Cohort profile

Cohort Profile: The ONtario Population Health and Environment Cohort (ONPHEC)

Hong Chen[1,2,3]*, Jeffrey C Kwong[1,2,3,4], Ray Copes[1,2], Paul J Villeneuve[2,5], Mark S Goldberg[6,7], Sherry L Ally[1], Scott Weichenthal[8], Aaron van Donkelaar[9], Michael Jerrett[10], Randall V Martin[9,11], Jeffrey R Brook[12], Alexander Kopp[3] and Richard T Burnett[13]

[1]Public Health Ontario, Toronto, ON, Canada, [2]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada, [3]Institute for Clinical Evaluative Sciences, Toronto, ON, Canada, [4]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada, [5]CHAIM Research Centre, Carleton University, Ottawa, ON, Canada, [6]Department of Medicine, McGill University, Montreal, QC, Canada, [7]Division of Clinical Epidemiology, Research Institute of the McGill University Health Centre, Montreal, QC, Canada, [8]Air Health Effects Science Division, Health Canada, Ottawa, ON, Canada, [9]Department of Physics and Atmospheric Science, Dalhousie University, Halifax, NS, Canada, [10]Department of Environmental Health Sciences, Jonathan and Karin Fielding School of Public Health, University of California, Los Angeles, CA, USA, [11]Harvard-Smithsonian Centre for Astrophysics, Cambridge, MA, USA, [12]Air Quality Research Division, Environment Canada, Toronto, ON, Canada and [13]Population Studies Division, Health Canada, Ottawa, ON, Canada

*Corresponding author. Public Health Ontario, 480 University Avenue, Suite 300, Toronto, Ontario M5G 1V2, Canada.
E-mail: hong.chen@oahpp.ca

Why was the cohort set up?

Chronic diseases, including cardiovascular disease, diabetes mellitus, cancers and neurological disorders, constitute the most important global burden on disease, accounting for 68% of deaths and 54% of disability-adjusted life years (DALYs) each year.[1,2] With increasing life expectancy, the impacts of chronic diseases will continue to rise.[3] To reduce the substantial burden of chronic diseases, intervening on their major risk factors is the most cost-effective approach.[4] Among various modifiable risk factors, environmental factors are increasingly being recognized as playing important roles in the development of a number of chronic diseases.[5]

The World Health Organization (WHO) has estimated that approximately 24% of DALYs and 23% of deaths are attributable to five environmental risk factors: outdoor air pollution; indoor air pollution; exposure to lead; climate change; and lack of access to clean water, sanitation and hygiene.[6] Environmental exposures may also yield beneficial effects on human health. For example, living in greener neighbourhoods may contribute to improvements in psychological well-being,[7–9] increased physical activity[10] and reduced obesity[11] and overall mortality.[12]

Many environmental risk factors, such as ambient air pollution and noise, are ubiquitous and can lead to large public health impacts. However, to quantify their health impacts, large cohorts are usually required, as the effect sizes of these exposures at the individual level are relatively small compared with risk factors such as tobacco smoking (for example, smoking increases the risk of lung cancer mortality by 15-fold).[13] Recent advances in the volume and variety of electronic health records, and the rate at which they can be merged and analysed in the era of Big Data, provide an opportunity to create very large cohorts

to acquire new insights into the environmental burden of chronic diseases.[14,15]

The **ON**tario **P**opulation **H**ealth and **E**nvironment **C**ohort (ONPHEC) is a large, retrospective cohort in the province of Ontario, Canada (population of 10.8 million in 1996),[16] created in 2014 by linking multiple large-scale health administrative databases. Comprising virtually the entire Canadian-born population of Ontario who were 35 years or older in 1996 ($\sim$ 4.9 million), with a follow-up until 2014, the primary objectives of ONPHEC are to investigate the independent and combined effects of environmental stressors (such as air pollution, traffic-related noise) on the incidence of chronic diseases and their interactions with 'healthy' environmental factors (e.g. green areas). Secondary objectives of ONPHEC include the following:

- to determine the effects of environmental exposures on the long-term survival of adults with selected chronic diseases;
- to quantify the burden of disease from adverse environmental factors;
- to evaluate the impacts of changes in environmental exposure on disease risk such as moving to less polluted neighbourhoods or vice versa; and
- to identify population subgroups who are most susceptible to the effects and who are disproportionately exposed to environmental stressors.

Ontario represents an ideal setting to examine these impacts, for multiple reasons. First, information on the incidence of major chronic diseases such as acute myocardial infarction (AMI), heart failure, diabetes and dementia is available for virtually the entire population of Ontario, using multiple chronic disease databases, some of which have been populated starting as early as the mid 1980s.[17–19] Long-term historical data for environmental exposures, such as ambient air pollution and meteorological conditions, are also available with complete spatial coverage across Ontario.[20,21] With a retrospective cohort design, ONPHEC provides an efficient and statistically powerful way to investigate the impact of environmental exposures on the onset of chronic diseases.

Second, Canada routinely conducts population-based health surveys to collect data on health behaviour and determinants of health, such as smoking, obesity and diet, using a representative sample of the population.[22,23] ONPHEC augments health administrative data with these survey data, thus allowing for a two-stage cohort design to combine the strengths of health administrative data (that is large statistical power) and population-based health surveys (that is rich information on person-level risk factors for major chronic diseases). This design allows for indirect control of important risk factors such as smoking and obesity, that are often unavailable in large observational studies, especially those using administrative databases.[24]

Third, Ontario has generally good environmental conditions, but there are also large spatial variations in their exposures. This provides an opportunity to develop exposure-response functions at relatively low levels of environmental exposures, which will have important public health implications globally. For example, average levels of ambient air pollution in Ontario are considerably lower than in many cities in the USA and Europe, and are well below the WHO air quality guidelines.[25,26]

Fourth, Ontario is the most populous and ethnically diverse province in Canada, with a population that represents more than 200 ethnic origins.[27] With demographic characteristics comparable to the USA and many European countries, findings from ONPHEC will be highly generalizable to populations in many other regions.

Who is in the cohort?

