A Spatial Assessment of Environmental Risk Factors for Lung Cancer in Canada: The Role of Air Pollution, Radon and Neighborhood Socioeconomic Status

by

Perry Wesley Hystad

A THESIS SUMBITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Health Care and Epidemiology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

June 2013

Abstract

In this dissertation I examined whether three exposures associated with the physical and social residential environment – specifically, ambient air pollution, radon and neighborhood socioeconomic status (SES) – are risk factors for the development of lung cancer in Canada. Throughout this dissertation I used the National Enhanced Cancer Surveillance System (NECSS), a large population-based case-control study conducted in eight Canadian provinces, including 3,280 incident lung cancer cases and 5,073 population controls.

In the first section of this dissertation, I developed methods to estimate ambient air pollution, both nationally and retrospectively, and applied these to 20 years of residential histories in the NECSS study. Epidemiological analyses showed that the odds of lung cancer incidence associated with a 10-unit increase in $PM_{2.5}$ ($\mu g/m^3$), NO_2 (ppb) and O_3 (ppb) were 1.29 (95% CI = 0.95-1.76), 1.11 (1.00-1.24), and 1.09 (0.85-1.39) respectively, indicating that ambient air pollution exposure is associated with lung cancer development in Canada.

In the second section, I used maps of radon concentration and potential in combination with the NECSS residential histories to estimate ecological radon exposures. A 50 Bq/m^3 increase in average health region radon concentration was associated with a 7% (-6-21%) increase in the odds of lung cancer and for every 10 years that individuals lived in high radon geological potential zones, the odds of lung cancer increased by 11% (1-23%). This study also indicated that risk mapping may be used to target population health prevention efforts for radon.

In the third section, I developed methods to estimate long-term exposure to neighborhood SES and applied these to the residential histories of the NECSS study. The odds of lung cancer cases residing in the most versus least deprived long-term neighborhood SES quintiles were significantly elevated and in the city sub-analysis remained significant (OR:

1.38 (1.01-1.88)) after adjusting for smoking and other lung cancer risk factors. Smoking behavior was the predominant partial-mediating pathway of the neighborhood effect.

Collectively, this dissertation contributes to the methodological literature on spatial exposure assessment and spatial epidemiology, as well as to the etiological evidence linking air pollution, radon and neighborhood SES to lung cancer risk.

Preface

This dissertation is composed of five research chapters (Chapters 2-6), each of which has been written as a stand-alone manuscript for publication in peer-reviewed journals. Three have been published (Chapters 2-4) and two have been submitted for publication (Chapters 5, 6). I developed the overall research plan for this dissertation, conducted all data analyses and prepared all manuscripts with the guidance of my committee and co-authors. This research was approved by the UBC Behavioral Research Ethics Board (certificate #H09-00772).

A large number of data sources and methodologies were used throughout this research; as such there are contributions from a large number of co-authors. Below are summaries of my contributions to each research chapter and those of each co-author.

CHAPTER 2: Creating National Air Pollution Models for Population Exposure Assessment in Canada

Hystad P., Setton E., Cervantes A., Poplawski K., Deschenes S., Brauer M., Martin, R., van Donkelarr, A., Lamsal, L., Jerrett, M., Demers, P. (2011). Creating National Air Pollution Models for Population Exposure Assessment in Canada. *Environmental Health Perspectives*.119:1123-1129.

I initiated this study, developed the modeling approach, conducted analyses and prepared the manuscript (contribution > 85%). This research was conducted as a component of Carex Canada, and Eleanor Setton, the lead of the environmental portion of the project, was involved in the conceptualization of modeling approaches and data formatting. Alejandro Cervantes, Karla Poplawski and Steeve Deschenes were research assistants for Carex Canada and compiled and formatted a number of the predictor variables that were used in the final national models. Michael Brauer and Paul Demers were involved in the conceptualization of

the modeling approaches and edited drafts of the final manuscript. Aaron van Donkelaar developed the satellite based fine particulate matter ($PM_{2.5}$) surface for Canada and Lok Lamsal developed the satellite based nitrogen dioxide (NO_2) surface. Randall Martin contributed to the development of both the $PM_{2.5}$ and NO_2 satellite estimates and edited drafts of the final manuscript. Michael Jerrett collected the Toronto land use regression data used in the national model evaluation and also edited drafts of the final manuscript.

CHAPTER 3: Spatiotemporal Air Pollution Exposure Assessment for a Canadian Population-Based Lung Cancer Case-Control Study

Hystad, P., Demers, P., Johnson, K.C., Brook, J., van Donkelaar, A., Lamsal., L., Martin, R., Brauer, M. (2012). Spatiotemporal air pollution exposure assessment for a Canadian population-based lung cancer case-control study. *Environmental Health*. 11:22.

I initiated this study, developed the modeling approach, conducted analyses and prepared the manuscript (contribution > 90%). Paul Demers and Michael Brauer were involved in the conceptualization of the historical modeling approach and edited drafts of the final manuscript. Jeff Brook provided the national Ozone surface and edited drafts of the final manuscript and Aaron van Donkelaar, Lok Lamsal, and Randall Martin provided the $PM_{2.5}$ and NO_2 Satellite derived surfaces and edited drafts of the final manuscript. Kenneth Johnson implemented the NECSS case-control study and edited drafts of the final manuscript.

CHAPTER 4: Long-Term Residential Exposure to Air Pollution and Lung Cancer Risk

Hystad, P., Demers, P.A, Johnson, K.C., Carpiano, R.M., Brauer, M. Long-Term Residential Exposure to Air Pollution and Lung Cancer Risk. *Epidemiology*. Accepted January 25th 2013.

I initiated this study, developed the modeling approach, conducted analyses and prepared the manuscript (contribution > 90%). Paul Demers, Richard Carpiano, and Michael Brauer were involved in the conceptualization of the project and edited drafts of the final manuscript. Kenneth Johnson implemented the NECSS case-control study and edited drafts of the final manuscript.

CHAPTER 5: Geographic Variation in Radon and Associated Lung Cancer Risk in Canada: Results from a Population-Based Lung Cancer Case-Control Study

Hystad, P., Brauer, M., Demers, P.A., Johnson, K., Setton, E., Cervantes, A., Poplawski, K., Whitehead, A., McFarlane, A., Nicol, A.M. Spatial Variation in Radon and Associated Lung Cancer Risk in Canada: Results from a Population-Based Lung Cancer Case-Control Study. Submitted for Publication.

I initiated this study, developed the modeling approach, conducted analyses and prepared the manuscript (contribution > 85%). Michael Brauer, Paul Demers, and Anne-Marie Nicol were involved in the conceptualization of the project and edited drafts of the final manuscript. Kenneth Johnson implemented the NECSS case-control study and edited drafts of the final manuscript. Eleanor Setton, Alejandro Cervantes and Karla Poplawski helped format the national radon survey and edited drafts of the final manuscript. Allan Whitehead and Alana McFarlane created the national radon potential map and edited drafts of the final manuscript.

CHAPTER 6: Neighborhood Socioeconomic Status and Individual Lung Cancer Risk: Evaluating Long-Term Exposure Measures and Mediating Mechanisms

Hystad, P., Carpiano, R.M., Demers, P.A, Johnson, K.C., Brauer, M. Neighborhood Socioeconomic Status and Individual Lung Cancer Risk: Evaluating Long-Term Exposure Measures and Mediating Mechanisms. Submitted for Publication.

I initiated this study, developed the modeling approach, conducted analyses and prepared the manuscript (contribution > 90%). Paul Demers, Richard Carpiano, and Michael Brauer were involved in the conceptualization of the project and edited drafts of the final manuscript. Kenneth Johnson implemented the NECSS case-control study and edited drafts of the final manuscript.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AOD	Aerosol Optical Depth
CHRONOS	Canadian Regional and Hemispheric O_3 and NO_x System
CD	Census Divisions
CSD	Census Subdivisions
CT	Census Tracts
EQDB	Environmental Quality Database
GIS	Geographic Information Systems
HPA	Hypothalamic-Pituitary-Adrenal Axis
HR	Hazard Ratio
IARC	International Agency for Research on Cancer
IDW	Inverse Distance Weighting
IQR	Interquartile Range
LUR	Land Use Regression
MISR	Multiangle Imaging Sectroradiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
NAPS	National Air Pollution Surveillance
NECSS	National Enhanced Cancer Surveillance System
NO_2	Nitrogen Dioxide
NO_x	Nitrogen Oxides
OMI	Ozone Monitoring Instrument
OR	Odds Ratio
O_3	Ozone
PAF	Population Attributable Fraction
PAH	polycyclic Aromatic Hydrocarbons

$PM_{2.5}$	Fine Particulate Matter
PM_{10}	Inhalable Particulate Matter
RR	Relative Risk
SES	Socioeconomic Status
SD	Standard Deviation
SHS	Second Hand Smoke
SIC	Standard Industrial Classification
SO_2	Sulphur Dioxide
TEOM	Tapered Element Oscillating Microbalances
TSP	Total Suspended Particulate Matter
VOC	Volatile Organic Compound

Acknowledgements

First, I must express my heartfelt gratitude to my amazing supervisors, Dr. Michael Brauer and Dr. Paul Demers, whose unwavering support, direction and expertise were invaluable throughout this research process. Both have been everything that one could ask for in a supervisor and more.

I must also thank my other committee members – Dr. Richard Carpiano's expertise in social epidemiology was extremely important to this dissertation and to furthering my interest and approach to studying the social environment and health. This research also could not have been conducted without Dr. Ken Johnson's work and extensive experience with the NECSS case-control study.

Special thanks must also be made to the University of British Columbia, the Bridge Program, the Roman M. Babicki family, the Michael Smith Foundation for Health Research, and the Canadian Institutes of Health Research for their generous financial support.

I was fortunate to start my PhD with a wonderful cohort and to meet many friends and co-workers along this long road. I am grateful for their encouragement and friendship, as well as for the late-night life-discussions that routinely occurred. Sharing this experience with you has made it all that much more rewarding.

I must also thank my family and friends for their unconditional love and support – as well as for forcing me *not* to work on my research.

Finally and most importantly, I would like to thank my wife for her unwavering support and continued encouragement through my studies – *It is not so much where we stand, as in what direction we are moving*. I love you.

Chapter 1

Introduction

Worldwide, lung cancer is one of the most commonly diagnosed cancers with approximately 1.4 million deaths annually, corresponding to 18% of all cancer deaths (Jemal et al., 2011). In Canada, there are approximately 25,600 new lung cancer cases and 20,200 lung cancer deaths annually, representing 14% of new cancers and 27% of all cancer deaths (CCS, 2012). The five year survival rate for lung cancer patients in Canada remains low at 13% for males and 19% of females (CCS, 2012).

The majority of lung cancer cases (between 80-90%) can be attributed to cigarette smoking (Danaei et al., 2005; Doll and Peto, 1981; Parkin et al., 2005). Lung cancer risk increases with both the intensity and duration of smoking, with latency periods of 20-40 years (Parkin et al., 2005; Tyczynski et al., 2003). Average and heavy smokers, respectively, have a ten-fold and twenty-fold increase in the risk of developing lung cancer compared with never-smokers (IARC, 2004). Second hand smoke (SHS) exposure is also a risk factor for lung cancer, with excess risks estimated at 15% and 25% for exposure to SHS at home and work, corresponding to population attributable fractions (PAF) for non-smokers of 4.2% and 9.7%, respectively (Sisti and Boffetta, 2012).

Lung cancer is a multi-factorial disease, and in addition to tobacco smoke there are a number of known and suspected lung cancer risk factors for both smokers and non-smokers. Lung cancer in never smokers is rising in Canada and now ranks as the tenth most common cause of cancer death (CCS, 2012). Major non-tobacco related risk factors include: cooking and heating smoke from indoor biomass combustion (Lim and Seow, 2012); a wide-range of occupational carcinogens, such as asbestos, arsenic, chromate and silica (Siemiatycki et al.,

2004); ionizing radiation (Boice JD, 1990); diets low in fruits and vegetables and high is meat consumption (Donaldson, 2004; Sandhu et al., 2001); low physical activity (Mao et al., 2003); arsenic in drinking water (Celik et al., 2008); residential radon exposure (Krewski et al., 2005); and ambient air pollution exposure (Chen et al., 2008). Importantly, despite the large number of known or suspected lung cancer risk factors besides tobacco smoke, only a small proportion of lung cancer cases can be explained in non-smokers, highlighting the need for further research in this area (Sisti and Boffetta, 2012).

Two of the most prominent environmental risk factors for lung cancer are ambient air pollution and radon exposure. The International Agency for Research on Cancer (IARC) recently classified diesel exhaust, a component of outdoor air pollution, as a group 1 human carcinogen (IARC, 2012a) and is evaluating ambient air pollution (as a mixture) in the upcoming volume 109 monograph (IARC, 2012b). It has been estimated that fine particulate matter ($PM_{2.5}$) air pollution may cause approximately 5% of the global mortality from lung cancer (Cohen et al., 2005). Radon is also a recognized group 1 human carcinogen (IARC, 2012c) and is a leading cause of lung cancer after tobacco smoking (Samet et al., 2009). For both of these risk factors, very little research has been conducted to-date in Canada.

In addition to environmental risks, large socioeconomic status (SES) gradients exist for lung cancer incidence in Canada (and worldwide), even after accounting for smoking behaviors (Mao et al., 2001). The environmental risk factors mentioned previously may contribute to this remaining SES gradient, or additional contextual SES conditions may influence an individual's lung cancer risk. Modern epidemiology tends to focus on individual-level risk factors in isolation from contextual factors (Diez-Roux, 2003) – and the study of lung cancer is an illustrative example. On the other hand, laws prohibiting SHS in public places have recently been effectively applied at a population level in many developed

countries to reduce smoking rates. While there has been an increase in interest surrounding residential social conditions as important upstream health determinants for many chronic health conditions and behaviors (Yen et al., 2009), little research has examined neighborhood SES and lung cancer risk. Advancing our understanding of the complex linkages between neighborhood SES and lung cancer risk may offer new opportunities for upstream interventions to both address the SES disparities seen for lung cancer and reduce the overall burden of lung cancer.

The purpose of this dissertation is to examine whether these three exposures associated with the physical and social residential environment – ambient air pollution, radon and neighborhood SES – are risk factors for lung cancer development in Canada. Throughout this dissertation, a population-based case-control study is used for epidemiological analyses, which includes 3,280 histologically-confirmed lung cancer cases and 5,073 population controls collected between 1994 and 1997 in eight Canadian provinces. This case-control study focused on assessing environmental risk factors for cancer and therefore collected comprehensive information on individual characteristics, including residential histories that form the basis of the spatiotemporal exposure assessments conducted for each risk factor. A major gap in the existing literatures on these potential risk factors is exposure misclassification and a substantial portion of this dissertation is therefore dedicated to conducting comprehensive, long-term exposure assessments that include residential histories and temporal and spatial exposure variability. This dissertation builds upon methods employed in previous studies by developing comprehensive prediction models to estimate long-term exposures to ambient air pollution, radon and neighborhood SES, and by estimating associations within a large population-based lung cancer case-control study after controlling for each environmental risk factor as well as a broad set of potential individual and geographic confounding factors.

The following is a brief review of the pertinent literature on lung cancer and air pollution, radon and neighborhood SES exposures. The state of evidence linking each risk factor to lung cancer is reviewed, along with mechanisms and limitations of the literature. Subsequent research chapters provide further literature reviews and rationale for each specific study.

1.1. Air Pollution and Lung Cancer

Ambient air pollution consists of a mixture of gases and particles that arise from multiple sources and is typically classified into common and hazardous air contaminants. Common air contaminates include fine and inhalable particulate matter ($PM_{2.5}$ and PM_{10}), nitrogen dioxide (NO_2), ozone (O_3), and sulphur dioxides (SO_2), while hazardous air pollutants include a wide-range of substances, such as heavy metals, volatile organic compounds (VOCs) and polycyclic aromatic hydrocarbons (PAHs). Many hazardous air pollutants are known or suspected human carcinogens (IARC, 2012c).

The overall health burden associated with air pollution is large. Globally, it is estimated that ambient (outdoor) $PM_{2.5}$ air pollution accounts for 3.2 million deaths and 3.1% of disability-adjusted life-years (Lim et al., 2012). In terms of lung cancer specifically, it is estimated that $PM_{2.5}$ causes approximately 5% of the global mortality from lung cancer (Cohen et al., 2005). Canada has relatively low ambient air pollution concentrations compared to the rest of the world; however, health effects associated with air pollution are seen even at very low concentrations and no safe level of air pollution has been identified (Pope et al., 2011).

1.1.1. Summary of the literature

There is growing evidence for a causal association between air pollution exposure and lung cancer development. IARC recently classified diesel exhaust, a component of outdoor air

pollution, as a group 1 carcinogen (carcinogenic to humans) (IARC, 2012b) and is now evaluating ambient air pollution (as a mixture) in the upcoming volume 109 monograph (IARC, 2012c). A meta-analysis of air pollution and epidemiologic studies of lung cancer published up to 2006 reported pooled RR estimates per 10 $\mu g/m^3$ increases in $PM_{2.5}$, NO_2 and SO_2 of 1.21 (1.10-1.32), 1.11 (0.99-1.24), and 1.12 (0.98-1.29), respectively (Chen et al., 2008). In terms of O_3, four studies were identified that examined lung cancer mortality and two studies that examined incidence, although not enough information was available to calculate a pooled estimate (due primarily to the small sample sizes of existing studies).

Since this systematic review, evidence has continued to accumulate to support an association between lung cancer and exposure to ambient air pollution. An analysis of never-smokers in the American Cancer Society Cancer Prevention Study II cohort, based on 26 years of follow-up, reported a RR of 1.19 (0.97-1.47) per 10 $\mu g/m^3$ increase of $PM_{2.5}$ (Turner et al., 2011b). Similarly, extended follow-ups of the Harvard Six Cities Study from 1974-1998 (Laden et al., 2006) and from 1974-2009 (Lepeule et al., 2012) observed a RR of 1.27 (0.9-1.69) and 1.37 (1.07-1.75) per 10 $\mu g/m^3$ increase of $PM_{2.5}$, respectively. In a large cohort of men in the U.S. trucking industry, RR's for $PM_{2.5}$, PM_{10}, NO_2 and SO_2 of 1.07 (0.97-1.17), 1.05 (0.95-1.25), 1.09 (0.95-1.25) and 1.07 (0.96-1.20) per inter-quartile range increase in estimated residential exposure (excluding long-haul drivers) were observed (Hart et al., 2011). The Danish Diet, Cancer and Health cohort specifically examined traffic-related pollution and observed a RR for the highest compared with the lowest quartile of nitrogen oxides (NO_x) exposure of 1.30 (1.05-1.61) and a RR for living within 50m of a major road (>10,000 vehicles/day) of 1.21 (0.95–1.55) (Raaschou-Nielsen et al., 2011). Alternatively, a Dutch cohort did not observe associations with black smoke exposure (RR: 1.03 (0.88-1.20) per 10 $\mu g/m^3$) or with NO_2 and $PM_{2.5}$ (Beelen et al., 2008). One of the few Asian cohort studies

conducted was in Japan, and reported RR's associated with a 10-unit increase ($\mu g/m^3$) in area-level $PM_{2.5}$, SO_2 and NO_2 of 1.24 (1.12–1.37), 1.10 (1.05–1.14), and 1.09 (1.05–1.14), respectively (Katanoda et al., 2011). There have been two case-control studies conducted since the Chen et al. (2008) systematic review. A hospital-based case-control study in Northern Spain compared individuals living near industry to those that did not and reported an OR of 1.49 (0.93-2.39) for all lung cancer subtypes combined and a statistically significant association for small cell carcinomas (López-Cima et al., 2011). A case-control study was also conducted in the city of Windsor, Ontario that reported an OR for men exposed to NO_2 concentrations >28.5 $\mu g/m^3$ compared to < 21 $\mu g/m^3$ of 5.49 (1.04-29.0); however, this included only 9 cases and 5 controls. For women an OR of 1.33 (0.26-5.64) was observed based on 9 cases and 12 controls. Lower NO_2 concentration classes had more statistical power and also showed increased though non-significant associations (Band et al., 2011).

1.1.2. Biological mechanisms

The associations observed between ambient air pollution and lung cancer are supported by several biological mechanisms. For particles, there is a general consensus that once deposited in the lungs, particles generate reactive oxygen/nitrogen species that trigger an inflammation-related response that damages cellular proteins, lipids, membranes, and DNA, which can then lead to lung cancer development (Donaldson and Stone, 2003; Ghio et al., 2000). This process can be related to the surface composition and structure of particles or to their overall shape, size and composition (Knaapen et al., 2004). Gaseous air pollutants (e.g. O_3 and NO_2) are also potent oxidants, either through direct effects on lipids and proteins or indirectly through the activation of intracellular oxidant signaling pathways (Rahman and MacNee, 2000). Generally, the influence of NO_2 on oxidant pathways is thought to be less

potent than O_3 (Brunekreef and Holgate, 2002); however, it is important to note that the use of NO_2 in epidemiological studies may be acting as a marker for other traffic-related pollutions, many of which are known carcinogens.

1.1.3. Limitation of existing research

Evidence has continued to accumulate to support an association between lung cancer and exposure to ambient air pollution, yet several uncertainties remain. A wide-range of risk estimates has been reported for air pollution exposure and lung cancer risk, especially for pollutants other than $PM_{2.5}$. Chen et al. (2008) identified significant heterogeneity in estimates from existing studies of NO_2 and O_3, due partially to a lack of statistical power as few studies had been conducted. A growing interest in the literature has also emerged examining intra-urban gradients in air pollution (Health Effects Institute, 2010), primarily from traffic-related emissions; however, relatively few studies have been conducted that examine traffic emissions and lung cancer risk (Beelen et al., 2008; Raaschou-Nielsen et al., 2011, 2010). Other common air pollutants, particularly O_3, also remain of interest although only limited studies have been conducted (Beeson et al., 1998; Jerrett et al., 2005; Pope et al., 2002).

Potential confounding remains an important issue for air pollution and lung cancer studies, due to the possible association between air pollution exposure and SES – and the large number of lung cancer risk factors that are associated with SES (Sidorchuk et al., 2009). The spatial association between within-city air pollution exposures and SES is especially problematic as a number of studies have demonstrated that neighborhoods of low SES have higher levels of ambient air pollution (Buzzelli and Jerrett, 2004; Gunier et al., 2003; Neumann et al., 1998; Perlin et al., 2001). This spatial association requires epidemiological studies of lung cancer to include covariates on individual SES and detailed smoking histories.

In addition, covariates should also be included for other potential confounding factors, including second hand smoke exposure, diet, physical activity, and occupational exposures to lung carcinogens as these are risk factors for lung cancer that may also be associated with SES and hence air pollution exposures.

Recent literature has also strengthened the evidence that lung cancer is a heterogeneous cancer such that different histological subtypes have different etiological factors (Pesch et al., 2012). The few studies that have examined air pollution exposure and lung cancer risk by histological subtypes have reported no clear associations (Barbone et al., 1995; Chen et al., 2009; Katsouyanni et al., 1991; Liaw et al., 2010) but it is probable that risks associated with air pollution exposure may vary by histological subtype, given the differences seen in the risks from smoking (smoking is most strongly related to squamous cell carcinoma, followed by small cell carcinoma, with adenocarcinoma the most common subtype in never smokers (Pesch et al., 2012)). In addition, evidence from occupational (Villeneuve et al., 2011b) and animal studies (Nagy et al., 2005) also suggest that it is probable that risks associated with air pollution exposure may vary by histological subtype.

An overarching uncertainty in the lung cancer and air pollution literature is exposure misclassification and the effects this may have on epidemiological findings. Exposure misclassification is a particular concern for lung cancer epidemiology because of the long latency periods associated with lung cancer development and the potential for changes in air pollution levels as well as residential mobility during this time. Residential mobility data are therefore required to conduct long-term air pollution exposure assessments, but due to the difficulties in obtaining this information, residential location at study entry or at time of diagnosis remains the most common method for estimating air pollution exposure. To fully capitalize on residential histories, corresponding air pollution concentration estimates are also

required over the spatiotemporal study domain. To date, the association between air pollution and lung cancer has been examined using a variety of study periods and exposure assessment approaches – the most common being aggregating air pollution monitoring levels within cities or defined areas at time of study-entry (Dockery et al., 1993; Katanoda et al., 2011; Laden et al., 2006; Pope et al., 2002). While more robust methods are now being developed for estimating both between and within city air pollution concentrations over large geographical areas (for use in large multicity studies) as well as historical concentrations (Beelen et al., 2007; Paciorek et al., 2009; Hart et al., 2009; Yanosky et al., 2008), much more work is needed in this area.

1.2. Radon and Lung Cancer

Radon is a colorless, odorless, naturally occurring radioactive gas released from the breakdown of Uranium in soils. Uranium is present throughout the earth's crust and radon exposure can therefore be widespread in areas with certain geological characteristics. Exposure to high levels of radon occurs when it enters buildings (primarily through cracks in the foundation) and accumulates due to poor air exchange. Radon is recognized as a human carcinogen by the International Agency for Research on Cancer (IARC, 2012c), greatly increases the risk of lung cancer in smokers due to synergistic effects (Lantz et al., 2013) and is a major cause of lung cancer for individuals who have never smoked (Darby, 2005; IARC, 2012c; Krewski et al., 2005; Samet et al., 2009). Globally, it is estimated that approximately 100,000 lung cancer deaths are associated with radon exposure annually (Lim et al., 2012).

1.2.1. Summary of the literature

The initial evidence that radon may be associated with lung cancer came from studies of miners. In a pooled analysis of eleven miner studies, including 65,000 male miners and

2,700 lung cancer deaths, it was found that approximately 40% of all lung cancer deaths may be attributable to radon exposure, with 70% of lung cancer deaths in never-smokers, and 39% of lung cancer deaths in smokers (Lubin et al., 1995). This evidence led to the concern that residential radon exposure may be a risk factor for lung cancer in the general population.

To date, there have been 22 case-control studies of residential radon and lung cancer risk conducted, including nine studies in North America (of which one was in Canada), 13 in Europe, and two in China. These studies have generally reported positive associations between lung cancer risk and residential radon exposure, but some have reported results indicating no association. Due to the small sample sizes of these individual case-control studies, three pooled analyses have been conducted of the North America, European and Chinese data. The North American pooled analysis of seven studies reported an 11% (0-28%) increase in lung cancer risk per 100 Bq/m^3 increase in residential radon concentrations (Krewski et al., 2005). For subjects who had resided in only one or two houses in the 5–30 year exposure window and who had residential radon measurements for at least 20 years, there was an 18% (2-43%) increase in lung cancer risk per 100 Bq/m^3. The meta-analysis of 13 European studies found an 8.4% (3.0-15.8%) increase in lung cancer risk per 100 Bq/m^3 increase in measured radon (Darby et al., 2005). After correction for the dilution caused by uncertainties measuring and extrapolating radon concentrations, the increase in risk was estimated to be 16% (5-31%) per 100 Bq/m^3 increase in radon exposure. The pooled results of the two Chinese studies showed similar outcomes with an increased lung cancer risk of 13.3% (1-36%) per 100 Bq/m^3 of measured radon and 32% (7-91%) per 100 Bq/m^3 for subjects residing in the current home for 30 years or more (Lubin et al., 2004). For these existing studies, the US Environmental Protection Agency estimates that, at a radon level of 148 Bq/m^3 the lifetime risk of radon induced lung cancer death for never-smokers is 7 per 1000, compared with 62 per 1000 for

ever-smokers (EPA, 2012).

Only one residential radon study has been conducted in Canada and was limited in geographic scope. This case-control study was conducted in Winnipeg from 1983-1990 and included 738 individuals with histologically confirmed lung cancer and 738 controls. Radon dosimeters were placed in all residences in which the study subjects had reported living within the Winnipeg metropolitan area for at least 1 year. After adjusting for cigarette smoking and education, no increases in the odds of lung cancer or any of the histological subtypes of lung cancer were observed with increasing radon measurements (Létourneau et al., 1994).

An alternative exposure assessment approach which has the goal of increasing study population size and scope has recently been used by two epidemiological studies conducted in the US and Denmark that used maps and spatial prediction models to estimate long-term residential radon concentrations in larger population samples. Within the Cancer Prevention Study-II cohort, average radon measurements were assigned to 811,961 participants (3,493 lung cancer deaths) based on their zip code at study entry and average county-level radon concentrations (n=2,754) (Turner et al., 2011a). Both short-term and long-term indoor radon monitoring data were used along with a variety of geological, soil, meteorological, and housing data to predict county-level mean residential radon concentrations. In the fully adjusted model, a 100 Bq/m^3 increase in radon was associated with a 15% (1-31%) increase in lung cancer mortality and participants with mean radon concentrations above the EPA guideline value (148 Bq/m^3) experienced a 34% (7-68%) increase in risk (Turner et al., 2011a). In the Danish Diet, Cancer and Health cohort, 57,053 persons were recruited during 1993-1997 and followed until 2006, resulting in 589 lung cancer cases. Residential addresses from 1971-2006 were used with predictive radon models to estimate long-term radon exposure. The predictive radon model included nine explanatory variables (geographic

location, soil type, residential type, floor level, basement and building materials) and was able to predict 40% of measured radon concentrations in model evaluation (Andersen et al., 2007). An incidence rate ratio of 1.04 (0.69–1.56) per 100 Bq/m^3 increase in radon was reported, and 1.67 (0.69–4.04) among non-smokers (Bräuner et al., 2012).

1.2.2. Biological mechanisms

The specific mechanisms by which radon influences lung cancer development are well characterized. The decay of radon results in polonium-218 and polonium-214 that emit alpha radiation. Alpha radiation is classified as high linear energy radiation, which has low penetration distance but transfers high energy resulting in a large number of ionizing events (Sethi et al., 2012). Radon decay products may attach to particulates or aerosols in the air and get inhaled into the lungs, and subsequently cause DNA damage when radiation is released, either by a direct hit to DNA or by the creation of reactive oxygen species (Narayanan et al., 1997). The radiation damage by alpha particles is also seen to extend beyond the directly irradiated cells to surrounding cells, resulting in damage and contributing to tumor genesis (Azzam and Little, 2004). These mechanisms of radon carcinogenesis have been clearly demonstrated in experimental studies on animals (IARC, 2012c).

1.2.3. Limitations of existing research

While there have been numerous case-control studies conducted on residential radon and lung cancer risk, the majority of these studies suffer from limited statistical power arising from small sample sizes, small geographic coverage, and exposure misclassification from extrapolating short-term residential measurements to estimate long-term exposures, potentially for different households. Numerous ecological studies have been conducted, but causal inference is severely limited when individual-level data on confounding factors are not

available (Stidley and Samet, 1994). Only two studies, reviewed previously, have used alternative exposure methods to estimate ecological radon measures within an epidemiological study of individual data (Bräuner et al., 2012; Turner et al., 2011a).

In Canada, there is also a need for more studies on residential radon exposure and associated lung cancer risks. While there is good evidence that radon is causally associated with lung cancer, national-level studies are important to estimate attributable disease burden and to develop population health policy. In Canada, less than 30% of the population is able to describe radon as a health hazard, and only 5% of individuals have tested their homes for radon (Statistics Canada, 2010). Currently, there are no official radon maps for Canada, despite most developed countries having published radon risk maps that indicate high/low areas based on radon measurements or radon potential information (Tollefsen et al., 2011). Geographical targeting of population prevention initiatives could increase the awareness of radon as a health hazard and also increase the cost-effectiveness of radon prevention options, which are often criticized for being too costly (Gray et al., 2009).

