Fine particulate matter components and interstitial lung disease in rheumatoid arthritis

Naizhuo Zhao [1], Ziyad Al-Aly [2,3], Boyang Zheng [4], Aaron van Donkelaar [5,6], Randall V. Martin [5,6], Christian A. Pineau [4,7] and Sasha Bernatsky [4,7]

[1]Division of Clinical Epidemiology, McGill University Health Centre, Montreal, QC, Canada. [2]Clinical Epidemiology Center, Research and Development Service, VA St Louis Health Care System, St Louis, MO, USA. [3]Dept of Medicine, Washington University in St Louis, St Louis, MO, USA. [4]Division of Rheumatology, McGill University Health Centre, Montreal, QC, Canada. [5]Dept of Energy, Environmental and Chemical Engineering, Washington University in St Louis, St Louis, MO, USA. [6]Dept of Physics and Atmospheric Science, Dalhousie University, Halifax, NS, Canada. [7]Dept of Medicine, McGill University, Montreal, QC, Canada.

Corresponding author: Sasha Bernatsky (sasha.bernatsky@mcgill.ca)

Shareable abstract (@ERSpublications)
Interstitial lung disease (ILD) in rheumatoid arthritis is associated with long-term exposure to ambient fine particulate matter ($PM_{2.5}$). Among the major $PM_{2.5}$ chemical components, ammonium contributed the most to ILD risk. https://bit.ly/3DzWuTG

Cite this article as: Zhao N, Al-Aly Z, Zheng B, *et al*. Fine particulate matter components and interstitial lung disease in rheumatoid arthritis. *Eur Respir J* 2022; 60: 2102149 [DOI: 10.1183/13993003.02149-2021].

Received: 3 Aug 2021
Accepted: 28 Nov 2021

Abstract

Background Exposure to ambient fine particulate matter with an aerodynamic diameter <2.5 $\mu g \cdot m^{-3}$ ($PM_{2.5}$) is a risk factor for pulmonary and systemic autoimmune diseases; however, evidence on which $PM_{2.5}$ chemical components are more harmful is still scant. Our goal is to investigate potential associations between major $PM_{2.5}$ components and interstitial lung disease (ILD) onset in rheumatoid arthritis (RA).

Methods New-onset RA subjects identified from a US healthcare insurance database (MarketScan) were followed for new onset of RA-associated ILD (RA-ILD) from 2011 to 2018. Annual concentrations of ambient $PM_{2.5}$ chemical components (*i.e.* sulfate, nitrate, ammonium, organic matter, black carbon, mineral dust and sea salt) were estimated by combining satellite retrievals with chemical transport modelling and refined by geographically weighted regression. Exposures from 2006 up to 1 year before ILD onset or end of study were assigned to subjects based on their core-based statistical area or metropolitan division codes. A novel time-to-event quantile-based g (generalised)-computation approach was used to estimate potential associations between RA-ILD onset and the exposure mixture of all seven $PM_{2.5}$ chemical components adjusting for age, sex and prior chronic obstructive pulmonary disease (as a proxy for smoking).

Results We followed 280 516 new-onset RA patients and detected 2194 RA-ILD cases across 1 394 385 person-years. The adjusted hazard ratio for RA-ILD onset was 1.54 (95% CI 1.47–1.63) per every decile increase in all seven exposures. Ammonium, mineral dust and black carbon contributed more to ILD risk than the other $PM_{2.5}$ components.

Conclusion Exposure to components of $PM_{2.5}$, particularly ammonium, increases ILD risk in RA.

Introduction

Rheumatoid arthritis (RA) is a potentially disabling systemic autoimmune disorder that affects up to 80 million people worldwide [1]. Interstitial lung disease (ILD) is a severe extra-articular RA manifestation that contributes greatly to morbidity and mortality [2, 3]. Although it is increasingly recognised that fine particulate matter with an aerodynamic diameter <2.5 $\mu g \cdot m^{-3}$ ($PM_{2.5}$) in air pollution is an environmental risk factor associated with some pulmonary and systemic autoimmune diseases [4, 5], knowledge about a potential association between ambient $PM_{2.5}$ and ILD in RA (RA-ILD) is scant. Ambient $PM_{2.5}$ is composed of different chemical components (*e.g.* organic matter, black carbon, mineral dust and mineral salt) [6]. A few studies suggested that individual chemical components of $PM_{2.5}$ may be more closely related to adverse health effects than aggregated $PM_{2.5}$ [7, 8]. Ambient air pollution has been linked to

subclinical ILD in the non-RA setting [9]. However, for many diseases, including RA-ILD, it remains unclear which $PM_{2.5}$ chemical components are the most harmful.

