplemental insurance status]) were defined using the last value before or at the time of the index creatinine. Comorbidities (hypertension, diabetes, cardiovascular disease, stroke, hyperlipidemia, and cancer) were considered present if at least two International Classification of Disease (ICD)–9/10 codes for that condition were present at or before the date of the index creatinine (Supplemental Data). The number of hospitalizations per patient over follow-up was calculated. We assigned neighborhood sociodemographic characteristics of wealth, education, and income for each patient based on census tract of residence using the 2012 American Community Survey (ACS) 5-year Data [2008–2012] [29–31]. Measures included: (1) median value of owner-occupied housing units, (2) percentage of residents over age 25 with a Bachelor's degree or more, and (3) median household income.

Individual covariates that presented  $> 5\%$  missing data, including obesity and insurance, were adjusted with an additional "missingness" variable rather than by imputation, which is limited to the assumption of randomness (missing completely at random) [32, 33].

### Statistical analyses

**Spatial analyses.** We explored spatial patterns of CKD intensity and clustering using the 2012 EHR data for two patient populations: (1) individuals residing in the 5-county Twin Cities area, we included, only 5 of the 7-county areas as these counties had sufficient patient density for analysis and (2) patients residing in a sub-region of the 5-county Twin Cities area encompassing a high density of patients (Supplemental Fig. 1). The targeted sub-region captures 93% of the Twin Cities 5-county area patients. We did not examine the area beyond the 7-county Twin Cities region to avoid spurious spatial results that can arise when patient locations are sparsely distributed near census tract margins.

We evaluated the spatial variation in CKD risk by estimating the spatial intensity of both CKD patients and non-CKD patients (e.g., controls). Spatial intensity represents the number of CKD cases or controls per unit area, which we defined with a Gaussian weighted kernel. We estimated an optimal kernel bandwidth using a cross-validation approach; however, upon consideration of the geographic coverage of the Fairview System (7-county Twin Cities area) and after sensitivity testing of multiple distances (1.0, 1.25, 1.50, 1.75, and 2.5 km), we selected diameter of 2 km for both case and control bandwidths to provide sensible smoothing. Under a null association, we expect constant intensity across our study region. A spatial relative risk was estimated by taking the log ratio of the spatial intensity of CKD cases and the spatial intensity of controls, with a value of zero representing an equal case-control spatial distribution [34, 35]. As a sensitivity test, we estimated a non-log transformed CKD relative risk and performed estimates for the Twin Cities sub-region, and found our distribution of estimates to be very similar.

To evaluate CKD spatial clustering in the 5-county Twin Cities region and sub-region, we applied a K-function test statistic that takes the difference in the K-function for controls from the K-function of CKD cases [36, 37]. A difference in K-function value of zero indicates the null hypothesis of equal spatial risk, a positive departure from 0 indicates clustering of cases more than controls, and a negative value indicates clustering of controls over cases. For our study, clustering is when a CKD patient (or control) is more likely to reside close to another CKD patient (or control), excluding those within the same household. In this study, we evaluated clustering at a distance of up to 10 km to capture most neighborhood-level effects. The significance of clustering was assessed by simulating 95% confidence envelopes using a Monte Carlo random re-assignment of the case and control spatial labels while maintaining the same underlying population at-risk ( $N = 500$  iterations). A calculated p-value represents the overall departure from null [37].

**Cross-sectional analyses.** We used a multilevel model with a random intercept at the census tract level to estimate the  $PM_{2.5}$  association with CKD prevalence from 2012 to 2014. Since CKD prevalence ( $eGFR < 60$  ml/min/1.73 m<sup>2</sup>) was a common outcome, we estimated the prevalence ratio (PR) and 95% confidence interval using Poisson regression with robust error variance [38–40]. We adjusted for demographic (age, race, sex), clinical (obesity, smoking history, history of hypertension, diabetes, cardiovascular disease, hyperlipidemia, and cancer), insurance status, and census tract characteristics of education, and income. Potentially unaccounted for spatial variation was checked by estimating a semivariogram of residuals for each regression model.

**Table 1.** Baseline characteristics of Fairview cohort and of those with and without CKD (eGFR < 60 ml/min/1.73 m<sup>2</sup>) (2012–2014).