1.3. Neighborhood Socioeconomic Status and Lung Cancer

Over the last decade, there has been a resurgence of interest in the role that place-based social conditions, in particular residential neighborhoods, have in shaping individual health outcomes. This resurgence has been driven by several interrelated trends within population health, public health and epidemiology, including: the realization that solely-individual determinants of health are insufficient; a growing interest in explaining health inequalities; and the need to consider prevention that targets contextual environments rather than only the individual (Diez Roux and Mair, 2010). Given the large socioeconomic status (SES) gradients in lung cancer incidence in Canada (and worldwide), even after accounting for smoking

behaviors (Mao et al., 2001), a better understanding of the complex linkages between neighborhood SES and lung cancer risk may offer new opportunities for upstream interventions to address the smoking and non-smoking related SES disparities seen for lung cancer as well as for reducing the overall burden of lung cancer.

1.3.1. Summary of the literature

Neighborhood SES is used to capture inequality in area-level SES contextual conditions over and above residents SES (e.g. disadvantaged neighborhoods have fewer resources and present residents few opportunities) (Ross and Mirowsky, 2008) and is one of the most common neighborhood social context measures used throughout the literature. SES is a complex multidimensional construct comprising diverse socioeconomic factors capturing economic resources, power, and/or prestige (Braveman, 2005); however, neighborhood SES is often measured using single or composite measures from readily available census data (e.g. education, income, employment, and ethnicity characteristics) that are aggregated to census administrative boundaries (which likely do not correspond to actual neighborhood boundaries).

Despite the use of often crude indicators of neighborhood SES, there is now growing recognition of the independent influence of neighborhood SES on a broad range of health outcomes (e.g. Bird et al., 2010; Major et al., 2010; Diez-Roux et al., 2001; Wight et al., 2009). A systematic review and meta-analysis of neighborhood effects on individual mortality found that significantly higher mortality was present among individuals living in areas with low SES, after accounting for individual SES measures (Meijer et al., 2012b). The specific outcomes included in this analysis were all-cause mortality (26 studies), mortality from cardiovascular/ischemic heart diseases (13 studies), and mortality from all cancers or cause-

specific cancer mortality (11 studies). Importantly, only three studies were identified that examined cancer incidence. Given the importance of cancer to the total chronic disease burden, the influence of neighborhood SES on cancer risk represents a large and important gap in the neighborhood context literature. Here, I review the literature examining neighborhood context and lung cancer risk, and also include other cancer-sites due to the small number of lung cancer studies available.

A total of four studies were found that examined the influence of neighborhood SES on lung cancer risk after accounting for important individual-level characteristics (at least income or education, and smoking behavior). Analysis of a Swedish cohort study reported no association between neighborhood median income and lung cancer incidence, although dose-response gradients were seen for population density (Chaix et al., 2006). In Denmark, a population-based cohort study found that lung cancer incidence was significantly decreased in areas with the lowest unemployment (HR: 0.88; 95% CI: 0.84-0.92) (Meijer et al., 2012a). A similar registry-based cohort study in Helsinki, Finland, found no associations between neighborhood SES characteristics and lung cancer risk for men 25-64 years of age; however, men over 64 years and living in neighborhoods with the highest percent of manual workers compared to the lowest had a RR of 1.32 (0.99-1.75) (Martikainen et al., 2003). In Canada, a case-control study in Montreal examined SES and lung cancer risk using 1,203 lung cancer cases and 1,513 controls with neighborhood context derived from census data for residences at the time of study entry (Matukala et al., 2012). Strong associations were observed between neighborhood income and lung cancer incidence after controlling for a number of individual-level variables; however, once comprehensive adjustment for smoking was included (i.e. smoking status, smoking pack-years and years since cessation) the association between the lowest versus the highest categories of neighborhood mean income disappeared (OR: 0.97;

95% CI: 0.51-1.86), suggesting that smoking was a key mediator of the neighborhood SES and lung cancer association.

Additional literature also exists supporting associations between neighborhood SES and other cancer sites, although many of these studies examine mortality rather than incidence (e.g. Major et al., 2010; Reitzel et al., 2012; Sundquist et al., 2012; Waitzman and Smith, 1998). It is important to examine cancer incidence, rather than mortality, as even for lung cancer (with low survival rates), there are differences in health care treatment by neighborhood SES (Earle et al., 2000).

The small number of non-lung cancer studies identified show associations between neighborhood SES and cancer incidence; although directions of the effect vary by cancer site. Generally prostate cancer incidence was found to be higher in lower SES areas (Meijer et al., 2012a; Sanderson et al., 2006), while breast cancer incidence was higher in higher SES areas (Robert et al., 2004; Webster et al., 2008). These results, however, were not always consistent, as a large Danish cohort study found no association between neighborhood unemployment and breast cancer (Meijer et al., 2012a). One study examined colorectal cancer incidence and observed higher risks (RR: 1.19; 95% CI: 1.08-1.32) for those living in the most deprived neighborhood SES quintile compared to the least (Doubeni et al., 2012). In the Health and Retirement Study, neighborhood features for adults aged 55 years or older were examined against all cancers, and after adjustment for a number of individual characteristics living in higher-crime areas was associated with greater chances of developing cancer for both men (OR: 1.31; 95% CI: 1.10-1.56) and women (OR: 1.25; 95% CI: 1.04-1.52) (Freedman et al., 2011). While these studies were conducted for cancer sites with very different etiology than lung cancer, they suggest that neighborhood SES may play a role in the development of cancer.

1.3.2. Pathways linking neighborhood socioeconomic status and lung cancer

Unlike ambient air pollution and radon exposures, which are well defined and relatively well measured exposures, neighborhood SES is used to represent a latent neighborhood construct that may influence lung cancer risk through a number of different pathways. A recent review of these pathways highlights that most research has focused on neighborhood influences on behaviors and psychosocial processes (Diez-Roux and Mair, 2010); however, environmental (and occupational) exposures may also play a role in the neighborhood SES effect. While a comprehensive review of these pathways is outside the scope of this literature review, I provide an overview of the pathways that could potentially link neighborhood SES to lung cancer risk.

1.3.2.1. Health behaviors

Neighborhood SES may influence lung cancer risk through a number of health behaviors. In particular, the effects of neighborhood SES may be mediated through smoking behaviors, due to the large risk posed by smoking to lung cancer (Pesch et al., 2012), and the fact that smoking, as well as smoking cessation, has been associated with neighborhood SES (e.g., Duncan et al., 1999; Giskes et al., 2006; Miles, 2006; Pickett and Pearl, 2001). For example, in a longitudinal study in Brisbane, Australia, after adjustment for individual-level socioeconomic factors, the probability of quitting smoking was lower for residents of disadvantaged neighborhoods (9.0–12.8%) compared to advantaged neighborhoods (20.7–22.5%) (Turrell et al., 2012). In a review of the pathways linking neighborhood context and smoking, two general categories were identified: *place-based practices* (including social capital, contagion, crime, disorder and related stressors) and *place-based regulations* (including smoking cessation policies, tobacco retailing and availability, tobacco advertising,

and urban renewal) (Pearce et al., 2012). Second hand smoke exposure may also play a mediating role in neighborhood influence, but has not been examined to date.

More broadly, neighborhood SES may influence all health-behaviors (including smoking, diet, and physical activity) through: (1) differential resources and opportunity structures within neighborhoods, such as the availability of healthy food retailers, or physical activity opportunities (e.g., Black et al., 2011; Boone-Heinonen et al., 2011; Frohlich et al., 2002); (2) prevalent norms and attitudes within neighborhoods regarding health-behaviors (e.g., Annear et al., 2009; Curry et al., 1993; Karasek et al., 2012); (3) local social networks and community belonging that support or hinder health-behaviors (e.g., Carpiano, 2007, 2008) as well as health-behaviour change (Hystad and Carpiano, 2012); and (4) disordered and stressful neighborhood environments that may lead to unhealthy coping behaviors (e.g., alcohol consumption) and worse general health (e.g., Ellaway & Macintyre, 2009; Weden et al., 2008).

1.3.2.2. Psychosocial influences and stress

Disordered and stressful neighborhood environments may also be associated with chronic psychosocial stress leading to allostatic load. Specifically, the persistent activation of the hypothalamic-pituitary-adrenal (HPA) axis in the chronic stress response has been shown to impair the immune response and may contribute to the development and progression of some types of cancer (Reiche et al., 2004). Applications of the social stress model generally treat SES as an attribute of the individual; however, neighborhood SES is also important to an individuals' health (Matheson et al., 2006). At the neighborhood level, individuals are differentially exposed to stressors, and equipped with different resources to combat these stressors as a function of their SES. The combination of greater exposure to stressors and

fewer resources to cope (the differential vulnerability hypothesis) may result in a deterioration of mental and physical health statuses (Elliott, 2000). Biomarkers of allostatic load have been found to be increased among residents of lower SES neighborhoods (Bird et al., 2010; Finch et al., 2010) and there is a growing interest among researchers in the role played by neighborhood SES in the stress process and resulting implications to health (Elliott, 2000).

1.3.2.3. Environmental and occupational exposures

There is an increasing need to disentangle physical environmental influences (e.g. air pollution, radon, noise, etc.) on health from other spatially clustered health determinants (e.g. stress, physical activity, smoking, etc.), as well as in assessing their possible interactions (Clougherty and Kubzansky, 2010; Gee and Payne-Sturges, 2004; Morello-Frosch and Jesdale, 2006; O'Neill et al., 2007, 2003). The potential for confounding, effect modification and mediation from spatially clustered physical and social contextual characteristics is an important area of research and has not yet received adequate attention. Air pollution is most advanced in this regard, due to potential clustering of high air pollution with low SES (Clougherty and Kubzansky, 2009); however, even for radon health effects there may be important socioeconomic gradients due to the synergistic effects with smoking and higher smoking rates in low socioeconomic groups.

The environmental justice paradigm has been developed for some time and presents a framework to examine how neighborhood SES may increase lung cancer risk through increased exposure to environmental hazards (e.g., Havard et al., 2009; Jerrett et al., 2001). This framework suggests that population subgroups of lower SES may be exposed to higher levels of environmental hazards as well as be more susceptible to associated health effects. A large number of ecological studies have demonstrated that low SES neighborhoods have

higher levels of ambient air pollution, primarily from traffic and industrial emissions (Buzzelli and Jerrett, 2004; Gunier et al., 2003; Neumann et al., 1998; Perlin et al., 2001), but the association is highly dependent on the scale of analysis and cannot be generalized to all situations (Havard et al. 2009). In addition to ambient air pollution, it is also hypothesized that individuals residing in low SES neighborhoods are disproportionately exposed to lower quality housing and therefore are exposed to more diverse environmental hazards (Adamkiewicz et al., 2011).

How occupational exposures may contribute to the neighborhood SES effect has received little attention, despite occupation being a major determinant of health and a measure itself of neighborhood SES. In the European Prospective Investigation into Cancer and Nutrition (EPIC study), occupational exposures explained 14% of the individual socioeconomic inequalities in lung cancer incidence after adjustment for smoking and fruits and vegetables consumption (Menvielle et al., 2010). This study highlights the potential importance of occupational exposures to socioeconomic gradients in lung cancer; however, whether such a relationship applies to neighborhood SES has yet to be examined. For example, living in low SES neighborhoods may shape occupational opportunities towards working class or "blue collar" jobs that are associated more frequently with exposure to cancer causing substances (Evans and Kantrowitz, 2002). While Martikainen et al. (2003) demonstrated the important connection between neighborhood context, occupational composition (i.e. proportion of manual labor workers) and lung cancer risk, much more research is needed to understand the relationship between individual and neighborhood SES, occupations and lung cancer risk.

1.3.3. Limitations of existing research

The influence of neighborhood SES on lung cancer risk (and cancer risk in general) represents a large and important gap in the neighborhood context literature. This gap may be due to two important methodological challenges that face neighborhood context and cancer research: (1) neighborhood exposure assessment methods; and (2) the challenge of testing the numerous pathways through which neighborhood SES may influence cancer risk.

First, latency periods of 20-50 years are associated with lung cancer development and measures of long-term neighborhood SES context are therefore needed to examine their influence on lung cancer. A review of neighborhood studies of older adults found that ten of the thirty-three identified studies accounted for respondents' length of residence in a neighborhood, and only one study accounted for changes in neighborhood environments (Yen et al., 2009). Several studies (primarily conducted in Scandinavia) have examined contextual effects over the life-course using linked administrative databases and reported mixed findings (Carson et al., 2007; Clarke et al., 2013; Curtis et al., 2004; Leyland and Næss, 2009; Ohlsson and Merlo, 2011); however, these studies tended to use larger geographic areas than neighborhoods. For example, Ohlsson and Merlo (2011) examined the amount of variance that was present for four health outcomes (one being all cancer mortality) at the district level over a 35 year period in Scania, Sweden. Overall, only a small proportion of the variance (<6.5%) for all health outcomes was explained at the district level. Alternatively, Naess et al. (2008) examined area effects in Oslo Norway using 1960, 1970, 1980, and 1990 election area measures and found evidence for a cumulative effect for cardiovascular disease. Other studies have included neighborhood context at multiple time points and observed lagged effects, but much of this research has examined only early life exposures and outcomes (Lloyd et al., 2010; Sampson et al., 2008; Wheaton and Clarke, 2003). The inclusion of longitudinal

measurements of neighborhood context has therefore been identified by scholars as an important next step for furthering research on neighborhood context and health effects (Clarke et al., 2013; Kawachi and Subramanian, 2007; Murray et al., 2010; Sampson et al., 2002), which is especially relevant for cancer.

The second major challenge facing neighborhood studies of cancer are the difficulties in identifying pathways through which neighborhood SES may influence cancer risk. Though studies have found that neighborhood SES is associated with the onset of cancer after controlling for individual-level socioeconomic characteristics (e.g. Freedman et al., 2011; Meijer, Bloomfield, et al., 2012; Webster et al., 2008), there have been few studies that examine whether such associations remain after accounting for a comprehensive set of individual risk factors and whether these factors are mediators of the neighborhood SES effect on cancer risk. The NIH-AARP diet and Health Study found that 26.4% of the excess incidence of colorectal cancer incidence in low SES neighborhoods was mediated by higher prevalence's of behavioral risk factors (Doubeni et al., 2012). Similarly, Kim et al., (2010) found that neighborhood SES was not associated with colon cancer among higher educated woman, with mediation by red meat intakes and body mass index. Rectal cancer in all women was inversely related to higher neighborhood SES, and mediation was observed by multivitamin use and body mass index. These finding correspond to the literature on neighborhood context and health in general, which shows that there is no one neighborhood effect mechanism, but rather a multitude of potential pathways linking neighborhood context to individual health outcomes (Diez-Roux, 2003).

Overall, no studies were found that examine long-term exposure to neighborhood SES and its influence on lung cancer risk. Similarly, very little is known surrounding the potential

pathways by which long-term neighborhood SES may operate, including its influence on smoking, other health-behaviors and occupational and environmental exposure pathways.

1.4. Data Sources and Study Population

Throughout this dissertation, I used the lung cancer component of the National Enhanced Cancer Surveillance System (NECSS), a population based case-control study that was a collaborative project between Health Canada and the provincial cancer registries. This project was specifically focused on assessing environmental cancer risk factors and Johnson et al. (1998) describe the purpose and overall study design of the NECSS. Here, I use the lung cancer component of the NECSS, which includes 3,280 lung cancer cases and 5,073 population controls collected in eight Canadian provinces (British Columbia, Alberta, Saskatchewan, Manitoba, Ontario, Prince Edward Island, Nova Scotia and Newfoundland). Briefly, between 1994 and 1997 all lung cases were identified by provincial cancer registries within 1-3 months of initial diagnosis and randomly sampled for inclusion into the study. Population controls were selected from a random sample of individuals within each province, frequency matched on sex and five-year age categories to the overall collection of NECSS cancer cases (~20,000 cases including 19 types of cancer). Recruitment methods for controls depended on data availability and accessibility by province and included provincial health insurance plans in five provinces, random digit dialing in two, and property assessment data in one. A research questionnaire was mailed to selected cases and controls and active follow-up was conducted. The response rate for contacted lung cancer cases was 61.7% and for population controls was 67.4%. The research questionnaire collected comprehensive information regarding individual characteristics, lifetime occupational exposures and residential histories. Residential histories were geo-coded to 6-digit postal codes and are the

basis of the spatiotemporal exposure assessments conducted throughout this thesis. Figure 1 illustrates the location of the residential histories reported for population controls and lung cancer cases, and while not all provinces participated in the NECSS, due to residential mobility national exposure assessments were required.

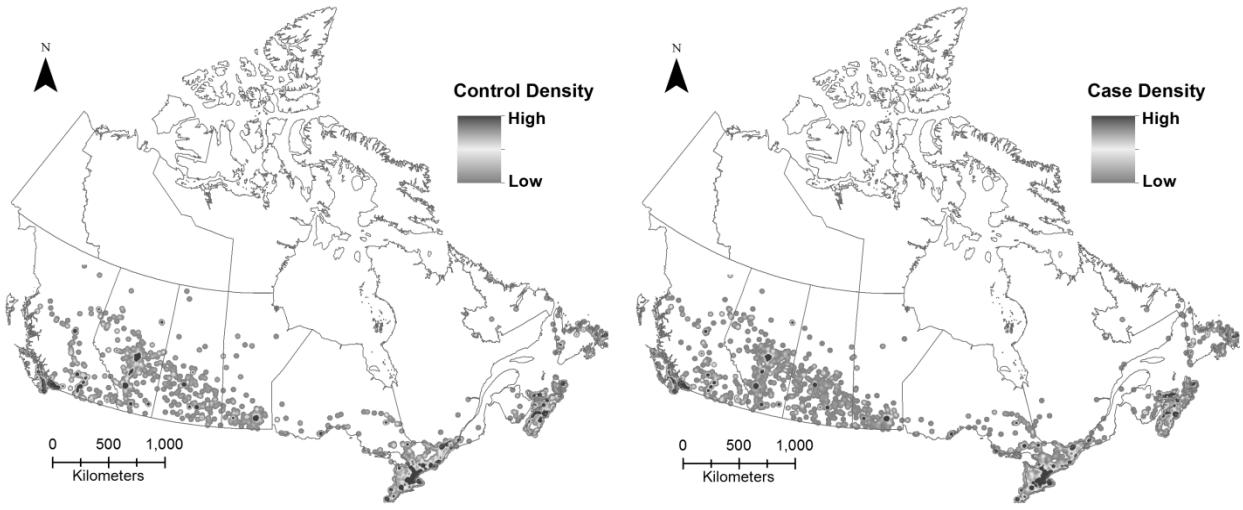

Figure 1. Locations reported in the residential histories for 1975-1994 for population controls and lung cancer cases within the National Enhanced Cancer Surveillance System.

1.5. Dissertation Objectives

The objective of this dissertation was to examine whether three exposures associated with the physical and social residential environment – specifically, ambient air pollution, radon and neighborhood SES – are risk factors for developing lung cancer in Canada. These exposures cover the broad spectrum of important risk factors associated with place; with air pollution and radon being traditional environmental hazards, while neighborhood SES represents a measure of social contextual conditions that may be important upstream health determinants. For air pollution, no large study of lung cancer incidence has been conducted that conducts comprehensive historical exposure assessments using residential histories,

examines multiple pollutants and exposure sources, controls for a wide-range of potential individual-level confounding factors, examines risk by histological subtypes. For radon, there is a need for epidemiological studies linking residential radon exposure to lung cancer risk in Canada, which will help estimate the importance of residential radon exposures, raise awareness of the risk posed by radon exposure (which is very low in Canada) and help guide population prevention efforts. In terms of neighborhood SES, no studies have been completed that examine the influence of long-term neighborhood SES on lung cancer risk, as well as the potential pathways by which long-term neighborhood SES may operate.

The specific objectives of the dissertation are:

1. To create national and retrospective air pollution exposure assessment methods for ambient air pollution.
2. To apply air pollution exposure assessment methods to a population-based lung cancer case-control study and estimate the associations between ambient air pollution exposure and lung cancer risk in Canada.
3. To create and apply maps of radon concentration and potential to a population-based lung cancer case-control study and estimate the associations between ecological radon exposure and lung cancer risk in Canada.
4. To create methods to estimate long-term neighborhood socioeconomic status and apply these to a population-based lung cancer case-control study and estimate the associations between exposure to long-term neighborhood SES and lung cancer risk in Canada.

1.6. Dissertation Structure

This dissertation follows a manuscript format and subsequent research chapters have been published or submitted for publication in peer reviewed journals. Figure 2 illustrates the framework for this dissertation and the corresponding research objectives. Due to the complexity of national and retrospective air pollution exposure assessment methods, two independent exposure papers are presented. The first focuses on creating national (spatial) land use regression (LUR) models for Canada and estimating current population exposures (Chapter 2), and the second extends these national models and methods using spatiotemporal techniques to estimate historical air pollution concentrations (Chapter 3). These air pollution models are then used in Chapter 4, which presents the epidemiological analyses of ambient air pollution and lung cancer incidence using the NECSS case-control study. Next, the radon analysis is presented in Chapter 5, including radon risk and potential mapping methods and epidemiological results from the NECSS study. Finally, Chapter 6 includes the exposure assessment methods for neighborhood SES as well as epidemiological results from applying these to the NECSS case-control study. All epidemiologic analyses mutually adjust for each of the physical and social environmental risk factors examined in this dissertation. Conclusions are then presented in Chapter 7, including a summary of the dissertation findings and their implications, limitations of the research, and recommendations for further research directions.

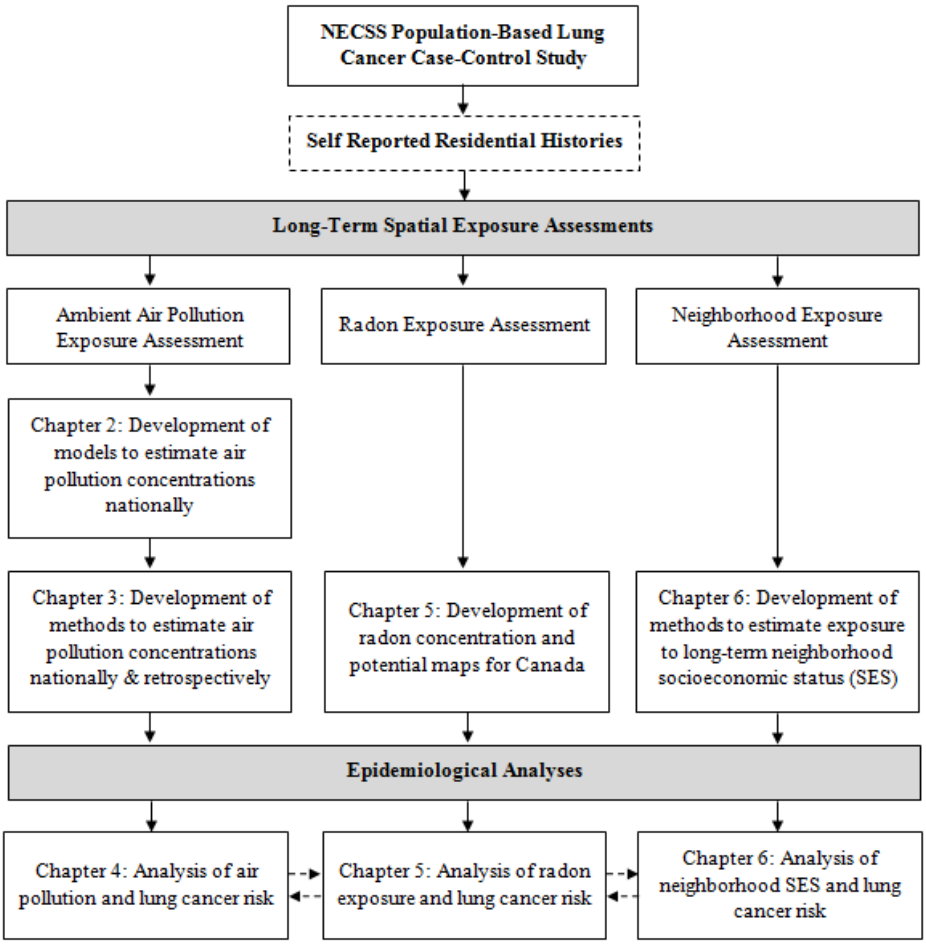

Figure 2. Dissertation framework and corresponding research chapters and objectives.

Chapter 2

Creating National Air Pollution Models for Population Exposure Assessment in Canada

Hystad P, Setton E, Cervantes A, Poplawski K, Deschenes S, Brauer M, et al. 2011. Creating National Air Pollution Models for Population Exposure Assessment in Canada. *Environmental Health Perspectives*.119:1123-1129.

Abstract

Background: Population exposure assessment methods that capture local-scale pollutant variability are needed for large-scale epidemiological studies and surveillance, policy, and regulatory purposes. Currently, such exposure methods are limited.

Methods: We created 2006 national pollutant models for fine particulate matter [PM with aerodynamic diameter ≤ 2.5 μm ($PM_{2.5}$)], nitrogen dioxide (NO_2), benzene, ethylbenzene, and 1,3-butadiene from routinely collected fixed-site monitoring data in Canada. In multiple regression models, we incorporated satellite estimates and geographic predictor variables to capture background and regional pollutant variation and used deterministic gradients to capture local-scale variation. The national NO_2 and benzene models are evaluated with independent measurements from previous land use regression models that were conducted in seven Canadian cities. National models are applied to census block-face points, each of which

represents the location of approximately 89 individuals, to produce estimates of current population exposure.

Results: The national NO_2 model explained 73% of the variability in fixed-site monitor concentrations, $PM_{2.5}$ 46%, benzene 62%, ethylbenzene 67%, and 1,3-butadiene 68%. The NO_2 model predicted, on average, 43% of the within-city variability in the independent NO_2 data compared with 18% when using inverse distance weighting of fixed-site monitoring data. Benzene models performed poorly in predicting within-city benzene variability. Based on our national models, we estimated Canadian ambient annual average population-weighted exposures (in micrograms per cubic meter) of 8.39 for $PM_{2.5}$, 23.37 for NO_2, 1.04 for benzene, 0.63 for ethylbenzene, and 0.09 for 1,3-butadiene.

Conclusions: The national pollutant models created here improve exposure assessment compared with traditional monitor-based approaches by capturing both regional and local-scale pollution variation. Applying national models to routinely collected population location data can extend land use modeling techniques to population exposure assessment and to informing surveillance, policy, and regulation.

2.1. Introduction

Predicting air pollution concentrations at resolutions capable of capturing local-scale pollutant gradients over large geographical areas is becoming increasingly important for multi-city and national health studies, in population exposure assessment, and in support of policy, surveillance and regulatory initiatives. Currently, fixed-site government monitors are the foundation of these activities; however, such monitors may fail to fully capture local-scale pollutant variability due to siting criteria. In addition, the number of monitors and their spatial distribution may be limited, as is the case in Canada. At present, few methodologies are available to adequately capture local-scale pollutant variability at a national scale when monitor density, distribution and/or siting is sub-optimal.

A number of approaches may be used to model air pollution over large areas, including interpolation of fixed-site government monitoring data, dispersion modeling, satellite remote sensing, land use regression (LUR), and proximity and deterministic methods; each approach, however, has inherent limitations that restrict its use for producing local-scale pollution estimates. Interpolation of fixed site air pollution monitoring data has typically been used to predict pollution concentrations across large areas (Beelen et al. 2009), with recent interest directed towards kriging methods and spatial smoothing with geographic covariates (Beelen et al. 2009; Hart et al. 2009; Yanosky et al. 2008). Fixed site monitors may not capture entire populations and measurements typically represent regional and between-city pollution differences due to monitor siting criteria, which prevent monitors from being placed in close proximity to major roads and other pollution sources. Dispersion models also exist for large geographical areas and have been incorporated into regulatory and epidemiological studies of air pollution (Cyrys et al. 2005; Nafstad et al. 2003). Importantly, the resolutions of pollutant

estimates from dispersion models over large geographical areas are typically restricted, for example, to one or three square kilometres (Jerrett et al. 2005). Satellite remote sensing is a new methodology available to predict air pollution concentrations over large geographic areas and a number of studies have evaluated different remotely sensed concentrations of $PM_{2.5}$ (e.g. van Donkelaar et al. 2010) and gaseous pollutants (Martin, 2008) and found moderate to good associations with ground level monitoring data. Currently, the resolution of satellite data limits their use to representing regional pollution concentrations but indicators of local air pollution may be used in concert to improve the spatial resolution of predictions (Liu et al. 2009). LUR approaches have been extensively used to predict within-city pollutant concentrations of NO_2 and $PM_{2.5}$ (see Hoek et al. 2008 for a review), but to a lesser extent for volatile organic compounds (VOC's), although the approach is well-suited to modeling pollutants which exhibit significant spatial variation, especially traffic-related VOCs (Atari and Luginaah 2009; Mukerjee et al. 2009; Smith et al. 2006; Su 2010; Wheeler et al. 2008). The city-by-city approach in which LUR models are created is costly and integration and interpretation across multiple city models is difficult. Simple proximity and deterministic approaches have also been widely used as surrogates for exposure to vehicle and industrial sources, specifically in epidemiological studies; yet, such measures in isolation are often poor surrogates for exposure. To-date, few population exposure assessment approaches have incorporated multiple sources of data, specifically satellite pollutant estimates, LUR modeling of geographic characteristics and information on proximity and pollution gradients, to estimate local-scale air pollution concentrations at a national scale.

Here we report a modeling initiative to produce 2006 national fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2), benzene, ethylbenzene and 1,3-butadiene models for Canada that capture local-scale pollutant variability and apply these models to routinely collected

population location data to calculate population exposures. This research is part of Carex Canada, a national surveillance initiative estimating the number of Canadians potentially exposed to known or suspected environmental and occupational carcinogens (Carex Canada, 2011). This research adds to the literature on air pollution modeling and exposure assessment by: (1) creating national LUR models from fixed site monitoring data; (2) incorporating various predictor datasets and methods to capture the different scales of pollution sources; and (3) extending LUR modeling techniques to population exposure assessment and to informing surveillance, policy and regulation.

2.2. Methods

Pollutant Modeling Approach

Models were developed in two stages using different predictor variables and methodology to capture background, regional and local-scale pollution variation. First, for each national air pollution surveillance (NAPS) fixed-site monitoring station we derived satellite-based estimates ($PM_{2.5}$ and NO_2 only) and geographic variables (e.g. road length, population density, proximity to large emitters, etc.) using a geographic information system (GIS). We developed LUR models using forward stepwise regression and retained variables that corresponded to hypothesized effect directions and maximized the sums of squares explained and Akaike's information criterion, and evaluated spatial autocorrelation using the Moran's *I* statistic. We sought to develop parsimonious models rather than traditional predictive models that maximize prediction but make interpretation of individual variable contributions difficult. Only variables significant at the $p<0.05$ level were included in the final models. As expected, NAPS monitoring locations in Canada did not display sufficient variability to estimate model coefficients for important local-scale parameters, such as

proximity to major roadways, due to monitor siting. Local-scale predictors were therefore under-powered in the LUR modeling approach.

In the second stage, we conducted comprehensive literature reviews to identify deterministic factors to represent local-scale gradients in pollutant concentrations associated with specific sources (i.e. highways, major roads, gas stations). For each pollutant, we identified concentrations near these selected sources in relation to local background levels, and developed deterministic multipliers with distance decay rates (together referred to as gradients in this paper) to apply to the background and regional concentrations predicted by our LUR models. All statistical analyses were conducted using SAS 9.1 (SAS Institute Inc., Cary, North Carolina).

