Fine particulate air pollution, nitrogen dioxide, and systemic autoimmune rheumatic disease in Calgary, Alberta

Sasha Bernatsky, Audrey Smargiassi, Markey Johnson, Gilaad G. Kaplan, Cheryl Barnabe, Larry Svenson, Allan Brand, Stefania Bertazoni, Marie Hudson, Ann E. Clarke, Paul R. Fortin, Steven Edworthy, Patrick Bélisle, Lawrence Joseph

A Department of Medicine, McGill University, Montreal, Quebec, Canada
B Division of Clinical Epidemiology, McGill University Health Centre, Montreal, Quebec, Canada
C Département de Santé Environnementale et de Santé au Travail, Université de Montréal, Montreal, Quebec, Canada
D Institut National de Santé Publique du Québec, Montréal, Canada
E Air Health Science Division, Health Canada, Ottawa, Ontario, Canada
F Department of Medicine, University of Calgary, Calgary, Alberta, Canada
G Department of Community Health Sciences, University of Calgary, Canada
H Surveillance and Assessment, Alberta Ministry of Health, Canada
I Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Albert, Canada
J University of Alberta, School of Public Health, Edmonton, Alberta, Canada
K Division of Geography, University of Calgary, Calgary, Alberta, Canada
L Division of Rheumatology, Jewish General Hospital, Montréal, Quebec, Canada
M Division of Allergy and Clinical Immunology, McGill University Health Centre, Montreal, Quebec, Canada
N Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

Article info

Article history:
Received 3 February 2015
Received in revised form 5 May 2015
Accepted 7 May 2015

Keywords:
Systemic autoimmune rheumatic diseases
Air pollution

Abstract

Objective: To estimate the association between fine particulate (PM$_{2.5}$) and nitrogen dioxide (NO$_2$) pollution and systemic autoimmune rheumatic diseases (SARDs).

Methods: Associations between ambient air pollution (PM$_{2.5}$ and NO$_2$) and SARDs were assessed using land-use regression models for Calgary, Alberta and administrative health data (1993–2007). SARD case definitions were based on ≥2 physician claims, or ≥1 rheumatology billing code; or ≥1 hospitalization code (for systemic lupus, Sjogren’s Syndrome, scleroderma, polymyositis, dermatomyositis, or undifferentiated connective tissue disease). Bayesian hierarchical latent class regression models estimated the probability that each resident was a SARD case, based on these case definitions. The sum of individual level probabilities provided the estimated number of cases in each area. The latent class model included terms for age, sex, and an interaction term between age and sex. Bayesian logistic regression models were used to generate adjusted odds ratios (OR) for NO$_2$ and PM$_{2.5}$ pollutant models, adjusting for neighbourhood income, age, sex, and an interaction between age and sex. We also examined models stratified for First-Nations (FN) and non-FN subgroups.

Results: Residents that were female and/or aged > 45 had a greater probability of being a SARD case, with the highest OR estimates for older females. Independently, the odds of being a SARDs case increased with PM$_{2.5}$ levels, but the results were inconclusive for NO$_2$. The results stratified by FN and non-FN groups were not distinctly different.

Conclusion: In this urban Canadian sample, adjusting for demographics, exposure to PM$_{2.5}$ was associated with an increased risk of SARDs. The results for NO$_2$ were inconclusive.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Systemic autoimmune rheumatic diseases (SARDs) are complex chronic inflammatory disorders (such as systemic lupus, SLE) characterized by multi-system inflammation (Callaghan et al.,...
2. Methods

We examined associations between local-scale ambient NO\textsubscript{2} and PM\textsubscript{2.5}. In Calgary, Alberta using land use regression (LUR) models and administrative health data. The LUR modelling approach has been widely used for research estimating individual exposure to ambient air pollution (Johnson et al., 2013). Land-use regression models are built by including predictor variables, such as traffic, topography, and other geographic variables, in a multivariable regression using monitored pollution levels as the outcome. Subsequently, levels of pollution may then be estimated for any point, using the parameter estimates derived from the regression model. This method has been identified as a preferred approach to estimating small area variations in air pollution effectively, when household level monitoring data are not available (Ryan and LeMasters, 2007; Health Effects Institute, 2010).