ONPHEC is a retrospective cohort that was created in 2014. It comprises all residents of Ontario who on 1 April 1 1996: (i) were alive and were between 35 and 100 years of age; (ii) had a recorded birth date and sex; (iii) had a valid health card number; and (iv) were Canadian-born individuals. To better characterize long-term exposure to environmental factors, we excluded individuals who had resided in Ontario for fewer than 5 years on cohort entry, yielding a total of population of 4 854 759 individuals. We chose 1 April 1 1996 as the date of cohort inception to allow for a 5-year look-back window to establish residential history and to accrue a reasonably long follow-up period until 2014 (the end of follow-up). The study population was constructed using the Ontario Registered Persons Database, a registry of all Ontario residents who have ever had a health card number.[28] Because Ontario has a single-payer health insurance system, the registry is an ideal sampling frame that covers the entire target population. Figure 1 outlines the creation of ONPHEC.

Another important feature of ONPHEC is that a representative sample of this cohort also participated in the 1996/97 cycle of the National Population Health Survey (NPHS) or the 2000/01, 2003, 2005, or 2007/08 cycles of the Canadian Community Health Survey (CCHS), and they agreed to share and link their responses to health administrative databases via their health card number. These surveys are conducted routinely by Statistics Canada and provide detailed data on variables such as smoking, alcohol use, obesity and household income.[22,23] These surveys cover all household residents aged 12 years or older, excluding individuals living on First Nations Reserves,
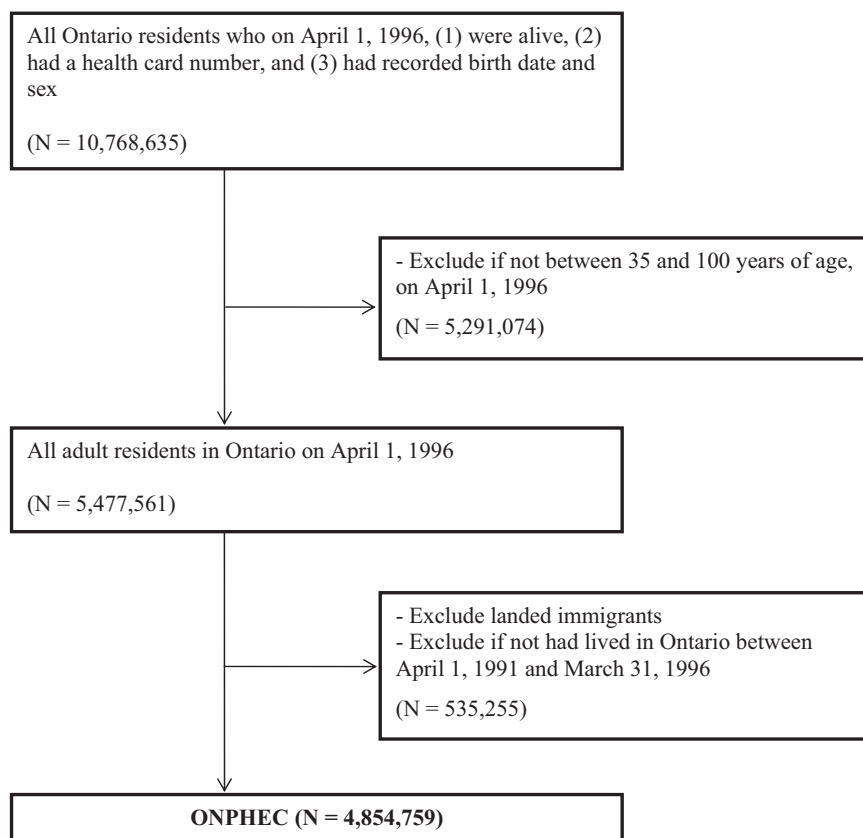
Figure 1. Creation of the **ON**tario **P**opulation **H**ealth and **E**nvironment **C**ohort (ONPHEC).

Canadian military bases or certain remote regions of Québec.[22,23] Ontario response rates for these surveys ranged from 73.4% to 92.8%.[22,23]

How often have they been followed up?

Because ONPHEC was created using the population registry, once included in the cohort, individuals remain in it until death or termination of Ontario health insurance (due to moving out of the province). Follow-up is continuous, with annual updates of the health administrative databases that track important information such as vital status, clinical events and residential postal codes for all cohort members. This is achieved through deterministic record linkage across multiple health administrative databases, using a unique, encoded identifier, thereby protecting individual privacy. To date, ONPHEC has linked data across 20 health administrative, chronic disease and environmental databases (Figure 2). A description of each database is provided in Supplementary Data, available at *IJE* online. As new data become available for these datasets, the follow-up for ONPHEC will be further extended beyond 2014.

All the health administrative and health survey data are held and linked at the Institute for Clinical Evaluative Sciences (ICES), located at Sunnybrook Health Sciences

Centre in Toronto, Ontario.[17] Under Ontario's Personal Health Information Protection Act, ICES researchers are allowed to link encoded population-based health databases for conducting research following stringent privacy and security policies and practices. The institutional review board at Sunnybrook Health Sciences Centre has approved all ongoing studies under ONPHEC.

What has been measured?