Air Quality Data

Annual average concentrations of $PM_{2.5}$ (177 monitoring stations), NO_2 (134 monitors), and benzene, ethylbenzene and 1,3-butadiene (53 monitors) were calculated using data from unique NAPS monitoring sites operating during 2006 (see Figure 3). Continuous monitoring data from a given monitor were included if at least 50% of hourly observations were available for a 24hr period and at least 50% of days were available in a month. Monthly averages from filter based $PM_{2.5}$ measurements required a minimum of 3 out of 5 valid measurements per month. 2006 annual averages were not calculated for individual monitors unless there were at least six months of complete data with one valid month per quarter.

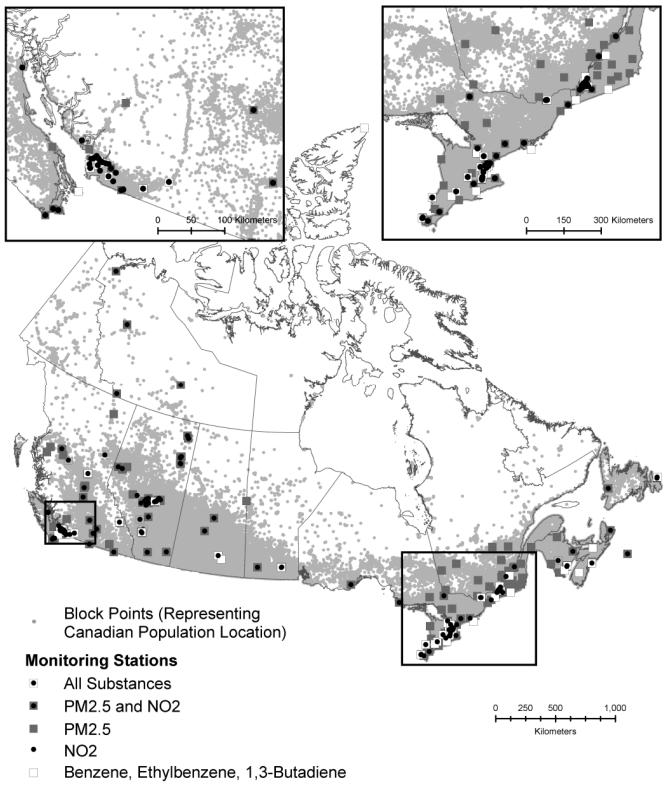

Figure 3. Location of NAPS monitors used to create national $PM_{2.5}$, NO_2, benzene, ethylbenzene and 1,3-butadiene models.

NAPS includes different monitor types for $PM_{2.5}$, including tapered element oscillating microbalances (TEOM), Dichotomous samplers, Partisols, and beta-attenuation mass (BAM) monitors. Multiple monitors are often present at one location, and our comparative analysis found differences in levels measured by TEOMS (TEOMS are known to under-predict $PM_{2.5}$ due to nitrate evaporation (T. Dann, personal communication)). We therefore selected other monitor types when they were available at the same location. Those stations with only TEOMs available were adjusted based on yearly calibration between collocated Dichotomous and TEOM monitors during 2006 (n=14, Dich=1.640+1.089*(TEOM), R^2=0.89, $p<0.001$). NO_2, benzene, ethylbenzene and 1,3-butadiene were measured using standard methods (NAPS, 2004).

Predictor Variables

$PM_{2.5}$ and NO_2 Satellite Data

Canada-wide concentrations of $PM_{2.5}$ and NO_2 were estimated using satellite atmospheric composition data combined with local, coincident scaling factors from a chemical transport model (GEOS-Chem, 2011). Ground-level $PM_{2.5}$ estimates were derived from aerosol optical depth (AOD) data from the Terra satellite, in combination with output from GEOS-Chem simulations to estimate the relationship between aerosol optical depth over the atmospheric column and ground-level $PM_{2.5}$ (van Donkelaar et al. 2010). Ground-level NO_2 concentrations were estimated from tropospheric NO_2 columns retrieved from the Ozone Monitoring Instrument (OMI) and also used GEOS-Chem to calculate the relationship between the NO_2 column and ground-level concentration (Lamsal et al. 2008). Both $PM_{2.5}$ and NO_2 were estimated at a 0.1x0.1 degree resolution (~10x10km). Estimates for $PM_{2.5}$ are from 2001 to 2006 data to ensure sufficient observations. NO_2 estimates used data from 2005 and 2006 since OMI measurements began in late 2004.

Geographic Data

We modeled regional pollutant variation using geographic predictor variables potentially relevant to pollutant sources, emissions and dispersion. To capture varying spatial influences of predictors, all variables were calculated in a GIS for circular buffer distances ranging from fifty meters to fifty kilometres. Classes of variables included: population density derived from census street block points; 1km land use classifications for LandSat; DMTI Spatial high resolution (30m) land use classifications (DMTI, 2011); large industrial emissions sources from the Canadian National Pollutant Release inventory (NPRI, 2010); small point source locations extracted from the Dun and Bradstreet Selectory database of businesses in Canada (D&B, 2011); length of, and distance to, specific road classifications using the DMTI Spatial

road network (DMTI, 2011) (freeway, highway, major road, minor road); length and density of railroads; elevation; and meteorological variables (precipitation and temperature). Any geographic variables with over 30% zero values (i.e., with no predictive features in proximity to a monitor) were re-coded as binary (i.e. present/absent). In total, 10 variable classes and 270 buffer-specific variables were explored in the LUR models.

Deterministic Gradients

Gradients were developed focusing on mobile sources and gas stations. We conducted a comprehensive literature review of published studies to identify: (a) the distance from sources at which pollutant concentrations typically return to background levels; and (b) an expected ratio of near source pollutant levels compared to background pollutant levels for each source and pollutant. We searched PubMed, Web of Science and Google Scholar using a range of keywords to identify studies with measurements of pollutant gradients. Studies varied widely in terms of location, date, methods used, duration of measures, number of samples, and definition of near source and background. We developed linear gradients using the steepest portion of the exponential decay curves typically found in the literature, as the tails of the decay functions were very sensitive to local parameters. Gradients were also selected to represent Canadian conditions. Table 1 summarizes the gradients developed for Canada and applied to the LUR models.

We used a GIS to identify the distance of each NAPS monitor from the nearest highway, major road, local urban road, and gas station, using DMTI road network data and Dun and Bradstreet commercial data for point sources. If a monitor was close enough to one of these features for the source to influence pollutant levels, we modified the corresponding LUR model results (not including point source industrial variables) to account for the deterministic gradients. For example, based on our review of the literature, we assumed that

NO_2 concentrations at the side of a highway would be 1.65 times higher than LUR-based background concentrations but consistent with background levels 300m from the highway, resulting in a distance decay rate of 0.33% per meter that was applied to estimate NO_2 levels within the 300m gradient buffer.

Table 1. $PM_{2.5}$, NO_2, benzene, ethylbenzene and 1,3-butadiene gradients determined from the literature and incorporated with national LUR model predictions.

Substance	Source	Increase at Source	Gradient Distance
$PM_{2.5}$	Highway	1.25[a]	75m[c]
	Major roads	1.1[a]	75m[c]
NO_2	Highway	1.65[a]	300m[b]
	Major roads	1.2[a]	100m[b]
Benzene	Gas stations	6.5[d]	100m[d]
	Highways/ Major roads	3.25[e]	50m[f]
	Local road	1.5[e]	50m[f]
Ethyl-benzene	Highway	3.7[g]	300m[h]
	Major road	2.2[g]	300m[h]
	Local road	1.4[g]	300m[h]
1,3-Butadiene	Highway	4[i]	75m[i]

Key References: [a]Smargiassi et al (2005); [b]Su et al (2009), Gilbert et al (2003), Gilbert et al (2007), Beckerman et al (2008), Roord Knape et al (1998); [c]Roorda-Knape et al (1998), Tiitta et al (2002), Beckerman et al (2008), Hitchens et al (2000); [d]Karakitsios et al (2007); [e]Thorsson and Eliasson (2006), Hellen et al (2005), Vardoulakis et al (2002), Parra et al (2009); [f]Thorsson and Eliasson (2006), Beckerman et al (2008), Venkatram et al (2009); [g]Wang and Zhao (2008), Roukos et al (2009), Parra et al (2009); [h]Wang and Zhao (2008); [i]Venkatram et al (2009)

Model Evaluation

We use three approaches for model evaluation. Due to the small number of NAPS monitoring stations for $PM_{2.5}$, NO_2, benzene, ethylbenzene and 1,3-butadience, we did not leave out a percentage for independent post-model evaluation as we wanted to capture the greatest range of model predictors possible. Therefore, we first evaluated all LUR models using a bootstrap approach to determine the sensitivity of model prediction and parameter

estimates to monitor sampling. Random selection of monitors was conducted, with replacement, and variable coefficients and model R^2 values were recorded from the new full sample. This was repeated for ten thousand iterations to estimate the 95% confidence interval for overall model prediction and individual variable coefficients. Next, we conducted a leave-one-out analysis where each LUR model was repeatedly parameterized on N - 1 data points and then used to predict the excluded monitor measurement. The mean differences between the predicted and measured values were used to estimate model error.

Finally, we evaluated the NO_2 and benzene LUR models, with and without gradients, against independent data (35 – 201 monitoring sites per city) previously collected for LUR models in seven Canadian cities (for a full description of data collection and modeling see: Allen et al. 2010; Atari and Luginaah 2009; Crouse et al. 2009; Henderson et al. 2007; Jerrett et al. 2007; Su et al. 2010). Briefly, in each city monitoring occurred for two-week periods and data from fixed site monitors, monitoring during yearly average concentration periods, or multiple measurement periods, were used to estimate yearly averages (see Appendix 1, Table 1 for the city-specific data used for model evaluation). These pollution measurements were collected at much higher spatial densities than NAPS and from monitors that were located to specifically capture spatial pollutant gradients. Consequently, these data were reasonable for use as a "gold standard" to determine how well the two national NO_2 and benzene models (the LUR models and the LUR models with gradients) predicted within-city variation. In addition, we compared the city-specific data to estimates based on inverse distance weighting (IDW) of annual average NO_2 and benzene concentrations measured at NAPS monitors (with and without deterministic gradients). Due to NAPS monitor density in Canada kriging could not be applied.

Population Exposure Assessment

The national pollutant models were applied to each of the 478,831 Statistics Canada street block centroid locations in 2006 to estimate population exposures. First, we applied the LUR models to each block point to derive a unique predicted pollutant concentration for each point – representing the average exposure level for 89 (+/-158) individuals. We used a GIS to identify the distance of each block centroid to the nearest highway, major road, local urban road, and gas stations, and adjusted the corresponding LUR model estimate when the street block point was located within an associated gradient. We then estimated population weighted exposures to $PM_{2.5}$, NO_2, benzene, ethylbenzene and 1,3-butadiene in the Canadian population as a whole, and estimated uncertainty using the 95% confidence limits for LUR model predictions and assumed that concentrations at each pollutant source were within +/- 50% of the gradient values shown in Table 1. Fifty percent was selected for all gradients as there was insufficient literature to examine uncertainty for specific gradients and sources.

2.3. Results

National LUR Model Results

Table 2 summarizes the national LUR model results. The $PM_{2.5}$ model predicted 46% of $PM_{2.5}$ variation and was dominated by satellite predictions, which alone explained 41% of $PM_{2.5}$ variation. The NO_2 model predicted 73% of NO_2 variation and length of all roads within 10km was the dominant predictor, explaining 55% of NO_2 variation. This variable was only moderately correlated (r=0.56) to NO_2 predictions from satellite data, which further explained 4% of NO_2 variation in the final model. The models for benzene, ethylbenzene and 1,3-butadiene had similar predictive results, explaining 62%, 67% and 68% of pollutant variability respectively. Data from one monitor was removed as an outlier from the benzene and ethylbenzene models (St. John Baptiste, located in Montreal east city) and from the 1,3-

butadiene model (Sarnia, located in southern Ontario near the Detroit-Windsor border), which were associated with the highest pollutant concentration for each substance.

Table 2. National LUR model results for $PM_{2.5}$, NO_2, benzene, ethylbenzene, and 1,3-butadiene.

Variable	Distance[a]	Value	SE	p
$PM_{2.5}$ Model [R^2=0.46, RMSE=1.529]				
Intercept	-	2.802	0.497	<.0001
Satellite $PM_{2.5}$ (ln µg/m^3)	-	2.392	0.263	<.0001
NPRI emissions (tonnes)	5km	1.63e-3	5.95e-4	0.007
Industrial landuse (m^2)	1km	1.03e-6	4.18e-7	0.014
NO_2 Model [R^2=0.73, RMSE=5.470]				
Intercept	-	13.179	1.374	<.0001
Satellite NO_2 (ppb)	-	1.4903	0.355	<.0001
Industrial landuse (m^2)	2km	3.21e-6	5.73e-7	<.0001
Road length (m)	10km	7.42e-6	9.04e-7	<.0001
Summer rainfall (mm)	-	-0.010	0.002	<.0001
Benzene Model[b] [R^2=0.62, RMSE=0.298]				
Intercept	-	0.346	0.069	<0.001
Major road length (m)	10km	1.18e-6	2.56e-7	<0.001
NPRI emissions (present)	10km	0.526	0.089	<0.001
Ethylbenzene Model[c] [R^2=0.67, RMSE=0.193]				
Intercept	-	0.152	0.039	<0.001
Population (count)	10km	6.74e-7	7.25e-8	<0.001
NPRI emissions (present)	2km	0.272	0.071	<0.001
1,3-Butadiene Model[d] [R^2=0.68, RMSE=0.034]				
Intercept	-	0.011	0.009	0.208
Road length (m)	750m	3.89e-6	7.93e-7	<0.001
Highway (present)	500m	0.041	0.012	0.002
Commercial landuse (m^2)	10km	1.60e-9	5.97e-10	0.010

[a] Circular buffer distance variables derived from.
[b] One outlier removed with benzene concentration of 3.55µg/m^3.
[c] One outlier removed with ethylbenzene concentration of 2.57µg/m^3.
[b] One outlier removed with 1,3-butadiene concentration of 0.82µg/m^3.
*Variable descriptions: Satellite $PM_{2.5}$ / NO_2=Satellite derived estimates of $PM_{2.5}$ (ln µg/m^3) and NO_2 (ppb). Landuse=Area (in meters squared) of specific land use types (industrial, commercial) within the associated buffer distance. Road length=Length (in meters) of different road classifications (all, major, highways) within the associated buffer distance. Summer rainfall=millimetres of summer (May-Sept.) rainfall recorded at nearest meteorological station. NPRI Emissions (tonnes)=amount of annual emissions (in tonnes) of model substance released from industries reporting to the National Pollutant Release Inventory (NPRI). NPRI Emissions (present)=Presence of NPRI facilities releasing model substance to air. Population(count)= Number of individuals residing within associated buffer distance.

Spatial Autocorrelation of National LUR Models

Spatial autocorrelation of the LUR model residuals was examined using Moran's *I* in ArcGIS. Spatial autocorrelation was present in the $PM_{2.5}$ LUR model residuals (Moran's I = 0.33, $p<0.001$), indicating a moderate amount of spatial autocorrelation remained that was not explained by the $PM_{2.5}$ model predictors. Clustering of positive residuals (model under predicting by an average of 2.57 $\mu g/m^3$) occurred in the rural interior of British Columbia. An indicator variable for British Columbia substantially reduced the spatial autocorrelation (Moran's I=0.03 p=0.04). Sensitivity analysis using a summer only $PM_{2.5}$ model indicated no spatial autocorrelation (Moran's I=0.04 p=0.01), supporting our hypothesis of wood burning as the primary source of model under prediction in this region. No significant spatial autocorrelation existed in LUR model residuals for NO_2 (Moran's I = 0.03, p=0.44), benzene (Moran's I = -0.20, p=0.13), ethylbenzene (Moran's I = -0.00, p=0.87) and 1,3-butadiene (Moran's I = 0.09, p=0.32).

Incorporating Gradients with National LUR Models

Deterministic gradients were added to LUR models as we could not estimate the effects of local-scale pollution sources from NAPS data alone. Figure 4a illustrates the final $PM_{2.5}$ model (LUR plus gradients) for Canada as a whole, and for southern Ontario and the city of Toronto (city locations represent locations of LUR data used for evaluation). Figure 4b illustrates the final national NO_2 model (LUR plus gradients) for Canada as a whole, and for south-western British Columbia and the city of Vancouver. These maps illustrate the spatial resolution of the final national pollutant models; however, for population exposure assessment the LUR model results and deterministic gradients were applied to street block point locations, as shown in Figure 5, which illustrates the final national benzene model (LUR plus gradients) calculated at the block point level.

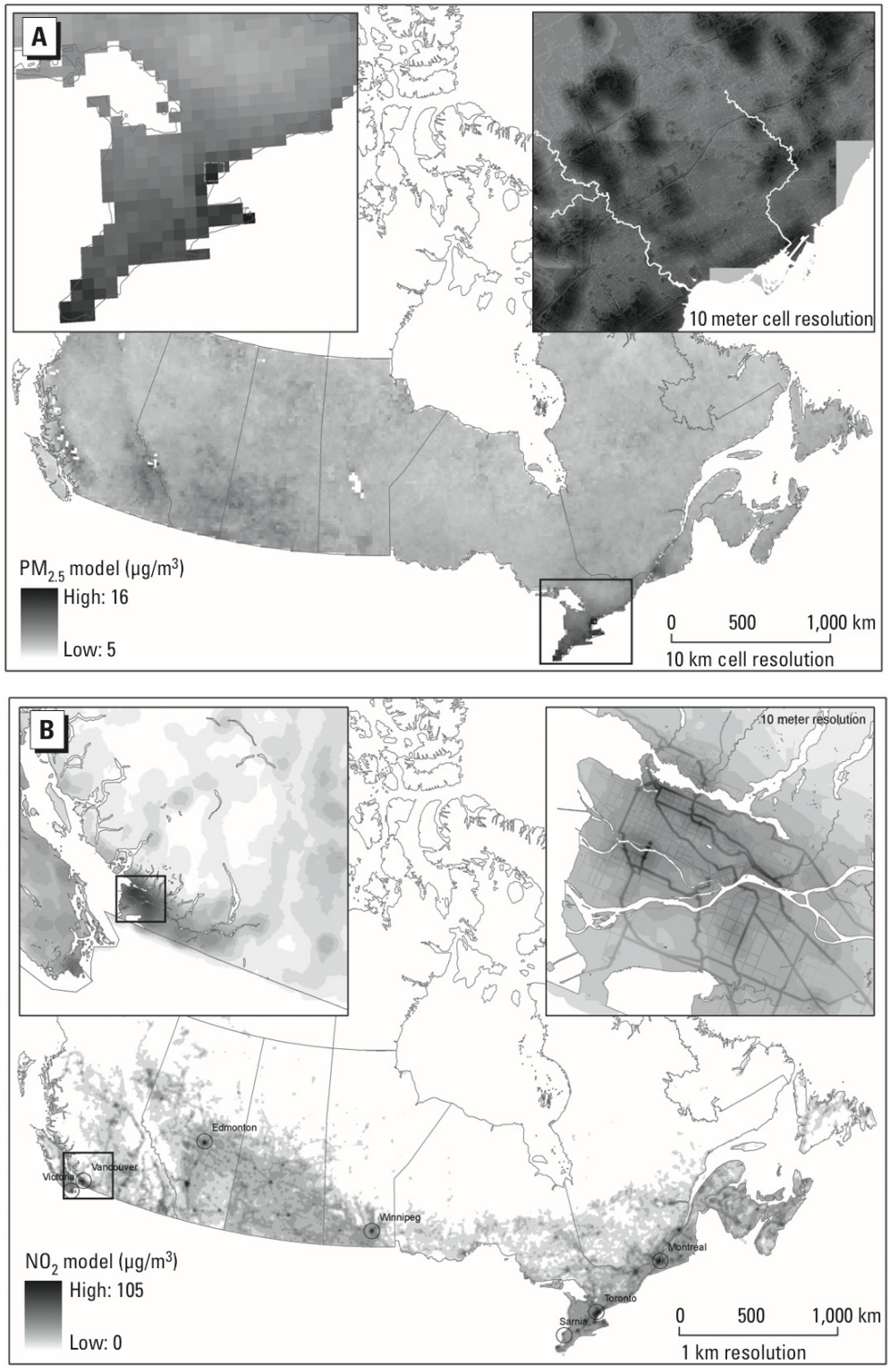

Figure 4. National annual average models for $PM_{2.5}$ (2a) and NO_2 (2b) incorporating satellite-derived pollutant estimates, geographic land use variables and deterministic gradients.

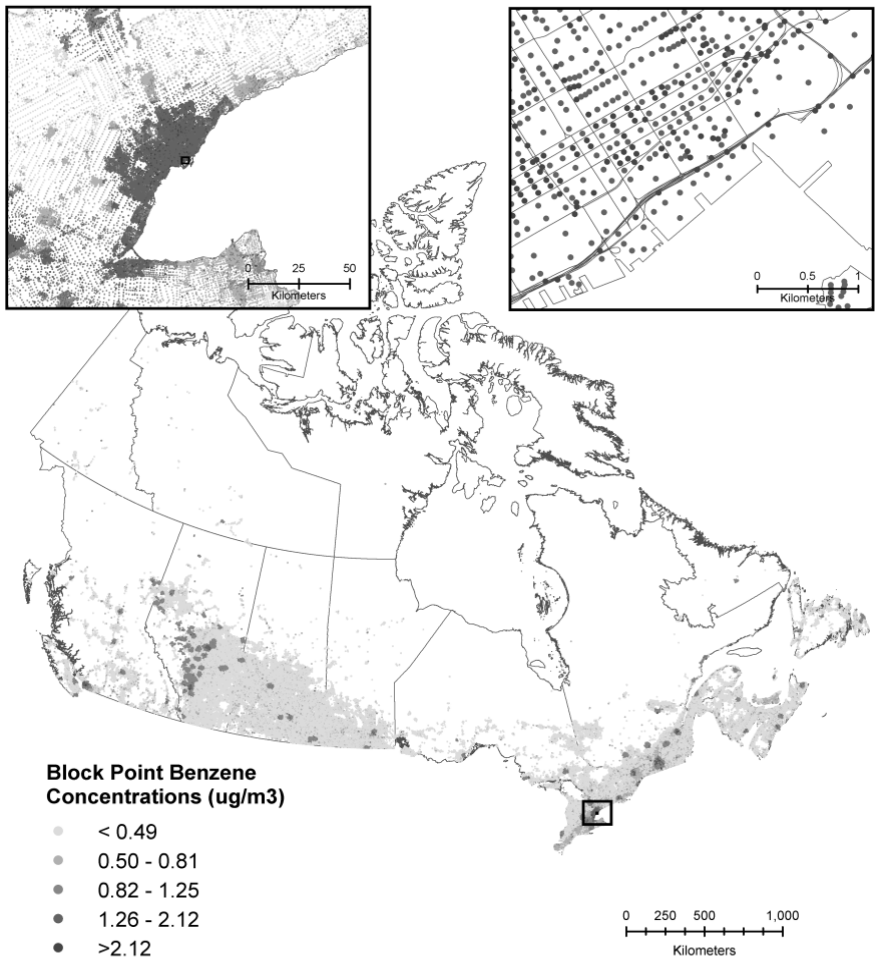

Figure 5. National benzene land use regression model plus gradients (illustrating the city of Toronto) calculated for each street block point in Canada (n=478,831).

Evaluation of National Pollutant Models

Evaluation of LUR Models using Bootstrap and Leave-One-Out Analyses

The distribution of all model coefficients for each pollutant resulting from bootstrap analysis showed normal distributions. The NO_2 model was the least sensitive to monitor selection with a bootstrap R^2 95% CI of 65%-81%. Models for $PM_{2.5}$, benzene, ethylbenzene, and 1,3-butadiene demonstrated larger uncertainty to monitor selection with R^2 95% CI of 33%-59%, 44%-80%, 49%-85%, and 53%-82% respectively. Variable coefficients for industrial NPRI proximity variables were extremely sensitive to monitor selection. The leave-

one-out analyses indicated no significant bias in any LUR model, as demonstrated by the mean (+/-SD) error: $1.07e^{-3}$ (5.61) for NO_2; $-6.35e^{-3}$ (1.59) for $PM_{2.5}$; -0.04 (0.32) for benzene; -0.01 (0.04) for 1,3-butadiene; and -0.04 (0.22) for ethylbenzene.

Evaluation of NO_2 and Benzene Models Using City-Specific Data

On average the national NO_2 LUR plus gradient model predicted 43% of the within-city NO_2 variation (based on the city-specific data evaluation), compared with 22% predicted based on IDW of NAPS monitors plus gradients (Table 3). National LUR, LUR plus gradients, IDW, and IDW plus gradients models over-predicted the city-specific NO_2 measurements, with average city-specific intercepts of 4.56, 7.45, 8.51 and 11.56 $\mu g/m^3$ respectively. City-specific scatter-plots of measured and modeled NO_2 concentrations are illustrated in the Appendix 1, Figure 1.

For benzene, all modeling methods performed poorly in explaining within-city benzene variation. The LUR plus gradients model explained on average only 16% of within-city variability in benzene concentrations compared with 11% based on IDW plus gradients (Table 3). In the evaluation using the Montreal city-specific benzene concentrations 4 outliers were removed (all concentrations $>2\mu g/m^3$), and one outlier (4.10 $\mu g/m^3$) was removed in the Toronto evaluation. Benzene models also over predicted city-specific concentrations, based on city-specific intercepts of modeled versus measured concentrations: see Appendix 1, Figure 2. Sarnia, a high density industrial community with 46 NPRI emitters, had poor NO_2 and benzene model evaluations.

Table 3. Evaluation of national NO_2 and benzene models, as well as inverse-distance weighted (IDW) estimates from fixed-site monitors, against independent city-specific measurement data.

Substance	n[a]	R^2 (RMSE)			
		LUR[b]	LUR+G[c]	IDW[d]	IDW+G[e]
NO_2					
Edmonton	50	0.60 (3.67)	0.41 (4.59)	0.10 (5.52)	0.01 (5.92)
Montreal	135	0.41 (4.28)	0.48 (4.04)	0.31 (4.63)	0.41 (4.29)
Sarnia	34	0.42 (4.21)	0.49 (4.04)	0.12 (5.15)	0.19 (5.12)
Toronto	196	0.18 (7.69)	0.36 (6.78)	0.13 (7.93)	0.32 (6.99)
Victoria	40	0.19 (3.95)	0.37 (3.70)	0.23 (3.86)	0.26 (3.98)
Vancouver	114	0.31 (6.41)	0.42 (5.93)	0.31 (6.43)	0.36 (6.24)
Winnipeg	49	0.54 (3.65)	0.51 (3.86)	0.08 (5.17)	0.02 (5.43)
Average	618	0.39 (4.84)	0.43 (4.71)	0.18 (5.53)	0.22 (5.42)
Benzene					
Montreal[f]	131	0.33 (0.24)	0.26 (0.25)	0.11 (0.28)	0.05 (0.29)
Sarnia	37	0.02 (0.57)	0.04 (0.56)	0.00 (0.57)	0.03 (0.56)
Toronto[g]	44	0.03 (0.19)	0.22 (0.17)	0.00 (0.19)	0.34 (0.16)
Winnipeg	94	0.08 (0.25)	0.10 (0.25)	0.00 (0.26)	0.01 (0.26)
Average	306	0.12 (0.31)	0.16 (0.31)	0.03 (0.33)	0.11 (0.32)

[a] Number of within-city measurement locations.
[b] National land use regression model.
[c] National land use regression model plus gradients.
[d] Inverse distance weighting interpolation of NAPS fixed site monitoring data.
[e] Inverse distance weighting interpolation of NAPS fixed site monitoring data plus gradients.
[f] 4 outliers removed with highest city concentrations (>2µg/m^3).
[g] 1 outlier removed with highest city concentration (4.10µg/m^3).

Canadian Population Exposure Assessment

The final LUR models and gradients were applied to all 478,831 street block centroid locations to conduct population exposure assessments. Estimated mean (95%CI) population exposures (µg/m^3) to ambient $PM_{2.5}$, NO_2, benzene, ethylbenzene and 1,3-butadiene in Canada based on the LUR models were 8.10 (5.84-10.43), 22.40 (13.14-33.51), 0.94 (0.57-1.31), 0.38 (0.25-0.52) and 0.086 (0.035-0.138) respectively. Estimates for the same pollutants based on the national LUR plus gradients models were 8.39 (6.00-11.13), 23.37 (14.01-35.73), 1.04 (0.59-1.49), 0.63 (0.35-1.10) and 0.089 (0.036-0.146) respectively. Wide ranges of exposure

levels were estimated in Canada for all substances; see Appendix 1, Figure 3 for population exposure distributions.

2.4. Discussion

We created national pollutant models from fixed-site monitoring data that incorporate satellite, geographic and deterministic components and demonstrated that these models can improve exposure assessment over large geographic areas compared to approaches based solely on interpolation of fixed site monitoring data. We also demonstrated how these models can be used to calculate current population exposure assessment.

The national LUR models explained 73% of pollution variation in NAPS measurements for NO_2, and lesser degrees for $PM_{2.5}$ (46%), Benzene (62%), ethylbenzene (67%) and 1,3-butadiene (68%). The NO_2 and $PM_{2.5}$ models were least sensitive to monitor selection, while models for VOC's were more sensitive – likely due to the smaller number of monitors on which LUR estimates were based (n=53). The predictive performance of the $PM_{2.5}$ model (R^2=0.46, RMSE=1.53$\mu g/m^3$) was consistent with other large-scale modeling studies based on different monitoring methodologies and data inputs (Beelen et al. 2009; Hart et al. 2009; Liao et al. 2006; Ross et al. 2007).

The national LUR models generally captured regional patterns in pollutant concentrations, corresponding to NAPS monitor siting criteria, but were less effective at identifying small scale geographic predictor variables. For example, only 35 NAPS monitors were located within 500 meters of a major road and only seven monitors were within 500 meters of a major industrial emission source. Such small sample sizes greatly reduce the models' power to capture these specific pollutant sources. Some city-specific LUR methods have used location-allocation methods to more fully represent the true spatial variation in

pollution levels and to capture the range of predictor variables (Jerrett et al. 2005). Models based on fixed site monitor data may therefore need additional approaches to represent local-scale pollutant variability not captured by fixed site monitors. This was indeed the case with the Canadian NAPS network, but larger regulatory networks, such as those in the United States, may better represent the range of predictor variables needed to build local-scale LUR models.

To address the lack of local-scale geographic variability in the NAPS data we incorporated deterministic gradients based on proximity to specific sources (i.e. vehicles and gas stations). The final NO_2 LUR plus gradient model improved prediction of within-city pollutant variation considerably compared to the LUR model alone and interpolation methods – on average the final model predicted 43% of within-city NO_2 variation compared to 18% using IDW. Both the national benzene model and IDW predicted within-city benzene poorly, which may be due to the small number of NAPS monitors on which the model was based, the relatively small variation in within-city benzene levels, or the inability of gradients to capture local benzene concentrations. Similar to the NO_2 model, the evaluation of the benzene model with Sarnia data was poor, reflecting the difficulty in capturing unique high-density industrial conditions in a national model.

Gradients were based on literature reviews and a clear limitation was the lack of methodological consistency among published data of pollutant level increases near specific sources and the distance required for pollutant levels to return to background. To improve reliability of gradients we used linear functions to represent the decreases in pollutant levels found in the initial portions of the exponential decay curves found in the literature. The methodology used here could be augmented as new gradients become available or with other modeled data, for example, with results from EPA's near-road research program (EPA, 2011).