We used comprehensive (physician billing and hospitalization) administrative provincial health data from Alberta (1993–2007) and focussed our analyses on Calgary, since LUR estimates for PM\textsubscript{2.5} and NO\textsubscript{2} were available for this region of Alberta, only. In Canada, each province maintains linkable databases on all residents who are recipients of a comprehensive health plan which covers physician visits and hospitalizations. In the Alberta health administrative databases, information on First Nations (FN) status is also available, at the individual level. The FN variable is defined by whether or not an individual’s health premiums are paid by the First Nations and Inuit Health Branch (Health Canada) at any time point since 1994, thus indicating Treaty Status as per the Indian Act (Alberta Health and Wellness, 2004).

The databases record one physician billing code per visit (in Alberta, up to three diagnosis codes are allowed, but use of more than one code is infrequent), and all hospitalizations (with up to 25 diagnostic codes per hospitalization in Alberta). The SARD case definition was based on three algorithms: two or more physician billing claims with the International Classification of Diseases (ICD)-9 code 710 (within 2 years but at least 2 months apart); at least one such billing code by a rheumatologist; or at least one hospitalization with an ICD-10 diagnostic code corresponding to a SARD diagnosis (M32.1, M32.8–32.9, M33–M34, M35.0, M35.8–35.9, M36.0). These diagnostic codes include SLE, Sjogren's Syndrome, scleroderma, polymyositis, dermatomyositis, and undifferentiated connective tissue disease. Within the ICD-9 classification system, in provinces where billing data are only limited to three digits, it is impossible to differentiate between these conditions (SLE, Sjogren's Syndrome, scleroderma, polymyositis, dermatomyositis, and undifferentiated connective tissue disease). Rheumatoid arthritis (and systemic vasculitis, by the way) fall under separate ICD-9 categories (714.x and 446.x respectively) and we were unable to include these groups in the current study.

Residents of Calgary could meet one or more of the three SARDs case definitions. Bayesian hierarchical latent class regression models (which do not assume any case definition to be a perfect gold standard) (Bernatsky et al., 2005) were used to estimate the probability that any given Calgary resident was a SARD case, based on their results relative to the three case definitions. For example, a resident who never fulfilled any of the case definitions would have an estimated probability of being a SARD that was much smaller then someone who fulfilled all three case definitions, the former near zero (but not exactly zero) and the latter nearer to 1 (but not exactly 1). Residents with one or two case definitions would have estimated probabilities between these extremes. The latent class approach to case definition within administrative data (as opposed to assuming chart review or anything else as a perfect gold standard), where individuals are not assigned a case status per se, but rather are provided with an estimated probability of being a case, has been shown to useful across several diseases and jurisdictions (Prosser et al., 2008).

The individual level probabilities estimated were summed to estimate the total number of SARD cases according to groups characterized by age (dichotomized to less than or equal to 45 years of age, or older than 45), sex, and dissemination area, DA (the smallest standard geographic area for which all Canadian census data are disseminated, generally representing between 400 and 700 persons). The DA associated with the residence of each Calgary subject was obtained through overlay of Calgary postal code centroids on a geographic layer of the Calgary dissemination areas. To obtain SARD prevalence rates for each DA, for each age and sex group, we divided the estimated number of SARD cases by the appropriate regional Calgary population figures obtained from the Canadian Census 2006.

We then used the case definition results as outcome data in a logistic regression model that estimated odds ratio (OR), estimates for sex, age group, and pollution levels (explained below), across all dissemination areas in Calgary. The model also included an interaction term between age and sex. Around our point estimates of the odds ratios (ORs) generated from the model are Bayesian 95% credible intervals (CrI). Regional income information was obtained from the Canadian Census 2006 for Alberta.