Selected health outcomes

The primary outcomes of ONPHEC are cardiovascular disease, diabetes and neurodegenerative disease, which were selected based on previous evidence linking environmental exposures to these specific outcomes.[29–35] To ascertain their incidence, we used existing chronic disease databases including the Ontario Hypertension Database, the Ontario Congestive Heart Failure Database, the Ontario Myocardial Infarction Database and the Ontario Diabetes Database. More recently, we have developed province-wide databases for three major neurodegenerative diseases (dementia, parkinsonism and multiple sclerosis). Each of these databases has been populated using health administrative data that are routinely collected by the Ontario
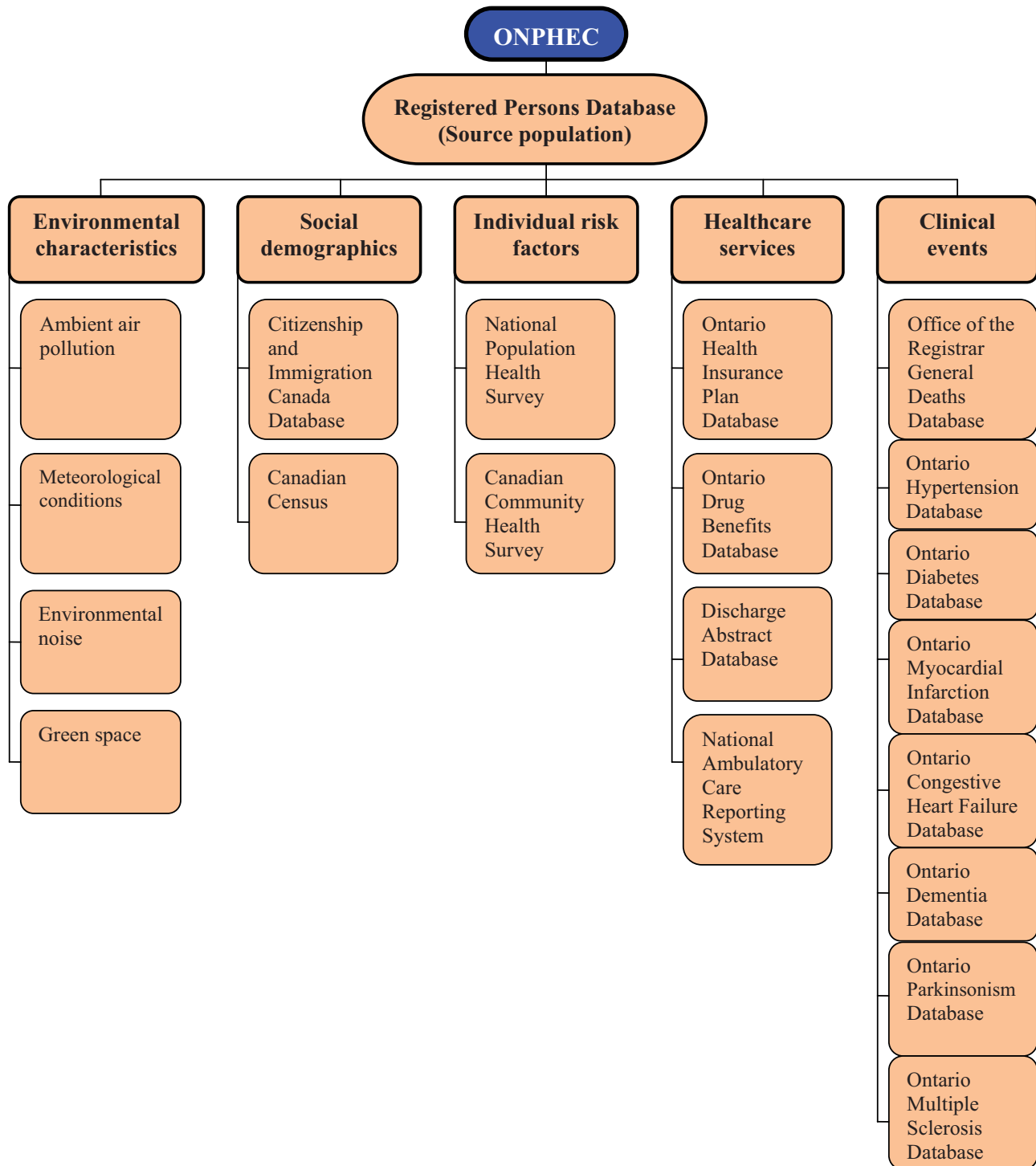
Figure 2. Data linkage to create ONPHEC cohort.

government for the purpose of health care system administration.[17] These chronic disease databases have been validated through chart review and found to have high sensitivity and specificity (for example, the Ontario Diabetes Database has a sensitivity of 86% and specificity of 97%).[19,36–40] Algorithms for populating each of these chronic disease databases are provided in Table 1. Additionally, record linkage to the Office of the Registrar

General Death Database provides cause-specific mortality information.

Selected characteristics of ONPHEC

At the individual level, using record linkage of administrative databases and health surveys we have obtained data such as age, sex, marital status, height, weight, smoking

Table 1. Case definitions of selected chronic diseases using Ontario health administrative data

Chronic disease	Case definition	Validation
Hypertension	Either one hospital admission with a hypertension diagnosis,[a] or an Ontario Health Insurance Program (OHIP) claim with a hypertension diagnosis followed within two years by either an OHIP claim or a hospital admission for hypertension	Sensitivity = 72% Specificity = 95%
Myocardial infarction	One hospital admission with myocardial infarction as a most responsible diagnosis,[b] excluding patients transferred from another acute care hospital or discharged with a total length of stay less than three days	Sensitivity = 89% Specificity = 93%
Congestive heart failure	Aged >40 years, and either one hospital admission with a congestive heart failure diagnosis,[c] or one OHIP claim/emergency department record with a congestive heart failure diagnosis followed within one year by a second record from either source or one hospital admission	Sensitivity = 85% Specificity = 97%
Diabetes mellitus	Either one hospital admission with diabetes mellitus diagnosis,[d] or an OHIP claim followed within two years by either an OHIP claim or a hospital admission with diabetes mellitus diagnosis	Sensitivity = 86% Specificity = 97%
Dementia	Aged >20 years, and either one hospital admission with a diagnosis of dementia,[e] or at least three OHIP claims for dementia within two years, or at least one prescription for drugs relating to dementia	Sensitivity = 75% Specificity = 99%
Parkinsonism	Aged >20 years, and either at least two OHIP claims for Parkinsonism within one year,[f] or at least one prescription for Parkinsonism and one OHIP claim within six months either before or after the drug prescription	Sensitivity = 78% Specificity = 100%
Multiple sclerosis	Aged >20 years, and either one hospital admission with a Multiple sclerosis diagnosis,[g] or five or more OHIP claims within two years	Sensitivity = 84% Specificity = 100%

[a]Hypertension: ICD-9 401-405; ICD-10 code I10-I13 or I15.
[b]Myocardial infarction: ICD-9 code 410; ICD-10 code I21.
[c]Congestive heart failure: ICD-9 code 428; ICD-10 code I50.
[d]Diabetes: ICD-9 code 250; ICD-10 code E10-E14.
[e]Dementia: ICD-9 code 046.1, 290.0-290.4, 294, 331.0, 331.1, 331.5; ICD-10 code G30, F00-F03.
[f]Parkinsonism: ICD-9 code 332.0-332.1; ICD-10 code G20, G21.0-G21.4, G21.8-G21.9, G22, F02.3.
[g]Multiple sclerosis: ICD-9 code 340; ICD-10 code: G35.

status, daily physical activity and selected pre-existing medical conditions (such as angina, arrhythmia). At the community level, we have derived contextual variables such as average education and unemployment rate, using information from the 1996, 2001, 2006 and 2011 Canadian censuses. Selected characteristics of ONPHEC at cohort entry are presented in Table 2. Additional individual-level information for cohort members who participated in the population-based health surveys is provided in Appendix B (available as Supplementary data at *IJE* online).