People are usually exposed to multiple air pollutants and chemical components simultaneously [10]. Concentrations of these different air pollutants or chemical constituents are often correlated in space, since they share common sources (*e.g.* industries and road traffic) [11]. Given this correlation, studies of the health effects of multiple $PM_{2.5}$ chemical components do not lend themselves well to a common parametric regression approach (*e.g.* a multi-exposure Cox proportional hazards model) due to the potential problem of collinearity.

g-computation belongs to the g-method family of "generalised" models which provide consistent estimates of contrasts (*e.g.* differences and ratios) of average potential outcomes under a less restrictive set of identification conditions than standard parametric regression methods. Quantile-based g-computation [10] represents a novel way to investigate the joint health effects of multiple exposures. Compared with other methods for assessing joint effects of multiple inter-correlated exposures, *e.g.* Bayesian kernel machine regression [12] and weighted quantile sum (WQS) regression [13], quantile-based g-computation is much more efficient, which is beneficial in analysing datasets with a large number of subjects and multiple exposure variables and covariates [10]. More importantly, current Bayesian kernel machine regression and WQS regression methods apply only to cross-sectional studies. By contrast, quantile-based g-computation can be used to fit time-to-event models to estimate marginal hazard ratios (HRs) for exposure mixtures. Therefore, in this study we used quantile-based g-computation in conjunction with conventional Cox proportional hazards models to investigate associations between RA-ILD onset and long-term exposure to a mixture of $PM_{2.5}$ chemical components, based on a general population cohort using administrative health data from the USA.

Methods

Study cohort

Our analyses were based on the Truven Health MarketScan Commercial Claims Database, which curates nonnominal longitudinal health claims (related to physician visits, emergency room encounters, hospitalisations and prescription drug dispensations) from all US healthcare insurance agencies willing to provide data. MarketScan has been used to estimate the prevalence of RA and ILD-RA [14]. The database records at least one diagnosis code for each physician billing claim and up to four diagnostic codes for each hospitalisation, coded using the International Classification of Diseases, Ninth/Tenth Revision, Clinical Modifications (ICD-9/10-CM) classification system. Besides standard demographic variables (*e.g.* age and sex), the database contains a geographic code for each enrolee, as either a core-based statistical area (CBSA) or metropolitan division code. A CBSA is a US geographic area that consists of one or more counties with an urban centre of at least 10 000 people. A metropolitan division is constituted by subdividing one of the 11 largest metropolitan statistical areas (*e.g.* Chicago–Naperville–Elgin, Los Angeles–Long Beach–Anaheim and New York–Newark–Jersey City). In the MarketScan database, people living in rural areas (*i.e.* not residing in a CBSA or metropolitan division) have no indicator of their geographic location and thus could not be included in our analyses.

New-onset RA subjects, defined by at least two physician billing claims with an RA diagnostic code (*i.e.* ICD-10 M05 or M06) within 2 years or at least one relevant hospitalisation diagnostic code, were identified in the MarketScan database from 1 January 2011 to 31 December 2018. Subjects were followed until ILD onset (at least two physician claims with relevant diagnostic codes of ICD-10 M05.1 or J84.1, at least 1 month apart), death or end of study (*i.e.* 31 December 2018). Subjects with incident ILD before RA diagnosis were excluded. Additionally, we adjusted for ICD codes (ICD-10 J41–J44.x) related to chronic obstructive pulmonary disease (COPD) for two reasons. First, COPD is a potential confounder in our analyses since it is associated with air pollution and because COPD may sometimes be mistaken for ILD within administrative coding [15]. Second, smoking is associated with ILD [16], and could be an effect modifier in the relationship between ILD and air pollution. However, smoking status is limited in MarketScan Commercial Claims data; instead, COPD has been used by investigators as a proxy for smoking, given COPD's strong association with tobacco use [17].

Exposure variables

The annual $PM_{2.5}$ chemical composition products (V4.NA.03 version) for 2006 to 2017 were obtained from the Atmospheric Composition Analysis Group (https://sites.wustl.edu/acag/datasets/surface-pm2–5). Concentrations of overall $PM_{2.5}$, sulfate, nitrate, ammonium, black carbon, organic matter, mineral dust and sea salt in the products were estimated by combining satellite retrievals of aerosol optical depth with GEOS-Chem chemical transport model calculations to relate aerosol optical depth with $PM_{2.5}$ composition.

Ground-based measurements were then incorporated using a geographically weighted regression to provide a spatially continuous dataset at approximately a 1 km×1 km resolution over North America [18]. In more detail, the geographically weighted regression was used to predict the bias of $PM_{2.5}$ and its components compared with ground-based measurements using predictor variables related to simulated composition, as well as land surface cover and local elevation features [18]. The overall cross-validated agreement (coefficients of variation) of the resultant all-composition $PM_{2.5}$, sulfate, nitrate, ammonium, black carbon, organic matter, mineral dust and sea salt to the *in situ* measurements for 2000–2016 are 0.70, 0.96, 0.86, 0.90, 0.59, 0.57, 0.60 and 0.80, respectively, over North America [18]. The average ambient $PM_{2.5}$ and its chemical composition concentrations from 2006 to 1 year before ILD onset or end of study for each CBSA or metropolitan division were calculated based on the gridded $PM_{2.5}$ chemical composition products and assigned to each subject. The spatial variations of overall $PM_{2.5}$ and its seven components for 2006 are shown in supplementary figure S1. Similar spatial variations of the air pollutants can be seen for the other 11 years (*i.e.* 2007–2017).

Similar to others [19, 20], we considered the role of ozone, another common ambient air pollutant, in our models. Ozone is a powerful oxidant and toxic air pollutant [21]. County-level daily ambient ozone concentrations were retrieved from the Centers for Disease Control and Prevention (www.data.cdc.gov). These ground-based daily measurements were averaged for each year between 2006 and 2017, aggregated to each CBSA or metropolitan division area, and assigned to each subject as was done for $PM_{2.5}$ and its chemical components.