Long-term exposure to ambient NO\textsubscript{2} and PM\textsubscript{2.5} was estimated using the LUR model results developed for Calgary dissemination block areas. The LUR is based on data from two-week summer and winter air monitoring campaigns (held in August 2010 and January 2011). NO\textsubscript{2} LUR model data were based on data from 46 sites in summer and 47 sites in winter, while PM\textsubscript{2.5} was based on data from 25 sites in summer and 29 sites in winter. NO\textsubscript{2} was modelled using a geographically weighted land-use regression model to address spatial non-stationarity and autocorrelation (Bertazzon et al., 2011; Elikan et al., 2011). PM\textsubscript{2.5} was modelled using traditional linear regression LUR methods, because the PM\textsubscript{2.5} spatial non-stationarity and autocorrelation do not have a lot of impact on PM\textsubscript{2.5} estimates. Some examples of important predictors of NO\textsubscript{2} and PM\textsubscript{2.5} in the LUR models included traffic
industrial sources (based on data from the National Pollutant Release Inventory, as well as road network and land-use information provided by the City of Calgary and Desktop Mapping Technologies Incorporated) and wind (meteorological data was obtained from three Calgary weather stations operated by Environment Canada). Mean annual ambient PM$_{2.5}$ and NO$_2$ levels were calculated as the average of the summer and winter LUR model estimates for each pollutant.

For our analyses, which occurred at the DA level, the dissemination block-level LUR pollution estimates were aggregated within their larger dissemination area (weighted according the block population within each DA, to best represent the exposure of the population within each DA); thus all residents within a given DA were assigned the same pollution exposure variables.

Since the risk of SARDs may be different (that is, sometimes higher) in First-Nations (FN) versus non-FN residents (Barnabe et al., 2012), we first ran models with all Calgary subjects, then we ran models where FN and non-FN residents were stratified.

### 3. Results

In the administrative data for the whole of Calgary, the average percentage of females was 50.2% (with the interquartile range across DAs being 48.6–51.7%), and the percentage of older individuals (> 45 years), across DAs, was 36.7% (interquartile range 29.6–43.9%). The average PM$_{2.5}$ levels across DAs was 6.38 µg/m$^3$ (median 6.29, interquartile range 6.03, 6.64) and the average of NO$_2$ across DAs was 16.30 ppb (median 16.33, interquartile range 14.44, 18.29). In Calgary, the estimated SARD prevalence, based on our methods, was about 3–4 cases per 10,000 persons.

Our hierarchical latent class logistic regression model accounted concurrently for age and sex (and an interaction term between them). Table 1 shows the results of models with all Calgary subjects, as well as models where FN and non-FN residents were stratified. We found strong evidence for effect modification, such that the OR for older females was higher by an order of magnitude compared to simply being older or being female. The effects were not distinctly different in the FN versus non-FN populations, but our precision was limited given that FN Calgary residents represent a relatively small number.

Independently of these effects, the odds of being a SARDs case appeared to be potentially associated with PM$_{2.5}$ levels. Results for NO$_2$ were inconclusive based on the large credible intervals (Table 2).

### 4. Discussion

Our findings of higher rates of SARDs diagnosis for certain demographic factors, particularly female sex and older age, is consistent with both our prior work, and the known epidemiology of these diseases (Broten et al., 2014). The evidence for a potential association of SARDs with ambient PM$_{2.5}$ is relatively novel; a few very few studies have examined the association of particulate matter and the onset of autoimmune rheumatic diseases. In one prior study, PM$_{2.5}$ levels were associated with a 60% increased risk of juvenile idiopathic arthritis in young children (Zeft et al., 2009) and in a 2009 analyses of the Nurses’ Health Study by Hart et al. (2009) suggested that pollution emissions from road traffic may be an environmental risk factor for rheumatoid arthritis, RA. However, Hart et al. have authored two recent additional studies with potentially conflicting results. In a Swedish cohort, they found that exposure to NO$_2$, but not particulate matter (based on PM$_{10}$), was associated with RA risk (Hart et al., 2013). However, in another follow up study of the Nurses’ Health Study (Hart et al., 2013), they failed to detect associations with either NO$_2$ or PM$_{2.5}$. Though not included in our current study, both JIA and RA are similar to the SARDs we studied, in terms of being rheumatic disease characterized by autoimmunity.