Among ONPHEC members, approximately 27% (~1.4 million) died during the follow-up period 1996 to 2014, with cardiometabolic disease as the most common underlying cause of death (Figure 3). Long-term trends of the incidence of six selected chronic diseases are displayed in Figure 4. The incidence of dementia among females aged > 65 years increased markedly from 14.6 per 1000 in 1996 to 45.3 per 1000 in 2012, whereas the incidence of some other diseases such as parkinsonism and AMI remained relatively constant over time.

Selected environmental exposure data

We have developed a variety of environmental exposure data,[10,21,41–50] using state-of-the-art methods, including satellite-based remote sensing[44] and land-use regression models.[51] These exposure data are assigned spatially to cohort members using their annual six-character postal code addresses during follow-up. Six-character postal codes in urban areas represent the centroid of the blocks in which the cohort members live. We included exposures to major air pollutants, including fine particulate matter (particles with aerodynamic diameter <2.5 μm, $PM_{2.5}$), measures of oxidative potential for particulate matter (mean glutathione depletion and mean ascorbic acid depletion), ultrafine particles (<0.1μm), nitrogen dioxide and ozone. $PM_{2.5}$ has been associated with various health effects, such as increased cardiovascular mortality.[30] Oxidative potential is a novel measure of air pollution exposure, which provides an assessment of regional differences in the ability of $PM_{2.5}$ to cause oxidative stress, a mechanism thought to play an important role in air pollution health effects.[52,53] We are currently evaluating if regional differences in $PM_{2.5}$

Table 2. Individual- and neighbourhood-level characteristics of the **ON**tario **P**opulation **H**ealth and **E**nvironment **C**ohort (ONPHEC)

Variable	ONPHEC Cohort (N = 4,854,759) [a]
Individual risk factors at time of entry, 1996	
Age at entry (years)	54 ± 14
Male (%)	47
Income quintile	
1 (lowest)	18
2	20
3	20
4	20
5 (Highest)	22
Urban residence	83
Pre-existing comorbidity	
Hypertension	35
Diabetes	11
Acute myocardial infarction	2
Coronary heart disease	4
Stroke	1
Congestive heart failure	2
Dementia	1
Other personal risk factors from population-based health surveys, 1996-2008 [b]	
Marital status (%)	
Married (including common law)	63
Single	11
Separated, widowed, or divorced	26
Body mass index (kg/m^2)[c]	26 ± 5
<18.5 kg/m^2 (%)	2
18.5-24.9 kg/m^2 (%)	41
25.0-29.9 kg/m^2 (%)	37
≥ 30 kg/m^2 (%)	17
Missing	3
Education (%)	
Less than high school	22
High school	18
Beyond high school	58
Missing	2
Smoking status (%)	
Never smoker	27
Current smoker	24
Former smoker	41
Missing	8
Type of drinker (%)	
Regular drinker	62
Occasional and former drinker	34
Never drinker	4
Total daily consumption of fruits and vegetables (%)	
Consumed fruits and vegetables less than 5 times per day	44
Consumed fruits and vegetables more than 5 times per day	29

(continued)

Table 2. Continued

Variable	ONPHEC Cohort (N = 4,854,759) [a]
Missing	27
Energy expenditure in daily physical activity (kcal/kg/day) (%)[d]	
≥3.0 (active)	22
1.5 - 2.9 (moderate)	25
<1.5 (inactive)	51
Missing	2
Neighbourhood-level risk factors, 1996 [e]	
Mean household income (in Can$ 1000)	53 ± 16
Percentage > 15 years of age with less than high school education	33 ± 10
Percentage > 15 years of age without employment	9 ± 4

Values are percent or mean ± standard deviation.

[a]Based on data linkage across health administrative and chronic disease databases.

[b]Among a representative sample of respondents (N≈100,000) to the 1996/1997 cycle of the NPHS or the 2000/2001, 2003, 2005, or 2007/2008 cycles of the CCHS.

[c]Body mass index is the weight in kilograms divided by the square of the height in meters.

[d]Average daily energy expenditure of participants in their leisure activities (such as walking, bicycling). For each activity, energy expenditure was estimated using frequency and time per session and the value of metabolic energy cost expressed as a multiple of the resting metabolic rate.

[e]At the Canadian census tract level. Census tracts are small, relatively stable geographic areas that usually have a population between 2,500 and 8,000 persons. The legal age to work in Canada is 14 years. It is a standard practice in Canada Census to collect and report labour force information among those who are aged 15 years or above.

oxidative potential may contribute to regional differences in PM$_{2.5}$-associated health effects, using ONPHEC. Additional information on air pollutant exposure data is provided in Appendix C (available as Supplementary data at *IJE* online).