Statistical methods

First, the risk of RA-ILD onset after RA diagnosis was assessed using single-exposure Cox proportional hazards models for overall $PM_{2.5}$, the seven $PM_{2.5}$ chemical components and ozone exposures separately, adjusting for age (in years), sex and coexistence of baseline COPD (at time of RA). Many previous studies have demonstrated that in North America (where $PM_{2.5}$ and its chemical composition levels are relatively low), the relationships between the $PM_{2.5}$ or its chemical composition concentration and hazard ratios are nearly linear [20, 22]. Hence, we did not categorise the continuous exposure variables to avoid unnecessary loss of information [23].

Next, we used quantile-based g-computation to estimate the marginal hazard ratio and 95% confidence interval for the exposure mixture of the seven individual $PM_{2.5}$ chemical component exposures, adjusting for the same covariates as those in the Cox proportional hazards models. The quantile-based g-computation approach was developed by combining WQS [13] and g-computation [24]. To address potential collinearity, the WQS method transforms each continuous exposure of interest (X) into an ordinal variable (X^q) and combines the ordinal exposure variables into a mixed effect index (S) to estimate the overall effect of increasing each exposure by one quantile using equation 1:

$$S = \sum_{i=1}^{n} w_i X_i^q \tag{1}$$

in which i denotes an exposure, n represents the number of exposures of interest, q is the number of quantiles of each exposure variable (10 in this study) and w is the weight for an exposure. All weights are forced to sum to 1 and have the same sign or be equal to 0. Considering that nonlinear effects of exposure variables would be examined in this study (see the following two paragraphs), we selected a relatively large value of q (*i.e.* 10) as per the suggestion of the developers of the quantile-based g-computation approach [25]. Under the directional homogeneity of the weights, the WQS regression model is expressed by equation 2:

$$Y = \beta_0 + \psi S + \beta_1^{\mathrm{T}} Z \tag{2}$$

where Y denotes the health outcome (*i.e.* the binary outcome of ILD onset in this study), β_0 is the model intercept, Z is a vector of potential confounders or effect modifiers (*i.e.* age, sex and coexistence of COPD in this study), β_1 represents the coefficient vector of the covariates and ψ is the coefficient of the mixed effect index. The coefficients of each ordinal exposure, w, in the index (usually called index weights) are obtained by the maximising likelihood method and are used to quantify the magnitudes of effects of individual exposures on the health outcome [13]. Although WQS regression has been widely applied (*e.g.* [26–28]), the assumption of directional homogeneity in WQS may lead to estimation biases and lack of convergence [10].

When the sample size is large, the WQS regression can be treated as a generalised linear model [13]. Variables in a generalised linear model do not need to adhere to directional homogeneity and generalised linear regression is often used to assess the effects of complex exposures in observational datasets [29]. g-computation (or g-formula) is usually used to estimate causal effects and can be fitted by a generalised linear model [30]. If the directional homogeneity assumption holds true, a quantile-based g-computation model is equivalent to a WQS model; otherwise, the coefficient of the mixed effect index (*i.e.* logarithm of the odds ratio or hazard ratio of the exposure mixture regarding the outcome) is estimated by the standard g-computation algorithm. Thus, quantile-based g-computation can be treated as a generalisation and extension of WQS, which eliminates the restriction of directional homogeneity [10]. Similar to the WQS regression, the index weights that are generated in quantile-based g-computation provide an estimation of the relative magnitude of associations regarding individual exposures and the outcome. However, this holds only if associations are in the same (positive or negative) direction. The index weights may go in either direction, suggesting that some exposures may have a positive association, while others a negative association, with the studied outcome.

Quantile-based g-computation uses Cox proportional hazards as the underlying model for time-to-event analysis to yield estimates of the effect of increasing all exposures by one quantile. Quantile-based g-computation can be extended to consider potential nonlinear effects of variables. In our preliminary assessment, we tried to use spline functions to model the nonlinear effects for each of the exposure variables and the mixed effect index was developed by equation 3:

$$S = \sum_{i=1}^{n} w_i f(X_i^q) \tag{3}$$

where $f(\cdot)$ is a spline function with a degree of 2 [25]. Comparing the Akaike Information Criterion (AIC) values [10], we found that the simple linear combination of the exposure variables (*i.e.* equation 1) provided a better fit (*i.e.* with a lower AIC) and avoided overfitting, compared with the use of the spline nonlinear function. Thus, we selected the linear additive strategy to structure the mixture effect index in the quantile-based g-computation model. Mathematical formulation details on the quantile-based g-computation approach have been elaborately demonstrated by Keil *et al.* [10]. Our specific quantile-based g-computation model was fitted by the "qgcomp" package in R version 4.0.4 (R Foundation for Statistical Computing, Vienna, Austria). The code of fitting a quantile-based g-computation model for the primary analysis is provided in the supplementary material.