	Overall <i>n</i> = 113,725	CKD <i>n</i> = 12,831 (11%)	No CKD <i>n</i> = 100,894 (89%)
Mean PM <sub>2.5</sub> (over 365 days from baseline eGFR), µg/m <sup>3</sup> Median [First Quartile, third Quartile]	10.1 [9.5, 10.7]	10.2 [9.7, 10.7]	10.1 [9.5, 10.6]
eGFR (ml/min/1.73 m <sup>2</sup> ), Median [First Quartile, third Quartile]	86 [72,101]	50 [41,56]	89 [77,104]
<i>Individual-level characteristics</i>			
Age, mean (SD)	50 ± 18	71 ± 15	48 ± 17
Black, <i>n</i> (%)	9399 (8%)	431 (4%)	8968 (9%)
Male, <i>n</i> (%)	50,982 (45%)	5031 (39%)	45,951 (46%)
Obesity (BMI ≥ 30 kg/m <sup>2</sup> ), <i>n</i> (%)	34,004 (35%)	4260 (40%)	29,744 (35%)
Ever smoker, <i>n</i> (%)	45,303 (40%)	5803 (45%)	39,500 (39%)
Private insurance <sup>a</sup> , <i>n</i> (%)	94,083 (88%)	9723 (78%)	84,360 (89%)
<i>Medical history</i>			
Hypertension, <i>n</i> (%)	39,838 (35%)	9382 (73%)	30,456 (30%)
Diabetes, <i>n</i> (%)	12,509 (11%)	3336 (26%)	9173 (9%)
Cardiovascular disease, <i>n</i> (%)	10,206 (9%)	4001 (31%)	6205 (6%)
Stroke, <i>n</i> (%)	2895 (3%)	1069 (8%)	1826 (2%)
Hyperlipidemia, <i>n</i> (%)	41,734 (37%)	8235 (64%)	33,499 (33%)
Cancer, <i>n</i> (%)	5886 (5%)	1523 (12%)	4363 (4%)
<i>Census tract median value of owner occupied housing units</i>			
Q1: <\$165,200	11,189 (10%)	1289 (10%)	9900 (10%)
Q2: \$165,200–188,100	14,800 (13%)	1917 (15%)	12,883 (13%)
Q3: \$188,100–231,300	38,386 (34%)	4497 (35%)	33,889 (34%)
Q4: ≥\$231,300	49,323 (43%)	5124 (40%)	44,199 (44%)
<i>Census tract % of Residents &gt; 25 years with a Bachelor's degree or more</i>			
Q1: <20.4%	13,652 (12%)	1627 (13%)	12,025 (12%)
Q2: 20.4–34.1%	29,394 (26%)	3581 (28%)	25,813 (26%)
Q3: 34.1–48.1%	36,533 (32%)	4152 (32%)	32,381 (32%)
Q4: ≥ 48.1%	34,123 (30%)	3469 (27%)	30,654 (30%)
<i>Census tract median household income</i>			
Q1: <\$35,935	13,489 (12%)	1395 (11%)	12,094(12%)
Q2: \$35,935–47,379	15,216 (13%)	1994 (16%)	13,222 (13%)
Q3: \$47,379–62,343	24,733 (22%)	2969 (23%)	21,764 (22%)
Q4: ≥ \$62,343	60,179 (53%)	6459 (50%)	53,720 (53%)

Cardiovascular disease includes congestive heart failure, acute myocardial infarction, ischemic heart disease, and peripheral vascular disease.

CKD chronic kidney disease defined as eGFR < 60 ml/min/1.73 m<sup>2</sup>, PM particulate matter, eGFR estimated glomerular filtration rate.

<sup>a</sup>Private insurance: not on Medicaid for patients <65-year-old and on Medicare with no supplemental private insurance for patients ≥65 years old.

**Longitudinal analyses.** We used a mixed effect Cox proportional hazards model with a random intercept at the census tract level to estimate the PM<sub>2.5</sub> association with CKD incidence. We adjusted for the same demographic, clinical, and census tract characteristics as the cross-sectional analysis and we adjusted additionally for the number of times the patient was hospitalized over follow-up and the number of eGFRs measures available.

For longitudinal and cross-sectional analyses, we did not adjust for tract wealth due to the high correlation between tract wealth and education. We also examined effect modification of the PM<sub>2.5</sub>—CKD relationship by hypertension, diabetes, sex, age (< vs. ≥65 years), and race (black vs. white). Exposures were performed for continuous (per 1 µg/m<sup>3</sup>) and categorical (per quartile) changes of PM<sub>2.5</sub>. All statistical analyses were conducted using R Statistical Software (version 3.3.1) and Stata [41, 42].

## RESULTS

### Baseline cohort characteristics

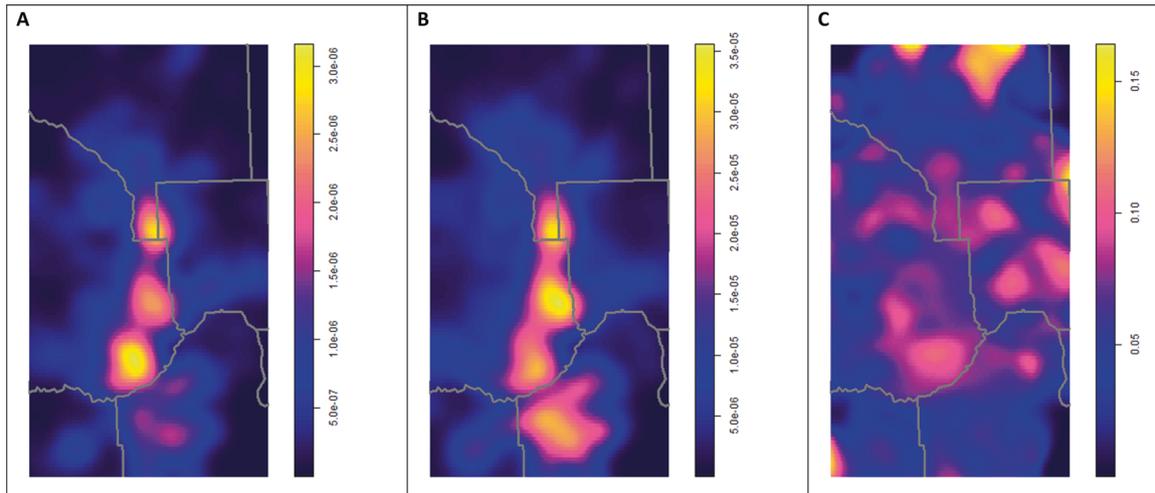
There were 345,042 patients identified from the EHR at baseline. We excluded 94,463 patients due to no primary care physician visit or who moved during the study period. We further excluded