In addition, we have information on various meteorological conditions such as air temperature and precipitation from all weather stations across Ontario during the study period.[21] Previous studies have demonstrated that both short-term (in days) and longer-term (months or years) variations in temperature increase morbidity and mortality.[33,54] Further, spatial variations in traffic-related noise in Toronto, the largest city in Ontario, were derived using exposure surfaces developed from continuous measures of noise from population-based surveys conducted in 2012-13.[49] Additionally, we have on hand different measures of green space using land use/classification data and satellite-based normalized difference vegetation index (NDVI).[55] NDVI was formulated as the ratio between the difference between the near-infrared region and red reflectance to the sum of the two measures, with greater values reflecting more green space.[47] The NDVI data covering the
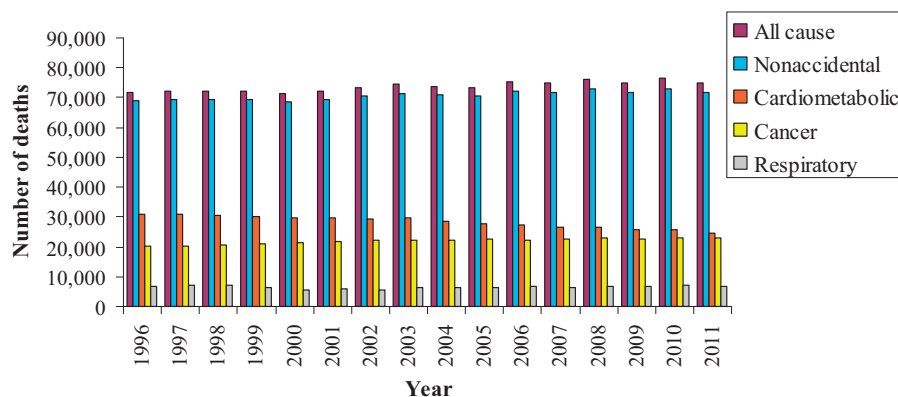
Figure 3. Annual number of deaths, by the underlying cause, in the **ON**tario **P**opulation **H**ealth and **E**nvironment **C**ohort (ONPHEC), 1996-2011.

entire province are available at 30-m and 500-m spatial resolutions. As new environmental exposure data become available, we will continue to link them to the cohort.

What statistical methods are used?

Novel statistical methods have been applied to the cohort. First, to examine the relationship between long-term exposure to environmental factors and incidence of chronic diseases, we employed multilevel spatial random-effects Cox proportional hazards models that were developed by one of us (R.T.B).[56] This model accounts for the possibility that health patterns among individuals living in the same or neighbouring communities are more similar than for individuals living farther apart, and that these patterns may not be completely explained by variables included in the model.[57]

Second, to assess the interactions between risk factors (e.g. environmental, behavioural) and to provide information that is more interpretable for public health decision makers, we have applied an additive hazards model.[58,59] With this modelling approach, interactions can be examined as a departure from additivity of the absolute effects (that is risk difference) of two exposures. Thus the coefficient of the interaction term can be interpreted as additional cases (per time) with each level of one exposure given a certain level of another exposure.[59]

Third, to control for important individual-level risk factors such as income and diet, which are often unavailable in large cohort studies using administrative databases, we apply an indirect adjustment method developed by ONPHEC investigators.[24] This is a flexible method that can readily accommodate any form of missing risk factors (continuous or categorical) and simultaneously adjust for multiple unmeasured variables. This method draws on information on the correlation between observed and missing risk factors from ancillary data sources (e.g. a similar population), and the relationship between the missing risk

factors and survival from existing literature. The indirect adjustment method depends largely on the representativeness of the ancillary information.[24] By drawing the ancillary data from a representative sample of the study cohort, ONPHEC allows for better control of unmeasured risk factors.

Fourth, we apply a flexible risk model to characterize the shape of exposure-response relations between environmental exposures and various chronic diseases.[60]

What have we found?

To date, we have examined the associations between ambient air pollution and the incidence of selected diseases. In one study, we followed ONPHEC members who participated in the population-based health surveys between 1996 and 2011. Among 62 012 adults, we observed that for every 10-μg/m^3 increase in exposure to PM$_{2.5}$, the incidence of diabetes increased by 11% [95% confidence interval (CI): 2-21%).[61] In another study, we reported that the onset of hypertension was associated with long-term exposure to PM$_{2.5}$ in Ontario, with an adjusted hazard ratio of 1.13 (95% CI: 1.05-1.22) per 10-μg/m^3 increase in PM$_{2.5}$.[20]

More recently, through record linkage of ONPHEC with the Enhanced Feedback For Effective Cardiac Treatment Study,[62] we followed up all newly admitted patients in Ontario having an AMI. We estimated that every 10-μg/m^3 increase in exposure to PM$_{2.5}$ was associated with a 22% increase in risk of non-accidental deaths, 43% increase in deaths from ischaemic heart disease and 64% increase in AMI-related deaths. These estimates translate to 12.4% of non-accidental deaths being attributable to PM$_{2.5}$ among AMI patients (Chen *et al.* 2015, submitted manuscript).