We repeated the aforementioned quantile-based g-computation analysis by adding ozone as an additional exposure variable. We also conducted a sensitivity analysis of the Cox proportional hazards models and the quantile-based g-computation models, excluding subjects with COPD (again, as a means of addressing heavy tobacco use, which was unmeasured in our dataset, and outcome ascertainment error related to COPD, which can mimic ILD). We conducted more sensitivity analyses of the quantile-based g-computation models that stratified subjects by sex and age (*i.e.* ⩽52 and >52 years old; 52 years being the mean age of the first ILD diagnosis of the RA patients).

Given that our follow-up period was relatively short (8 years) and ambient $PM_{2.5}$ levels were relatively stable in the contiguous USA over the period [18], we did not adopt time-varying exposures, so that the quantile-based g-computation models were more efficiently fitted. To explore whether this simplified assignment of exposure variables might generate large measurement biases, we performed Kendall's τ-tests in an attempt to detect any calendar-year trends in any of the $PM_{2.5}$ components in any CBSA or metropolitan division over time.

Results

We identified 280 516 new-onset RA patients (75.6% female) from 401 different CBSAs or metropolitan divisions. The distribution of the 280 516 RA patients is shown by figure 1. The mean±SD area of the 401 CBSAs or metropolitan divisions was 2636±3588 km^2. Among the RA patients, 2194 (74.3% female) developed ILD over a median (interquartile range) follow-up of 0.48 (1.17) years for a total of 1 394 385 person-years (incidence of 3.2 cases per 1000 patient-years). The mean±SD age at RA onset was 50.3±11.0 years. RA patients were exposed to overall $PM_{2.5}$ concentrations ranging from 3.0 to 12.4 $\mu g \cdot m^{-3}$. Detailed participant characteristics and exposures for subgroups of each covariate are shown in table 1. The distribution of population-weighted $PM_{2.5}$, $PM_{2.5}$ components and ozone exposures is shown in table 2. Concentrations of the $PM_{2.5}$ components were significantly inter-correlated; in particular,

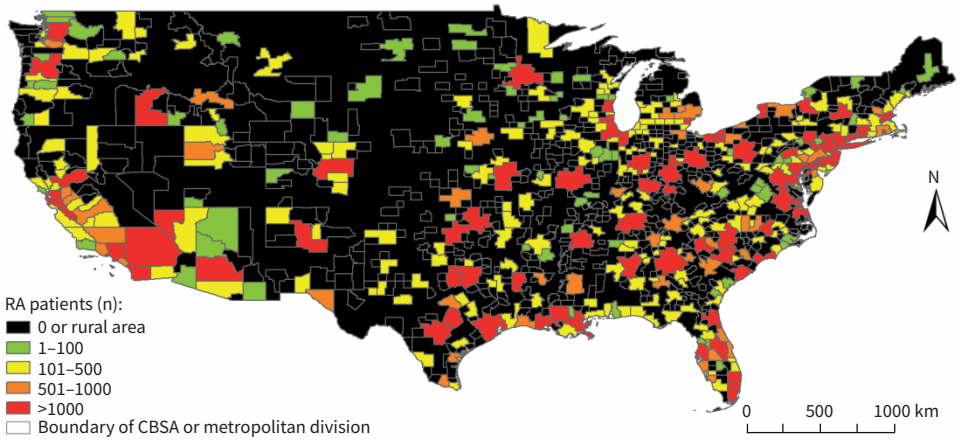

FIGURE 1 Distribution of identified new-onset rheumatoid arthritis (RA) patients in the USA. n-values are absolute numbers not rates and do not account for variations between states in terms of age, MarketScan enrolment or other factors. CBSA: core-based statistical area.

concentrations of mineral dust and sea salt were negatively correlated with those of the other $PM_{2.5}$ chemical components (supplementary table S1).

The single-pollutant Cox proportional hazards models (controlling for demographics but not concomitant air pollution exposure) showed that ambient overall $PM_{2.5}$ and ozone exposures were both associated with RA-ILD onset. In these partially adjusted models, most $PM_{2.5}$ chemical components had positive associations with RA-ILD onset, but mineral dust had a negative association with RA-ILD and sea salt had no clear association. The risk of developing RA-ILD increased with age (table 3). Similar results were obtained by removing COPD subjects from the single-pollutant Cox proportional hazards models (supplementary table S2).

Using quantile-based g-computation, we observed a significantly increased risk of RA-ILD onset with every decile increase in all seven $PM_{2.5}$ composition exposures (HR 1.54, 95% CI 1.47–1.63). This hazard ratio increased significantly after including ozone in the quantile-based g-computation model (table 4). Ammonium (index weight 0.59), mineral dust (0.15), black carbon (0.14), ozone (0.09) and sea salt (0.03) had positive effects on RA-ILD, while nitrate (0.57), organic matter (0.31) and sulfate (0.12) had negative effects. The negative index weights suggest that the $PM_{2.5}$ components of nitrate, organic matter and sulfate are not correlated with RA-ILD. Results from sensitivity analyses with COPD subjects removed were similar (table 4).