60,399 patients who lived outside the metro area and 2998 patients <18 years at the time of index creatinine. We included 113,725 patients for our baseline analyses.

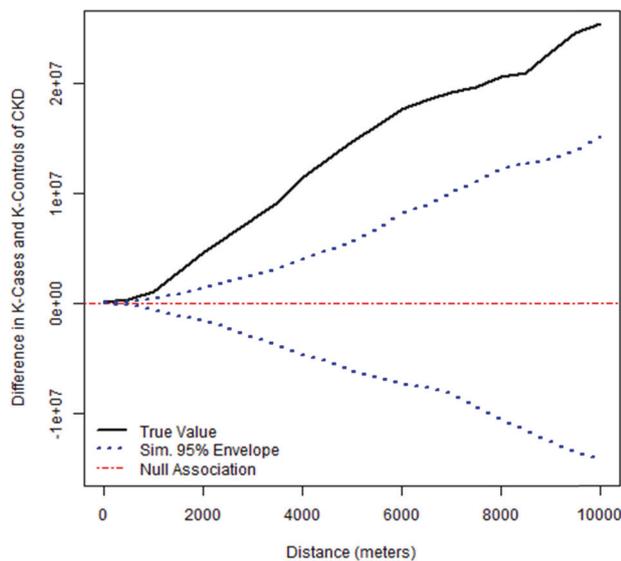
Patients were on average 50 years of age, 8% were black, 45% were males and 11% had CKD (eGFR < 60 ml/min/m<sup>2</sup>) (Table 1). Patients with CKD compared to non-CKD patients were older, less likely to be black and males, more obese, and more likely to have comorbidities, but had similar distribution across quartiles of SES census tract measures. This was true irrespective of whether the eGFR cutoff to define CKD was 60 or 45 ml/min/1.73 m<sup>2</sup> (Table 1, Supplemental Table 1). Tract income was moderately correlated with tract wealth and education ( $r = 0.33$  and  $0.30$ ). However, tract wealth was highly correlated with tract education ( $r = 0.68$ ).

### Spatial analyses

Within the Twin Cities sub-region, we visually interpret spatial heterogeneities in the intensity of both CKD cases and control patients (Fig. 1A, B). As anticipated, areas with high intensity correspond to neighborhoods with dense populations and a greater number of Fairview patients. In a corresponding step, we



**Fig. 1 Spatial intensity of CKD by County in the Twin Cities.** Spatial intensity of (A) CKD cases, (B) non-CKD patients (e.g., controls), and (C) log ratio of CKD case and control intensities (e.g., the spatial relative risk for CKD), using 2012 Fairview Health System records from the Twin Cities subregion, Minnesota. In panel (C), light colors denote regions with greater risk of CKD, while dark colors denote less risk of CKD. Gray lines denote county boundaries found in the inset of Fig. 1. Hotspots of CKD incidence are shown near-certain neighborhoods, such as: Maplewood, North St. Paul, downtown St Paul, Arden Hills, Bloomington, and portions west of Bredesen Park.



**Fig. 2 K-function by CKD status.** The difference in K-function for cases and controls of chronic kidney disease from the Fairview System in the 5-county Twin Cities Area.

estimated spatial variation in CKD risk by taking the log ratio of case and control intensities (Fig. 1C). A heterogeneous CKD risk is observed that is different from cases or control intensities alone. This indicates that elevated risk of CKD is specific to certain geographic areas and is not a function of patient density. Hotspots of CKD incidence are shown near-certain neighborhoods. It should be noted that high relative risk is observed in a few isolated spots near the boundaries of our study region; however, we consider these results to be artifacts of small case and control numbers and do not consider them valid risks. Evidence for spatial clustering of CKD cases compared to non-CKD patients was observed (Fig. 2). Our estimated difference in the K-function line was positive, indicating cases clustering more than controls, and outside the simulated 95% confidence envelopes, providing evidence for statistically significant CKD clustering for all distances up to 10 km ( $p < 0.01$ ). Both Figs. 1 and 2 provide evidence that the risk of CKD

varies significantly across the Twin Cities region and that the risk is not equally distributed with apparent geographic disparities.