We also observed that exposure to ambient cold temperatures had a greater impact on premature deaths compared with that from high temperatures in Ontario, and that the largest impact of cold temperature was seen on
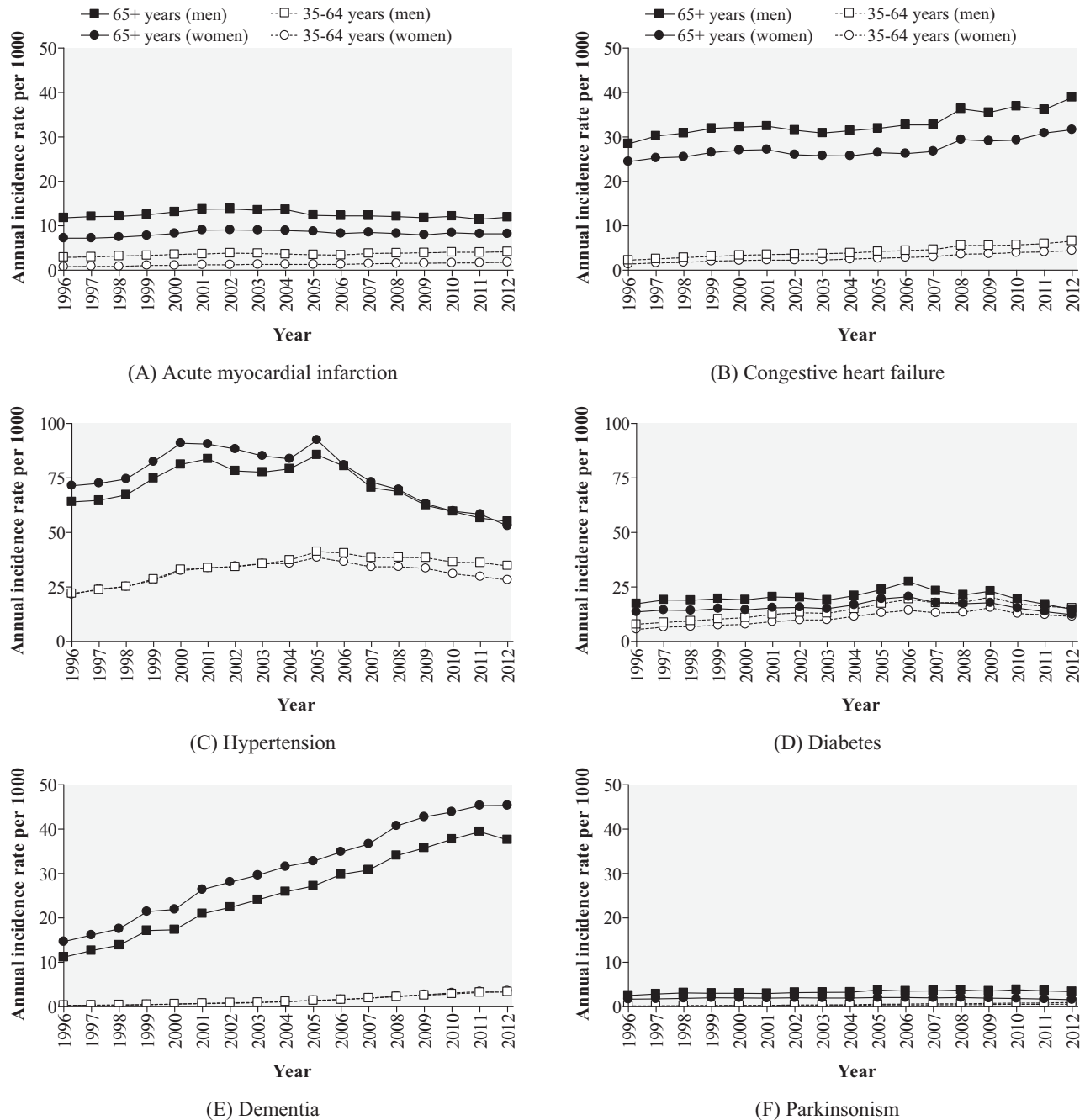
Figure 4. Annual incidence rates of six selected chronic diseases (A: acute myocardial infarction, B: congestive heart failure, C: hypertension, D: diabetes, E: dementia, F: parkinsonism) for the **ON**tario **P**opulation **H**ealth and **E**nvironment **C**ohort (ONPHEC), 1996-2012.

cardiovascular-related mortality, especially among individuals younger than 65 years.[21]

Furthermore, we are currently analysing the spatial and temporal trends of major neurodegenerative diseases, assessing the effectiveness of smog advisories on reducing cardiorespiratory morbidity and mortality, examining the combined effects of green space and air pollution on the long-term survival of patients living with cardiovascular diseases, quantifying the burden of low and high ambient temperatures on the risk of hospitalizations from

cardiovascular diseases and diabetes and evaluating the relationship between incident dementia and long-term exposure to air pollution.

What are the main strengths and weaknesses?

The major strengths of ONPHEC include its large size and inclusion of the entire Canadian-born population of Ontario who were 35 years or older in 1996 (∼4.9

million). This minimizes the potential for selection bias while greatly increasing generalizability. We also achieved virtually complete follow-up of all ONPHEC members by deterministic record linkage using unique, encoded identifiers. In addition, there are numerous population-based databases available in Ontario that contain incident diagnoses of major chronic diseases. Also, a wide range of environmental exposure data are available across Ontario. As new health information and environmental data become available, they can be incorporated into ONPHEC for future analyses. Furthermore, with a follow-up of 19 years, ONPHEC provides a time-efficient, low-cost way to study the environmental burden of chronic diseases. Last, given the generally good environmental conditions in Ontario and relatively large range in their exposures, ONPHEC offers a rare opportunity to investigate the health impacts of environmental exposures at relatively low levels.

This cohort has several limitations. First, similar to many large cohorts, ONPHEC does not have detailed information on some important personal risk factors such as smoking and body mass index (BMI) for all cohort members. However, with a unique two-stage cohort design in which health administrative data are supplemented with population-based health survey data, ONPHEC allows for better control of the potential influence of unmeasured personal risk factors on effect estimates. Second, the analysis of disease onset is limited to individuals who did not have a previous documented history of any of these events at baseline. Some of the cohort members may have experienced silent events. However, any resultant bias is expected to only slightly attenuate the risk estimates, because such misclassification of disease is unlikely to be differentially distributed by exposures to environmental risk factors such as air pollution. Third, as in many cohort studies, it is not possible to assess personal exposures to environmental risk factors. Other factors such as daily activity patterns may have an important influence on total personal exposure to environmental risk factors. However, a population survey reported that Canadian adults who resided in major cities during 1992-97 spent on average over 80% of their time each year at home (both indoors and outdoors).[63] Last, we were unable to track vital status and incidence of chronic diseases among individuals after they moved out of Ontario. Hence, we censored individuals if they became ineligible for health insurance.

Can I get hold of the data? Where can I find out more?

Collaborations are highly welcome, in particular collaborations that support linkage of ONPHEC with additional datasets on clinical outcomes, risk factors or environmental exposure data. Please contact the corresponding author with any inquiries: [hong.chen@oahpp.ca].

Supplementary Data

Supplementary data are available at *IJE* online.

Profile in a Nutshell

- The **ON**tario **P**opulation **H**ealth and **E**nvironment **C**ohort (ONPHEC) is a large, retrospective cohort in the province of Ontario, Canada, that was created in 2014 to investigate the independent and combined effects of various environmental exposures on the development and progression of chronic diseases.

- ONPHEC comprises virtually the entire Canadian-born population in Ontario who were 35 years or older in 1996 ($\sim$4.9 million) and combines multiple Big Data sources (e.g. large-scale health administrative databases, satellite-based environmental databases) and novel statistical methods.

- Follow-up extended until 2014 through individual-level record linkage to databases of hospitalizations, emergency department visits, physician office visits, prescription medications and mortality.

- A unique characteristic of ONPHEC is the two-stage cohort design: health administrative data are supplemented with population-based health survey data on major risk factors for chronic diseases from a representative sample of ONPHEC.