TABLE 1 Participant characteristics and exposures for subgroups of each covariate in the study cohort

	ILD (n (%)[#])	Person-years	Overall $PM_{2.5}$ ($\mu g \cdot m^{-3}$)	Black carbon ($\mu g \cdot m^{-3}$)	Mineral dust ($\mu g \cdot m^{-3}$)	Ammonium ($\mu g \cdot m^{-3}$)	Nitrate ($\mu g \cdot m^{-3}$)	Organic matter ($\mu g \cdot m^{-3}$)	Sulfate ($\mu g \cdot m^{-3}$)	Sea salt ($\mu g \cdot m^{-3}$)	Ozone (ppb)
Sex											
Male	564 (0.8)	342 464	8.30	0.71	0.64	0.76	0.94	2.88	2.00	0.37	39.37
Female	1630 (0.8)	1 052 876	8.31	0.71	0.67	0.74	0.93	2.89	2.00	0.37	39.47
Age											
⩽52 years	860 (0.6)	675 557	8.31	0.71	0.68	0.74	0.92	2.89	1.99	0.37	39.53
>52 years	1334 (0.9)	719 782	8.31	0.71	0.65	0.75	0.93	2.89	2.00	0.37	39.38
COPD											
No	121 (1.5)	37 302	8.55	0.74	0.64	0.78	0.95	2.98	2.11	0.36	39.41
Yes	2073 (0.8)	1 358 037	8.30	0.71	0.66	0.75	0.93	2.89	2.00	0.37	39.45

ILD: interstitial lung disease; $PM_{2.5}$: fine particulate matter with an aerodynamic diameter <2.5 $\mu g \cdot m^{-3}$; COPD: chronic obstructive pulmonary disease. [#]: denominator is the number of rheumatoid arthritis subjects in the subgroup.

TABLE 2 Distribution of population-weighted concentrations of ambient fine particulate matter with an aerodynamic diameter <2.5 μg·m^{-3} ($PM_{2.5}$) and its major chemical components over the 401 core-based statistical areas and metropolitan divisions in the USA for 2006–2017

	Minimum	Decile								Maximum
		10%	20%	30%	40%	50%	60%	70%	80%	90%	
$PM_{2.5}$ (μg·m^{-3})	3.01	6.10	7.15	7.65	8.39	8.68	8.98	9.21	9.68	10.30	12.40
Ammonium (μg·m^{-3})	0.10	0.31	0.47	0.59	0.68	0.75	0.84	0.92	1.02	1.16	1.86
Black carbon (μg·m^{-3})	0.22	0.43	0.54	0.59	0.67	0.72	0.79	0.80	0.87	0.94	1.43
Mineral dust (μg·m^{-3})	0.11	0.27	0.35	0.41	0.48	0.54	0.66	0.78	0.92	1.22	4.29
Nitrate (μg·m^{-3})	0.06	0.37	0.45	0.54	0.60	0.79	0.98	1.10	1.43	1.81	2.65
Organic matter (μg·m^{-3})	1.04	2.01	2.35	2.52	2.72	2.92	3.03	3.20	3.38	3.69	5.03
Sea salt (μg·m^{-3})	0.00	0.11	0.15	0.18	0.21	0.24	0.29	0.41	0.63	0.78	2.91
Sulfate (μg·m^{-3})	0.28	0.85	1.50	1.83	2.01	2.18	2.27	2.40	2.61	2.77	4.32
Ozone (ppb)	28.66	35.80	37.32	38.19	38.67	39.35	40.00	40.68	41.46	42.47	51.12

With the sex or age subgroups, similar positive associations between the mixture of $PM_{2.5}$ component exposures and RA-ILD and index weights (*i.e.* in the positive direction: index weight of ammonium larger than those of mineral dust, black carbon, ozone and sea salt; in the negative direction: index weight of nitrate larger than those of organic matter and sulfate) can be observed (supplementary table S3).

The Kendall's τ-tests showed that only 34.7% of the 401 CBSAs or metropolitan divisions had more than two (of the seven) $PM_{2.5}$ component concentrations time-series with statistically significant calendar-year trends ($p<0.01$).

Discussion

A few studies have demonstrated that long-term exposures to nitrogen dioxide, ozone and particulate matter with an aerodynamic diameter <10 μg·m^{-3} (PM_{10}) were associated with the increased risk of ILD or idiopathic pulmonary fibrosis (IPF, the most common form of ILD) in the general population [31–33], although clear associations between overall $PM_{2.5}$ and ILD/IPF were not clearly demonstrated in studies from Taiwan [31], North Italy [32], and Pennsylvania and New Jersey [33]. In our study, we observed a significant positive association between $PM_{2.5}$ exposure and RA-ILD incidence. More importantly, we quantified differential effects of $PM_{2.5}$ chemical components on RA-ILD onset and found that ammonium had the largest positive index weight among the seven $PM_{2.5}$ chemical components investigated (table 4). This finding complements other work suggesting that the health effects of $PM_{2.5}$ may not be entirely related to its total concentration, but also to the characteristics of its chemical constituents [34]. The composition of ambient $PM_{2.5}$ may vary greatly across different countries/areas, which should be taken into account when interpreting any study of its effects on health; a failure to demonstrate associations between a composite measure like total ambient $PM_{2.5}$ and a particular outcome does not exclude the possibility that a component of $PM_{2.5}$ may indeed be a culprit.