#### CKD prevalence

At baseline,  $PM_{2.5}$  quartiles were  $<9.5$ ,  $9.5\text{--}10.1$ ,  $10.1\text{--}10.7$ , and  $\geq 10.7 \mu\text{g}/\text{m}^3$  respectively. There was no cross-sectional association between  $PM_{2.5}$ , for each  $1 \mu\text{g}/\text{m}^3$  and by quartiles, with CKD prevalence ( $eGFR < 60 \text{ ml}/\text{min}/1.73 \text{ m}^2$ ) (Table 2). However, when defining CKD as having an  $eGFR$  of  $<45 \text{ ml}/\text{min}/1.73 \text{ m}^2$ , higher quartiles of  $PM_{2.5}$  were associated with greater odds of CKD compared to the first quartile after adjustment for demographic, clinical, and SES census tract characteristics (Supplemental Table 2). For example, the fourth quartile of  $PM_{2.5}$  was associated with 18% greater odds [1.05, 1.33] of CKD compared to the first quartile after adjustment. Similarly, a  $1 \mu\text{g}/\text{m}^3$  higher annual average  $PM_{2.5}$  was associated with 7% greater odds [1.03, 1.11] of CKD even after adjustment. We found no evidence of effect modification of the  $PM_{2.5}$ -CKD association with hypertension, diabetes, age, or sex.

#### CKD incidence

A total of 20,289 patients had no CKD at baseline and had at least three outpatient creatinine measures following index  $eGFR$ . Of those 3042 patients had incident CKD ( $eGFR < 60 \text{ ml}/\text{min}/1.73 \text{ m}^2$ ) over a median of 8.5 years (Supplemental Table 3). The overall CKD incidence was 2 per 100 person-years (Supplemental Table 4). We observed increased CKD incidence rates for both continuous and categorical measures of baseline  $PM_{2.5}$ . Each  $1 \mu\text{g}/\text{m}^3$  increase in baseline annual average  $PM_{2.5}$  was associated with increased risk of CKD after adjustment for demographic, clinical, and SES characteristics (HR: 1.78 [1.65, 1.89]). Comparing the second, third and fourth quartile of baseline  $PM_{2.5}$  exposure to the first quartile, we observe an HR of CKD incidence of 1.72, 2.15, and 2.49, respectively (Table 3). Similar trends were observed when CKD was defined based on having two consecutive  $eGFR < 60$  or  $<45$  or having an  $eGFR < 45 \text{ ml}/\text{min}/1.73 \text{ m}^2$  (Supplemental Tables 5–7). The magnitude of the association between  $PM_{2.5}$  and CKD incidence was highest when defining CKD incidence as having two consecutive  $eGFR < 45 \text{ ml}/\text{min}/1.73 \text{ m}^2$  and comparing the fourth quartile of  $PM_{2.5}$  to the first quartile, with an HR of 3.83 [2.04, 7.19]. There was no effect modification by hypertension, diabetes, race, age, or sex for any of the models.

**Table 2.** The prevalence ratio (PR) of chronic kidney disease (CKD) associated with annual PM<sub>2.5</sub> using the Fairview Health System cohort (2012–2014)<sup>a</sup>.

	Model 1	Model 2	Model 3	Model 4
PR for each 1 µg/m <sup>3</sup> increase of PM <sub>2.5</sub>	1.11 [1.09, 1.13]	1.02 [0.99, 1.03]	1.01 [0.99, 1.02]	0.99 [0.97, 1.01]
By quartiles of PM <sub>2.5</sub>				
Quartile 1 (<9.5 µg/m <sup>3</sup> )	Reference	Reference	Reference	Reference
Quartile 2 (9.5–10.1 µg/m <sup>3</sup> )	1.37 [1.30, 1.44]	1.08 [1.02, 1.13]	1.05 [1.01, 1.11]	1.02 [0.97, 1.07]
Quartile 3 (10.1–10.7 µg/m <sup>3</sup> )	1.36 [1.29, 1.44]	1.05 [0.99, 1.11]	1.02 [0.98, 1.08]	0.99 [0.95, 1.05]
Quartile 4 (>10.7 µg/m <sup>3</sup> )	1.36 [1.29, 1.44]	1.07 [1.01, 1.12]	1.04 [0.99, 1.09]	1.00 [0.96, 1.06]

Model 1: crude model.

Model 2: age, race, sex, obesity, smoking history, insurance.

Model 3: Model 2 + history of cardiovascular disease, stroke, hyperlipidemia, and cancer, tract % of residents >25 years with a Bachelor's degree or more, tract average household income

Model 4: Model 3 + history of hypertension and diabetes.

No interaction between PM<sub>2.5</sub> (continuous) and hypertension ( $p = 0.60$ ) or diabetes ( $p = 0.11$ ) or race ( $p = 0.09$ ) or age ( $p = 0.98$ ) or sex ( $p = 0.085$ ).