- ONPHEC collects information on incidence and mortality from major chronic diseases, including cardiovascular diseases, diabetes and neurodegenerative diseases.

- ONPHEC contains a variety of environmental exposure datasets developed by the investigators using state-of-the-art methods such as satellite-based remote sensing.

- Collaborations are highly welcome. Please contact the corresponding author with any enquiries: [hong.chen@oahpp.ca].

Acknowledgements

References

1. World Health Organization. *Global Status Report on Noncommunicable Diseases 2014.* http://www.who.int/nmh/publications/ncd-status-report-2014/en/ (6 November 2015, date last accessed).

2. Murray CJ, Vos T, Lozano R *et al*. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;**380**:2197–223.

3. U.S. National Institute on Aging and World Health Organization. *Global Health and Aging.* http://www.who.int/ageing/publications/global_health.pdf (6 November 2015, date last accessed).

4. World Health Organization. *Preventing Chronic Diseases: A Vital Investment.* http://www.who.int/chp/chronic_disease_report/en/ (6 November 2015, date last accessed).

5. Lim SS, Vos T, Flaxman AD *et al*. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;**380**:2224–60.

6. Pruss-Ustun A, Corvalan CF. *Preventing Disease Through Healthy Environments: Towards an Estimate of the Environmental Burden of Disease.* http://www.who.int/quantifying_ehimpacts/publications/preventingdisease/en/ (6 November 2015, date last accessed).

7. Lee AC, Maheswaran R. The health benefits of urban green spaces: a review of the evidence. *J Public Health* 2011;**33**:212–22.

8. Dadvand P, Nieuwenhuijsen MJ, Esnaola M *et al*. Green spaces and cognitive development in primary schoolchildren. *Proc Natl Acad Sci U S A* 2015;**112**:7937–42.

9. Reklaitiene R, Grazuleviciene R, Dedele A *et al*. The relationship of green space, depressive symptoms and perceived general health in urban population. *Scand J Public Health* 2014;**42**:669–76.

10. McMorris O, Villeneuve PJ, Su J, Jerrett M. Urban greenness and physical activity in a national survey of Canadians. *Environ Res* 2015;**137**:94–100.

11. Dadvand P, Villanueva CM, Font-Ribera L *et al*. Risks and benefits of green spaces for children: a cross-sectional study of associations with sedentary behavior, obesity, asthma, and allergy. *Environ Health Perspect* 2014;**122**:1329–35.

12. Gascon M, Triguero-Mas M, Martinez D *et al*. Residential green spaces and mortality: A systematic review. *Environ Int* 2015;**86**:60–7.

13. Doll R, Peto R, Boreham J, Sutherland I. Mortality from cancer in relation to smoking: 50 years observations on British doctors. *Br J Cancer* 2005;**92**:426–29.

14. Mooney SJ, Westreich DJ, El-Sayed AM. Commentary: Epidemiology in the era of big data. *Epidemiology* 2015;**26**:390–94.

15. Crouse DL, Peters PA, van Donkelaar A *et al*. Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: a Canadian national-level cohort study. *Environ Health Perspect* 2012;**120**:708–14.

16. Statistics Canada. *1996 Census of Canada.* https://www12.statcan.gc.ca/english/census96/data/popdwell/Table.cfm?T=101 (23 November 2015, date last accessed).

17. Tu JV, Chu A, Donovan LR *et al*. The Cardiovascular Health in Ambulatory Care Research Team (CANHEART): using big data to measure and improve cardiovascular health and healthcare services. *Circ Cardiovasc Qual Outcomes* 2015;**8**:204–12.

18. Tu K, Chen Z, Lipscombe LL. Prevalence and incidence of hypertension from 1995 to 2005: a population-based study. *CMAJ* 2008;**178**:1429–35.

19. Yeung DF, Boom NK, Guo H, Lee DS, Schultz SE, Tu JV. Trends in the incidence and outcomes of heart failure in Ontario, Canada: 1997 to 2007. *CMAJ* 2012;**184**:E765–73.

20. Chen H, Burnett RT, Kwong JC *et al*. Spatial association between ambient fine particulate matter and incident hypertension. *Circulation* 2013;**129**:562–9.

21. Chen H, Wang J, Li Q *et al*. Assessment of the impact of cold and hot temperatures on mortality in Ontario, Canada: population-based study. *CMAJ Open* 2016;**4**:E48–58.

22. Statistics Canada. *Canadian Community Health Survey (CCHS).* Version cited 20 April 2015, http://www.statcan.gc.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SDDS=3226&lang=en&db=imdb&adm=8&dis=2 (6 May, date last accessed).

23. Statistics Canada. *National Population Health Survey (NPHS).* Version cited 24 October 2013, http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3236 (4 April 2013, date last accessed).

24. Shin HH, Cakmak S, Brion O *et al*. Indirect adjustment for multiple missing variables applicable to environmental epidemiology. *Environ Res* 2014;**134**:482–87.

25. World Health Organization. *Ambient (Outdoor) Air Pollution Database, by Country and City.* Version cited July 2015, http://www.who.int/phe/health_topics/outdoorair/databases/cities/en/ (27 November 2015, date last accessed).

26. World Health Organization. *Air Quality Guidelines - Global Updated 2005.* http://www.who.int/phe/health_topics/outdoorair_aqg/en/ (23 August 2012, date last accessed).

27. Statistics Canada. *2001 Census of Canada.* http://www12.statcan.ca/english/census01/home/Index.cfm (1 June 2012, date last accessed).

28. Chan B. Supply of physicians' services in Ontario. *Hosp Q* 2000;**3**:17.

29. Block ML, Elder A, Auten RL *et al*. The outdoor air pollution and brain health workshop. *Neurotoxicology* 2012;**33**:972–84.

30. Brook RD, Rajagopalan S, Pope CA 3rd *et al*. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. *Circulation* 2010;**121**:2331–78.

31. Chen H, Goldberg MS, Villeneuve PJ. A systematic review of relation between long-term exposure to ambient air pollution and chronic diseases. *Rev Environ Health* 2008;**23**:243–96.