Population exposure assessment was conducted using the national models and census street block points. The population weighted average exposures to $PM_{2.5}$, NO_2, benzene, ethylbenzene and 1,3-butadiene were 8.39, 23.37, 1.04, 0.63 and 0.089 ($\mu g/m^3$) respectively. The uncertainty of population exposure estimates were driven primarily by LUR model uncertainty. While the national LUR models' results are similar to city-specific LUR models in their predictive capacity and error, we are unaware of any LUR models that have been applied to estimate exposure uncertainty. While these exposures are low compared to other countries, there are particular locations in Canada where exposures are relatively high; for example, the 90th percentiles of exposures (in $\mu g/m^3$) are 9.78 for $PM_{2.5}$, 34.81 for NO_2, 1.61 for benzene, 1.01 for ethylbenzene, and 0.14 for 1,3-butadiene. The ability of the national models to capture local-scale pollutant variability allows for more realistic exposure assessments and assessments that can potentially identify high-risk populations. Future work will refine approaches for using the national models to calculate population exposure assessments, incorporate socioeconomic information from census to examine environment injustice issues, and integrate national models into a risk assessment framework that incorporates exposures from other sources and micro-environments.

This study faced a number of challenges and limitations to creating national pollutant models from fixed site monitors and applying these models to estimate Canadian population exposures. Firstly, the NAPS monitors in Canada are centered in large metropolitan areas and modeled relationships will therefore be weighted towards these areas. This is appropriate for population exposure assessment, as these locations represent the majority of Canadians, but in rural areas the models could be adjusted or a background concentration used. This is particularly relevant to the benzene, ethylbenzene and 1,3-butadiene models, which were based on data from monitors located almost exclusively in large urban areas or sited near large

industrial sources. Secondly, we had limited data on pollutant sources and source strengths such as traffic volumes. In addition, we did not model emissions from wood burning stoves and forest fires, which may have caused us to under predict $PM_{2.5}$ concentrations in the interior of British Columbia. Thirdly, parsimonious LUR models were created since the specificity of model variables may be important for informing surveillance and regulation. This, however, leads to models that do not capture the complex interactions between geographic characteristics and pollutant sources and even the simplest LUR predictors (e.g. major roads or NPRI within 10km) capture complex mixes of geographic variables and pollutant sources. Fourthly, we compared model estimates to city-specific measurements for NO_2 and benzene collected in different years and using a variety of different methodologies. Nevertheless, these measurements represent the best data on within-city pollutant variability available. Fifthly, applying LUR model results to approximately half a million block points is currently extremely computationally and time intensive. Lastly, the geographic accuracy of street block centroids may introduce errors into the gradient portions of the models, and therefore the exposure assessment, particularly between rural and urban areas. These errors, however, are likely spatially random within rural and urban areas across Canada.

2.5. Conclusion

National exposure models were required by Carex Canada to produce population exposure assessments that captured both between and within-city pollution variability. We created national $PM_{2.5,}$ NO_2, benzene, ethylbenzene, and 1,3-butadiene models from fixed site monitoring data and found that a combination of data sources and methods to capture background, regional and local-scale pollution variation improved exposure assessment over traditional IDW interpolation approaches. The national pollutant models were applied to street

block points representing the locations of the Canadian population to determine population exposure estimates. Estimates of average population exposure levels in Canada are $PM_{2.5}$ 8.39, NO_2 23.37, benzene 1.04, ethylbenzene 0.63 and 1,3-butadiene 0.09 ($\mu g/m^3$). The modeling approach developed here uses readily available data and could be reproduced over time, for example, every five years with the Canadian census. This would provide updated population exposure assessments and a long-term surveillance capacity for monitoring trends in population exposures, for identifying potential susceptible populations and geographic locations with elevated exposures, and for evaluating the impacts of policies and regulatory changes on exposure levels.

These models can also be used for current national-level epidemiological studies; however, due to the detailed predictor variables present in the models further work needs to examine how these types of models can be modified to estimate historical exposures (due to the lack of detailed GIS data available).

Chapter 3

Spatiotemporal Air Pollution Exposure Assessment for a Canadian Population-Based Lung Cancer Case-Control Study

Hystad, P., Demers, P., Johnson, K.C., Brook, J., van Donkelaar, A., Lamsal., L., Martin, R., Brauer, M. (2012). Spatiotemporal air pollution exposure assessment for a Canadian population-based lung cancer case-control study. Environmental Health. 11:22.

Abstract

Background: Few epidemiological studies of air pollution have used residential histories to develop long-term retrospective exposure estimates for multiple ambient air pollutants and vehicle and industrial emissions. We present such an exposure assessment for a Canadian population-based lung cancer case–control study of 8353 individuals using self-reported residential histories from 1975 to 1994. We also examine the implications of disregarding and/or improperly accounting for residential mobility in long-term exposure assessments.

Methods: National spatial surfaces of ambient air pollution were compiled from recent satellite-based estimates (for $PM_{2.5}$ and NO_2) and a chemical transport model (for O_3). The surfaces were adjusted with historical annual air pollution monitoring data, using either spatiotemporal interpolation or linear regression. Model evaluation was conducted using an independent ten percent subset of monitoring data per year. Proximity to major roads, incorporating a temporal weighting factor based on Canadian mobile-source emission

estimates, was used to estimate exposure to vehicle emissions. A comprehensive inventory of geocoded industries was used to estimate proximity to major and minor industrial emissions.

Results: Calibration of the national $PM_{2.5}$ surface using annual spatiotemporal interpolation predicted historical $PM_{2.5}$ measurement data best ($R^2 = 0.51$), while linear regression incorporating the national surfaces, a time-trend and population density best predicted historical concentrations of NO_2 ($R^2 = 0.38$) and O_3 ($R^2 = 0.56$). Applying the models to study participants residential histories between 1975 and 1994 resulted in mean $PM_{2.5}$, NO_2 and O_3 exposures of 11.3 $\mu g/m^3$ (SD = 2.6), 17.7 ppb (4.1), and 26.4 ppb (3.4) respectively. On average, individuals lived within 300 m of a highway for 2.9 years (15% of exposure-years) and within 3 km of a major industrial emitter for 6.4 years (32% of exposure-years). Approximately 50% of individuals were classified into a different $PM_{2.5}$, NO_2 and O_3 exposure quintile when using study entry postal codes and spatial pollution surface, in comparison to exposures derived from residential histories and spatiotemporal air pollution models. Recall bias was also present for self-reported residential histories prior to 1975, with cases recalling older residences more often than controls.

Conclusions: We demonstrate a flexible exposure assessment approach for estimating historical air pollution concentrations over large geographical areas and time-periods. In addition, we highlight the importance of including residential histories in long-term exposure assessments.

3.1. Introduction

Exposure to ambient air pollution is a suspected risk factor for lung cancer (Beelen et al. 2008b; Katanoda et al. 2011; Laden et al. 2006; Nafstad et al. 2003; Pope III et al. 2002; Raaschou-Nielsen et al. 2011). Due to the long latency periods associated with lung cancer, epidemiological analyses are particularly challenging, especially for air pollution where spatial and temporal variation in both residential mobility and air pollution concentrations may produce significant exposure misclassification if not properly incorporated into the exposure assessment approach.

Residential mobility data are required for accurate long-term air pollution exposure assessments, but due to the difficulties in obtaining this information, residential location at study entry or at time of diagnosis is often used to estimate lifetime or long-term exposure estimates in epidemiological studies. Given that approximately half of all individuals move within a five year period (Canadian Census 2006) and that residential mobility varies depending on socio-economic factors (Hurley et al. 2005; Kan 2007; Oishi 2010; Urayama et al. 2009), there is potential for exposure misclassification and bias in studies that ignore or improperly account for residential mobility. While there is growing recognition of the need for spatiotemporal epidemiology approaches and life-time residential histories in exposure assessment (Meliker and Sloan 2011), mainly in cancer epidemiology (Behren et al. 2008; Gallagher et al. 2010), little is known regarding the potential exposure misclassification and bias resulting from self-reported residential histories, the most common form of attaining residential histories in epidemiological studies (Boscoe 2011), and from the assumption of residential stationarity in air pollution epidemiology.

Incorporating residential histories into air pollution exposure assessments requires corresponding air pollution concentration estimates that cover the spatiotemporal domain of the study period. To date, the association between air pollution and lung cancer has been examined using a variety of study periods and exposure assessment approaches. The most common approaches have aggregated air pollution monitoring levels within cities or defined areas (Dockery et al. 1993; Katanoda et al. 2011; Laden et al. 2006; Pope III et al. 2002), estimated ambient air pollution levels at residential addresses using fixed-site monitoring data or dispersion models (Beelen et al. 2008b; Beeson et al. 1998; Bellander et al. 2001; Nafstad et al. 2003; Raaschou-Nielsen et al. 2011), or used proximity to roads and industrial sources as exposure surrogates (Beelen et al. 2008a; Vineis et al. 2006). In terms of national retrospective exposure assessment studies, few are available that examine multiple pollutants and exposure sources (Beelen et al. 2007; Hart et al. 2009).

Here we develop a comprehensive spatiotemporal exposure assessment approach for Canada and apply it to a population-based case-control study of 8353 individuals who provided lifetime self-reported residential histories. This approach expands from the national spatial-only exposure assessment methods presented in Chapter 2. For the exposure period 1975 to 1994, we assign fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2) and ozone (O_3) air pollution exposures, as well as exposures to vehicle and industrial emissions. The implications of disregarding and/or improperly accounting for residential histories in long-term exposure assessments are also examined. The exposure assessment methods developed produce annual spatiotemporal exposure estimates and will allow subsequent epidemiologic analyses to examine latency periods, to include both urban and rural populations, and to study the contributions of multiple ambient pollutants and local vehicle and industrial emissions to lung cancer risk in Canada.

3.2. Methods

The Lung Cancer Case-Control Study

We utilize the lung cancer component of the National Enhanced Cancer Surveillance System (NECSS), which includes 3280 histological-confirmed lung cancer cases and 5073 population controls collected between 1994 and 1997 in the provinces of British Columbia, Alberta, Saskatchewan, Manitoba, Ontario, Prince Edward Island, Nova Scotia and Newfoundland. Due to residential mobility, study participants are located in all provinces of Canada requiring national-level exposure assessment. Johnson et al. (1998) describe the overall recruitment methodology for the NECSS. Briefly, cases were identified through provincial cancer registries and mailed a research questionnaire. The response rate for contacted lung cancer cases was 61.7%. Population controls were selected from a random sample of individuals within each province, with an age/sex distribution similar to that of all cancer cases (strategies for recruiting population controls varied by province depending on data availability and accessibility). Provincial cancer registries collected information from sampled controls using the same protocol as for the cases. The response rate for contacted population controls was 67.4%.

Residential histories at the 6-digit postal code level are the basis of the air pollution exposure assessment reported here. In urban areas a 6-digit postal code typically incorporates one side of a city block, but represents substantially larger areas in rural locations (e.g. greater than 100km^2 in remote locations of Canada). Residential histories were converted to postal codes by the Public Health Agency of Canada and geocoded using DMTI Inc. 1996 postal codes. While lifetime residential histories were collected, the exposure period was restricted to 1975 to the start of study enrolment (1994), due to the presence of recall bias in earlier reported histories (explained in more detail in the discussion section) as well as the lack of

information on postal code locations, air pollution monitoring data and geographic information prior to 1975.

Air Pollution Exposure Assessment Approach

A multi-staged approach was required to assign ambient air pollution concentrations to residential histories from 1975 to 1994. The spatiotemporal exposure assessment included three steps. First, national spatial surfaces were created from recent satellite-based estimates (for $PM_{2.5}$ and NO_2) and a chemical transport model (for O_3). Second, all National Air Pollution Surveillance (NAPS) monitoring data were compiled and formatted for the study period, including 120 NO_2 stations and 1030 measurement-years, 187 O_3 stations and 1440 measurement-years, 177 TSP stations and 1826 measurement-years, and 25 $PM_{2.5}$ stations and 141 measurement-years. Due to the small number of $PM_{2.5}$ measurements available, and no measurements made prior to 1984, a random effect model was used to estimate $PM_{2.5}$ based on TSP measurements and metropolitan indicator variables. Finally, the spatial pollutant surfaces were calibrated yearly to estimate average annual concentrations between 1975 and 1994. Two approaches were used for calibration: the first estimated historical annual averages using smoothed inverse distance weighting (IDW) interpolation of the ratios of spatial co-located historical NAPS and surface estimates, while the second used linear regression models.

Exposure to vehicle emissions was estimated using proximity to highways and major roads, adjusted based on historical vehicle emissions in Canada. Exposures to industrial emissions were calculated based on proximity to major and minor industrial sources extracted from a comprehensive database of industrial facilities in Canada operating during the study exposure period. Estimates for different vehicle and industrial emission sources were not converted into concentrations and added to ambient concentration estimates as we want to examine each source and distance threshold separately in subsequent epidemiological

analyses. Specific components of the exposure assessment approach are described in detail below.

National Spatial Pollutant Surfaces

Spatial models of ambient $PM_{2.5}$, NO_2 and O_3 concentrations were developed to represent current spatial pollution patterns across Canada. A $PM_{2.5}$ surface was derived from Aerosol Optical Depth (AOD), using data from the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Multiangle Imaging Sectroradiometer (MISR) satellite instruments, and was combined with a chemical transport model (GEOS-Chem; www.geos-chem.org) to estimate the relationship between aerosol optical depth and surface $PM_{2.5}$ (for full details see (van Donkelaar et al. 2010)). Estimates for $PM_{2.5}$ represented a composite estimate developed from 2001 to 2006 and included locations with greater than 100 valid measurements to ensure estimate representativeness. The NO_2 surface was estimated from tropospheric NO_2 columns retrieved from the Ozone Monitoring Instrument (OMI) and also used GEOS-Chem to calculate the relationship between the NO_2 column and surface NO_2 (Lamsal et al. 2008). NO_2 estimates used data from 2005 to 2007 as OMI measurements began in late 2004. Both $PM_{2.5}$ and NO_2 were estimated at a 0.1x0.1 degree resolution (~10x10km). The O_3 surface was created from the Canadian Regional and Hemispheric O_3 and NO_x System (CHRONOS) (Environment Canada 2011). This model is reinitialized every 24 hours with meteorology and is fused with the O_3 observations across Canada and the U.S. on an hourly basis using an optimal interpolation approach based upon a least square combination of the CHRONOS and measured O_3 data that minimized the error variance. This surface was created at a 21km resolution and represents average summer (May through September) concentrations from 2004 to 2006. Figure 6 illustrates the $PM_{2.5}$, NO_2 and O_3 pollutant surfaces used to represent current spatial concentrations across Canada. Next, these surfaces were calibrated

with NAPS monitoring data to estimate historical annual spatial exposure surfaces.

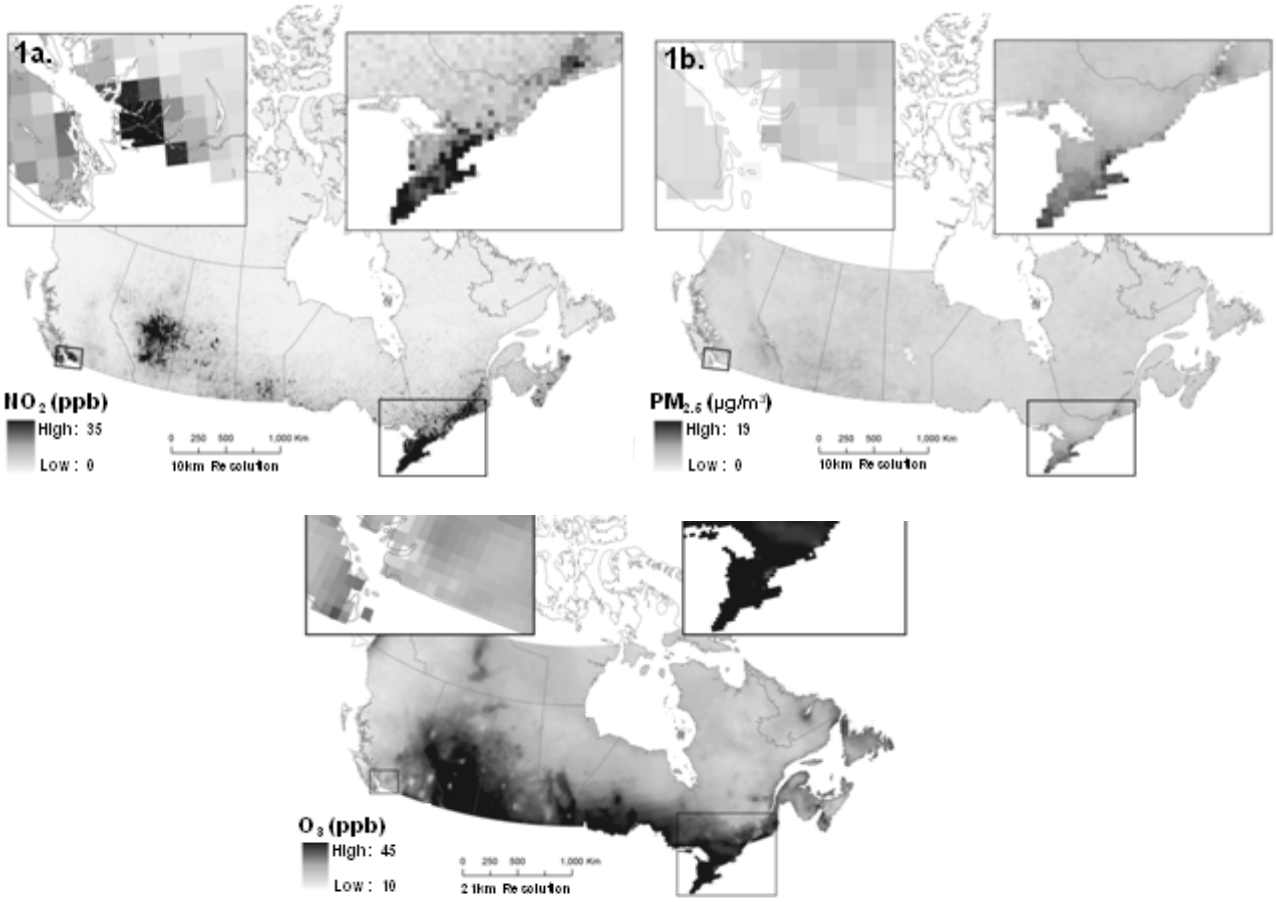

Figure 6. National pollutant surfaces created from recent satellite estimates (for $PM_{2.5}$ and NO_2) and a dispersion model (for O_3). Insets represent higher population density locations in Canada (south western BC and southern Ontario and Quebec).

Air Pollution Monitoring Data

The NAPS monitoring network began measurements of TSP in 1970, NO_2 and O_3 in 1975 and $PM_{2.5}$ and PM_{10} in 1984. Figure 7 illustrates the location of all NAPS monitors in Canada, 1975 TSP monitoring stations with 50km buffers (for reference of historical monitor spatial coverage) and all study participant residential postal codes between 1975 and 1994.

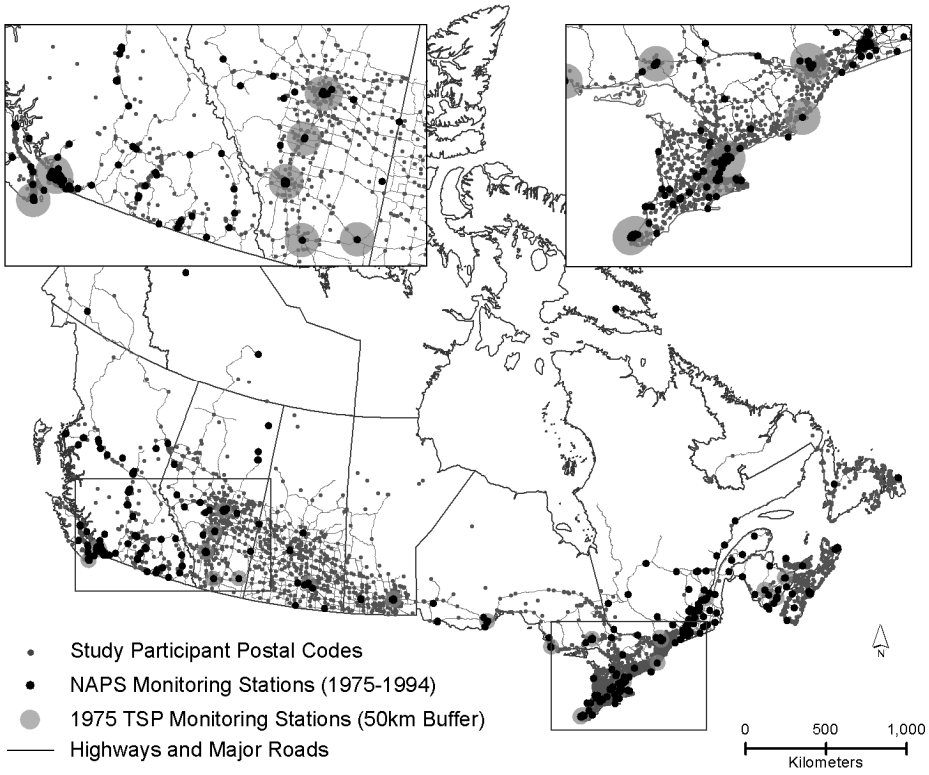

Figure 7. Location of all national air pollution surveillance monitors in Canada and study participant residential postal codes between 1975 and 1994. Insets represent higher population density locations in Canada (south western BC and southern Ontario and Quebec).

NAPS monitoring data were first formatted into monthly averages for all pollutants. Continuous monitoring data were included if at least 50% of daily hourly observations were available and at least 50% of days were available in a month. Monthly averages from dichotomous samplers ($PM_{2.5}$) required a minimum of 3 of 5 valid monthly measurements. Yearly averages were not calculated unless there were at least six months of complete data with one month per season, and summer O_3 averages unless there were 3 months of data available. Appendix 2, Figure 1 illustrates historical annual average pollutant concentrations from available NAPS monitoring stations that were in operation for all years. Temporal trends show a large decrease in TSP concentration during the study period (51% from 1970 to 1994),

a decrease in NO_2 (28% from 1975 to 1994) and $PM_{2.5}$ (32% from 1984 to 1994), and an increase in O_3 (19% from 1975 to 1994). Importantly, the changes in pollutant concentrations were not uniform across geographic areas in Canada.

Modeling Historical $PM_{2.5}$ Concentrations from TSP

Due to the lack of historical spatial and temporal $PM_{2.5}$ measurement coverage, we used collocated $PM_{2.5}$ and TSP measurements between 1984 and 2000 to create predictive models of historical $PM_{2.5}$ concentrations. The overall approach to estimating $PM_{2.5}$ is similar to that used by Lall et al. (Lall et al. 2004) to estimate metropolitan area specific $PM_{2.5}$ and PM_{10} relationships with TSP across the U.S. We used random effect models (GLIMMIX procedure in SAS 9.3) to account for the clustering of annual measurements over time at each NAPS station. Table 4 summarizes the final $PM_{2.5}$ model incorporating TSP concentrations ($\mu g/m^3$) and census metropolitan area (CMA) indicator variables. The R^2 and RMSE for the $PM_{2.5}$ model was 0.67 and 2.31. Figure 8 illustrates the measured and predicted $PM_{2.5}$ concentrations. The resulting $PM_{2.5}$ model was applied to all valid TSP monitoring stations; the nearest CMA core within 100km was used to determine the CMA model coefficient for the $PM_{2.5}$ model, otherwise no CMA variable was included in the model. Figure 2 in the Appendix 2 maps the CMA's used in the model and areas covered by the 100km buffers.

Table 4. Model used to predict historical $PM_{2.5}$ using TSP measurements and census metropolitan area indicator variables.

Variables	Estimate	SE	p
$PM_{2.5}$ Model [R^2=0.67[1], RMSE=2.31[1]]			
Intercept	1.93	2.30	0.42
TSP	0.13	$1.78e^{-2}$	<0.001 *
CMA Indicator			
Calgary	0.44	2.63	0.87
Edmonton	-1.82	2.69	0.50
Halifax	7.71	3.02	0.01*
Hamilton	4.76	3.02	0.12
Montreal	6.01	2.42	0.01*
Ottawa	4.86	2.94	0.10
Quebec	3.17	2.60	0.23
St. Johns	5.72	3.81	0.13
Saint John	3.28	30.7	0.29
Toronto	5.63	2.60	0.03*
Vancouver	6.50	2.47	0.01*
Victoria	2.48	2.73	0.36
Windsor	5.63	2.56	0.03*
Winnipeg	1.00	-	-

[1]R^2 and RMSE estimated by regressing the predictions from the fixed-effects terms against measured values.
*Significant at $p<0.05$.

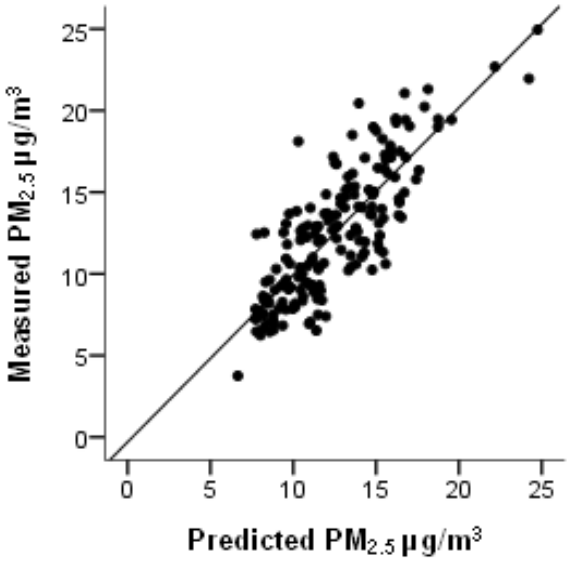

Figure 8. Predicted $PM_{2.5}$ concentrations using TSP concentrations and metropolitan indicator variables against NAPS $PM_{2.5}$ measurements.

Calibrating Spatial Pollutant Surfaces Using Historical Data

Two approaches were used to extrapolate current $PM_{2.5}$, NO_2 and O_3 surfaces to estimate annual concentrations between 1975 and 1994. Both approaches were developed using 90% of the monitoring data available for each year, while retaining 10% for model evaluation. Model performance was assessed using adjusted R^2 and root-mean-square error (RMSE).

The first approach calibrates the current spatial surfaces using annual NAPS monitoring data and smoothed inverse distance weighting (IDW) interpolation of the ratio's of spatial co-located historical NAPS and surface estimates. The yearly calibrations were performed using the following equation:

Equation 1 $$Yearly\ Historical\ Surface_j = Surface_{x,y} \times \left[\frac{\sum_{k=1}^{Nnaps} \left(\frac{1}{(d_{x,y,k})} \times \frac{\overline{NAPS_k^J}}{Surface_k} \right)}{\sum_{k=1}^{naps} \frac{1}{(d_{x,y,k})}} \right]$$

Where for each year between 1975 and 1994 the annual historical surface for pollutant *j* is equal to the current spatial surface of pollutant *j* ($Surface_{x,y}$) at coordinates *x,y* multiplied by the IDW interpolation of the ratio's of spatial co-located historical NAPS and surface estimates. $d_{x,y,k}$ is the distance (km) from NAPS monitoring station *k* to location *x,y*. $\overline{NAPS_k^J}$ and $Surface_k$ are coincidently sampled pollutant concentrations of *j* at station *k*. A smooth interpolation option (smooth factor=0.2) was included in the IDW interpolation (not shown in equation 1 for simplicity), which uses three ellipses in the interpolation method: points that fall outside the smaller ellipse but inside the largest ellipse are weighted using a sigmoid function (ESRI 2012). The smoothed IDW function was used to reduce abrupt

changes in the yearly calibration surfaces as these do not reflect spatial patterns of pollution change.

The second approach uses linear regression to model annual concentrations. Predictor variables include the spatial pollutant surfaces, a time-trend and historical population density data (no other detailed land use variables are available historically). Population location data were derived from the 1971, 1976, 1981, 1986, 1991, and 1996 Canadian census; between census years were assigned the nearest census. The annual population density variables were calculated in a GIS for various buffer distances (1km to 50km's) around each NAPS monitor. Roads and industry were not included in the models as we want to separately evaluate exposure to these sources and lung cancer risk. We used random effect models (GLIMMIX procedure in SAS 9.3) to account for the clustering of annual measurements over time at each NAPS station and selected predictor variables that maximized model fit. We estimated R^2 and RMSE statistics by predicting the measurement data with the fixed-effect coefficients using ordinary least squares regression.

Exposure to Vehicle Emissions

Exposures to vehicle emissions were estimated using proximity measures to highways (freeways and major highways) and major roads (freeways, highways, and arterial and collector roads). The 1996 DMTI Inc. road network was used to derive proximity measures for all case and control residential years, due to the lack of historical national road networks. The average distance to each road class was calculated separately as well as the number of years residing within 50, 100 and 300m of a highway and/or major road. These proximity distances were selected as vehicle related pollutant gradients, such as for NO_2 and volatile organic compounds, are highest within 50 and 100m of a major road but remain significantly elevated to 300m (Zhou and Levy 2007).

Emissions from vehicles have changed significantly over time due to increases in vehicle kilometres travelled and improved vehicle emission controls (Kahn 1996; Sawyer et al. 2000). Exposure indicators for years residing near highways and major roads were therefore weighted to account for these changes. Appendix 2, Figure 3 shows the decrease in the total NO_x emissions from on-road mobile sources in Canada (used here to represent primary vehicle emissions), including heavy and light duty diesel and gasoline vehicles, from 1980 to 2007 and extrapolated levels to 1970. NO_x emissions estimates were compiled by Environment Canada using the latest emission estimation methodologies and statistics available as of March 2008. Emission factors were developed using MOBILE6.2C and the number of vehicle kilometres travelled. MOBILE6.2C is a vehicle emissions modeling software specific to Canada and accounts for the vehicle fleet profile, vehicle emission standards, and fuel characteristics (NPRI 2011). Given the NO_x emissions trends documented in the United States from 1970 to 1980 (EPA 2005), linear extrapolation was used to estimate NO_x emissions from 1980 to 1970. The ratio of resulting 1994 and 1975 NO_x emission estimates suggest that living near a major road in 1975 is equivalent to 1.26 "1994" years due to changes in vehicle emissions (the ratio also accounts for changes in vehicle numbers). A weighting factor (1+0.013*(1994-proximity exposure year)) was therefore used to adjust proximity-based vehicle exposures to account for decreases in the magnitude of vehicle emissions over the study period.

Exposure to Industrial Emissions

A comprehensive inventory of industrial emissions sources was compiled as part of the NECSS within the Environmental Quality Database (EQDB) (Argo 1998; Argo 2007; Johnson et al. 1998). Locations of industrial manufacturing facilities and activities in approximately fifty standard industrial classifications (SIC) from 1970 to 1994 are included in the database

along with operational time periods. Approximately 7800 sources with a 4 digit SIC are included and 8200 municipal waste facilities. Major industries, including metal smelters, pulp and paper mills, petroleum product companies, foundry and steel plants, aluminum smelters, non-hydro power plants, and petrochemical companies, contain pollutant discharge estimates while minor industrial sources have no emission records. The distance between an industrial source and a subjects' postal code has been validated to +/-150m in urban locations (Argo 2007). The EQDB has been used in conjunction with the NECSS to examine leukemia and chlorination by-products (Kasim et al. 2006) and residential proximity to industrial plants and Non-Hodgkin's Lymphoma (Johnson et al. 2003). We calculate exposure to major industrial emissions and to minor sources within 1, 2 and 3km buffers from residential postal codes. These distances were selected to ensure specificity of proximity based exposure assessments for multiple industries and substances. Similar distance thresholds have been used previously in small area health studies (Aylin et al. 1999; Ramis et al. 2009). To be considered exposed, and to calculate the number of years exposed to each proximity category, at least 1 industrial facility had to be operating within the associated buffer distance.