<sup>a</sup>CKD defined as eGFR <60 ml/min/1.73 m<sup>2</sup>. CKD is a common outcome (prevalence = 11%), we, therefore, estimated the prevalence ratio. We used a multilevel Poisson regression model with robust error variance and with a random intercept at the census tract level.

### Sensitivity analysis

Spatial variograms of residuals from the CKD prevalence model revealed no evidence for unexplained spatial dependence in our regression analyses, eliciting confidence in the variance of the risk estimates. It also indicates that spatial patterns of CKD are likely explained by air pollution concentrations and individual or community characteristics. When restricting our spatial evaluation to the Twin Cities sub-region, we observed similar clustering trends of CKD cases compared to the full study area, although the degree of clustering is greater (Supplemental Fig. 2). We did not observe any association between short-term PM<sub>2.5</sub> exposure using 7 and 30-day mean exposures and prevalent CKD.

### DISCUSSION

We found that among patients seen at a healthcare system servicing the 7-county Twin Cities in Minnesota, higher levels of annual PM<sub>2.5</sub> were associated with greater CKD prevalence (eGFR <45 ml/min/1.73 m<sup>2</sup>). Similarly, patients living in census tracts with higher levels of PM<sub>2.5</sub> at baseline were associated with a greater incidence of CKD. We also observed clustering of CKD cases indicating geographic heterogeneity of disease in the 7-county Twin Cities area, but these spatial relationships disappeared after controlling for environmental and sociodemographic factors. These findings demonstrate the presence of neighborhood disparities in CKD and hint that improvements in air quality specifically at CKD hotspots may reduce CKD.

Previous epidemiological studies have shown an association between PM<sub>2.5</sub> and incident CKD. In a large cohort study, Bowe et al. studied the association between annual average PM<sub>2.5</sub> and risk of kidney disease among 2,482,737 veterans over a median of 8.5 years. The authors found, a 10 µg/m<sup>3</sup> increase in PM<sub>2.5</sub> was associated with an increased risk of eGFR <60 ml/min/1.73 m<sup>2</sup> (HR: 1.2 [1.1, 1.3]), incident eGFR decline ≥30% (HR: 1.3[1.2, 1.4]), and incident end-stage renal disease (ESRD) (HR: 1.3 [1.2–1.4]) [11] In a recent study of 10,997 participants from the atherosclerosis risk in communities project, each 1 µg/m<sup>3</sup> increase in annual average PM<sub>2.5</sub> was associated with: (1) greater risk of incident CKD (follow up eGFR <60 ml/min/1.73 m<sup>2</sup> or ≥25% decrease in eGFR relative to baseline); (2) higher CKD related hospitalization or ESRD (HR: 1.1 [1.0, 1.1]); and (3) increased albuminuria (percentage difference: 7% [3%, 11%]) [18] Our results further solidify that PM<sub>2.5</sub> is associated with incident CKD as has been shown in veterans and the ARIC study. Additionally, we observed that the strength of the association between PM<sub>2.5</sub> with CKD was greatest when CKD was defined as having two consecutive eGFR <45 ml/min/1.73 m<sup>2</sup>. This may reflect that air pollution impacts are greatest among patients

with advanced CKD and using two consecutive eGFR measures confirms a severe diagnosis.

Conversely, the association between PM<sub>2.5</sub> and CKD prevalence and eGFR in cross-sectional studies is unclear. Blum et al. [18] found no significant association between baseline PM<sub>2.5</sub> and eGFR<sup>16</sup>, which differed from our results indicating an association between annual PM<sub>2.5</sub> and prevalent CKD, eGFR <45 ml/min/1.73 m<sup>2</sup> (18% greater odds). Two previous cross-sectional studies looking at the Medicare population and veterans also showed an association between PM<sub>2.5</sub> and CKD prevalence. Among 669 older veteran men living in the Boston area, one year PM<sub>2.5</sub> was associated with lower eGFR [12]. A 2.1 interquartile range higher 1 year PM<sub>2.5</sub> was associated with a 1.9 ml/min/1.73 m<sup>2</sup> [−3, −0.8] lower eGFR. Similarly, among 1.1 million persons from the 2010 Medicare sample, every 4 µg/m<sup>3</sup> increase in PM<sub>2.5</sub> was associated with a 3% [1.02, 1.05] greater prevalence of CKD (ICD9 codes). These differences could be due to the underlying population assessed (age, race, clinical comorbidities).