TABLE 3 Hazard ratios (95% CIs) from the single-pollutant Cox proportional hazards models for time to rheumatoid arthritis-associated interstitial lung disease onset[#]

	Exposure variable	Age	Sex	COPD
Overall $PM_{2.5}$	1.50 (1.45–1.55)	1.02 (1.01–1.02)	0.96 (0.87–1.06)	1.76 (1.46–2.11)
Ammonium	1.38 (1.36–1.40)	1.02 (1.01–1.02)	0.99 (0.90–1.09)	1.73 (1.43–2.08)
Black carbon	1.26 (1.23–1.28)	1.02 (1.01–1.02)	0.96 (0.87–1.06)	1.79 (1.49–2.15)
Mineral dust	0.97 (0.96–0.98)	1.02 (1.01–1.02)	0.97 (0.88–1.07)	1.86 (1.55–2.24)
Nitrate	1.01 (1.01–1.02)	1.02 (1.01–1.02)	0.96 (0.86–1.06)	1.86 (1.54–2.34)
Organic matter	1.01 (1.01–1.02)	1.02 (1.01–1.02)	0.96 (0.87–1.05)	1.85 (1.54–2.23)
Sea salt	1.00 (0.98–1.01)	1.02 (1.01–1.02)	0.96 (0.87–1.06)	1.87 (1.55–2.24)
Sulfate	1.17 (1.16–1.18)	1.02 (1.01–1.02)	0.97 (0.88–1.07)	1.71 (1.43–2.07)
Ozone	1.03 (1.01–1.04)	1.01 (1.01–1.02)	0.96 (0.87–1.06)	1.87 (1.55–2.45)

[#]: per 0.1 μg·m^{-3} (1 μg·m^{-3}) increase in fine particulate matter with an aerodynamic diameter <2.5 μg·m^{-3} ($PM_{2.5}$) components (overall $PM_{2.5}$) and 1 ppb increase in ozone, adjusting for age (years), sex (male as reference) and coexistence of chronic obstructive pulmonary disease (COPD).

TABLE 4 Adjusted hazard ratios (95% CIs) and index weights[#] from the quantile-based g-computation models for time to rheumatoid arthritis-associated interstitial lung disease onset with an increase in all exposures by one decile

COPD subjects	Ozone exposure	HR (95% CI)	Index weight							
			Ammonium	Mineral dust	Black carbon	Ozone	Sea salt	Nitrate	Organic matter	Sulfate
Included	Excluded	1.54 (1.47–1.63)	0.66	0.18	0.16		0.00	−0.61	−0.27	−0.12
Included	Included	2.21 (2.08–2.34)	0.59	0.15	0.14	0.09	0.03	−0.57	−0.31	−0.12
Excluded	Excluded	1.63 (1.55–1.72)	0.66	0.18	0.15		0.01	−0.61	−0.27	−0.12
Excluded	Included	2.30 (2.16–2.45)	0.59	0.15	0.13	0.09	0.04	−0.58	−0.31	−0.11

The number of overall subjects was 280 516, of which 7835 had chronic obstructive pulmonary disease (COPD). [#]: the positive and negative weights should not be compared with each other; the weights are only compatible with other weights in the same (*i.e.* positive or negative) direction.

Many studies have demonstrated harmful effects of mineral dust on respiratory health [35]; however, a significant negative association between mineral dust exposure and RA-ILD onset was seen with our single-exposure Cox model. This may have been due to confounding since the concentration of mineral dust in ambient $PM_{2.5}$ was negatively associated with most of the other $PM_{2.5}$ chemical components in the metropolitan statistical areas of the USA (including ammonium). Clearly, a single-pollutant Cox model cannot capture the real effect of the components of particulate matter.

With the quantile-based g-computation approach, we observed the second largest weight index for mineral dust, preceded only by ammonium (table 4). There is strong biological plausibility for both ammonia and mineral dust as potential triggers for RA-ILD, since both are very strong triggers of pulmonary inflammation [36, 37]. In the current study, we observed negative index weights from nitrate, organic matter and sulfate. Setting a negative direction for the effects of these three exposure variables ensured convergence of the quantile-based g-computation regression [10], but it does not necessarily indicate that nitrate, organic matter and sulfate are significantly associated with RA-ILD. In quantile-based g-computation, index weights are fixed (*i.e.* no confidence interval can be generated) [10], which is a shortcoming of the method, since statistical significance cannot be estimated. Considering there were already eight exposures (seven $PM_{2.5}$ chemical components+ozone) in our quantile-based g-computation model, adding further interaction terms might induce overfitting [10]. Future exploration of interactions between individual $PM_{2.5}$ chemical components may be possible using a nonparametric Bayesian procedure and high-performance computers [12].

In this study, we assigned air pollution estimates to subjects based on their CBSA or metropolitan division codes, since we did not have access to postal codes or census tracts, unlike other studies [19, 20, 28]. While potential misclassification of exposure is a limitation, we expect this to be nondifferential, meaning that our estimates would have tended to be biased towards the null instead of finding any association. Assigning exposures based on residential postal code or census tract may also be misleading since people are typically mobile in their cities as they engage in work and other activities [38]. Thus, a CBSA's or metropolitan division's average concentration of air pollution may be closer to a person's actual exposure than the concentration at the centroid point of the person's residential postal code. Additionally, unlike Pope *et al.* [39], in which air pollution exposures were assigned at the metropolitan statistical area level, we assigned $PM_{2.5}$ exposures to subjects living in the 11 largest metropolitan statistical areas based on their metropolitan division codes, which can reduce the exposure misclassification error generated by the over-large metropolitan areas.