3.3. Results

Residential Histories

The NECSS questionnaire asked participants to list each place in Canada that they had lived for at least one year. A total of 8176 individuals (98%) reported at least one full 6-digit postal code and 6918 individuals (83%) reported at least 15 years of residential histories from 1975 to 1994. On average, individuals reported 2.3 (SD=1.6) different residences from 1975 to 1994; 1617 individuals lived only in rural areas and 4222 individuals lived only in urban areas of Canada. Urban areas were defined using Statistics Canada community size classifications

(urban core, urban fringe, urban areas outside of CMA, rural fringe, and rural areas outside of CMA). In total, 77% of the studies exposure-years occurred in urban areas.

Importantly, while no significant difference (p=0.54) was found in the number of geocoded residential-years between cases and controls for the 1975 to 1994 exposure period, cases tended to report older addresses more often than controls. Recall bias was especially evident for residential histories prior to 1975, as shown in Figure 9.

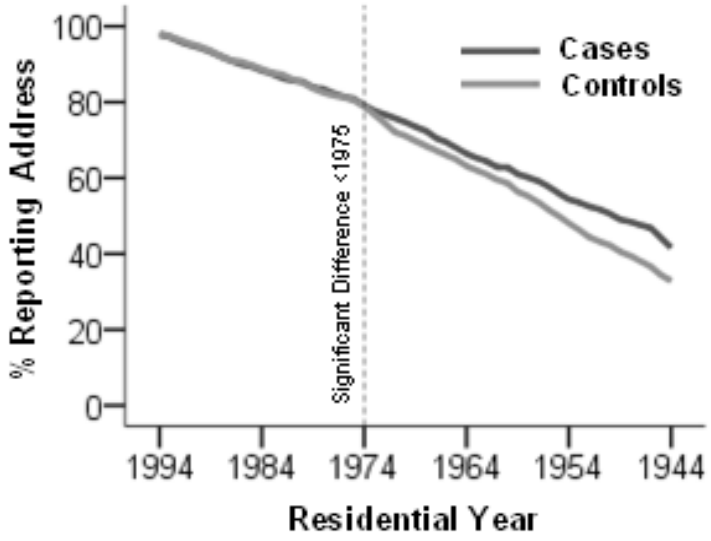

Figure 9. Percent of cases and controls reporting addresses at the 6-digit postal code level from 1994 (start of case-control study enrollment) and 1944.

Ambient Exposure Assessments

The first approach to calibrating current pollution surfaces used IDW interpolation to create annual surfaces between 1975 and 1994. Figure 10 illustrates the resulting $PM_{2.5}$ exposure surfaces for 1975, 1980, 1985, 1990 and 1994, $PM_{2.5}$ measurement locations with 50km buffers, the average $PM_{2.5}$ exposure surface between 1975 and 1994, and the location of the case-control study subjects. Twenty annual exposure surfaces were created from1975 to

1994, but only five are shown here. The study population residential years indicates the locations of all yearly residential histories during the twenty year exposure period summed within a 50km grid. The temporally adjusted surfaces for NO_2 and O_3 are provided in Figures 4 and 5 of the Appendix 2.

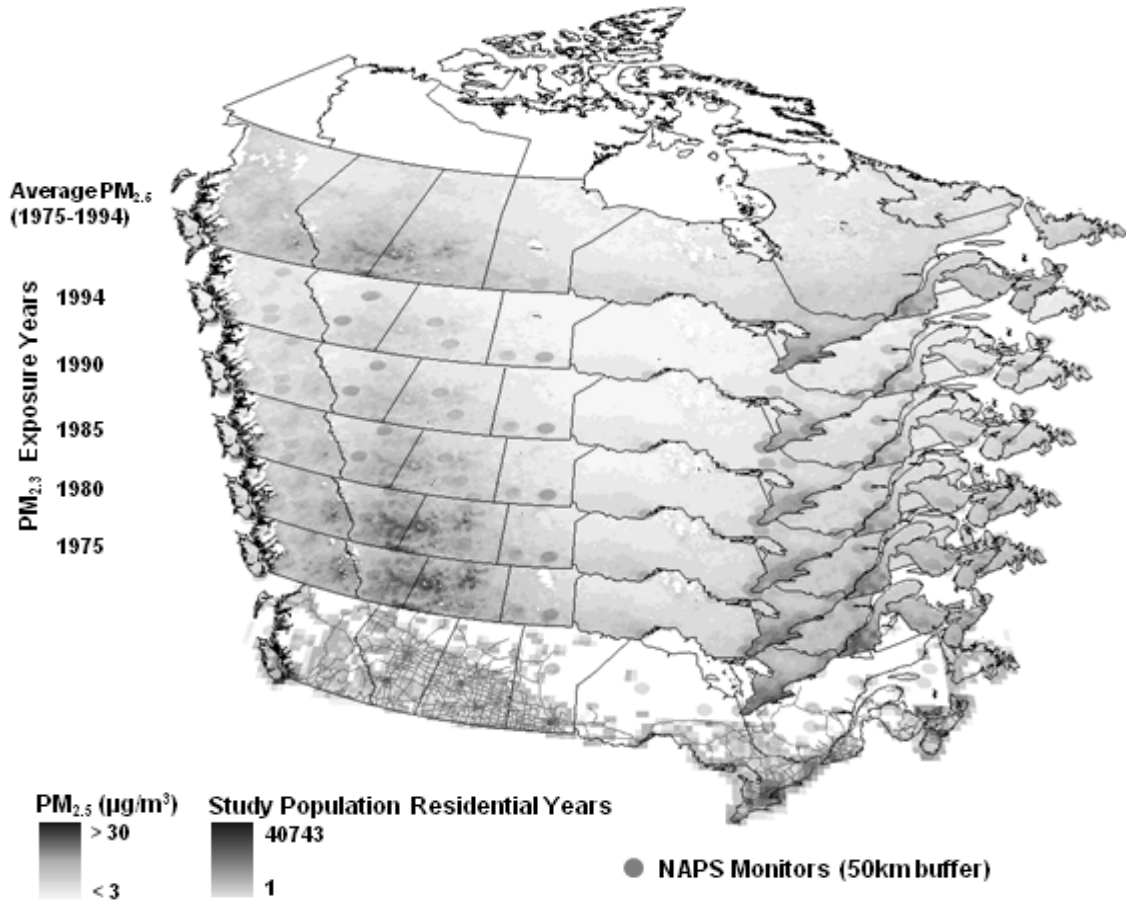

Figure 10. Example of annual $PM_{2.5}$ exposure surfaces created using the IDW interpolation calibration approach for all years between 1975 and 1994. The study population residential years represents all residential locations during the study period summed within a 50km grid.

The performance of the linear regression models was moderate for all three pollutants ($PM_{2.5}$ R^2=0.33, NO_2 R^2=0.36 and O_3 R^2=0.47) as described in Table 5. Population density within 10km of monitoring stations was most strongly associated with $PM_{2.5}$, while population

density with 5km was most strongly associated with NO_2 (positively associated) and O_3 (negatively associated). A linear time-trend did not improve the O_3 model and was therefore not included in the final model.

Table 5. Results of historical $PM_{2.5}$, NO_2 and O_3 linear regression models.

Model	Distance	Value	SE	p
$PM_{2.5}$ Model [R^2=0.33, RMSE=3.57]				
Intercept	-	1.18	1.16	0.31
Satellite $PM_{2.5}$	-	0.46	0.11	<0.001
Population Density	10km	$3.94e^{-6}$	$2.89e^{-7}$	<0.001
Years <1994	-	0.29	$9.28e^{-3}$	<0.001
NO_2 Model [R^2=0.36, RMSE=7.00]				
Intercept	-	10.88	1.07	<0.001
Satellite NO_2	-	1.67	0.46	<0.001
Population Density	5km	$2.6e^{-5}$	$5.11e^{-6}$	<0.001
Years <1994	-	0.28	0.028	<0.001
O_3 Model [R^2=0.47, RMSE=5.13]				
Intercept	-	6.85	1.66	<0.001
O_3 Dispersion Model	-	0.73	0.06	<0.001
Population Density	5km	$-2.0e^{-5}$	$2.5e^{-6}$	<0.001

Evaluation of the two historical calibration approaches are shown in Table 6, which summarizes the R^2 and RMSE of model evaluations using the 10% sample of monitoring data withheld each year. The fused IDW interpolation of $PM_{2.5}$ had the best performance (R^2=0.51), while the NO_2 and O_3 linear models had the best performance (R^2=0.38 and R^2=0.56). Model performance tended to decrease for older measurements, but not substantially. Appendix 2, Figure 6 presents the scatter plots for each model evaluation.

Table 7 presents the exposure assessment results using both historical calibration methods and air pollution exposures derived from NAPS monitoring data within 50km of residential postal codes. To ensure accurate exposure assessment, results are presented for individuals with at least 15 complete exposure-years between 1975 and 1994. Exposures for

different time-periods (e.g. 1975-1980, 1975-1985, and 1975-1990) were also calculated to examine different latency periods (data not shown).

Table 6. Evaluation of IDW interpolation and linear regression models to predict annual historical air pollution.

				IDW Interpolation		Linear Models	
	Year	**Stations**	**N**	**R^2**	**RMSE**	**R^2**	**RMSE**
NO_2	All	120	1030	0.22	6.66	0.38	5.92
	1994-1990	94	349	0.30	5.66	0.36	5.42
	1989-1985	88	300	0.20	6.61	0.44	5.54
	1984-1980	62	226	0.13	6.72	0.40	5.62
	1979-1975	52	155	0.17	8.75	0.29	8.07
$PM_{2.5}$	All	177	1826	0.51	2.96	0.30	3.53
	1994-1990	106	446	0.64	1.96	0.32	2.70
	1989-1985	113	480	0.57	2.30	0.36	2.81
	1984-1980	124	476	0.34	3.79	0.12	4.36
	1979-1975	123	424	0.43	3.32	0.26	3.77
O_3	All	187	1440	0.39	5.29	0.56	4.48
	1994-1990	158	582	0.53	4.92	0.65	4.25
	1989-1985	125	409	0.36	5.41	0.54	4.57
	1984-1980	80	286	0.25	4.67	0.28	4.57
	1979-1975	48	163	0.22	6.33	0.60	4.50

Table 7. Ambient exposure estimates derived from NAPS monitors within 50km of residential postal codes and spatiotemporal exposure models.

Pollutant	**N***	**Mean**	**SD**	**Min**	**IQR**	**Max**
NAPS Measurements ≤50km						
TSP ($\mu g/m^3$)	4027	60.0	16.9	22.3	21.4	114.1
Modeled $PM_{2.5}$ ($\mu g/m^3$)[a]	4027	17.0	2.5	11.9	3.4	25.7
NO_2 (ppb)	3649	23.4	6.0	6.0	7.6	37.8
O_3 (ppb)[b]	4382	21.0	3.9	7.0	5.3	32.6
Spatiotemporal IDW Interpolation						
$PM_{2.5}$ ($\mu g/m^3$)	6833	11.3	2.6	3.6	3.9	19.0
NO_2 (ppb)	6919	15.3	8.8	1.1	14.5	43.4
O_3[b] (ppb)	6919	23.2	3.7	12.9	4.6	35.4
Linear Regression Models						
$PM_{2.5}$ ($\mu g/m^3$)	6833	9.1	1.9	4.7	2.2	16.1
NO_2 (ppb)	6919	17.7	4.1	13.1	5.0	35.1
O_3[b] (ppb)	6919	26.4	3.4	18.1	4.7	37.2

*Number of individuals with ≥15 complete exposure-years.
[a] Modeled using TSP and CMA indicator variables as described previously in Table 5.
[b] Summer (May through September) O_3.

Exposure to Vehicle and Industrial Emissions

Proximity measures used to represent exposure to vehicle emissions are summarized in Table 8. Individuals lived within 50, 100 and 300m of a highway for a mean of 0.5 (SD=2.9), 1.1 (SD=4.0) and 2.9 (SD=6.3) years, respectively. Exposure years increased slightly when weighted by temporal emission changes. The average mean distance from study participants' postal codes to the nearest highway was 3.9km. When residential histories were restricted to urban areas (where proximity is a more accurate measure of exposure than in rural areas), the distance to highways and major roads decreased substantially. Over half of the study population was exposed to emissions from a major road at some point during the study period (i.e. had lived at least one year within 300m of a major road).

Table 8. Proximity measures to highways and major roads.

Proximity Measure	**# of People Exposed**[a]	***# of Years Exposed (Mean ±SD)***	**# of Weighted**[b] **Years Exposed (Mean ±SD)**
Highways			
≤ 50m	341	*0.5 (2.9)*	0.7 (3.9)
≤ 100m	647	*1.1 (4.0)*	1.5 (5.4)
≤ 300m	1640	*2.9 (6.3)*	4.0 (8.5)
Major Roads			
≤ 50m	1438	*2.3 (5.5)*	3.2 (7.6)
≤ 100m	2283	*4.0 (6.9)*	5.5 (9.5)
≤ 300m	4517	*10.1 (8.8)*	13.8 (12.1)

[a] *Number of individuals living >1 year within 50/100/300m of a highway or major road.*
[b] *Weighted to account for temporal changes in vehicle emissions.*

The number of years study participants lived within 1, 2 and 3km of a major and minor industry are summarized in Table 9, as are aggregated emission estimates for major industrial sources. Proximity to specific emission sources (e.g. oil refineries, smelters, and pulp and paper mills) were also calculated (data not shown). Individuals lived within 1, 2 and 3km of a major industrial source for a mean of 1.6 (SD=5.3), 4.3 (8.3) and 6.4 (9.5) years respectively.

Over half of the study population (n=5942) lived within 3km of a minor industrial source for at least one year between 1975 and 1994.

Table 9. Proximity measures to major and minor industrial sources.

Proximity Measure	# of People Exposed[a]	# of Years Exposed (Mean ±SD)	# of Facilities (Mean ±SD)	Emissions[b] (tonnes) (Mean ±SD)
Major Industries				
≤ 1km	838	1.6 (5.3)	6.2 (5.5)	$4.5e^5$ ($3.6e^7$)
≤ 2km	1995	4.3 (8.2)	13.3 (11.6)	$4.5e^5$ ($3.5e^7$)
≤ 3km	2743	6.4 (9.5)	21.3 (18.6)	$1.9e^3$ ($1.6e^4$)
Minor Industries				
≤ 1km	4137	11.4 (11.2)	32.6 (59.3)	-
≤ 2km	5515	16.7 (10.0)	115.7 (163.2)	-
≤ 3km	5942	18.9 (9.0)	218.0 (303.8)	-

[a] Number of individuals living >1 year within 1/2/3km of a major or minor industrial source.
[b] Summary of facility emissions > 0 tonnes. Only available for major industries.

Disregarding Residential Histories and Exposure Error

A total of 3305 study participants (40%) lived at their study entry address for the entire twenty year exposure period, while 622 (7.6%) participants lived for 15-19 years, 970 (11.9%) for 10-14 years, 1433 (17.5%) for 5-9 years, and 1756 (23%) for less than 5 years. Correlation between ambient air pollution exposures derived from study entry residential addresses only, in place of exposures derived from residential histories and spatiotemporal air pollution models, were relatively high for $PM_{2.5}$ r=0.70, NO_2 r=0.76 and O_3 r=0.72. However, when examining exposure misclassification based on incorrectly assigned exposure quintiles, 50%, 49% and 46% of individuals where classified into a different $PM_{2.5}$, NO_2 and O_3 quintile. When temporal variation is removed from the exposure assessment (i.e. historical exposures are derived from residential histories applied to the current spatial pollution surfaces) 17%, 15% and 14% of individuals where classified into a different $PM_{2.5}$, NO_2 and O_3 exposure quintile. Similar results were found for proximity based exposures, for example, 30% of

individuals classified as not exposed to highway emissions based on their address at study entry were actually exposed when residential histories were used for exposure assessment.

3.4. Discussion

Incorporating residential mobility in chronic air pollution studies is fundamental to accurate exposure estimates. Boscoe (2011) presents a review of environmental health studies that have incorporated residential histories. In our study, only 40% of participants lived at their study entry residence for the entire 20 year exposure period; on average, 2.3 (SD=1.6) different residences per subject were reported. Recall bias was present for self-reported residential histories prior to 1975, with cases recalling older residences more often than controls. This has important implications for environmental epidemiology using self-reported residential histories as many environmental exposures have decreased substantially over time. Consequently, exposure assessment based on a greater proportion of older residential histories in cases compared to controls will result in an upward bias, rather than non-differential bias typically assumed from exposure misclassification. Studies that incorporate self-reported residential histories, particularity long-term residential histories - in this case over twenty years, may need to account for reporting bias in epidemiological analysis.

This study also demonstrated the importance of estimating air pollution exposures from residential histories, both in terms of including different residential locations as well as the corresponding spatiotemporal air pollution concentration estimates. Exposure quintiles based on residential addresses at study entry had approximately 50% correspondence to exposure quintiles developed from residential histories and spatiotemporal air pollution surface. These results address one of the research opportunities suggested by Meliker and Sloan (Meliker and Sloan 2011): "indentifying circumstances under which it is worthwhile to

compile and incorporate extensive space–time data histories of mobility or environmental contaminants". Epidemiological studies of diseases with long latency periods (in this case lung cancer) and/or that examine spatially and temporally varying exposures (in this case ambient air pollution) are clearly such circumstances.

Despite the fact that the Canadian NAPS monitoring network is one of the longest-standing national air pollution monitoring programs worldwide and now covers the majority of urban centers in Canada, its limited spatiotemporal coverage necessitated the creation of national models that capture both urban and rural populations. We were able to use NAPS data within 50km of residential postal codes to assign exposures to 63%, 70% and 54% of exposure-years for TSP, O_3 and NO_2. Very limited spatial and temporal $PM_{2.5}$ monitoring data were available (only 40% of exposure-years between 1984 and 1994 could be assigned) and we therefore estimated historical $PM_{2.5}$ using TSP and metropolitan area indicator variables. The resulting models predicted $PM_{2.5}$ variability well; the ratio for modelled $PM_{2.5}$/TSP (0.32, SD=0.12) is very similar to that found in US metropolitan areas ($PM_{2.5}$/TSP=0.30, SD=0.11) (Lall et al. 2004).

National spatial pollutant surfaces were compiled and calibrated with historical NAPS data to assign ambient pollutant concentrations to all study participants' residential postal codes between 1975 and 1994. The two approaches used to calibrate spatial pollutant surfaces differ in their approach to account for temporal and spatial change; IDW interpolation accounted for the heterogeneity in pollution level changes across Canada during the exposure period, while linear regression models incorporated a linear time-trend and population density as a spatial predictor. The interpolation approached better represented historical $PM_{2.5}$ concentrations, potentially due to the larger spatial scale of $PM_{2.5}$, while the linear regression

models better represented historical NO_2 and O_3 concentration, which have finer spatial resolutions.

The creation of national spatiotemporal models allowed for the inclusion of all study participants, regardless of geographic location and NAPS monitor coverage. This was important as 42884 (23%) of exposure-years occurred in rural areas. The mean $PM_{2.5}$, NO_2 and O_3 exposure estimates derived from the spatiotemporal models were 11.3µg/m^3 (SD=2.6), 17.7ppb (4.1), and 26.4ppb (3.4) respectively. The magnitude of these exposures are less than those used in other studies, for example, the widely cited ACS study ($PM_{2.5}$: 17.7µg/m^3 (3.0), NO_2 21.4ppb (7.1); and O_3 45.5ppb (7.3)) (Pope III et al. 2002). This is likely due to the inclusion of rural study participants as well as lower ambient pollution levels in Canada. The ability to incorporate rural areas in the exposure assessment added to the variability in the studies exposure estimates, particularly for NO_2 and O_3, as the majority of historical NAPS measurements in Canada represent pollutant concentration in large urban areas.

The results of the retrospective air pollution modeling approach conducted here are comparable to other such studies; however, the majority of retrospective air pollution exposure assessments have been conducted solely for urban areas. For example, Bellander et al. (Bellander et al. 2001) used emission data, dispersion models, and geographic information systems (GIS) to assess exposure to NO_2, NO_x and SO_2 ambient air pollution during 1960, 1970 and 1980 in Stockholm, Sweden. Model evaluation using historical data was not possible, but the model was found to have high correlation (r=0.96) with aggregated 1994-1997 data from 16 monitors. In terms of national models, Hart et al. (Hart et al. 2009) developed U.S. nationwide models of annual exposure to PM_{10} and NO_2 from 1985 to 2000. Generalized additive models were used to predict spatial surfaces from monitoring data and GIS–derived covariates (e.g. distance to road, elevation, proportion of low-intensity

residential, high-intensity residential, and industrial, commercial land use). Model performance (R^2) for PM_{10} and NO_2 was 0.49 and 0.88 respectively. Another national retrospective study was conducted as part of the Netherlands Cohort Study on Diet and Cancer (Beelen et al. 2007). Ambient air pollution exposures were estimated using regional (IDW monitor interpolation), urban (regression modelling), and local (road proximity) components. This approach explained 84%, 44%, 59% and 56% of the variability in averaged monitor data between 1976 and 1997 for NO_2, NO, BS and SO_2, respectively. The density of monitors in the Netherlands and the use of aggregated monitoring data may explain the higher model performance than seen in this study.

The exposure assessment approach presented here capitalizes on study participants' lifetime residential histories and incorporates comprehensive modelling approaches to estimate exposures to ambient air pollution and to vehicle and industrial emissions. Nevertheless, there are several limitations to this approach that may lead to exposure misclassification. Due to privacy concerns, residential addresses were coded using a standard geographic reference of 6-digit postal codes. Using a set geographic reference reduced error from changing postal codes over time; however, the spatial accuracy of postal codes varies substantially between urban and rural areas of Canada. Proximity analyses for exposures to vehicle and industrial emissions will therefore be more accurate in urban areas. The ambient air pollution exposure assessment relies on the accuracy of NAPS monitoring data, and historical monitor locations, especially in rural areas, may have been sited to capture local pollution problems. Unfortunately, no historical data were available to evaluate the representativeness of NAPS monitoring data. Due to sparse temporal and spatial $PM_{2.5}$ monitor coverage, we created historical models based on TSP monitoring data and CMA indicator variables. While the model had good prediction, it was created from a limited number of

monitoring stations from 1984 to 2000. Nevertheless, several studies have estimated $PM_{2.5}$ successfully from TSP (Katanoda et al. 2011; Lall et al. 2004). The accuracy of the final spatiotemporal $PM_{2.5}$, NO_2 and O_3 surfaces is also determined from the initial concentration surface as well as fusion with historical NAPS monitoring data or predictions incorporating a linear time-trend and population density. Some anomalies exist in the current spatial surfaces, for example, high $PM_{2.5}$ concentrations in mountainous regions and $PM_{2.5}$ and NO_2 in certain locations in the Prairies; however, few study participants lived in these locations and exposure misclassification is therefore limited. All historical monitors were used to adjust annual spatial pollution surfaces, which resulted in urban monitor ratios extrapolated to rural areas. Few rural monitors exist and it was not possible to restrict to rural monitors when adjusting the spatial pollution surfaces in rural areas. Exposure to vehicle emissions was based on proximity measures to a national 1996 road network and a clear limitation was the lack of historical road databases. Industrial emissions were based on a comprehensive database on industrial locations from 1970 to 1994; however, emission estimates were only available for major industries, which restricted the examination of specific industrial chemicals when minor industries were included.

3.5. Conclusions

We conducted a comprehensive air pollution exposure assessment for a population based lung cancer case-control study of 8353 individuals using self-reported residential histories between 1975 and 1994. Incorporating residential histories was an important component of the exposure assessment approach, and necessitated the creation of national spatiotemporal air pollution models. Due to the lack of historical air pollution measurements, as well as differences in data availability between urban and rural areas, a number of

modelling approaches were used to assign annual ambient $PM_{2.5}$, NO_2 and O_3 concentrations, as well as proximity measures for vehicle and industrial emissions, to study participants' residential addresses. The exposure assessment methods developed here will allow subsequent epidemiological analyses to examine latency periods associated with lung cancer, include both urban and rural populations, and study the contributions of multiple ambient pollutants and local vehicle and industrial emissions to lung cancer risk in Canada. In addition, this exposure assessment has demonstrated the importance of including residential histories in long-term exposure assessments, as well as the need to carefully examine self-reported residential histories for recall bias.

Chapter 4

Long-Term Residential Exposure to Air Pollution and Lung Cancer Risk

Hystad, P., Demers, P.A, Johnson, K.C., Carpiano, R.M., Brauer, M. Long-Term Residential Exposure to Air Pollution and Lung Cancer Risk. *Epidemiology*. Accepted January 25th 2013.

Abstract

Background: There is accumulating evidence that air pollution causes lung cancer. Still, questions remain about exposure misclassification, the components of air pollution responsible, and the histological subtypes of lung cancer that might be produced.

Methods: We investigated lung cancer incidence in relation to long-term exposure to three ambient air pollutants and proximity to major roads, using a Canadian population-based case-control study. We compared 2,390 incident, histologically-confirmed lung cancer cases with 3,507 population controls in eight Canadian provinces from 1994-1997. We developed spatiotemporal models for the whole country to estimate annual residential exposure to fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2) and ozone (O_3) over a 20-year exposure period. We carried out a sub-analysis in urban centers, using exposures derived from fixed-site air pollution monitors, and also considered traffic-proximity measures. Hierarchical logistic regression models incorporated a comprehensive set of individual and geographic covariates.

Results: The increase in lung cancer incidence (expressed as fully adjusted odds ratios (OR)) was 1.29 (95% confidence interval=0.95-1.76) with a 10-unit increase in $PM_{2.5}$ ($\mu g/m^3$), 1.11 (1.00-1.24) with a 10-unit increase in NO_2 (ppb), and 1.09 (0.85-1.39) with a 10-unit increase in O_3 (ppb). The urban monitor-based sub-analyses generally supported the national results, with larger associations for NO_2 (OR =1.34; 1.07-1.69) per 10 ppb increase. No dose-response trends were observed, and no clear relationships were found for specific histological cancer subtypes. There was the suggestion of increased risk among those living within 100 m of highways, but not among those living near major roads.

Conclusions: Lung cancer incidence in this Canadian study was increased most strongly with NO_2 and $PM_{2.5}$ exposure. Further investigation is needed into possible effects of O_3 on the development of lung cancer.

4.1. Introduction

Evidence is accumulating for a causal association between exposure to ambient air pollution and lung cancer (Chen et al. 2008; Katanoda et al. 2011; Pope et al. 2011; Raaschou-Nielsen et al. 2011; Lepeule et al. 2012); however, several uncertainties remain. Air pollution exposure misclassification is a particular concern, due to the long latency period for lung cancer, temporal changes in air pollution levels, and the likelihood of substantial residential mobility during biologically relevant exposure periods. To date, few studies of lung cancer have incorporated historical exposure assessments (Beeson et al. 1998; Nafstad et al. 2003; Nyberg et al. 2000; Raaschou-Nielsen et al. 2011; Vineis et al. 2006) or examined different air pollutants and emission sources (Beeson et al. 1998; Nafstad et al. 2003; Nyberg et al. 2000; Vineis et al. 2006), especially beyond urban settings (Vineis et al. 2006; Beelen et al. 2008). In addition, little research has examined air pollution exposure and lung cancer risk by histological subtypes (Barbone et al. 1995; Chen et al. 2009; Katsouyanni et al. 1991; Liaw et al. 2010), due to the need for large sample sizes. Given the variation in risks associated with cigarette smoking and lung cancer histology (Pesch et al. 2012), as well as evidence from occupational (Villeneuve et al. 2011) and animal studies (Nagy et al. 2005), it is probable that risks associated with air pollution also vary by histological subtype.

The present study builds upon prior work to partially address these uncertainties by identifying associations between three ambient air pollutants and proximity to traffic emissions, and lung cancer incidence. Specifically, we use a Canadian population-based case-control study that includes comprehensive individual and geographic information on potential confounding factors such as cigarette smoking, second-hand smoke exposure, occupational hazards, and residential radon exposures, as well as complete 20-year residential histories

from 1975-1994. Spatiotemporal models were developed and applied to annual residential histories in both urban and rural locations to estimate long-term exposures to fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2) and ozone (O_3) (Hystad et al. 2012). An urban sub-analysis was also conducted using exposures derived from the nearest fixed-site air pollution monitors within 50 km, as well as proximity measures to highways and major roads.

4.2. Method

Study Design

The National Enhanced Cancer Surveillance System is a population-based, multi-cancer-site case-control study that includes 3,280 histologically confirmed lung cancer cases, and 5,073 population controls collected between 1994 and 1997 in eight of Canada's ten provinces. Johnson et al. (1998) describe the recruitment methodology and study design of the overall National Enhanced Cancer Surveillance System project. Between 1994 and 1997 cases were identified and randomly sampled for inclusion in the study by provincial cancer registries within 1-3 months of initial diagnosis. Population controls were selected from a random sample of people within each province, frequency matched on sex and five-year age categories to the overall collection of National Enhanced Cancer Surveillance System cancer cases (~20,000 cases including 19 types of cancer). Recruitment methods for controls depended on data availability and accessibility by province, and included provincial health insurance plans in five provinces, random digit dialing in two, and property assessment data in one. A research questionnaire was mailed to selected cases and controls and active follow-up was conducted. The response rate was 62% for contacted lung cancer cases and 67% for population controls. The research questionnaire collected comprehensive information regarding personal characteristics, lifetime occupational exposures, and residential histories. Residential histories

were geo-coded to 6-digit postal codes and are the basis of the air pollution exposure assessment. Due to residential mobility, postal codes were located in all provinces of Canada, requiring national-level exposure assessment.

National Air Pollution Exposure Assessment

Long-term exposures to ambient $PM_{2.5}$, NO_2, and O_3, and proximity to highways and major roads, were estimated from residential histories from 1975-1994. Residential histories were available prior to 1975; however, few air pollution measurements and geographic data were available for these years, and recall bias was present for residential histories prior to 1975 (cases tended to report more residences than controls (Hystad et al. 2012)). To ensure reliable exposure assessment, only persons with complete twenty-year residential histories in Canada during this period were included in the final analysis, which reduced the study to 2,390 cases and 3,507 controls. Various exposure periods were examined (e.g. 1975-1980/85/90), but ambient pollution exposures for all periods were highly correlated with the 1975-1994 period ($r \geq 0.96$).

The spatiotemporal air pollution exposure assessment approach is described in detail elsewhere (Hystad et al. 2012). Briefly, a multi-staged approach was used to assign annual concentrations of $PM_{2.5}$ and NO_2, and summer (May to September) O_3, to residential histories. First, national spatial surface estimates of each pollutant were created from recent satellite-based estimates at a 10x10 km resolution (for $PM_{2.5}$ (van Donkelaar et al. 2010) and NO_2 (Lamsal et al. 2008)) and from a 21x21 km resolution chemical transport model (for O_3 (Environmental Canada, 2011). Next, all fixed-site National Air Pollution Surveillance monitoring data were formatted to annual averages for the study period. Since $PM_{2.5}$ measurements were not available prior to 1984, a random-effects linear regression model was

used to estimate pre-1986 $PM_{2.5}$ based on total suspended particulate (TSP) measurements (as these were measured beginning in 1974) and metropolitan variables (Model R^2=0.67, root mean square error=2.31 $\mu g/m^3$). This approach is similar to others studies that have estimated $PM_{2.5}$ from TSP (Katanoda et al. 2011; Lall et al. 2004). Finally, yearly calibration of the national spatial pollutant surfaces was conducted by calculating a ratio of measured to surface estimates at each National Air Pollution Surveillance monitoring station. Smoothed inverse-distance-weighted interpolation was conducted using the ratios, and the resulting surface applied to adjust the spatial pollutant surface for each year in the 1975-1994 study period.