Findings from our study are consistent with those reported worldwide as far as greater risk of CKD as PM<sub>2.5</sub> concentrations increase. In China, a national renal biopsy study found that each 10 µg/m<sup>3</sup> increase in annual PM<sub>2.5</sub> concentration was associated with 14% greater odds [1.1, 1.2] of a patient developing membranous nephropathy in high PM<sub>2.5</sub> regions (>70 µg/m<sup>3</sup> annual average) [43]. This association was attenuated in areas where PM<sub>2.5</sub> levels were <60 µg/m<sup>3</sup> (OR: 1.0[0.99, 1.0]). In Taiwan, a study among the general population showed that a 10 µg/m<sup>3</sup> increase in PM<sub>2.5</sub> was associated with a 6% higher risk [1.0, 1.1] of developing CKD (eGFR < 60 ml/min/1.73 m<sup>2</sup>). It is important to note that, PM<sub>2.5</sub> concentrations vary greatly between countries. For instance, the annual average of PM<sub>2.5</sub> concentrations in China and India exceed 50 µg/m<sup>3</sup> compared to the US, UK, and Japan where the levels are between 10 and 14 µg/m<sup>3</sup> [44]. Therefore, it would be of value not only to further understand the association of PM<sub>2.5</sub> and CKD but also to identify clusters and areas where this association is more apparent. In the 7-county Twin Cities area, a closer look at the military and industrial facilities is needed to identify if PM<sub>2.5</sub> is contributing to this clustering, especially since CKD risk is greatest in these historically industrial areas.

The mechanism by which PM<sub>2.5</sub> affects the kidneys is unclear. Several hypotheses have been proposed [45]. PM<sub>2.5</sub> can trigger autonomic system imbalance and proinflammatory responses and can activate endothelial dysfunction and arterial vasoconstriction. PM<sub>2.5</sub> may also increase blood pressure and decrease nitrous oxide excretion and can induce further insulin resistance. All these possible pathways could in turn cause vascular damage, glomerulosclerosis, tubulointerstitial damage, and intraglomerular

**Table 3.** Hazards rate and 95% CI of incident CKD (first eGFR <60 ml/min/1.73 m<sup>2</sup>) associated with annual PM<sub>2.5</sub> using the Fairview Health System cohort (2012–2019).

	Model 1 (HR, 95% CI)	Model 2 (HR, 95% CI)	Model 3 (HR, 95% CI)	Model 4 (HR, 95% CI)
For each 1 µg/m <sup>3</sup> increment of PM <sub>2.5</sub>	1.70 [1.62, 1.79]	1.78 [1.68, 1.89]	1.77 [1.67, 1.88]	1.78 [1.67, 1.89]
By quartiles of PM <sub>2.5</sub>				
Quartile 1 (<9.9 µg/m <sup>3</sup> ) n = 6785	Reference	Reference	Reference	Reference
Quartile 2 (9.9–10.3 µg/m <sup>3</sup> ) n = 3,865	1.74 [1.55, 1.95]	1.75 [1.54, 1.99]	1.73 [1.53, 1.97]	1.72 [1.51, 1.95]
Quartile 3 (10.3–10.8 µg/m <sup>3</sup> ) n = 5,966	2.06 [1.86, 2.28]	2.08 [1.87, 2.33]	2.16 [1.93, 2.42]	2.15 [1.92, 2.41]
Quartile 4 (>10.8 µg/m <sup>3</sup> ) n = 4002	2.39 [2.15, 2.67]	2.56 [2.27, 2.89]	2.51 [2.21, 2.85]	2.49 [2.20, 2.82]

Model 1: crude model.

Model 2: age, race, sex, obesity, smoking history, insurance.

Model 3: Model 2 + history of cardiovascular disease, stroke, hyperlipidemia, and cancer, tract % of residents >25 years with a Bachelor's degree or more, tract average household income.

Model 4: Model 3 + history of hypertension and diabetes.

No interaction with hypertension ( $p = 0.27$ ) and diabetes ( $p = 0.83$ ) or race ( $p = 0.52$ ) or age ( $p = 0.62$ ) or sex ( $p = 0.53$ ).

hypertension which could lead to CKD and ESRD. We observed in models including both patient hypertension and diabetes (model 4), a slight attenuation of the magnitude of the association in all our analyses, indicating potential mediation. Future work should focus on discerning the exact mechanism through which air pollution affects CKD and the role of co-morbidities.

Our study has several limitations. First, we studied a single healthcare system in the 7-county Twin Cities which may not be representative of the general population and other states. However, the Fairview health system covers 161 zipcodes, nine counties, and represents a population of almost 2.7 million people, therefore providing a good representation of most of Minnesota. Second, we used an indicator variable to account for missingness for obesity (9% missing) and insurance (5% missing) [32, 33]. Third, we have a relatively small cohort of patients assessed for CKD incidence and that contributes to the wide CI. However, our results were consistent across all definitions of CKD incidence, indicating a toxicological phenomenon. Fourth, we were unable to measure the association of PM<sub>2.5</sub> with albuminuria since most patients in our cohort did not have urine albumin assessed. Fifth, we only looked at one source of modeled PM<sub>2.5</sub> air pollution, but the selection of the EPA Downscaler model was critical to support the fine spatial and daily temporal resolution needed for our geographic study area. Furthermore, the downscaler model has been shown to provide a better prediction performance compared to both modeled (e.g., CMAQ) and interpolated (e.g., kriging) surfaces [46]. Sixth, by including patients with >3 creatinine measures to study the association of PM<sub>2.5</sub> with incident CKD we are subject to selection bias. Finally, we used ICD codes to determine comorbidities. In order to minimize misclassification bias by ICD codes, we used at least two ICD codes to confirm diagnoses. We also adjudicated 100 patient charts and found that ICD code diagnoses were consistent with clinic notes (*data not shown*). Finally, our study included individual and neighborhood characteristics at baseline and did not have the capacity to longitudinally evaluate changes over time. Strengths of our study include using an EHR data set from routine clinical practice that, unlike structured cohort studies, reflects the population of this healthcare system and the population it is serving and assessing spatial clustering in the Twin Cities in addition to accounting for clustering by tracts in the regression analyses.