Our own team has shown the association between road-traffic density and SARD onset in Montreal (Labrecque et al., 2010) as well as suggested links between PM$_{2.5}$ levels and SLE activity in our Montreal SLE cohort (Bernatsky et al., 2011). Very recently proinflammatory effects of particulate matter (diesel exhaust nanoparticles) were demonstrated in an in-vitro study of scleroderma skin cells (Mastrofrancesco et al., 2014), suggesting a possible mechanism for PM-mediated effects. In addition, we have recently shown the association between PM$_{2.5}$ and SARDs

---

### Table 1

Odds ratio (OR) estimates for systemic autoimmune rheumatic disease in Calgary, for PM$_{2.5}$ air pollution, adjusted for demographics (older age, sex, income) and stratified by First Nations status.

<table>
<thead>
<tr>
<th></th>
<th>OR</th>
<th>95% CI</th>
<th>CrI</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Calgary residents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female sex</td>
<td>8.14</td>
<td>6.15</td>
<td>10.3</td>
</tr>
<tr>
<td>Mean income***</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>Age &gt; 45</td>
<td>4.78</td>
<td>3.67</td>
<td>5.89</td>
</tr>
<tr>
<td>Interaction term between age and sex PM$_{2.5}$</td>
<td>24.6</td>
<td>19.3</td>
<td>31.4</td>
</tr>
<tr>
<td>First Nation residents only</td>
<td>1.10</td>
<td>1.01</td>
<td>1.22</td>
</tr>
<tr>
<td>Female sex</td>
<td>7.31</td>
<td>2.73</td>
<td>22.6</td>
</tr>
<tr>
<td>Mean income</td>
<td>1.01</td>
<td>0.88</td>
<td>1.11</td>
</tr>
<tr>
<td>Age &gt; 45</td>
<td>2.66</td>
<td>0.84</td>
<td>9.24</td>
</tr>
<tr>
<td>Interaction term between age and sex PM$_{2.5}$</td>
<td>51.0</td>
<td>17.0</td>
<td>180.4</td>
</tr>
<tr>
<td>Non-First Nation Calgary residents only</td>
<td>1.40</td>
<td>0.72</td>
<td>2.60</td>
</tr>
<tr>
<td>Female sex</td>
<td>7.98</td>
<td>6.06</td>
<td>10.9</td>
</tr>
<tr>
<td>Mean income</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>Age &gt; 45</td>
<td>4.88</td>
<td>3.63</td>
<td>6.49</td>
</tr>
<tr>
<td>Interaction term between age and sex PM$_{2.5}$</td>
<td>25.4</td>
<td>19.6</td>
<td>35.6</td>
</tr>
</tbody>
</table>

*** Mean income values for each Calgary dissemination area were obtained from Canadian Census data.

---

### Table 2

Odds ratio (OR) estimates for systemic autoimmune rheumatic disease in Calgary, for nitrogen dioxide (NO$_2$) levels, adjusted for demographics (older age, sex, income) and stratified by First Nations status.