Figure 11 illustrates the average spatiotemporal pollutant surfaces from 1975-1994 and the location of study participants' residential histories (sum of residential postal-code locations within a 50-km grid). These maps represent pollution concentrations that would be assigned if there were no residential mobility; in practice, the exposure assessment was conducted using yearly pollutant concentrations and residential histories.

Urban Fixed-Site Monitor Exposure Assessment

An urban sub-analysis was conducted using air pollution exposures derived solely from fixed-site National Air Pollution Surveillance measurements. As mentioned, the spatial and temporal coverage of $PM_{2.5}$ monitors is limited prior to 1986 and TSP measurements and modeled $PM_{2.5}$ are thus examined in the urban analysis. Annual average pollutant concentrations were calculated for postal codes using the nearest National Air Pollution Surveillance monitor (within 50 km) with at least six months of complete measurements and one month per season for TSP and NO_2, and at least three summer months for O_3. Cumulative averages were calculated for people with at least 18 years of complete monitor coverage from 1975-1994.

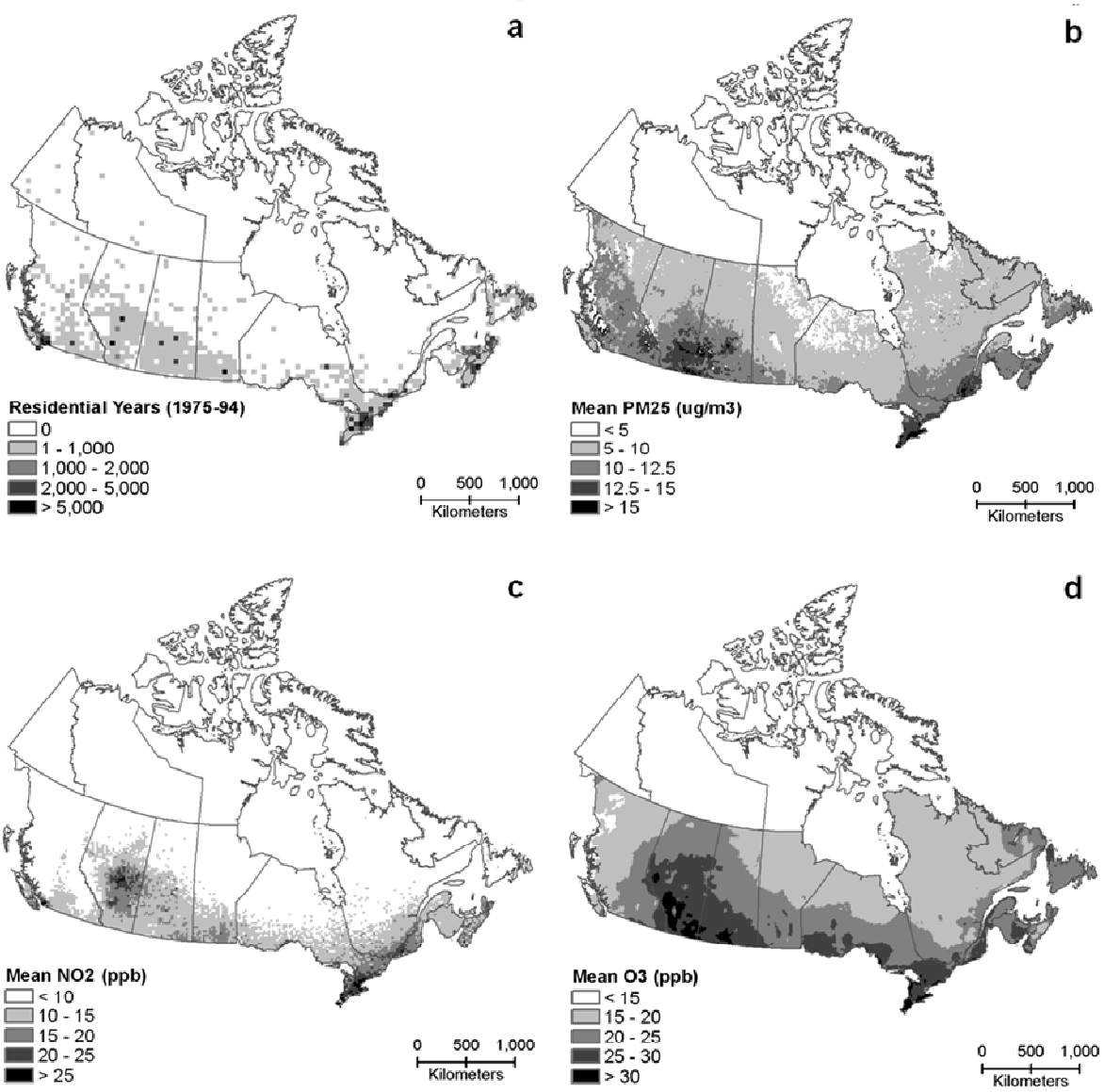

Figure 11. Location of study participant residential histories (a), and average $PM_{2.5}$ ($\mu g/m^3$) (b), NO_2 (ppb) (c), and O_3 (ppb) (d) concentrations for the period 1975 to 1994.

Proximity Measures to Highways and Major Roads

Proximity measures to major roads were used to estimate exposure to vehicle emissions. The 1996 (DMTI Spatial, Inc.) road network was applied to derive proximity measures for all residential years, due to the lack of historical national road networks. We calculated the

number of years residing within 50, 100 and 300 m of a highway or major road. Since emissions from vehicles have decreased significantly over the study period, proximity indicators were weighted to account for these changes using annual motor vehicle emission estimates (Hystad et al. 2012). Analyses of proximity to highways and major roads were also restricted to participants residing in urban areas (defined as >30,000 residents) due to large spatial errors associated with rural postal-code locations.

Outcomes

Histologically confirmed lung cancer incidence is the primary outcome variable of this study. We also examined specific histological subtypes, which for the 2,390 lung cancer cases with complete residential histories included: 669 (28%) squamous cell carcinoma; 756 (32%) adenocarcinoma; 363 (15%) small cell carcinoma; 213 (9%) large cell carcinoma; and 389 (16%) other or unspecified carcinomas (which are not included in subsequent analyses due to the heterogeneity of this category).

Covariates

We include a comprehensive set of individual and geographic-level variables in the multivariate models. Individual-level covariates include age, sex, educational attainment, average household income during the five years prior to study interview, smoking pack-years, years since quitting smoking, person-years of residential and occupational second-hand smoke exposure (defined by the number of smokers in the home multiplied by number of residential years and the number of smokers in the immediate work environment multiplied by number of occupational years), average alcohol and meat consumption per week, years working with daily or weekly exposure to dust, odors and hazardous substances, and exposure to specific occupational lung hazards (arsenic, asbestos, asphalt, benzene, mustard gas, welding, and wood dust). Geographic covariates included study province (to account for the study design),

ecological radon risk (defined using mean residential radon concentrations by Health Regions (Health Canada, 2013)), and neighborhood contextual deprivation variables (described in Appendix 3). Coding for all individual and geographic variables is provided in Appendix 3, Table 1.

Statistical Analyses

Analyses were conducted using two-level random-intercept logistic regression models (GLIMMIX, SAS version 9.3; SAS Institute, INC, Cary, NC). The random intercept was defined from Statistics Canada 1986 census division boundaries (n=188), representing regional areas in Canada, and assigned to each person's longest residential location to account for residual geographic patterns. We report ORs and 95% confidence intervals (95% CI) for 10-unit increases in ambient pollutant concentrations and for exposure quintiles. Only the national models were stratified by major lung cancer histological subtypes, given the reduced sample sizes for the urban subset analysis. National models were also stratified to examine pollutant interactions by a priori variables (smoking status, education and sex) that may modify the relationship between air pollution and lung cancer (Pope et al. 2002; Yorifuji et al. 2010; Beelen et al. 2008; Raaschou-Nielsen et al. 2011). Variance inflation factors were also used to quantify the severity of multicollinearity in the regression modes.

4.3. Results

Characteristics of Case and Control Subjects

Table 10 provides descriptive statistics and ORs (adjusted for age, sex and study province) for selected subject characteristics (descriptive statistics for all individual and geographic variables are shown in Appendix 3, Table 1).

Table 10. Descriptive statistics and odds ratios (ORs) for the association between lung cancer incidence and select subject characteristics

Variable	Cases (n=2390) No. (%)	Controls (n=3507) No. (%)	OR[a] (95% CI)
Age; Mean (SD)	63.5 (8.2)	59.0 (12.6)	NE[b]
Sex			
Female	1152 (48)	1719 (49)	NE[b]
Male	1238 (52)	1788 (51)	NE[b]
Smoking Pack Years			
Non Smoker	130 (6)	1337 (38)	1.00
1 - 19	319 (14)	1169 (34)	3.3 (2.6-4.2)
20 -29	467 (20)	392 (11)	15.1 (12.0-19.1)
30 - 39	519 (22)	247 (7)	27.9 (21.7-35.7)
40 - 49	446 (19)	149 (4)	39.3 (28.9-51.8)
50 - 59	205 (9)	69 (2)	40.6 (28.8-57.4)
≥ 60	235 (10)	79 (2)	44.4 (31.9-61.8)
Years Since Cessation of Smoking			
Non Smoker	130 (6)	1337 (38)	1.00
>35	29 (1)	177 (5)	1.3 (0.8-2.0)
26-35	70 (3)	312 (9)	2.0 (1.4-2.7)
16-25	158 (7)	383 (11)	4.4 (3.3-5.7)
11-15	168 (7)	223 (6)	7.5 (5.7-10.0)
6-10	268 (11)	208 (6)	13.6 (10.4-17.8)
2-5	276 (12)	143 (4)	23.1 (17.4-30.8)
Current Smoker	1273 (54)	715 (20)	22.6 (18.3-28.0)
Median Household Income			
> $100,000	47 (2)	137 (4)	1.00
$50,000 - $99,999	283 (12)	630 (18)	1.3 (0.0-1.9)
$30,000 - 49,000	474 (20)	840 (24)	1.4 (1.0-2.1)
$20,000 - 29,999	398 (17)	548 (16)	1.7 (1.2-2.4)
$10,000 - 19,999	366 (15)	363 (10)	2.6 (1.6-3.3)
< $10,000	133 (6)	100 (3)	3.2 (2.1-5.0)
Prefer not to Report	689 (29)	889 (25)	1.8 (1.2-2.5)
Education			
> High school	590 (25)	1373 (39)	1.00
High school	406 (17)	607 (17)	1.5 (1.3-1.8)
< High school	1379 (58)	1514 (43)	1.8 (1.6-2.1)

[a] OR adjusted for age, sex and study province.
[b] Not estimated, frequency matched to all cancer cases in NECSS study

Study subjects were approximately evenly divided by sex, and lung cancer cases were slightly older than population controls. Cases had a higher number of smoking pack-years, less education, lower income, higher alcohol and meat consumption, higher residential and

occupational second-hand smoke exposures, and more occupational exposures to dust, odors, and hazardous substances. Only 130 (6%) of lung cancer cases were never-smokers compared with 1,337 (38%) of population controls. Cases lived in regions with higher average indoor radon measurements and resided longer in the most socioeconomic deprived neighborhoods. Table 11 summarizes study-participant air pollution exposures from the national spatiotemporal models and correlations between pollutants

Table 11. Distribution of ambient air pollution exposures and pollutant exposure correlations.

					Spearman Correlation		
Pollutant	Mean (SD)	Median	IQR	Range	$PM_{2.5}$	NO_2	O_3
$PM_{2.5}$ ($\mu g/m^3$)	11.9 (3.0)	12.1	4.5	3.8-19.6	1.00	-	-
NO_2 (ppb)	15.4 (9.0)	13.8	14.3	1.1-44.9	0.73	1.00	-
O_3 (ppb)	20.3 (4.9)	21.2	6.2	6.6-33.8	0.25	0.11	1.00

National Analyses

Table 12 summarizes lung cancer odds ratio with exposure to $PM_{2.5}$, NO_2, and O_3 derived from the national spatiotemporal models. Adjusted for all individual and geographic variables, the OR for a 10 $\mu g/m^3$ increase in $PM_{2.5}$ was 1.29 (95% CI: 0.95-1.76), and for a 10 ppb increase in NO_2 and O_3 was 1.11 (1.00-1.24) and 1.09 (0.85-1.39), respectively. For NO_2, all exposure quintiles were elevated relative to the lowest (<7.1 ppb) but there was no dose-response relationship. Although variance inflation factors for all three pollutant exposures were less than 2.5, the high positive correlation between $PM_{2.5}$ and NO_2 exposures (r=0.73) and the complex spatial patterns of these pollutant relationships limit the interpretation of joint models. We did, however, examine joint models for NO_2 and O_3 to explore the independent associations between each pollutant and lung cancer incidence, since O_3 is typically decreased in high NO_2 locations. In the joint national model, the NO_2 OR was slightly increased to 1.14 (1.02-1.28) and the O_3 OR doubled to 1.20 (0.92-1.56).

Table 12. Odds Ratios (ORs) for the association between lung cancer incidence and $PM_{2.5}$, NO_2 and O_3 exposures, as derived from national spatiotemporal models.

Pollutant	Cases[a]	Cont.[a]	Unadjusted[b]	Individual Covariates[c]	Individual + Geographic Covariates[d]
$PM_{2.5}$					
All lung (per 10 µg/m^3)	2154	3264	0.82 (0.66-1.02)	1.24 (0.92-1.67)	1.29 (0.95-1.76)
Q1 [<9.0]	378	718	1.00	1.00	1.00
Q2 [9.0 -10.9]	470	598	1.25 (1.05-1.50)	1.26 (1.00-1.59)	1.26 (0.99-1.59)
Q3 [11.0 - 12.8]	462	619	1.13 (0.94-1.35)	1.32 (1.04-1.67)	1.35 (1.06-1.71)
Q4 [12.9-14.7]	445	646	1.05 (0.87-1.26)	1.35 (1.05-1.72)	1.39 (1.08-1.79)
Q5 [>14.7]	399	683	0.86 (0.70-1.05)	1.14 (0.87-1.49)	1.19 (0.90-1.57)
Histology (per 10 µg/m^3)					
Squamous cell	643	3264	0.64 (0.46-0.89)	1.24 (0.91-1.68)	1.09 (0.70-1.70)
Adenocarcinoma	816	3264	0.91 (0.67-1.24)	1.22 (0.81-1.83)	1.27 (0.84-1.90)
Small cell	383	3264	0.98 (0.64-1.51)	1.56 (0.87-2.81)	1.70 (0.92-3.13)
Large cell	226	3264	0.89 (0.52-1.51)	1.08 (0.48-2.44)	1.11 (0.48-2.54)
NO_2					
All lung (10 ppb)	2154	3264	0.97 (0.92-1.02)	1.09 (0.99-1.21)	1.11 (1.00-1.24)
Q1 [<7.1]	373	720	1.00	1.00	1.00
Q2 [7.1-11.4]	454	604	1.36 (1.12-1.65)	1.57 (1.22-2.01)	1.64 (1.28-2.11)
Q3 [11.4-16.0]	455	631	1.20 (1.00-1.48)	1.54 (1.19-2.00)	1.63 (1.26-2.12)
Q4 [16.0-25.5]	452	649	1.11 (0.91-1.35)	1.66 (1.27-2.15)	1.79 (1.37-2.36)
Q5 [>25.5]	420	660	1.06 (0.87-1.30)	1.49 (1.13-1.97)	1.59 (1.19-2.13)
Histology (per 10 ppb)					
Squamous cell	653	3264	0.88 (0.78-0.98)	1.00 (0.87-1.15)	0.99 (0.85-1.16)
Adenocarcinoma	828	3264	1.03 (0.94-1.14)	1.13 (0.99-1.30)	1.17 (1.01-1.35)
Small cell	390	3264	0.98 (0.84-1.14)	1.07 (0.88-1.3)	1.10 (0.89-1.37)
Large cell	230	3264	0.96 (0.80-1.15)	1.03 (0.77-1.37)	1.08 (0.79-1.46)
O_3					
All lung (per 10 ppb)	2154	3264	1.15 (0.96-1.37)	1.09 (0.86-1.38)	1.09 (0.85-1.39)
Q1 [<15.3]	455	615	1	1	1.00
Q2 [15.3-20.2]	421	659	1.19 (0.98-1.46)	1.13 (0.86-1.47)	1.10 (0.84-1.45)
Q3 [20.3-22.0]	417	686	0.99 (0.77-1.26)	0.93 (0.68-1.29)	0.90 (0.65-1.25)
Q4 [22.0-24.4]	427	660	1.07 (0.83-1.38)	1.00 (0.72-1.40)	0.97 (0.69-1.37)
Q5 [>24.4]	434	644	1.10 (0.85-1.43)	1.15 (0.81-1.62)	1.13 (0.79-1.61)
Histology (per 10 ppb)					
Squamous cell	653	3264	1.21 (0.91-1.62)	1.13 (0.80-1.62)	1.19 (0.82-1.71)
Adenocarcinoma	828	3264	1.03 (0.79-1.34)	1.07 (0.77-1.48)	1.04 (0.74-1.44)
Small cell	390	3264	1.14 (0.80-1.63)	1.07 (0.68-1.71)	1.07 (0.65-1.75)
Large cell	230	3264	1.09 (0.70-1.7)	0.92 (0.49-1.71)	0.89 (0.57-1.38)

[a] Case and control numbers are for the final models including all individual and geographic characteristics.
[b] Unadjusted model includes age, sex and study province.
[c] Unconditional logistic regression model with random effect for census division lived in the longes, adjusted for age, sex, cigarette smoking pack years, years since quite smoking, educational attainment, household income, average weekly alcohol and meat consumption, residential and occupational second hand smoke exposure, years working in occupations with dust or odors from industry, and years working with potential lung hazards.
[d] Unconditional logistic regression model with random effect for census division lived in the longest, adjusted for all individual variables, study province (to account for study design), ecological radon exposure, and years living in the lowest quintile of neighborhood median household income, percent without a high school diploma percent of households >30 years old dwellings.

We also examined the influence of urban residence using a community-size category based on the longest residence during the exposure period. A community-size variable was not included in the national models due to high correlation with NO_2 (r=0.73) and to a lesser degree with $PM_{2.5}$ (r=0.55). When the urban-size category was included in the national models, the fully adjusted OR per 10 unit increase in NO_2 was 1.14 (0.99-1.31) and for $PM_{2.5}$ was 1.26 (0.90-1.77). No change was seen when average population density within 5 and 10 km of residential postal codes (over the 20-year exposure period) was added. There were weak associations between population density within 5 and 10 km and lung cancer incidence [ORs of 1.06 (0.83-1.15) and 1.10 (0.86-1.40) for the highest versus lowest population density categories].

Table 13 presents stratified models for smoking status, smoking pack-years, educational attainment, and sex. No consistent patterns were observed for any of the national $PM_{2.5}$, NO_2 and O_3 exposures. For example, compared with current smokers, larger ORs for lung cancer were seen among former smokers for $PM_{2.5}$ and O_3, but smaller ORs for NO_2. The small number of never-smokers in this study makes interpretation of these models difficult. For all three pollutants, higher ORs were seen in men.

Table 13. Stratification of lung cancer and national pollutant models by smoking status, education and sex.

Stratification Variable	Cases No.	Controls No.	National Exposure Odds Ratios[a]		
			$PM_{2.5}$	NO_2	O_3
Smoking Status					
Never smoker	120	1261	0.95 (0.38-2.34)	0.98 (0.72-1.34)	1.24 (0.59-2.59)
Former	885	1351	1.45 (0.96-2.19)	1.11 (0.96-1.28)	1.10 (0.79-1.52)
Current	1149	652	1.17 (0.75-1.84)	1.20 (1.03-1.39)	0.85 (0.59-1.23)
Smoking Pack-Years					
Never smoker	120	1261	0.95 (0.38-2.34)	0.98 (0.72-1.34)	1.24 (0.59-2.59)
1-20	296	1121	1.53 (0.85-2.76)	1.33 (1.09-1.63)	0.91 (0.55-1.50)
20-40	928	599	1.24 (0.76-2.01)	1.07 (0.91-1.27)	1.23 (0.84-1.80)
>40	810	283	1.66 (0.84-3.28)	1.11 (0.87-1.41)	1.06 (0.63-1.80)
Education					
<High school	1223	1388	1.49 (0.96-2.31)	1.04 (0.89-1.22)	1.30 (0.91-1.85)
High school	381	567	1.97 (0.86-4.51)	1.66 (1.28-2.16)	0.77 (0.40-1.49)
>High school	550	1309	0.99 (0.54-1.83)	1.07 (0.86-1.32)	1.00 (0.62-1.63)
Sex					
Male	1117	1654	1.59 (1.05-2.40)	1.22 (1.06-1.40)	1.12 (0.80-1.58)
Female	1037	1610	1.12 (0.69-1.81)	1.02 (0.87-1.21)	1.08 (0.73-1.60)

[a] Adjusted for all individual variables and geographic variables (age, sex, cigarette smoking pack years, years since quite smoking, educational attainment, household income, average weekly alcohol and meat consumption, residential and occupational second hand smoke exposure, years working in occupations with dust or odors from industry, years working with potential lung hazards, study province, ecological radon exposure, and years living in the lowest quintile of neighborhood median household income, percent without a high school diploma and percent of households >30 years old dwellings.

Urban Fixed-Site Monitor Sub-Analyses

The urban analyses, based on exposures derived from the closest monitor within 50 km, are summarized in Table 14. In the fully adjusted model, a 10 $\mu g/m^3$ increase in TSP was associated with an OR of 1.04 (0.95-1.13). The largest difference from the national analysis was seen for NO_2: a 10 ppb increase in the monitor-based analysis was associated with an OR of 1.34 (1.07-1.69). It is likely that NO_2 exposures derived for the urban monitors are also capturing a component of $PM_{2.5}$, due to the correlation between the two pollutants. Appendix 3, Figure 1 illustrations the relationship between exposures derived from measured NO_2 and TSP (as $PM_{2.5}$ measurements were available only after 1984 and had poor spatial coverage).

Table 14. Odds ratios for the association between lung cancer incidence and $PM_{2.5}$, TSP, NO_2 and O_3 exposure as derived from national air pollution surveillance monitors within 50 km of residential postal codes.

Pollutant	Cases[a] No.	Controls[a] No.	Unadjusted[b]	Individual Covariates[c]	Individual + Geographic Covariates[d]
PM_{25} (Measured & Modelled)					
All lung (10 µg/m^3)	1200	1862	1.10 (0.88-1.39)	1.29 (0.80-2.07)	1.33 (0.82-2.15)
Q1 [<12.6]	219	385	1	1	1
Q2 [12.6 -14.2]	246	370	1.06 (0.85-1.34)	1.28 (0.93-1.76)	1.17 (0.80-1.72)
Q3 [14.2- 15.0]	247	366	1.09 (0.85-1.39)	1.00 (0.71-1.41)	0.96 (0.66-1.39)
Q4 [15.0-15.8]	254	356	0.85 (0.66-1.10)	0.92 (0.64-1.31)	1.03 (0.72-1.46)
Q5 [>15.8]	234	385	0.95 (0.74-1.24)	1.17 (0.81-1.71)	1.29 (0.94-1.78)
TSP					
All lung (10 µg/m^3)	1196	1859	1.06 (0.99-1.12)	1.05 (0.97-1.14)	1.04 (0.95-1.13)
Q1 [<43]	268	346	1	1	1
Q2 [43-52.8]	208	407	1.07 (0.81-1.42)	0.96 (0.64-1.42)	0.98 (0.65-1.47)
Q3 [52.8-61.4]	258	362	1.39 (1.01-1.91)	1.21 (0.78-1.87)	1.23 (0.79-1.90)
Q4 [61.4-67.3]	245	355	1.05 (0.75-1.46)	0.95 (0.60-1.49)	0.98 (0.62-1.55)
Q5 [>67.3]	217	389	1.37 (1.00-1.89)	1.33 (0.86-2.06)	1.29 (0.83-2.02)
NO_2					
All lung (10 ppb)	983	1550	1.05 (0.89-1.24)	1.34 (1.08-1.67)	1.34 (1.07-1.69)
Q1 [<19.1]	209	295	1	1	1
Q2 [19.1-22.8]	194	321	1.18 (0.89-1.56)	1.41 (0.92-2.14)	1.45 (0.95-2.22)
Q3 [22.8-24.6]	189	344	0.91 (0.65-1.27)	1.31 (0.87-1.99)	1.37 (0.90-2.08)
Q4 [24.6-28.8]	207	284	1.03 (0.76-1.39)	1.34 (0.87-2.05)	1.40 (0.91-2.16)
Q5 [>28.8]	184	306	1.04 (0.76-1.41)	1.63 (1.04-2.56)	1.60 (1.01-2.54)
O_3					
All lung (10 ppb)	1015	1478	1.15 (0.9-1.48)	1.11 (0.80-1.55)	1.11 (0.79-1.54)
Q1 [<17.8]	219	283	1	1	1
Q2 [17.8-19.4]	168	322	1.30 (0.99-1.71)	1.34 (0.94-1.90)	1.27 (0.89-1.81)
Q3 [19.4-21.8]	211	294	1.22 (0.91-1.63)	1.26 (0.87-1.83)	1.22 (0.84-1.78)
Q4 [21.8-23.8]	221	278	1.02 (0.75-1.39)	0.89 (0.59-1.34)	0.88 (0.58-1.33)
Q5 [>23.8]	196	301	1.33 (0.99-1.80)	1.36 (0.92-2.01)	1.33 (0.90-1.98)

[a] Case and control numbers are for the final models including all individual and geographic characteristics.
[b] Unadjusted model includes age, sex and study province.
[c] Unconditional logistic regression model with random effect for census division lived in the longest, adjusted for age, sex, cigarette smoking pack years, years since quite smoking, educational attainment, household income, average weekly alcohol and meat consumption, residential and occupational second hand smoke exposure, years working in occupations with dust or odors from industry, and years working with potential lung hazards.
[d] Unconditional logistic regression model with random effect for census division lived in the longest, adjusted for all individual variables, study province (to account for study design), ecological radon exposure, and years living in the lowest quintile of neighborhood median household income, percent without a high school diploma percent of households >30 years old dwellings.

Proximity to Vehicle Emissions

Table 15 summarizes ORs per 10 years living in proximity (50, 100 or 300 m) to a highway or major road, as well as weighted-proximity measures that capture the decrease in vehicle emissions over the exposure period. Few study participants lived within 50 m of highways, but increased ORs were observed for these participants, as well those living within 100 m of highways. No associations were seen for those residing near major roads.

Table 15. Adjusted ORs per 10 years living in proximity to a highway or major road for study participants residing in urban areas of Canada.

Exposure Measure	Exposed		Individual[b]	Individual + Geographic[c]	+ Ambient Pollutants[d]
	Cases[a]	Controls[a]			
Highways					
Years ≤50m	59	58	1.21 (0.76-1.94)	1.19 (0.74-1.91)	1.23 (0.76-1.98)
Years ≤50m (w)[e]	59	58	1.12 (0.80-1.58)	1.11 (0.78-1.56)	1.13 (0.80-1.60)
Years ≤100m	123	137	1.08 (0.82-1.43)	1.07 (0.81-1.42)	1.10 (0.83-1.46)
Years ≤100m (w)[e]	123	137	1.05 (0.86-1.29)	1.04 (0.85-1.28)	1.06 (0.87-1.31)
Years ≤300m	320	416	0.97 (0.83-1.13)	0.94 (0.81-1.10)	0.95 (0.82-1.12)
Years ≤300m (w)[e]	320	416	0.97 (0.87-1.09)	0.96 (0.86-1.07)	0.97 (0.86-1.08)
Major Roads					
Years ≤50m	331	427	1.05 (0.90-1.23)	1.00 (0.85-1.18)	1.00 (0.85-1.17)
Years ≤50m (w)[e]	331	427	1.04 (0.92-1.16)	1.00 (0.89-1.13)	1.00 (0.89-1.12)
Years ≤100m	507	717	1.02 (0.90-1.16)	0.99 (0.87-1.12)	0.98 (0.87-1.12)
Years ≤100m (w)[e]	507	717	1.01 (0.93-1.11)	0.99 (0.90-1.08)	0.99 (0.90-1.08)
Years ≤300m	1040	1485	0.99 (0.90-1.10)	0.96 (0.87-1.07)	0.96 (0.86-1.07)
Years ≤300m (w)[e]	1040	1485	0.99 (0.47-2.12)	0.97 (0.90-1.05)	0.97 (0.90-1.05)

[a] All analyses included 1265 cases and 1868 controls.
[b] Unconditional logistic regression model, adjusted for age, sex, cigarette smoking pack years, years since quite smoking, educational attainment, household income, average weekly alcohol and meat consumption, residential and occupational second hand smoke exposure, years working in occupations with dust or odors from industry, and years working with potential lung hazards.
[c] Unconditional logistic regression model with random effect for census division lived in the longest, adjusted for all individual variables, study province (to account for study design), ecological radon exposure, and years living in the lowest quintile of neighborhood median household income, percent without a high school diploma percent of households >30 years old dwellings.[b] Adjusted for all individual and geographic variables.
[d] Unconditional logistic regression model with random effect for census division lived in the longest, adjusted for all individual and geographic variables, and $PM_{2.5}$, NO_2 and O_3 exposures.
[e] Weighted be vehicle emissions to account for emission changes from 1975-1994.

4.4. Discussion

The present study aimed to enhance current understanding of the risks posed by air pollution to lung cancer incidence. We attempted to reduce exposure misclassification by conducting extensive spatiotemporal air pollution exposure assessments that incorporate long-term residential histories, and we examined associations with various pollutants and sources of exposure. We were also able to control for a comprehensive set of potential individual and geographic confounding factors.

Overall, our results support previous literature showing that ambient $PM_{2.5}$ air pollution is associated with increased lung cancer risk. In our national analysis, we found that a 10 $\mu g/m^3$ increase in $PM_{2.5}$ was associated with an OR of 1.29 (0.95-1.76). This estimate is similar to the effect size reported in a 2008 meta-analysis, with a pooled RR of 1.21 (1.10-1.32) per 10 $\mu g/m^3$ increase in $PM_{2.5}$ (Chen et al. 2008). An extended follow-up of the Harvard six cities study from 1974 to 2009 also found a 37% (7-75%) increase (Lepeule et al. 2012), and a recent analysis of never-smokers in the ACS cohort based on 26 years of follow-up found a RR of 1.19 (0.97-1.47) (all for a 10 $\mu g/m^3$ increase in $PM_{2.5}$) (Turner et al. 2011).

Unlike the relatively robust literature on $PM_{2.5}$ and lung cancer, there are fewer studies on the associations of the gaseous pollutants NO_2 and O_3 with lung cancer. We found an OR for a 10-unit increase in NO_2 of 1.11 (1.00-1.24) in the national analysis, and a substantially larger OR [1.34 (1.07-1.69)] in the urban monitor-based analysis. This higher estimate may be due to restricting the study to large urban areas, more accurate exposure assessment, or exposure assessment that captured both NO_2 and $PM_{2.5}$ influences (due to the high correlation between $PM_{2.5}$ and NO_2 and the lack of $PM_{2.5}$ monitoring data prior to 1984). Studies of NO_2 and lung cancer risk generally show positive associations ranging from 5-30% increases in

risk per 10 ppb increases in NO_2 (Katanoda et al. 2011; Nyberg et al. 2000; Vineis et al. 2006; Yorifuji et al. 2010); however, negative associations have also been observed [RR 0.86 (0.70-1.07) per 30 $\mu g/m^3$] (Beelen et al. 2008).