In conclusion, we found spatial clustering of CKD cases in the Twin Cities region and identified that higher levels of annual PM<sub>2.5</sub> were associated with increased CKD prevalence and incidence. Future research should focus on identifying areas of CKD case clusters and determine whether efforts to improve air quality in these hotspots will reduce the burden of disease. Exploring the association between CKD hotspots with current and historical industrial sites may further reveal relationships between local

hazards and kidney-related morbidity. While this has global importance, efforts should focus on settings worldwide where elevated air pollution levels put populations at increased risk [47].

## REFERENCES

- Murphy D, McCulloch CE, Lin F, Banerjee T, Bragg-Gresham JL, Eberhardt MS, et al. Trends in prevalence of chronic kidney disease in the United States. *Ann Intern Med.* 2016;165:473–81.
- Kazancioğlu R. Risk factors for chronic kidney disease: an update. *Kidney Int Suppl.* 2013;3:368–71. (2011)
- Bello AK, Alrukhaimi M, Ashuntantang GE, Basnet S, Rotter RC, Douthat WG, et al. Complications of chronic kidney disease: current state, knowledge gaps, and strategy for action. *Kidney Int Suppl.* 2017;7:122–9. (2011)
- Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS, de Jong PE, et al. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. *Lancet.* 2010;375:2073–81.
- Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int.* 2020; 97:829–38.
- <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html>. Accessed 28 Nov 2020.
- Bowe B, Xie Y, Yan Y, Al-Aly Z. Burden of cause-specific mortality associated with PM<sub>2.5</sub> air pollution in the United States. *JAMA Netw Open.* 2019;2:e1915834.
- Bowe B, Xie Y, Li T, Yan Y, Xian H, Al-Aly Z. Estimates of the 2016 global burden of kidney disease attributable to ambient fine particulate matter air pollution. *BMJ Open.* 2019;9:e022450.
- Bowe B, Xie Y, Xian H, Lian M, Al-Aly Z. Geographic variation and US county characteristics associated with rapid kidney function decline. *Kidney Int Rep.* 2017;2:5–17.
- Bragg-Gresham J, Morgenstern H, McClellan W, Saydah S, Pavkov M, Williams D, et al. County-level air quality and the prevalence of diagnosed chronic kidney disease in the US Medicare population. *PLoS ONE.* 2018;13:e0200612.
- Bowe B, Xie Y, Li T, Yan Y, Xian H, Al-Aly Z. Particulate matter air pollution and the risk of incident CKD and progression to ESRD. *J Am Soc Nephrol.* 2018;29:218–30.
- Mehta AJ, Zanobetti A, Bind MA, Kloog I, Koutrakis P, Sparrow D, et al. Long-term exposure to ambient fine particulate matter and renal function in older men: the veterans administration normative aging study. *Environ Health Perspect.* 2016; 124:1353–60.
- Hamanaka RB, Mutlu GM. Particulate matter air pollution: effects on the cardiovascular system. *Front Endocrinol.* 2018;9:680.
- Risom L, Møller P, Loft S. Oxidative stress-induced DNA damage by particulate air pollution. *Mutat Res.* 2005;592:119–37.
- Nemmar A, Karaca T, Beegam S, Yuvaraju P, Yasin J, Hamadi NK, et al. Prolonged pulmonary exposure to diesel exhaust particles exacerbates renal oxidative stress, inflammation and DNA damage in mice with adenine-induced chronic renal failure. *Cell Physiol Biochem.* 2016;38:1703–13.
- Lue SH, Wellenius GA, Wilker EH, Mostofsky E, Mittleman MA. Residential proximity to major roadways and renal function. *J Epidemiol Community Health.* 2013;67:629–34.