<table>
<thead>
<tr>
<th></th>
<th>OR</th>
<th>95% CI</th>
<th>CrI</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Calgary residents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female sex</td>
<td>7.54</td>
<td>5.83</td>
<td>9.57</td>
</tr>
<tr>
<td>Mean Income***</td>
<td>0.99</td>
<td>0.98</td>
<td>1.00</td>
</tr>
<tr>
<td>Age &gt; 45</td>
<td>5.14</td>
<td>3.81</td>
<td>6.40</td>
</tr>
<tr>
<td>Interaction term between age and sex NO$_2$</td>
<td>26.6</td>
<td>20.6</td>
<td>33.3</td>
</tr>
<tr>
<td>First Nation residents only</td>
<td>1.00</td>
<td>0.98</td>
<td>1.02</td>
</tr>
<tr>
<td>Female sex</td>
<td>7.43</td>
<td>2.72</td>
<td>24.4</td>
</tr>
<tr>
<td>Mean income</td>
<td>1.00</td>
<td>0.87</td>
<td>1.10</td>
</tr>
<tr>
<td>Age &gt; 45</td>
<td>2.58</td>
<td>0.71</td>
<td>10.5</td>
</tr>
<tr>
<td>Interaction term between age and sex NO$_2$</td>
<td>52.3</td>
<td>17.5</td>
<td>170.8</td>
</tr>
<tr>
<td>Non-First Nation Calgary residents only</td>
<td>1.03</td>
<td>0.89</td>
<td>1.20</td>
</tr>
<tr>
<td>Female sex</td>
<td>7.80</td>
<td>6.26</td>
<td>10.0</td>
</tr>
<tr>
<td>Mean income</td>
<td>0.99</td>
<td>0.98</td>
<td>1.00</td>
</tr>
<tr>
<td>Age &gt; 45</td>
<td>4.77</td>
<td>3.85</td>
<td>6.44</td>
</tr>
<tr>
<td>Interaction term between age and sex NO$_2$</td>
<td>25.2</td>
<td>20.3</td>
<td>32.7</td>
</tr>
</tbody>
</table>

*** Mean income values for each Calgary dissemination area were obtained from Canadian Census data.

* NO$_2$ level estimates were expressed in parts per billion.

b Bayesian credible intervals.
occurrence throughout the province of Alberta, based on satellite-derived estimates of ambient PM$_{2.5}$ (Bernatsky et al., 2014). The OR for the fourth versus the first quartile of PM$_{2.5}$ exposure in Alberta, where the effect of PM$_{2.5}$ was non-linear, was 1.13 (95% CI 1.01, 1.25) per 1 μg/m$^3$. In the province of Quebec, modelling PM$_{2.5}$ as a linear variable, we estimated an OR for SARDs of 1.055 (95% CI 1.051, 1.059) per increase of 1 μg/m$^3$.

Those earlier analyses of the effects of pollution on SARDs in Alberta were limited by the spatial scale and pollutants estimated by satellite remote sensing data. The analyses only assessed PM$_{2.5}$; although satellite remote sensing estimates are available for NO$_2$, they are not considered as accurate as the estimates for PM$_{2.5}$. Moreover, LUR estimates for air pollution provide more spatially refined, local scale estimates than other methods. In the earlier analyses based on satellite remote sensing, PM$_{2.5}$ levels were estimated at a much larger geographic scale (i.e., 10 km x 10 km). The trade-off is that the LUR based analyses were limited to the city of Calgary, versus using province-wide data. Though use of LUR to assign exposure is considered one of the best available methods, novel approaches combining LUR estimates with satellite data are possible, and we will explore this in the future.

In our current analyses, we were able to demonstrate a potential link between PM$_{2.5}$ and SARDs, but we were unable to draw conclusions regarding NO$_2$ and SARDs. It should be kept in mind that the original LUR estimates for PM$_{2.5}$ and NO$_2$ were produced to reflect averages over the dissemination block, but our analyses were performed based on average concentrations at the (larger) dissemination area-level. NO$_2$ is more spatially heterogeneous compared with PM$_{2.5}$, which means that as size of the geographic unit of analysis increases, there may be increasing non-differential misclassification of NO$_2$. This effect would be less pronounced for PM$_{2.5}$, which is more spatially homogeneous. This could have led to a "bias towards the null" in our NO$_2$ analyses. We note that the goodness-of-fit of the LUR model for NO2 was actually better than the model for PM$_{2.5}$, although some of the standard error terms for predictors in the NO2 LUR model were a bit larger than for PM$_{2.5}$ model predictors (Johnson et al., 2013).