In addition to NO_2, a number of studies have examined NO_x air pollution (primarily as a marker of traffic air pollution) with most reporting positive associations with lung cancer (Beelen et al. 2008; Nafstad et al. 2003; Raaschou-Nielsen et al. 2011; Raaschou-Nielsen et al. 2010; Nyberg et al. 2000). When we considered proximity to highways and major roads as a surrogate for traffic air pollution exposure, we found elevated risk of lung cancer incidence associated with living within 100 m of highways [OR 1.10 (0.83-1.46) per 10 year-residence], but not for major roads. Our results are similar to those from a Danish cohort [incidence rate ratio of 1.21 (0.95-1.55) for lung cancer associated with living within 50 m of a major road (> 10,000 vehicles per day)] (Raaschou-Nielsen et al. 2011) as well as those from a Dutch cohort [RR of 1.10 (0.74-1.62) for living within 100 m of a motorway or 50 m of a road with > 10,000 vehicles/day] (Beelen et al. 2008). Major roads in urban locations of Canada have similar traffic volumes; however, we did not see any associations between living near major roads and lung cancer incidence.

We found a trend of increasing lung cancer incidence with increasing O_3 concentrations [OR 1.09 (0.85-1.39) for a 10 ppb increase in the national models] with similar results in the urban analysis. In multi-pollutant models incorporating NO_2 and O_3, the O_3 OR increased substantially to 1.20 (0.92-1.56), suggesting that accounting for areas with low O_3 but high NO_2 may be important to further understand the association between long-term O_3 exposure and lung cancer risk. There are no other large studies we are aware of to compare with these findings.

Lastly, we did not observe clear patterns between air pollution exposures and specific histological subtypes. Generally, $PM_{2.5}$ exposure was most strongly associated with small cell and adenocarcinoma; NO_2 with adenocarcinoma; and O_3 with squamous cell carcinoma. The most persuasive association was for NO_2 and adenocarcinoma [OR 1.17 (1.01-1.35)]. Adenocarcinoma is the most common histological subtype among never smokers, but there is no consensus in the literature as to whether air pollution is associated more strongly with adenocarcinoma or other histological subtypes. Some studies have found air pollution to be more strongly associated with adenocarcinoma (Liaw et al. 2010; López-Cima et al. 2011; Chen et al. 2009), while others have found the strongest associations with other histological subtypes (Barbone et al. 1995; Katsouyanni et al. 1991; Raaschou-Nielsen et al. 2010).

This study relies on the accuracy of historical exposure assessments. A number of sensitivity analyses were conducted to examine how the ORs change with different historical exposure assessment methods (summarized in Figure 12). These methods included the spatiotemporal models (used in national models and described in methods); spatiotemporal models developed with a national ratio of historical pollutant concentrations to current levels (for $PM_{2.5}$ only); historical regression models that use satellite data, population density, and a time trend to predict historical concentrations (Hystad et al. 2012); the satellite or chemical transport model spatial surfaces without temporal adjustments; and exposures estimated only from fixed-site monitoring data within 50 km. Figure 12 demonstrates a relatively small degree of variability in the $PM_{2.5}$ and O_3 OR estimates, while the NO_2 urban monitor exposure assessment has a higher OR than the two national NO_2 models incorporating spatial and temporal variability. For all pollutant models, the a priori national spatiotemporal exposure assessments had the smallest standard errors.

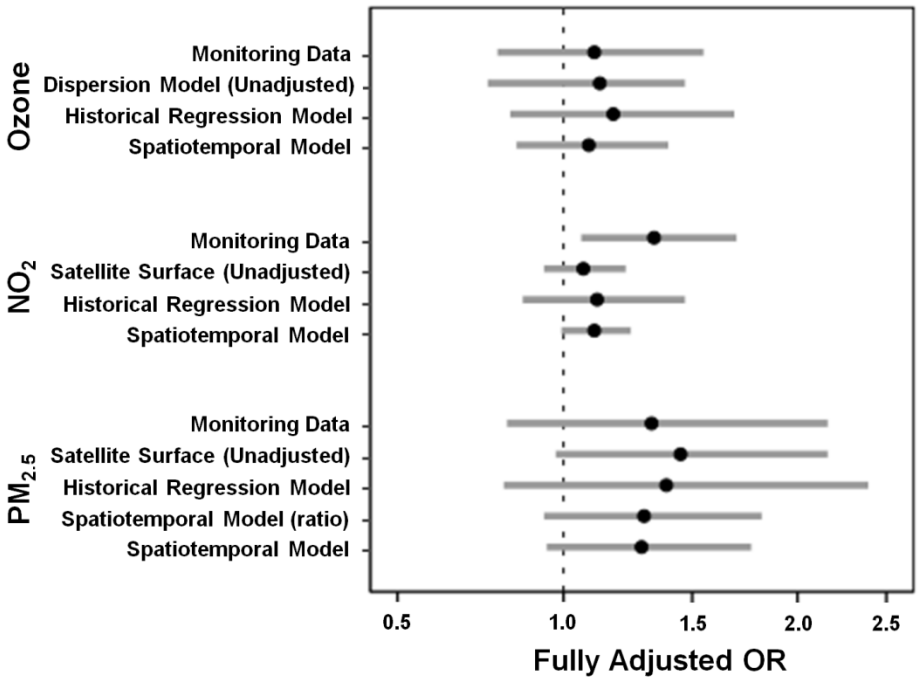

Figure 12. Sensitivity analyses using different air pollution exposure assessment approaches.

This study has a number of strengths that address important limitations in the current air pollution and lung cancer literature. First, we estimated long-term historical air pollution levels at six-digit residential postal codes. To reduce exposure misclassification, exposures were derived from twenty years of residential histories. This time-period was selected because, prior to 1975, cases tended to report more addresses than population controls, which would have incorporated bias into the study (Hystad et al. 2012). To further reduce bias, only people with complete twenty-year residential histories were included in the final analyses. We were able to examine the influence of residential-history completeness, and found that including study subjects with missing residential histories resulted in substantial attenuation of the OR estimates. For example, including subjects with 18 years (90%) of complete exposures in the national models resulted in ORs per 10 unit increase in $PM_{2.5}$, NO_2 and O_3 of 1.23 (0.92-1.65), 1.11 (1.00-1.22) and 1.05 (0.83-1.33). Attenuation was greater when subjects with 15 years

(75%) of complete exposures were included. Unlike other studies that assume participants have lived at their home residence for a certain amount of time, missing data in this study likely represents substantial exposure error as study participants self-reported their addresses and missing periods represent addresses they could not recall or residential locations outside of Canada.

Second, unlike most studies, which are restricted to single pollutants and city locations, we developed national models for multiple pollutants and were able to include participants in all areas of Canada. This type of exposure assessment has also been used in a recent national Canadian cohort analysis of $PM_{2.5}$ and non-accidental and cardiovascular mortality (Crouse et al. 2012). Third, unlike many prior studies, we had a large sample size (n=2,390 incident lung cancer cases and 3,507 population controls), which allowed us to examine the associations between air pollution and lung cancer histology. Fourth, a comprehensive set of individual and geographic level information was available for modeling important confounding variables. The inclusion of smoking information in particular had a large influence on study results. Smoking variables in the adjusted models substantially increased ORs, due to the small negative spatial association between smoking prevalence and air pollution exposures (Villeneuve et al. 2011) The inclusion of ecological radon exposures was also important, specifically in the NO_2 and $PM_{2.5}$ models, as high radon concentrations in Canada are located in areas that generally have lower NO_2 and $PM_{2.5}$ concentrations.

A number of study limitations also need to be considered. First, while this study has a relatively high response rate for cases (62%) and population controls (67%), response and recall bias cannot be ruled out. No difference in the completeness of self-reported residential histories was present between cases and controls when restricted to the 1975-1994 exposure period. Second, it is essential to note that populations are not distributed evenly across

geographic communities, and thus, a random sample of the population may not be a random sample of all places. The national enhanced cancer surveillance system was designed so each provincial cancer agency would sample and recruit study participants. A province variable was therefore included in the fully adjusted models to capture any differences between sampling strategies (health insurance plans were used in five provinces, random digit dialing in two, and property assessment data in one). This is not ideal, in that the province variable likely captured a portion of the air pollution variance. The province variable also had a large influence on histology results, suggesting possible classification or recruitment differences by province. In addition, a large portion of our study population was located in and around Toronto, Ontario, (see Figure 11a), which had the highest $PM_{2.5}$ exposures. Any response bias or exposure assessment error in this geographic area would have a large influence on our study results. A sensitivity analysis including all provinces but Ontario (1,399 cases and 2,050 controls) indicated that results changed only slightly for NO_2 [OR 1.12 (0.97-1.31) per 10 ppb increase] and O_3 [OR 1.12 (0.80-1.56) per 10 ppb increase], but were reduced for $PM_{2.5}$ [OR 1.15 (0.77-1.72) per 10 $\mu g/m^3$ increase]. The reduction for $PM_{2.5}$ is presumably due to the exclusion of the highest exposed (those living in Southern Ontario), which greatly reduced exposure variation in the analysis. The sensitivity to geographic variables is not as pronounced for NO_2 since those with the highest NO_2 exposure quintile lived in various large cities across Canada, rather than clustered in one region. We also included a random effect based on the census division of longest residence to account for unmeasured spatial structure in the data.

Third, the models were sensitive to sub-analyses, as seen with the monitor-based exposure-assessment results, which were substantially higher than the national NO_2 results. The difference in NO_2 results may be due to the various exposure assessment approaches, with the national models capturing inter- and intra-urban variation and the urban monitor-based

assessment capturing predominantly intraurban differences. NO_2 exposures derived from urban monitors may also be capturing a component of $PM_{2.5}$, since monitoring data for $PM_{2.5}$ were not available prior to 1984. Fourth, the OR estimates, primarily for $PM_{2.5}$, changed slightly with various coding schemes for smoking variables. For example, when a continuous smoking-pack-years-squared variable was included in the national model to account for non-linear associations between smoking and lung cancer, the OR associated with a 10 unit increase in $PM_{2.5}$ decreased to 1.23 (0.91-1.67). Fifth, all model results did not show dose-response gradients. This may have been due to the relatively small sample size and range of exposures for study participants, particularly in the urban monitor-based analyses. Sixth, due to privacy concerns, residential history locations were limited to six-digit postal codes, which are accurate in urban areas but can cover much larger regions in rural areas. Proximity analyses were therefore restricted to urban areas of Canada. Lastly, while we were able to estimate exposure from residential history, no information was available for other important micro-environments such as work locations.

4.5. Conclusions

In sum, we found increased risks of lung cancer incidence with residential exposures to ambient $PM_{2.5}$, NO_2, and O_3, as well as living within 100 m of highways. Results were most robust for NO_2 and $PM_{2.5}$. More research is needed to establish whether O_3 exposure is an independent risk factor for lung cancer.

Chapter 5

Geographic Variation in Radon and Associated Lung Cancer Risk in Canada: A Population-Based Case-Control Study

Submitted for publication: Hystad, P., Demers, P., Johnson, K., Brauer, M., Setton, E., Cervantes, A., Poplawski, K., McFarlane A., Whitehead, A., Nicole, A.M. Geographic Variation in Radon and Associated Lung Cancer Risk in Canada: A Population-Based Case-Control Study.

Abstract

Background: Radon is an important risk factor for lung cancer. Here we present radon concentration and potential maps and conduct the first national-level analysis of residential radon exposure and lung cancer risk in Canada.

Methods: We used 2,390 lung cancer cases and 3,507 population controls collected from 1994-1997 in eight Canadian provinces through the National Enhanced Cancer Surveillance System. Residential histories over a twenty year period were used to estimate ecological radon exposures from two geospatial mapping methods. The first uses a recent Health Canada survey of 14,000 residential radon measurements sampled by 121 Health Regions (HRs) across Canada. The second was based on radon potential areas estimated from geology, sediment geochemistry and aerial gamma-ray spectrometry measurements. Hierarchical logistic

regression analyses were used to estimate odds ratios for lung cancer incidence, after adjusting for a comprehensive set of individual and geographic covariates.

Results: Significant variation in average residential radon concentrations were found across HRs in Canada (range: 16-386 Bq/m^3). In multivariate models, a 50 Bq/m^3 increase in average HR radon concentration was associated with a 7% (95% CI:-6-21%) increase in the odds of lung cancer and exposure quartiles demonstrated a dose-response relationship. For every ten years that individuals lived in high radon potential zones, the odds of lung cancer incidence increased by 11% (95% CI:1-23%).

Conclusions: These results strengthen the evidence for radon exposure as a risk factor for lung cancer. The significant geographic variation in radon concentrations across Canada also supports the use of radon risk mapping to target population health prevention efforts.

5.1. Introduction

Radon is a colorless, odorless, naturally occurring gas released from the breakdown of Uranium in soils. Exposure to radon occurs primarily indoors, where levels can accumulate to high concentrations. The majority of lung cancer cases are due to tobacco smoke; however, radon increases the risk of lung cancer in smokers and is a major cause of lung cancer for individuals who have never smoked (Darby, 2005; IARC, 2012c; Krewski et al., 2005; Samet et al., 2009). In Canada, approximately 16% of lung cancers (3,261 cases annually) are estimated to be attributable to residential radon exposure (Chen et al., 2012).

While radon is recognized as being causally associated with lung cancer, national-level studies are important to estimate attributable disease burden and to develop population health policy. To date, only one residential radon epidemiological study has been conducted in Canada. This study was conducted in Winnipeg and reported no associations between residential radon concentrations and lung cancer (Létourneau et al., 1994). Similar to most epidemiological studies of residential radon, exposure was assessed using indoor residential measurements. These types of studies have limited statistical power arising from small sample sizes and exposure misclassification from extrapolating short-term measurements to estimate long-term exposures and the difficulty in accounting for residential mobility. Alternatively, two recent epidemiological studies conducted in the US and Denmark have used maps and spatial prediction models to estimate long-term residential radon concentrations in larger population samples, the approach we follow in this analysis. For the Cancer Prevention Study II cohort in the US, average radon measurements by zip codes were used to estimate individual radon exposure, and a 100 Bq/m^3 increase in radon was associated with a 15% (95% CI: 1-31%) increase in lung cancer mortality (Turner et al., 2011a). In the Danish Diet,

Cancer and Health cohort, information on geology and housing characteristics were used to predict radon concentrations at residential locations for 57,053 subjects (589 lung cancer cases) and an incidence rate ratio of 1.04 (95% CI: 0.69–1.56) per 100 Bq/m^3 increase in radon, and 1.67 (95% CI: 0.69–4.04) among non-smokers, was found (Bräuner et al., 2012).

Here we present national radon concentration and potential maps for Canada and apply these to a population based case-control study of 2,390 histological confirmed lung cancer incidence cases and 3,507 population controls from eight Canadian provinces to estimate lung cancer risk associated with ecological residential radon exposure.

5.2. Methods

Radon Mapping

Two distinct approaches were used to create radon concentration and potential maps for Canada. The first (Figure 13a) used a recently completed residential radon survey of three-month radon measurements collected from approximately 14,000 households across Canada (Health Canada, 2012). The sampling frame for this survey provides representative measures of residential radon concentrations by Health Regions in Canada. Participants were asked to place detectors on the lowest lived-in level of the home in which they spend a minimum of four hours per day, and we used these measurements without adjustment (i.e. we did not standardize to basement or first floor concentrations). The second approach (Figure 13b) used a map of geologic radon potential provided by Radon Environmental Management Corp. (REMC), created from geologic surveys, aerial gamma ray spectrometry measurements, and stream and lake sediment geochemistry (Radon Environmental, 2012).

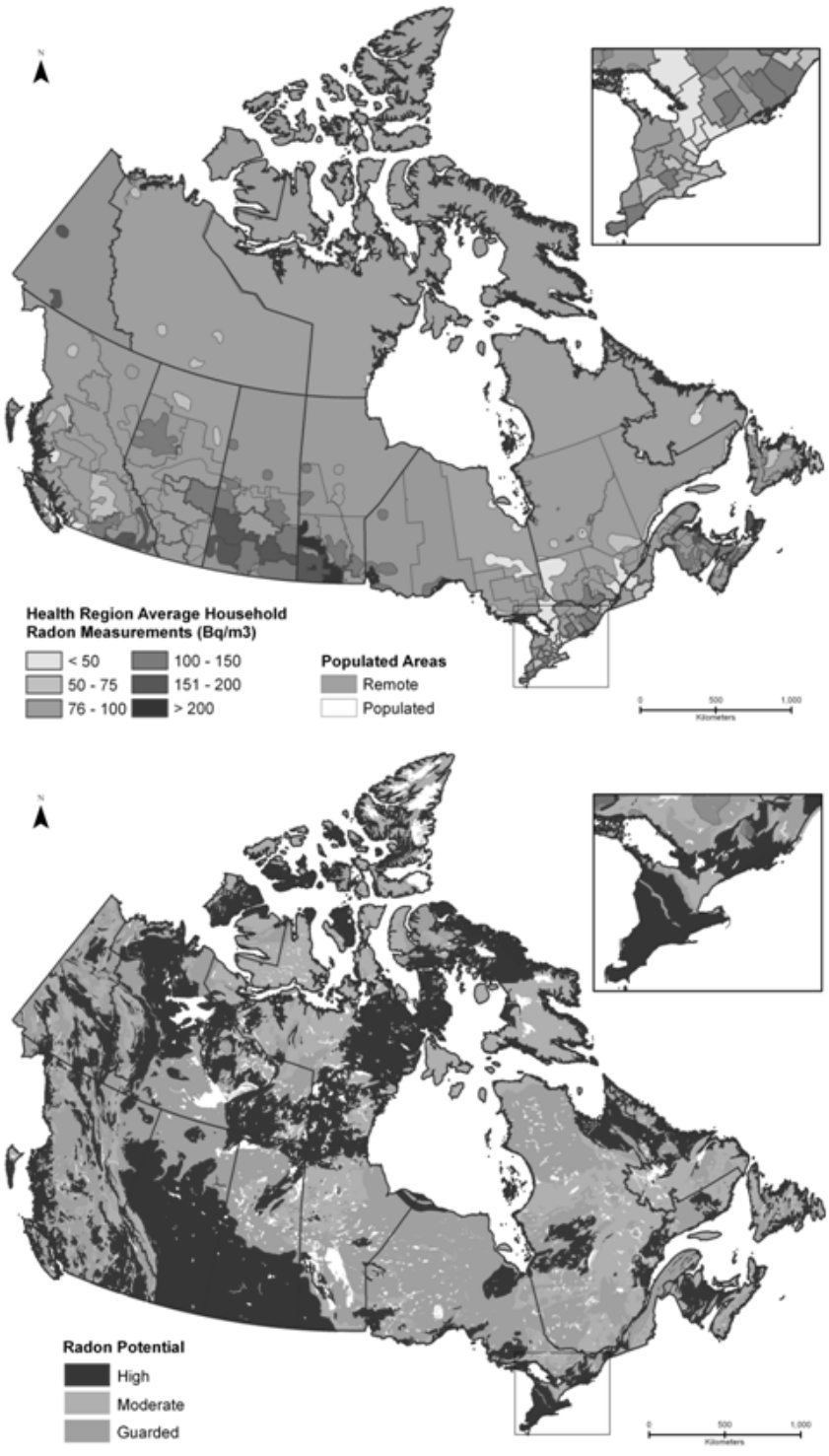

Figure 13. Results of the cross-Canada residential radon survey presented as average Health Region radon concentrations (a); and radon potential areas (b) developed from geology, soil uranium geochemistry and aerial gamma ray spectrometry measurements.

The geological units used in Figure 13b were grouped into one of three zones according to a set of rules that defined how the various predictor measurements contributed to the radon potential ranking. The rank classes were selected so that each of the three classes contained approximately equal portions of the Canadian landmass. These measurements were prioritized by data quality and type, where direct measurements were given higher weighting than extrapolated data. The US Geological Survey's data used to create the US geological radon potential map (USGS, 2012) were also used to calibrate predictor datasets.

Population-Based Lung Cancer Case-Control Study

We used data from the lung cancer component of the National Enhanced Cancer Surveillance System (NECSS) (Johnson et al., 1998), a multi-site population based case-control study that collected 3,280 lung cancer cases, with histological classification, and 5,073 population controls from 1994-1997. Provincial cancer registries identified and sampled cases within 1-3 months of diagnosis and population controls were selected from a random sample of individuals within each province, frequency matched on sex and five-year age categories to the overall collection of NECSS cancer cases. Recruitment methods for controls depended on data availability and accessibility by province and included provincial health insurance plans in five provinces (British Columbia, Saskatchewan, Manitoba, Prince Edward Island, and Nova Scotia), random digit dialing in two (Alberta and Newfoundland), and property assessment data in one (Ontario). A research questionnaire was mailed to selected cases and controls and active follow-up was conducted. The response rate for contacted lung cancer cases was 61.7% and for population controls was 67.4%. The research questionnaire collected a comprehensive set of information on individual characteristics and lifetime occupational exposures and residential histories. Only study participants with 20 years of geocoded

residential histories in Canada from 1975-1994 were included in the final analysis. The time-period and exclusion criteria are meant to reduce exposure assessment error and bias that may result from incomplete self-reported residential histories. These exclusion criteria reduced the study sample to 2,390 incident histologically-confirmed lung cancer cases and 3,507 population controls. No significant differences in demographic, socioeconomic, or smoking characteristics were found between excluded and retained lung cancer cases and population controls.

Radon Exposure Assessment

Residential histories of study subjects were geo-coded to 6-digit postal codes and are the basis of the estimation of radon exposure. Figure 14 illustrates kernel density maps of the residential locations of lung cancer cases and population controls between 1975 and 1994. As expected from a population-based study, the geographic distribution of population controls mirrors the geographic distribution of the population in the study provinces, while slight differences can be seen in the geographic distribution of lung cancer cases.

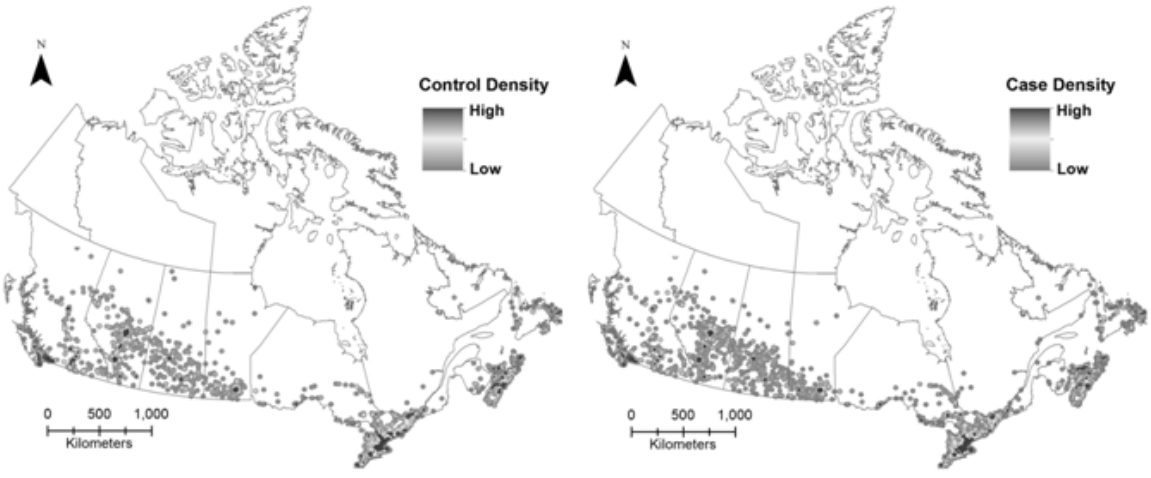

Figure 14. Kernel density maps (radius=25km's) of residential locations between 1975- 1994 for population controls and lung cancer cases.

Ecological measures of radon exposure were developed using the two radon maps. It was assumed that geographic variation in radon concentration and potential were constant during the 20 year exposure period. For the first exposure assessment approach we assigned average health region radon measurements (Figure 13a) to all postal codes located in each health region. We then calculated mean health region radon measures over the 20 year exposure period for each individual. For the second approach we assigned the radon potential zone (Figure 13b) that each residential postal code was located in and for each individual calculated the number of years living in high, moderate and guarded radon potential zones for the 20 year exposure period.

Covariates

A comprehensive set of individual and geographic variables were available for inclusion in the multivariate models and covariates that may be potential confounding factors were identified a priori. Individual variables included in the final analyses were age, sex, educational attainment, household income, smoking pack-years, years since quit smoking, residential second hand smoke exposure (defined by the number of smokers in the home multiplied by number of residential years), occupational second hand smoke exposure (defined by the number of smokers in the immediate work environment multiplied by number of occupational years), average weekly alcohol serving, average weekly meat servings, years working with dust and odors in the immediate work environment, and years working with hazardous substances. Geographic variables included were study province, long-term fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2) and ozone (O_3) air pollution exposure estimates (Hystad et al., 2012), an urban population size category based on where individuals had lived the longest during the exposure period, and neighborhood contextual deprivation

indicators (based on census tract and census sub-division data), including the number of years residing in the most deprived quintile of median household income, percent residence without a high school diploma, and percent rental dwellings.

Statistical Analyses

We conducted analyses using hierarchical logistic regression (GLIMMIX, SAS version 9.3; SAS Institute, INC, Cary, NC). Random intercepts were defined from Statistics Canada census division (CD) boundaries, representing regional areas in Canada, and assigned to individuals' longest residential location to account for residual geographic patterns. Individuals with missing covariates (cases=236, controls=243) were excluded from the final multivariate models. We report ORs and 95% confidence intervals (95% CI) for 50 Bq/m^3 increases in ecological radon concentrations as well as for exposure quartiles. For the radon potential zone exposure model, we report ORs for ten-year increases in residing within high radon potential zones as well as categorized residential years. Models were also stratified by major lung cancer histological subtypes.

5.3. Results

Selected descriptive statistics for study participants are provided in Table 16 (for a full description see Appendix 3 Table 1). Within lung cancer cases, only six percent were never smokers compared to 38% in the population control group. Lung cancer cases also had much higher smoking pack-years than controls and quite smoking more recently. There were also several differences between socioeconomic and individual health-behaviors as well as occupational exposures. No differences were found for years living in an urban area or ambient air pollution exposures in unadjusted analyses. Cases had higher ecological radon exposures for all measures: the mean (±SD) health region radon concentration for cases was

81.3 (40.8) Bq/m^3 compared to 78.6 (39.5) Bq/m^3 for controls, and lung cancer cases lived an average of 9.0 years in high radon potential zones compared to 7.9 years for controls.

Table 16. Descriptive statistics of lung cancer cases and population controls.

Variable	**Cases[a] (n=2390)**	**Controls[a] (n=3507)**
Age; Mean (SD)	63.5 (8.2)	59.0 (12.6)
Sex		
Female	1152 (48%)	1719 (49%)
Male	1238 (52%)	1788 (51%)
Median Household Income		
> $100,000	47 (2%)	137 (4%)
$50,000 - $99,999	283 (12%)	630 (18%)
$30,000 - 49,000	474 (20%)	840 (24%)
$20,000 - 29,999	398 (17%)	548 (16%)
$10,000 - 19,999	366 (15%)	363 (10%)
< $10,000	133 (6%)	100 (3%)
Prefer not to Report	689 (29%)	889 (25%)
Education		
> High school	590 (25%)	1373 (39%)
High school	406 (17%)	607 (17%)
< High school	1379 (58%)	1514 (43%)
Smoking Status		
Never-smoker	130 (6%)	1337 (38%)
Former Smoker	969 (41%)	1446 (41%)
Current Smoker	1288 (54%)	718 (2%)
Smoking Pack Years; mean (SD)	34.9 (21.0)	12.7 (17.5)
Years Since Quit Smoking; mean (SD)	6.2 (8.9)	10.3 (13.2)
Residential SHS exposure; mean (SD)	56.7 (44.1)	36.1 (37.9)
Occupational SHS exposure; mean (SD)	74.0 (83.2)	56.7 (74.2)
Years working with industrial dust/odors; mean (SD)	12.9 (16.8)	10.3 (15.4)
Years working with hazardous substances; mean (SD)	12.2 (27.8)	9.4 (26.0)
$PM_{2.5}$ (μg/m^3); mean (SD)	11.9 (2.9)	11.9 (3.1)
NO_2 (ppb); mean (SD)	15.6 (8.9)	15.3 (9.1)
O_3 (ppb); mean (SD)	20.2 (5.0)	20.3 (4.9)
Average HR Radon Measurements (Bq/m^3); mean SD	81.3 (40.8)	78.6 (39.5)
Years in High Radon Potential Area; mean (SD)	9.0 (9.5)	7.9 (9.5)

[a] Descriptive statistics for lung cancer cases and population controls with 20 years of complete residential histories.

Table 17 presents crude ORs (95% CI) as well as individual and individual+geographic adjusted ORs for lung cancer incidence using both radon exposure measures. In the fully-adjusted (individual+geographic) model a 50 Bq/m^3 increase in health region average radon was associated with a non-significant increase in the odds of all lung cancer incidence (OR 1.07, 95% CI: 0.94-1.21), and a dose-response trend was present in the exposure quartiles. A significant increase in the odds of adenocarcinoma (OR 1.23, 95% CI: 1.04-1.45) was found. For every ten years residing in a high radon potential zone, the odds of all lung cancer increased by 11% (95% CI: 1-23%). In the radon potential zone model, the largest risks were observed for large cell and squamous cell carcinoma. No statistically significant interaction was seen by any smoking variables and Health Region radon levels and radon potential areas.

Table 17. Association between ecological radon exposures and lung cancer incidence, by histological subtype.

Exposures	Cases	Controls	Unadjusted[a]	Individual Covariates[b]	Individual + Geographic Covariates[c]
Average Health Region Radon Measurements					
All lung cancer (per 50 Bq/m^3)	2154	3264	1.12 (1.03-1.22)	1.05 (0.94-1.17)	1.07 (0.94-1.21)
Categorized Bq/m^3 (Quartiles)					
Q1 (< 51)	496	808	1	1	1
Q2 (51-75)	525	894	1.06 (0.88-1.27)	1.08 (0.86-1.37)	1.04 (0.81-1.34)
Q3 (76-102)	567	792	1.23 (1.01-1.49)	1.09 (0.85-1.41)	1.11 (0.84-1.45)
Q4 (>102)	566	770	1.26 (1.03-1.54)	1.19 (0.92-1.54)	1.18 (0.89-1.56)
Histological Subtype(per 50 Bq/m^3)					
Squamous cell	605	3264	1.07 (0.93-1.22)	0.96 (0.81-1.15)	0.92 (0.76-1.11)
Adenocarcinoma	756	3264	1.15 (1.02-1.30)	1.14 (0.75-1.74)	1.23 (1.04-1.45)
Small cell	358	3264	1.17 (1.00-1.37)	1.08 (0.88-1.33)	1.09 (0.85-1.41)
Large cell	213	3264	1.06 (0.87-1.29)	1.05 (0.82-1.35)	1.08 (0.77-1.51)
Residing in High Radon Potential Zone					
All lung cancer (per 10 years)	2154	3264	1.10 (1.03-1.18)	1.12 (1.02-1.22)	1.11 (1.01-1.23)
Categorized Years					
0	1025	1743	1	1	1
1-9	135	209	1.18 (0.93-1.49)	1.18 (0.88-1.59)	1.20 (0.89-1.55)
10-19	134	184	1.32 (1.04-1.68)	1.27 (0.93-1.73)	1.30 (0.95-1.78)
20	860	1128	1.24 (1.07-1.44)	1.27 (1.05-1.53)	1.25 (1.01-1.62)
Histological Subtype(per 10 years)					
Squamous cell	605	3264	1.14 (1.02-1.27)	1.16 (1.01-1.34)	1.14 (0.97-1.34)
Adenocarcinoma	756	3264	1.07 (0.97-1.18)	1.05 (0.93-1.18)	1.05 (0.91-1.21)
Small cell	358	3264	1.07 (0.93-1.23)	1.11 (0.93-1.33)	1.13 (0.92-1.40)
Large cell	213	3264	1.18 (0.99-1.40)	1.21 (0.99-1.49)	1.21 (0.94-1.56)

[a] Adjusted for age, sex and study province(to account for study design).
[b] Adjusted for age, sex, study province, cigarette smoking pack years, years since quit smoking, educational attainment, household income, average weekly alcohol and meat consumption, residential and occupational second hand smoke exposure, years working in occupations with dust or odors from industry, years working with lung hazards.
[c] Adjusted for all individual variables as well as ambient $PM_{2.5}$, NO_2, and O_3 air pollution exposures, urban size category of longest residences, and years living in the lowest quintile of median household income, percent residence without a high school diploma, and percent rental dwellings.