Since only one-third of the CBSAs or metropolitan divisions had more than two (of the seven) $PM_{2.5}$ component concentrations time-series with statistically significant calendar-year trends, we believe our time-fixed exposure approach is reasonable. However, our study was limited to residents in urban areas and thus results may not be generalisable to individuals living in rural areas. Also, we could not obtain reliable data regarding concentrations of other gaseous air pollutants (*e.g.* nitrogen dioxide and sulfur dioxide) for the USA during our study period. Additionally, variables such as race/ethnicity, income and education are unavailable in the Commercial Claims dataset. Given that Americans with commercial health insurance may be more likely to have middle or high incomes, our results may not be generalisable to low-income populations.

Although use of a large population-based database is a strength, use of billing code diagnoses has imperfect specificity for ILD ascertainment, which could result in a nondifferential misclassification of the

outcome. This could be a contributor to the relatively wide confidence intervals for some of our hazard ratio estimates. Another potential limitation is that MarketScan subjects are no longer identifiable after they change insurance status; thus, it is possible that some subjects actually had prevalent RA instead of true incident RA and a few might have already had prevalent ILD.

In conclusion, we identified positive associations between a mixture of individual $PM_{2.5}$ chemical components and RA-ILD onset, and quantified specific effects of individual chemical components on RA-ILD using quantile-based g-computation. Our findings lend weight to the argument that some components of $PM_{2.5}$ (*e.g.* ammonium and mineral dust) are of greater concern than others and that greater public health benefits may be gained by controlling emissions of more toxic components. It seems increasingly clear that efficient use of nitrogen fertilisers and keeping more nitrogen and ammonium in the soil are critical ways to limit $PM_{2.5}$ levels in the atmosphere and curb the burden of many chronic diseases, potentially including RA-ILD [40].

Acknowledgements: The authors would like to thank Cristiano S. Moura (Research Institute of the McGill University Health Centre, Montreal, QC, Canada) for his help with assembling the RA-ILD cohort.

Conflict of interest: S. Bernatsky reports that support for the present manuscript is from the Canadian Institutes of Health Research (CIHR) to the Research Institute of the McGill University Health Centre. All other authors have nothing to disclose.

Support statement: This work was funded by the Canadian Institutes of Health Research (CIHR) (PJT-159682). Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Cross M, Smith E, Hoy D, *et al.* The global burden of rheumatoid arthritis: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis* 2014; 73: 1316–1322.
2 Iqbal K, Kelly C. Treatment of rheumatoid arthritis-associated interstitial lung disease: a perspective review. *Ther Adv Musculoskelet Dis* 2015; 7: 247–267.
3 O'Dwyer DN, Armstrong ME, Cooke G, *et al.* Rheumatoid arthritis (RA) associated interstitial lung disease (ILD). *Eur J Intern Med* 2013; 24: 597–603.
4 Farhat SCL, Silva CA, Orione MAM, *et al.* Air pollution in autoimmune rheumatic diseases: a review. *Autoimmun Rev* 2011; 11: 14–21.
5 Sunyer J. Urban air pollution and chronic obstructive pulmonary disease: a review. *Eur Respir J* 2001; 17: 1024–1033.
6 Behera SN, Sharma M. Reconstructing primary and secondary components of $PM_{2.5}$ composition for an urban atmosphere. *Aerosol Sci Technol* 2010; 44: 983–992.
7 Franklin M, Koutrakis P, Schwartz J. The role of particle composition on the association between $PM_{2.5}$ and mortality. *Epidemiology* 2008; 19: 680–689.
8 Jia Y-Y, Wang Q, Liu T. Toxicity research of $PM_{2.5}$ compositions *in vitro*. *Int J Environ Res Public Health* 2017; 14: 232.
9 Sack C, Vedal S, Sheppard L, *et al.* Air pollution and subclinical interstitial lung disease: the Multi-Ethnic Study of Atherosclerosis (MESA) air–lung study. *Eur Respir J* 2017; 50: 1700559.
10 Keil AP, Buckley JP, O'Brien KM, *et al.* A quantile-based g-computation approach to addressing the effects of exposure mixtures. *Environ Health Perspect* 2020; 128: 047004.
11 Zhao N, Smargiassi A, Hatzopoulou M, *et al.* Long-term exposure to a mixture of industrial SO_2, NO_2, and $PM_{2.5}$ and anti-citrullinated protein antibody positivity. *Environ Health* 2020; 19: 86.
12 Bobb JF, Valeri L, Henn BC, *et al.* Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics* 2015; 16: 493–508.
13 Carrico C, Gennings C, Wheeler D, *et al.* Characterization of a weighted quantile sum regression for highly correlated data in a risk analysis setting. *J Agric Biol Environ Stat* 2015; 20: 100–120.
14 Hunter TM, Boytsov NN, Zhang X, *et al.* Prevalence of rheumatoid arthritis in the United States adult population in healthcare claims databases, 2004–2014. *Rheumatol Int* 2017; 37: 1551–1557.
15 Jiang X-Q, Mei X-D, Feng D. Air pollution and chronic airway diseases: what should people know and do? *J Thorac Dis* 2016; 8: E31–E40.
16 Hagmeyer L, Randerath W. Smoking-related interstitial lung disease. *Dtsch Arztebl Int* 2015; 112: 43–50.
17 Chang K-H, Hsu C-C, Muo C-H, *et al.* Air pollution exposure increases the risk of rheumatoid arthritis: a longitudinal and nationwide study. *Environ Int* 2016; 94: 495–499.