32. Hartig T, Mitchell R, de VS, Frumkin H. Nature and health. *Annu Rev Public Health* 2014;**35**:207–28.

33. Madrigano J, Mittleman MA, Baccarelli A *et al*. Temperature, myocardial infarction, and mortality: effect modification by individual- and area-level characteristics. *Epidemiology* 2013;**24**:439–46.

34. Turner LR, Barnett AG, Connell D, Tong S. Ambient temperature and cardiorespiratory morbidity: a systematic review and meta-analysis. *Epidemiology* 2012;**23**:594–606.

35. Correia AW, Peters JL, Levy JI, Melly S, Dominici F. Residential exposure to aircraft noise and hospital admissions for cardiovascular diseases: multi-airport retrospective study. *BMJ* 2013;**347**:f5561.

36. Hux JE, Ivis F, Flintoft V, Bica A. Determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care* 2002;**25**:512–16.

37. Tu JV, Austin PC, Walld R, Roos L, Agras J, McDonald KM. Development and validation of the Ontario acute myocardial infarction mortality prediction rules. *J Am Coll Cardiol* 2001;**37**:992–97.

38. Tu K, Campbell NRC, Chen ZL, Cauch-Dudek KJ, McAlister FA. Accuracy of administrative databases in identifying patients with hypertension. *Open Med* 2007;**1**:e18–26.

39. Butt DA, Tu K, Young J *et al*. A validation study of administrative data algorithms to identify patients with Parkinsonism with prevalence and incidence trends. *Neuroepidemiology* 2014;**43**:28–37.

40. Widdifield J, Ivers NM, Young J *et al*. Development and validation of an administrative data algorithm to estimate the disease burden and epidemiology of multiple sclerosis in Ontario, Canada. *Mult Scler* 2015;**21**:1045–54.

41. Jerrett M, Arain MA, Kanaroglou P *et al*. Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada. *J Toxicol Environ Health A* 2007;**70**:200–12.

42. Johnson M, Macneill M, Grgicak-Mannion A *et al*. Development of temporally refined land-use regression models predicting daily household-level air pollution in a panel study of lung function among asthmatic children. *J Expo Sci Environ Epidemiol* 2013;**23**:259–67.

43. Robichaud A, Menard R. Multi-year objective analyses of warm season ground-level ozone and PM$_{2.5}$ over North America using real-time observations and Canadian operational air quality models. *Atmos Chem Phys Discuss* 2013;**13**:13967–4035.

44. van Donkelaar A, Martin RV, Brauer M *et al*. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environ Health Perspect* 2010;**118**:847–55.

45. van Donkelaar A, Martin RV, Spurr RJ *et al*. Optimal estimation for global ground level fine particulate matter concentrations. *J Geophys Res Atmos* 2013;**118**:5621–36.

46. van Donkelaar A, Martin RV, Brauer M, Boys BL. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environ Health Perspect* 2015;**123**;135–43.

47. Villeneuve PJ, Jerrett M, Su JG *et al*. A cohort study relating urban green space with mortality in Ontario, Canada. *Environ Res* 2012;**115**:51–58.

48. Wheeler AJ, Smith-Doiron M, Xu X, Gilbert NL, Brook JR. Intra-urban variability of air pollution in Windsor, Ontario - measurement and modeling for human exposure assessment. *Environ Res* 2008;**106**:7–16.

49. Zuo F, Li Y, Johnson S *et al*. Temporal and spatial variability of traffic-related noise in the City of Toronto, Canada. *Sci Total Environ* 2014;**472**:1100–07.

50. Weichenthal S, Van RK, Goldstein A, Shekarrizfard M, Hatzopoulou M. Characterizing the spatial distribution of ambient ultrafine particles in Toronto, Canada: A land use regression model. *Environ Pollut* 2015;doi: 10.1016/j.envpol.2015.04.011.

51. Jerrett M, Arain A, Kanaroglou P *et al*. A review and evaluation of intraurban air pollution exposure models. *J Expo Sci Environ Epidemiol* 2004;**15**:185–204.

52. Boogaard H, Janssen NA, Fischer PH *et al*. Contrasts in oxidative potential and other particulate matter characteristics collected near major streets and background locations. *Environ Health Perspect* 2012;**120**:185–91.

53. Yang A, Wang M, Eeftens M *et al*. Spatial variation and land use regression modeling of the oxidative potential of fine particles. *Environ Health Perspect* 2015;**123**:1187–92.

54. Zanobetti A, O'Neill MS, Gronlund CJ, Schwartz JD. Summer temperature variability and long-term survival among elderly people with chronic disease. *Proc Natl Acad Sci U S A* 2012;**109**:6608–13.

55. Loveland T, Merchant J, Brown J, Ohlen D. Development of a land-cover characteristics database for the conterminous U. S. *Photogramm Eng Remote Sensing* 1991;**57**:1453–63.

56. Ma R, Krewski D, Burnett RT. Random effects Cox models: a Poisson modelling approach. *Biometrika* 2003;**90**:157–69.

57. Pope CA 3rd, Burnett RT, Thun MJ *et al*. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* 2002;**287**:1132–41.

58. Berrington de GA, Cox DR. Additive and multiplicative models for the joint effect of two risk factors. *Biostatistics* 2005;**6**:1–9.

59. Rod NH, Lange T, Andersen I, Marott JL, Diderichsen F. Additive interaction in survival analysis: use of the additive hazards model. *Epidemiology* 2012;**23**:733–37.

60. Burnett RT, Pope CA 3rd, Ezzati M *et al*. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environ Health Perspect* 2014;**122**:397–403.

61. Chen H, Burnett RT, Kwong JC *et al*. Risk of incident diabetes in relation to long-term exposure to fine particulate matter in Ontario, Canada. *Environ Health Perspect* 2013;**121**:804–10.

62. Tu JV, Donovan LR, Lee DS *et al*. Effectiveness of public report cards for improving the quality of cardiac care: the EFFECT study: a randomized trial. *JAMA* 2009;**302**:2330–37.

63. Leech JA, Nelson WC, Burnett RT, Aaron S, Raizenne ME. It's about time: a comparison of Canadian and American time-activity patterns. *J Expo Anal Environ Epidemiol* 2002;**12**:427–32.