17. Bowe B, Xie Y, Li T, Yan Y, Xian H, Al-Aly Z. Associations of ambient coarse particulate matter, nitrogen dioxide, and carbon monoxide with the risk of kidney disease: a cohort study. *Lancet Planet Health*. 2017;1:e267–e76.
18. Blum MF, Surapaneni A, Stewart JD, Liao D, Yanosky JD, Whitsel EA, et al. Particulate matter and albuminuria, glomerular filtration rate, and incident CKD. *Clin J Am Soc Nephrol*. 2020;15:311–9.
19. Clinical Data Repository. <https://www.ctsi.umn.edu/researcher-resources/clinical-data-repository> Accessed 6 Aug 2018.
20. Centers for Disease Control and Prevention. National Environmental Public Health Tracking Network. Web. Accessed 28 Oct 2020. [www.cdc.gov/ephracking](http://www.cdc.gov/ephracking).
21. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150:604–12.
22. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Suppl*. 2013;3:1–150.
23. Kidney Health Australia (KHA): chronic kidney disease (CKD) management in general practice—guidelines and clinical tips to help identify, manage and refer patients with CKD in your practice, 3rd ed. (2015). <https://kidney.org.au/health-professionals/ckd-management-handbook>.
24. NICE: Clinical guideline on chronic kidney disease in adults—assessment and management (2014, updated 2015). <https://www.nice.org.uk/guidance/cg182>.
25. Delanaye P, Schaeffner E, Ebert N, Cavalier E, Mariat C, Krzesinski JM, et al. Normal reference values for glomerular filtration rate: what do we really know? *Nephrol Dial Transpl*. 2012;27:2664–72.
26. Turin TC, Tonelli M, Manns BJ, Ravani P, Ahmed SB, Hemmelgarn BR. Chronic kidney disease and life expectancy. *Nephrol Dial Transpl*. 2012;27:3182–6.
27. Fox CS, Matsushita K, Woodward M, Bilo HJ, Chalmers J, Heerspink HJ, et al. Associations of kidney disease measures with mortality and end-stage renal disease in individuals with and without diabetes: a meta-analysis. *Lancet*. 2012;380:1662–73.
28. Mahmoodi BK, Matsushita K, Woodward M, Blankestijn PJ, Cirillo M, Ohkubo T, et al. Associations of kidney disease measures with mortality and end-stage renal disease in individuals with and without hypertension: a meta-analysis. *Lancet*. 2012;380:1649–61.
29. Steven Manson JS, Steven Ruggles DVR. IPUMS national historical geographic information system: Version 13.0 [Database]. Minneapolis: University of Minnesota. 2018. <https://doi.org/10.18128/D050.V13.0>.
30. US Census Bureau. Available online at <http://www.census.gov/>. Accessed 17 Oct 2019.
31. Oakes JMKJ, editors. Chapter two: the measurement of socioeconomic status. *Methods in social epidemiology*. 2nd ed. John Wiley & Sons, Incorporated, 2017. pp 23–42.
32. Chow W. A look at various estimators in logistic models in the presence of missing values. Santa Monica CRC, 417–20.
33. Anderson AB, HDM darotll RP, Wright JD, Anderson AB (eds). *Handbook of survey research*. New York: Academic Press; 1983. p. 415–92.
34. Kelsall JE, Diggle PJ. Non-parametric estimation of spatial variation in relative risk. *Stat Med*. 1995;14:2335–42.
35. Hazelton ML. Testing for changes in spatial relative risk. *Stat Med*. 2017;36:2735–49.
36. Diggle PJ, Gómez-Rubio V, Brown PE, Chetwynd AG, Gooding S. Second-order analysis of inhomogeneous spatial point processes using case-control data. *Biometrics*. 2007;63:550–7.
37. Bivand RS, Pebesma E, Gómez-Rubio V. *Applied spatial data analysis with R*. Springer, 2008.
38. McNutt LA, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol*. 2003;157:940–3.
39. Zou G. A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol*. 2004;159:702–6.
40. Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. *Am J Epidemiol*. 2004;160:301–5.
41. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing V, Austria. 2016. <https://www.R-project.org/>.
42. StataCorp. 2017. *Stata statistical software: release 15*. College Station TSL.
43. Xu X, Wang G, Chen N, Lu T, Nie S, Xu G, et al. Long-term exposure to air pollution and increased risk of membranous nephropathy in China. *J Am Soc Nephrol*. 2016;27:3739–46.
44. Ambient (outdoor) air pollution in cities database 2014. World Health Organization [http://www.who.int/phe/health\\_topics/outdoorair/databases/cities-2014/en/](http://www.who.int/phe/health_topics/outdoorair/databases/cities-2014/en/); 2014.
45. Afsar B, Elsurur Afsar R, Kanbay A, Covic A, Ortiz A, Kanbay M. Air pollution and kidney disease: review of current evidence. *Clin Kidney J*. 2019;12:19–32.
46. Berrocal VJ, Gelfand AE, Holland DM. Space-time data fusion under error in computer model output: an application to modeling air quality. *Biometrics*. 2012;68:837–48.
47. Yang X, Jiang L, Zhao W, Xiong Q, Yan X. Comparison of ground-based PM. *Int J Environ Res Public Health*. 2018;15.

## ACKNOWLEDGEMENTS

This research was supported by the National Institutes of Health's National Center for Advancing Translational Sciences, grant UL1TR002494. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health's National Center for Advancing Translational Sciences.

## CONFLICT OF INTEREST

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41370-021-00351-3>.

**Correspondence** and requests for materials should be addressed to L.G.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.