18 van Donkelaar A, Martin RV, Li C, *et al.* 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* 2019; 53: 2595–2261.

19 Cakmak S, Hebbern C, Pinault L, *et al.* Associations between long-term $PM_{2.5}$ and ozone exposure and mortality in the Canadian Census Health and Environment Cohort (CANCHEC), by spatial synoptic classification zone. *Environ Int* 2018; 111: 200–211.

20 Pappin AJ, Christidis T, Pinault LL, *et al.* Examining the shape of the association between low levels of fine particulate matter and mortality across three cycles of the Canadian Census Health and Environment Cohort. *Environ Health Perspect* 2019; 127: 107008.

21 Mudway IS, Kelly FJ. Ozone and the lung: a sensitive issue. *Mol Aspects Med* 2000; 21: 1–48.

22 Lavigne E, Talarico R, van Donkelaar A, *et al.* Fine particulate matter concentration and composition and the incidence of childhood asthma. *Environ Int* 2021; 152: 106486.

23 Bennette C, Vickers A. Against quantiles: categorization of continuous variables in epidemiologic research, and its discontents. *BMC Med Res Methodol* 2012; 12: 21.

24 Robins J. A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Math Model* 1986; 7: 1393–1512.

25 Keil AP. The qgcomp package: g-computation on exposure quantiles. 2020. https://rdrr.io/cran/qgcomp/f/vignettes/qgcomp-vignette.Rmd Date last accessed: 8 November 2021.

26 Czarnota J, Gennings C, Wheeler DC. Assessment of weighted quantile sum regression for modeling chemical mixtures and cancer risk. *Cancer Inform* 2015; 14: Suppl. 2, 159–171.

27 Deyssenroth MA, Gennings C, Liu SH, *et al.* Intrauterine multi-metal exposure is associated with reduced fetal growth through modulation of the placental gene network. *Environ Int* 2018; 120: 373–381.

28 Zhao N, Smargiassi A, Hudson M, *et al.* Investigating associations between anti-nuclear antibody positivity and combined long-term exposures to NO_2, O_3, and $PM_{2.5}$ using a Bayesian kernel machine regression approach. *Environ Int* 2020; 136: 105472.

29 Snowden JM, Rose S, Mortimer KM. Implementation of g-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol* 2011; 173: 731–738.

30 Cole SR, Richardson DB, Chu H, *et al.* Analysis of occupational asbestos exposure and lung cancer mortality using the g formula. *Am J Epidemiol* 2013; 177: 989–996.

31 Chen H-H, Yong Y-M, Lin C-H, *et al.* Air pollutants and development of interstitial lung disease in patients with connective tissue disease: a population-based case-control study in Taiwan. *BMJ Open* 2020; 10: e041405.

32 Conti S, Harari S, Caminati A, *et al.* The association between air pollution and incidence of idiopathic pulmonary fibrosis in Northern Italy. *Eur Respir J* 2018; 51: 1700397.

33 Winterbottom CJ, Shah RJ, Patterson KC, *et al.* Exposure to ambient particulate matter is associated with accelerated functional decline in idiopathic pulmonary fibrosis. *Chest* 2018; 153: 1221–1228.

34 Gao D, Ripley S, Weichenthal S, *et al.* Ambient particulate matter oxidative potential: chemical determinants, associated health effects, and strategies for risk management. *Free Radic Biol Med* 2020; 151: 7–25.

35 Moreno T, Trechera P, Querol X, *et al.* Trace element fractionation between PM_{10} and $PM_{2.5}$ in coal mine dust: implications for occupational respiratory health. *Int J Coal Geol* 2019; 203: 52–59.

36 Brautbar N, Wu MP, Richter ED. Chronic ammonia inhalation and interstitial pulmonary fibrosis: a case report and review of the literature. *Arch Environ Health* 2003; 58: 592–596.

37 Fubini B, Arean OC. Chemical aspects of the toxicity of inhaled mineral dusts. *Chem Soc Rev* 1999; 28: 373–381.

38 Yu X, Lvey C, Huang Z, *et al.* Quantifying the impact of daily mobility on errors in air pollution exposure estimation using mobile phone location data. *Environ Int* 2020; 141: 105772.

39 Pope AC, Ezzati M, Cannon JB, *et al.* Mortality risk and $PM_{2.5}$ air pollution in the USA: an analysis of a national prospective cohort. *Air Qual Atmos Health* 2018; 11: 245–252.

40 Subbarao GV, Searchinger TD. A “more ammonium solution” to mitigate nitrogen pollution and boost crop yields. *Proc Natl Acad Sci USA* 2021; 118: e2107576118.