Another potential issue is temporality, because LUR estimates static concentrations at a single time period, our analyses are based on the assumption that the geographic variation of NO$_2$ and PM$_{2.5}$ estimates was relatively constant for the 1993–2007 study period. Both NO$_2$ and PM$_{2.5}$ levels may have decreased or increased locally with the closure or opening of new industrial sources and other land-use changes such as expanding road networks and residential populations. However, there are currently no current fine scale pollution models available for estimating historic concentrations. Community-level trends suggest that NO$_2$ levels have decreased over time in most Canadian cities, while PM$_{2.5}$ levels have remained relatively constant since the late 1990s, with only a slight downward trend noted in major Canadian cities (Wood et al., 2012).

We used latent class regression models (which do not assume any case definition to be a perfect gold standard) to estimate the probability that any given Calgary resident was a SARD case, based on their results relative to the three case definitions. In the latent class approach to case definition, individuals are not assigned a case status per se, but rather are provided with an estimated probability of being a case, which been shown to be useful across several diseases and jurisdictions (Prosser et al., 2008). Our Bayesian latent class model in fact uses the information from the three case definitions to estimate, and adjust for, the imperfect nature of all case definitions, in terms of both sensitivity and specificity. A limitation of using the latent class approach, which does not identify cases per se, is that we could not with our methods distinguish between incident and prevalent SARD cases. Thus the temporal relationship between the pollutants and SARDS incidence cannot be specifically assessed. Future analyses in incident-only SARD cases, using other approaches, could be done to examine whether the effects we saw were the same in incident versus prevalent SARD cases. Non-differential misclassification of the exposure could indeed occur if persons who had SARDs diagnoses had the same migration rate as those without SARDs diagnoses, and this would have biased the results towards the null. Differential migration on the other hand, could cause biased results, if for example greater SARDs prevalence in urban areas (where PM$_{2.5}$ levels may be higher) is due to migration of persons from rural to urban areas after their SARDs diagnoses. Though we cannot completely rule out this potential bias, Labrecque (2013) studied migration rates in SLE and did not find rates were greater than control populations.

Residual confounding by variables incompletely captured in the current analyses, such as race/ethnicity or potentially socio-economic status, remain as potential limitations. Though the effects were not distinctly different in the FN versus non-FN populations, our precision was limited given that FN Calgary residents represent a relatively small number.

Future studies should address whether air pollutants from other sources are also associated with SARDs. Evaluating even smaller particulate matter generated by road traffic (especially submicron particles, PM$_{1}$.), would help to further characterize associations between particulate matter and SARDs. Estimates of SO$_2$ and PM$_{1}$ exposures are not widely available; sulphur dioxide (SO$_2$.), from industrial emissions would also be of interest. LUR model estimates for PM$_{1}$ and SO$_2$ are currently being developed in Calgary, so that future analyses can examine associations between these exposures of interest regarding SARDS prevalence. Also of potential interest would be the consideration of the effects of noise, as well as air pollution, on rheumatic diseases (De Roos et al., 2014). Importantly, environmental exposure to inhaled toxic substances may be able to induce citrullination in lung cells prior to any detectable onset of inflammatory responses, suggesting that this may be an important pathway linking environmental triggers to rheumatic disease risk (Valesini et al., 2015; Farhat et al., 2011).

In conclusion, after adjusting for demographics (regional income, age and sex), our new analyses in Calgary (using LUR, a different approach to exposure estimation) replicated our earlier findings (using satellite data) suggesting that PM$_{2.5}$ levels are associated with SARDs in Alberta. It should be kept in mind that this is an association, and that no causal inference can be made. We did not see a clear association with SARDs and NO$_2$ in the current analyses, but there remains additional work to do in this area.

**Conflict of interest**

The authors declare no conflict of interest.

**Funding**

This study was funded by the Canadian Institutes of Health Research, CAP-120355.

**References**


effects of particulate matter air pollution on human health. J. Med. Toxicol. 8 (2), 166–175.