5.4. Discussion

Epidemiological Results

We found associations between two separate measures of radon exposure and lung cancer incidence in this population-based case-control study that correspond closely to the existing literature on residential radon exposure and lung cancer risk. Combined analysis including seven North America studies found an 11% (95% CI: 0-28%) increase in lung cancer risk per 100 Bq/m^3 increase in residential radon concentrations, while for subjects who had resided in only one or two houses in the 5–30 year exposure window and who had residential radon measurements for at least 20 yr of this 25-yr period, there was an 18% (95% CI: 2-43%) increase in lung cancer risk per 100 Bq/m^3 (Krewski et al., 2005). The analysis of 13 European studies found an 8.4% (95% CI: 3-15.8%) increase in risk per 100 Bq/m^3 increase (Darby, 2005). After correction for the dilution caused by random uncertainties in measuring radon concentrations the increase in risk was estimated at 16% (95% CI: 5%-31%) per 100 Bq/m^3 increase in radon exposure. In our study, a 100 Bq/m^3 increase in average health region level radon concentrations was associated with a 13% (95% CI: -12%-46%) increase in lung cancer incidence. While not statistically significant, exposure quartiles demonstrated a dose-response relationship and the lack of statistical significance is not unexpected given the error associated with assigning individual radon exposures from aggregated data. While interaction between smoking and radon exposure is established, with higher radon risks for smokers (Darby, 2005; Krewski et al., 2005), we did not see significant effect modification by any smoking-related variable, which may be related to radon exposure assessment error. This is similar to a number of other residential radon studies (Turner et al., 2011a; Wilcox et al., 2008).

Spatial Variation in Radon and Population Health Prevention

We used radon concentration and potential maps to examine the spatial variation in radon across Canada and to determine the associated lung cancer risks. It is hoped that this new evidence can be used to enhance radon prevention.

In Canada, less than thirty percent of the population is able to describe radon as a health hazard, and only 5% of individuals have tested their homes for radon.(Statistics Canada, 2010) The geographic variation in radon risk across Canada (Figure 13) has important population health implications, although it is important to differentiate the individual and population-level utility of the radon maps presented here. The radon maps illustrate large variations in mean radon concentrations by health regions (range of 16-370 Bq/m^3); however, all areas of Canada had homes with high radon concentrations (even in the health region with the lowest mean radon concentration there were still homes that tested over the Health Canada guideline of 200 Bq/m^3). This supports Health Canada's policy that all individuals should get their homes tested for radon (Health Canada, 2012). However, from a population health perspective, areas of Canada that have much greater radon risk could be the target of focused radon prevention programs. Currently, there are no official radon maps for Canada, despite most developed countries having published radon risk maps that indicate high/low areas based on radon measurements or radon potential information (Tollefsen et al., 2011). Geographical targeting would also increase the cost-effectiveness of radon prevention options, which are often criticized for being too costly (Gray et al., 2009). The health region measurement map is recommended for informing population health prevention, as it incorporates actual indoor radon measurements and corresponds to the administrative boundaries in which public health activities are developed and implemented at the regional level. The radon potential map, however, could also be used for radon public awareness. Importantly, targeted prevention

efforts are only one component of a comprehensive radon strategy, as prevention measures in *all* new buildings is one of the most effective ways to reduce radon-related lung cancers (Bochicchio, 2011).

Limitations

The principal limitation of this study is the use of ecological radon measures to estimate individual-level radon exposures. Random error in exposure assessment is surely present, due to large variability in radon concentrations between homes in the same health region, which likely attenuated the ORs towards the null (Heid et al., 2004). Nevertheless, there is substantial radon variability between HRs in Canada that was captured by the residential radon survey. We also conducted a sensitivity analysis to examine how results changed if radon exposure was assigned based on average town/city measurements (if >5 measurements existed). Results showed slightly reduced ORs, with a 100 Bq/m^3 increase in town/city radon associated with an 8% (95% CI: -10% - 30%) increase in lung cancer incidence. This exposure approach uses a small sample of radon measurements to represent average city measurements even though the data are not representative at this scale (representing a trade-off between spatial scale and data representativeness). Conducting the epidemiological analysis by health regions also validates the health region radon map in terms of demonstrating an association with lung cancer incidence in this national case-control study. Radon potential was also used as an independent exposure method that was thought to have increased specificity. There were differences between the radon potential and health region measurement results, particularly for histological subtype analyses, but the overall conclusions from the two approaches are consistent. This is likely due to both methods capturing the very high radon locations in Canada; however, the correlation between individual assigned health

region measurements and radon potential exposures was only 0.24. This low correlation is not surprising given the known difficulties in predicting individual residential radon concentrations (Hauri et al., 2012), which depend predominantly on individual housing characteristics. While we were able to estimate exposure from long-term residential locations, no information was available for housing characteristics that may influence radon exposures.

A number of limitations related to the case-control study design also need to be highlighted. First, while this study has a relatively high response rates for cases (61.7%) and population controls (67.4%), response and recall bias cannot be ruled out. Second, different recruitment methods for population controls were conducted in each province that could incorporate bias into the study; however, it is unlikely that any recruitment differences would be related to radon exposures and stratified analyses by the three recruitment methods produced similar results. Third, no information was available for other environments, such as work locations, were radon exposure many occur.

5.5. Conclusions

We used radon concentration and potential maps to examine the spatial variation in radon across Canada and used these maps to examine associations with lung cancer incidence in a population-based case-control study conducted in eight provinces. In this study, increased odds of lung cancer were found for all radon exposure measures and effect sizes correspond closely with those reported in recent meta-analyses. These epidemiological results, along with the national radon concentration and potential maps, could be used as one component of a comprehensive radon strategy to geographically target prevention efforts to high risk areas of Canada.

Chapter 6

Neighborhood Socioeconomic Status and Individual Lung Cancer Risk: Evaluating Long-Term Exposure Measures and Mediating Mechanisms

Submitted for publication: Hystad, P., Carpiano, R., Demers, P., Johnson, K., Brauer, M. Neighborhood Socioeconomic Status and Individual Lung Cancer Risk: Evaluating Long-Term Exposure Measures and Mediating Mechanisms

Abstract

Background: Neighborhood socioeconomic status (SES) has been associated with a number of chronic diseases, but little information is available on the association between long-term neighborhood SES and lung cancer incidence.

Methods: Using data from a large Canadian population-based lung cancer case-control study, we: (1) investigate whether there is an association between lung cancer incidence and long-term neighborhood SES, derived from 20 years of residential histories and five national censuses; (2) compare long-term neighborhood estimates to single-point-in-time neighborhood SES measures; and (3) examine the extent to which the association between long-term neighborhood SES and lung cancer is mediated by a range of individual-level behaviors, and environmental and occupational exposures.

Results: The odds of lung cancer cases residing in the most versus least deprived long-term neighborhood SES quintiles were significantly elevated in analyses of national (OR: 1.36;

95% CI: 1.14-1.62) and city (OR: 1.46; 95% CI: 1.13-1.89) samples after adjustment for individual SES, and remained significant (OR: 1.38; 95% CI: 1.01-1.88) in the city sample after adjusting for smoking and other known and suspected lung cancer risk factors. No-dose response relationship was observed and the influence of neighborhood SES on lung cancer risk was completely restricted to the most socioeconomically deprived neighborhoods. Important differences were observed between long-term and single-point-in-time neighborhood SES measures, with the latter attenuating effect estimates by over 50 percent. Smoking behavior was the predominant mediating pathway of long-term neighborhood SES in both analyses, while occupational exposure to industrial dusts and odors was also a mediator in the national analysis.

Conclusions: We found associations between long-term neighborhood SES measures and lung cancer incidence after accounting for individual SES and smoking behaviour. More research is needed to further examine specific, modifiable pathways by which neighborhood context may influence lung cancer risk.

6.1. Introduction

Lung cancer represents approximately 18% of all cancer deaths worldwide (Jemal et al., 2011). In Canada, there are approximately 25,600 new lung cancer cases and 20,200 lung cancer deaths annually, representing 14% of new cancers and 27% of all cancer deaths (CCS, 2012). Strong social (and spatial) gradients in lung cancer are present in Canada and many other countries, with significantly higher incidence rates among disadvantaged populations as well as places (Sidorchuk et al., 2009). These differences remain after accounting for individual smoking behaviour as well as other established lung cancer risk factors (Mao et al., 2001; Sidorchuk et al., 2009).

Neighborhood context, especially socioeconomic status (SES), has an independent association with a number of health outcomes, including all-cause mortality (Meijer et al., 2012b) and cardiovascular disease (Kawakami, Li, & Sundquist, 2011; Major et al., 2010; Diez-Roux et al., 2001); however, the role of neighborhood SES in shaping lung cancer risk (and gradients) has not been adequately examined to date. Neighborhood influences on cancer incidence in general have only recently been examined in a relatively small number of studies (Freedman et al., 2011; Major et al., 2010; Meijer et al., 2012; Matukala Nkosi et al., 2012; Webster et al., 2008).

Overall, the lack of research on neighborhood SES and cancer may be due to two particular methodological challenges. First, long latency periods are typically associated with cancer development and properly examining these latency periods requires information on study participants' residential histories in which to link corresponding measures of neighborhood context over multiple time points – information that is not commonly available or collected. Studies have highlighted differential residential mobility patterns based on socio-

economic factors (Hurley et al. 2005) as well as important individual health-behaviours (Pearce & Dorling, 2010), which could bias neighborhood studies that do account for residential mobility. A review of neighborhood studies of older adults found that 10 of the 33 identified studies accounted for respondents' length of residence in a neighborhood, but only one study accounted for changes in neighborhood environments (Yen et al., 2009). Several studies have also examined area-level influences over the life course (Carson et al. 2007; Clarke et al. 2013; Curtis et al. 2004; Leyland & Næss 2009; Naess et al. 2008; Ohlsson & Merlo 2011) as well as neighborhood context at multiple time points (Lloyd et al., 2010; Sampson et al., 2008; Wheaton and Clarke, 2003), but findings are mixed and the majority of this literature has not examined cancer.

Second, is the challenge of testing numerous direct and mediating pathways through which neighborhood SES may influence cancer risk. Though studies have found that neighborhood SES is associated with the onset of cancer after controlling for individual-level socioeconomic characteristics (e.g. Freedman et al., 2011; Meijer, Bloomfield, et al., 2012; Webster et al., 2008) there have been few studies that examine whether such associations remain after accounting for a comprehensive set of individual risk factors and whether these factors are mediators of the neighborhood SES effect on cancer risk. In terms of lung cancer, the effects of long-term neighborhood SES may be particularly mediated through smoking behaviours, due to the large risk posed by smoking to lung cancer (Pesch et al., 2012), and the fact that smoking, as well as smoking cessation, has been associated with neighborhood SES (Duncan et al. 1999; Giskes et al. 2006). In addition to smoking, neighborhood SES may be associated with lung cancer risk through a number of other mediating pathways of known or suspected lung cancer risk factors, including: exposure to second hand smoke; alcohol, diet

and physical inactivity; occupational exposures; and exposure to environmental hazards such as ambient air pollution (Alberg et al., 2007).

A number of mechanisms may underlie how these behavioral, occupational and physical environmental factors affect the neighborhood SES and lung cancer association. In terms of individual health-behaviours, these may include: (1) differential resources and opportunity structures within neighborhoods, such as smoking outlets, availability of healthy food retailers, or physical activity opportunities (e.g., Black et al., 2011; Boone-Heinonen et al., 2011); (2) prevalent norms and attitudes within neighborhoods regarding health-behaviours (e.g., Annear et al., 2009; Karasek et al., 2012); (3) local social networks and community belonging that support or hinder health-behaviour (e.g., Carpiano, 2008; Hystad & Carpiano, 2012); and (4) disordered and stressful neighborhood environments that may lead to unhealthy coping behaviors, such as smoking and alcohol consumption, and worse general health (e.g., Ellaway & Macintyre, 2009). With regard to stressful neighborhood environments, neighborhood SES may shape exposure to chronic psychosocial stress leading to allostatic load. Allostatic load may contribute to the development and progression of some types of cancer (Reiche et al., 2004) and biomarkers of allostatic load have been found to be increased among residents of lower SES neighborhoods (Bird et al., 2010; Finch et al., 2010). In terms of physical environmental factors, an extensive body of literature demonstrates low SES neighborhoods may be disproportionately exposed to environmental hazards, particularly industrial and vehicle air pollution (Mohai et al., 2009). Little research has examined occupational exposures as a mediating pathway of neighborhood effects; however, living in low SES neighborhoods may shape occupational opportunities towards working class or "blue collar" jobs that are associated more frequently with exposure to cancer causing substances (Evans & Kantrowitz, 2002).

The present study aims to address these two important challenges and build upon prior research to better understand the nature of the neighborhood SES association with lung cancer incidence. Specifically, we examine the relationship between long-term neighborhood SES and lung cancer incidence by using data from a large Canadian population-based lung cancer case-control study and deriving long-term neighborhood SES measures from 20 years of residential histories and five censuses. The goals of this study are to: (1) investigate whether there is an association between long-term neighborhood SES and lung cancer incidence; (2) compare long-term neighborhood estimates to commonly used single point-in-time neighborhood measures; and (3) examine the extent to which the association between long-term neighborhood SES and lung cancer is mediated by a range of behavioral, environmental and occupational factors.

6.2. Methods

Study Population

We used the lung cancer component of the National Enhanced Cancer Surveillance System (NECSS), which includes 3,340 lung cancer cases and 5,039 population controls collected in eight Canadian provinces (Newfoundland, Nova Scotia, Prince Edward Island, Ontario, Manitoba, Saskatchewan, Alberta and British Columbia) between 1994 and 1997 (Johnson et al. 1998). Briefly, cases were identified and sampled by Provincial Cancer Registries with newly diagnosed histological confirmed primary lung cancer, and after obtaining physician consent, were mailed questionnaires and telephone follow up conducted if needed. The response rate for contacted lung cancer cases was 61.7%. Population controls without cancer were selected from a random sample of individuals within each province, with an age/sex distribution similar to that of all cancer cases collected in the NECSS. Recruitment

methods for controls depended on data availability and accessibility by province and included provincial health insurance plans in five provinces, random digit dialing in two, and property assessment data in one. Provincial Cancer Registries collected information from controls using the same protocol as for the cases. The response rate for contacted population controls was 67.4%.

Assessment of Long-Term Neighborhood Socioeconomic Status

Neighborhood SES characteristics were estimated from participants' residential postal codes between 1975 and 1994. Residential histories were collected prior to 1975, but limited census data were available for these years and recall bias was present for residential histories prior to 1975 (cases tended to report more residences than controls [Hystad et al., 2012]). To reduce exposure assessment error, the study was therefore limited to participants that had neighborhood census data for ≥18 residential years between 1975 and 1994. This reduced the sample size to 2,568 lung cancer cases and 3,821 population controls for subsequent analyses. Differences between excluded and retained lung cancer cases and population controls were examined and no significant differences in demographic, socioeconomic, or smoking characteristics were found.

Neighborhood SES characteristics were derived from 1971, 1981, 1986, 1991 and 1996 census data. Census tract (CT) geography was used to represent neighborhoods in Census Metropolitan Areas (CMA's) (hereafter referred to as cities), while census sub-division (CSD) geography was used in non-CMA locations, as CTs are only available for large cities in Canada. Statistics Canada defines neighborhoods using CTs, which encompass 2,500 to 8,000 individuals and are created from recognizable boundaries that maximize homogenous populations in terms of socioeconomic characteristics (Statistics Canada, 2012a). CSDs generally represent municipalities and represent much larger geographic areas than CTs.

Residential postal codes were linked to the geographic boundaries of the closest census year (e.g. 1978 residences received 1981 census data) using Geographic Information Systems (GIS) and Statistics Canada postal code conversion file (PCCF) software (Statistics Canada, 2012b).

Long-Term Neighborhood Socioeconomic Status Measures

Measuring neighborhood SES conditions across both the study time-period and geographic areas of Canada entailed a three step process. First, long-term neighborhood SES was measured using five census variables: household income, percent of adults' unemployed, percent of adults without a high school diploma, percent rental dwellings, and percent of residents that moved in the last five years. These variables were selected because they were measured in a comparable fashion in all census years and capture important constructs of neighborhood SES. We standardized the variables by creating z-scores for each census variable by census year as well as by city (for CT neighborhoods) and by province (for CSD neighborhoods). This method standardizes the large temporal changes in census data, particularly for household income and high school education. The z-scores also reflect the regional deviation of the neighborhood census value for each city or province within each census year, due to the large geographical differences in Canada. These Z-scores were each coded such that positive values represent increasingly improved neighborhood SES conditions while negative values indicate increasing neighborhood SES deprivation.

Second, we calculated a neighborhood SES index by computing the mean of all five neighborhood characteristic z-scores for each census year. If individuals were assigned CT measures, then these measures were used to represent neighborhood characteristics, and if CT measures were not available then CSD measures were used.

Lastly, the long term neighborhood SES index score for each participant was obtained by averaging annual z-scores over the twenty year exposure period. The Cronbach's alpha for

the five individual neighborhood census variables in the final national sample was 0.62 and in the city sample was 0.80. Quintiles were created from the final twenty year average z-scores and used in the subsequent analyses.

Individual-level Covariates

An extensive amount of individual-level information was collected by questionnaire for each study participant. The covariates included in our analyses are important socio-demographic characteristics and known or suspected risk factors for lung cancer:

1. *Socio-demographic variables* include age, sex, educational attainment, and household income during the five years prior to study interview.
2. *Smoking-related variables* include lifetime cigarette smoking pack-years, years since quit smoking, and person-years of residential second hand smoke exposure (computed as the number of smokers in the home multiplied by the number of years living in that home).
3. *Other health behaviours* include average alcohol drinks per week, average serving of meat and vegetables per week, and average moderate and rigorous physical activity per month.
4. *Occupational exposures* include person-years of occupational second hand smoke exposure (defined by the number of smokers in the immediate work environment multiplied by the number of years at the job), years working with daily or weekly exposure to industrial dusts or odors, and years working with occupational lung hazards (including arsenic, asbestos, cadmium, chromium, asphalt, welding, and wood dust).

5. *Environmental exposures* include individual estimates of exposure to a common traffic pollution, nitrogen dioxide (NO_2) (Hystad et al., 2012), the number of years residing within 100 meters of a major road, and average ecological-level radon exposures estimates (radon is a known lung cancer risk factor resulting from naturally occurring gas released from the breakdown of Uranium in soils) (Hystad et al., 2012). All environmental exposures were derived from residential histories over a twenty year period.

Statistical Methods and Analytic Procedures

Due to the differences in neighborhood definitions between city and non-city areas, two analyses were conducted for all models. The first is the national analysis, which includes all study participants and neighborhood SES estimates from both CTs (cities) and CSDs (non-cities), which has a sample size of 2,300 cases and 3,548 controls. The second is a subset of individuals living in cities with neighborhood measures derived only from CTs, which has a reduced sample of 1,116 cases and 1,682 controls.

Analyses of long-term neighborhood SES and lung cancer incidence were conducted using two-level random intercept logistic regression models (GLIMMIX, SAS version 9.3; SAS Institute, INC, Cary, NC). To account for residual geographic patterns, the two-level models used a random intercept defined from Statistics Canada 1986 census division (CD) boundaries (n=188) that represent regional areas in Canada, and assigned to individuals' longest residential location. Hierarchical models were not created to account for within-neighborhood correlation (i.e. by CTs or CSDs) as there were small census-year intra-class correlations (~0 for CTs and CSDs), reflecting the small number of individuals within each neighborhood. On average, participants reported 2.3 (SD=1.6) residences during the exposure period, therefore also reducing the effect of within-neighborhood clustering on parameter

estimates. This model was also estimated with different single point-in-time neighborhood SES measures to compare with the long-term neighborhood measure.

To assess mediation of the long-term neighborhood SES and lung cancer effect, we first assessed the direct association between long-term neighborhood SES and lung cancer incidence. Next, we introduced the health behaviours and occupational and environmental factors in incremental models to assess the degree to which the relationship between long-term neighborhood SES and lung cancer incidence changed. For continuous mediator variables, ORs are calculated for an inter-quartile range (IQR) increase. Finally, we assessed the association between the significant health behavior, occupational, and environmental factors and long-term neighborhood SES using linear and ordinal regression.

6.3. Results

Descriptive Statistics of Study Population and Neighborhood Measures

Descriptive characteristics of the national study participants are summarized in Table 18 as well as characteristics for individuals within the most and least deprived neighborhood SES quintiles. Study participants were approximately evenly distributed between men and women, and lung cancer cases were slightly older than population controls, due to frequency matching on all cancer cases in the NECSS. As expected, lung cancer cases had considerably higher smoking pack-years. In addition, cases had less education, lower income, higher weekly alcohol and meat consumption, lower vegetable consumption, lower moderate and physical activity levels, higher residential and occupational second hand smoke exposures, and higher occupational exposures to dust/odors and hazardous substances. Characteristics between individuals in the most and least deprived neighborhood SES quintiles also differed for both lung cancer cases and population controls, with detrimental health behaviors and

occupation and environmental exposures (expected for NO_2 air pollution exposures) clustering in low SES neighborhoods. Differences between the national and urban study participants were also seen for many individual health behaviours, and occupational and environmental exposures (data not shown). In terms of long-term neighborhood SES, lung cancer cases had significantly lower average long-term neighborhood SES index scores than population controls. Significant differences were also found for all of the specific neighborhood characteristics except for five year residential mobility and percent rental dwellings in the national analysis and five year residential mobility in the urban analysis (Table 19).

Table 18. Descriptive statistics for the national sample of lung cancer cases and population controls stratified by the most and least deprived quintile of long-term neighborhood SES.

Characteristics[1]	All Participants in National Analysis		Participants in Least Deprived Neighborhood Index Quintile		Participants in Most Deprived Neighborhood Index Quintile	
	Cases (n=2,568)	Controls (n=3,821)	Cases (n=469)	Controls (n=808)	Cases (n=599)	Controls (n=678)
Age; mean (±SD)	63.5 (8.2)	60.0 (12.6)	63.9 (7.7)	58.9 (12.5)	63.4 (8.1)	59.1 (12.3)
Men; n (%)	1329 (51.8%)	1934 (50.6%)	250 (52.3%)	416 (51.5%)	307 (51.3%)	340 (50.2%)
Median Household income						
<$10,000	143 (5.6%)	114 (3.0%)	15 (3.3%)	8 (1.0%)	47 (8.2%)	30 (4.6%)
$10,000-19,999	395 (15.4%)	401 (10.5%)	43 (9.5%)	43 (5.5%)	110 (19.1%)	86 (13.1%)
$20,000-29,999	420 (16.4%)	587 (15.4%)	64 (14.1%)	84 (10.7%)	99 (17.2%)	117 (17.8%)
$30,000-49,999	510 (19.9%)	913 (23.9%)	105 (23.1%)	177 (22.6%)	107 (18.6%)	160 (24.3%)
$50,000-$99,999	309 (12.0%)	705 (18.5%)	92 (20.3%)	233 (29.7%)	62 (10.8%)	90 (13.7%)
>$100,000	50 (2.0%)	158 (4.1%)	22 (4.9%)	69 (8.8%)	4 (0.7%)	17 (2.6%)
Preferred not to report	608 (23.7%)	824 (21.6%)	113 (24.9%)	170 (21.7%)	147 (25.5%)	159 (24.1%)
Years of education; mean (±SD)	10.7 (3.1)	11.9 (3.6)	11.7 (3.2)	13.1 (3.5)	10.2 (3.1)	11.2 (3.5)
Smoking pack years; mean (±SD)	35.0 (21.1)	12.5 (17.3)	34.1 (21.1)	11.3 (16.5)	37.3 (22.6)	13.7 (19.3)
Years since quit smoking; mean (±SD)	6.1 (8.9)	10.1 (13.1)	6.8 (9.9)	10.5 (13.5)	5.5 (8.3)	10.2 (13.6)
Alcohol servings/week; mean (±SD)	7.0 (12.4)	4.8 (9.1)	7.7 (11.6)	5.5 (8.7)	7.9 (14.4)	4.8 (9.1)
Meat servings/week; mean (±SD)	10.0 (8.6)	8.6 (8.9)	9.6 (6.5)	8.1 (6.8)	9.9 (8.6)	8.7 (14.5)
Vegetable servings/week; mean (±SD)	20.2 (14.2)	20.7 (14.4)	20.2 (10.8)	20.4 (10.9)	19.2 (11.5)	21.6 (22.9)
Moderate physical activity hrs/month; mean (±SD)	9.3 (13.6)	9.7 (13.5)	9.1 (13.7)	9.3 (13.3)	6.8 (12.4)	6.6 (11.6)
Rigorous physical activity hrs/month; mean (±SD)	1.6 (5.8)	2.5 (7.3)	1.5 (5.2)	2.3 (6.5)	1.1 (4.8)	1.8 (7.3)
Residential SHS exposure; mean (±SD)	55.8 (44.8)	35.7(37.5)	51.1 (42.9)	31.3 (33.7)	54.8 (44.4)	37.2 (37.8)
Occupational SHS exposure; mean (±SD)	73.3 (82.2)	55.9 (73.4)	71.3 (80.2)	60.0 (75.2)	73.9 (85.7)	55.3 (73.4)
Years working with industrial dust/odors; mean (±SD)	12.8 (16.6)	10.1 (15.1)	10.9 (15.7)	9.7 (15.3)	14.2 (17.4)	11.3 (16.0)
Years working with hazardous substances; mean (±SD)	12.2 (30.0)	9.2 (25.6)	11.0 (24.8)	8.9 (25.6)	11.5 (26.2)	9.8 (33.5)
Average HR Radon Measurements (Bq/m^3); mean ±SD	80.7 (40.4)	78.3 (39.5)	79.9 (39.7)	75.1 (36.8)	78.1 (37.9)	72.7 (33.6)
NO_2 (ppb); mean ±SD	15.7 (8.9)	15.3 (9.1)	18.4 (8.6)	18.5 (8.5)	16.9 (9.6)	16.9 (10.0)
Years living within 100m of Major Roads; mean ±SD	3.1 (6.5)	2.9 (6.3)	2.2 (5.7)	2.1 (5.4)	3.2 (6.6)	3.1 (6.6)

[1] Descriptive statistics for study participants with ≥18 years of residential histories from 1975-1994.

Table 19. Long-term neighborhood SES z-scores derived from residential histories between 1975 and 1994 for lung cancer cases and population controls.

Variable	Lung Cancer Cases	Population Controls	*p* value
National Analysis	N=2,467	N=3,675	
SES Index	-0.00 (0.47)	0.05 (0.46)	<0.01
Average household Income	0.08 (0.72)	0.17 (0.72)	<0.01
Percent residents who moved over last 5 years	-0.08 (0.73)	-0.08 (0.74)	0.78
Percent rental dwellings	-0.21 (0.94)	-0.18 (0.94)	0.26
Percent no high school diploma	0.11 (0.78)	0.20 (0.76)	<0.01
Percent unemployed	0.08 (0.58)	0.15 (0.54)	<0.01
City Analysis	N=1,224	N=1,802	
SES Index	0.03 (0.57)	0.13 (0.55)	<0.01
Average household Income	-0.01 (0.80)	0.15 (0.80)	<0.01
Percent residents who moved over last 5 years	0.12 (0.70)	0.16 (0.67)	0.09
Percent rental dwellings	0.04 (0.79)	0.15 (0.75)	<0.01
Percent no high school diploma	-0.03 (0.85)	0.08 (0.82)	<0.01
Percent unemployed	0.03 (0.67)	0.13 (0.63)	<0.01

Evaluating the Association between Long-term Neighborhood SES and Lung Cancer Risk

First, we examined the association between neighborhood SES and lung cancer incidence. Table 20 shows results from a series of models using the long-term and 1994 study-entry neighborhood SES measures.

In the unadjusted models, the OR for lung cancer among individuals with the most versus least deprived quintile of long-term neighborhood SES were 1.63 (95% CI: 1.38-1.92) in the national analysis and 1.66 (95% CI: 1.31-2.09) in the city analysis. These ORs were attenuated when individual income and education were added to the model, and further reduced in the fully-adjusted model that included all behavioral, occupational, and environmental covariates shown in Table 18. In the national analysis, the OR was no longer statistically significant (OR 1.18; 95% CI: 0.95-1.46), while in the city model analysis, the long-term neighborhood SES measure remained elevated and statistically significant for the most versus least deprived quintile (OR 1.38; 95% CI: 1.01-1.88). In neither the national nor urban analyses was a dose-response relationship observed and associations were completely restricted to the most deprived quintile of the neighborhood SES index.

Comparing Neighborhood SES Measures

How do these long-term neighborhood SES results differ from results using single point-in-time neighborhood SES measures – the most common approach in neighborhood health effects research? Table 20 shows that, overall, the unadjusted ORs for study-entry neighborhood SES were slightly lower than the long-term OR estimate and similar in the city analysis. Large differences, however, occurred in the fully adjusted models, in which case the study-entry measures were severely attenuated (greater than 50% compared to the long-term measures) and no longer statistically significant, while the long-term neighborhood measures remained elevated and statistically significant in the urban analysis. In the national analysis, the correlation between the long-term neighborhood SES index and point-in-time 1994 index was 0.78. In the city analysis, this correlation was 0.81. These results indicate relatively high correlations, but also that substantial differences in the neighborhood SES measures exist and that study-entry measures are not fully capturing long-term neighborhood SES. We also examined different latency periods, including 1975 and 1975-1985 derived neighborhood measures and found similar patterns of reduced OR estimates in all models compared to long-term neighborhood SES (data not shown).

Table 20. Odds ratios for lung cancer incidence by quintiles of the neighborhood SES index derived from long-term and study-entry residences.

	Long-Term (1975-1994) Neighborhood SES			Study Entry (1994) Neighborhood SES		
National Analysis	**Unadjusted[a]**	**SES-Adjusted[b]**	**Fully-Adjusted[c]**	**Unadjusted[a]**	**SES-Adjusted[b]**	**Fully-Adjusted[c]**
Q1 [most deprived]	1.63 (1.38-1.92)*	1.36 (1.14-1.62)*	1.18 (0.95-1.46)	1.47 (1.24-1.73)*	1.28 (1.07-1.53)*	1.05 (0.85-1.31)
Q2	1.12 (0.94-1.32)	0.94 (0.78-1.12)	0.91 (0.73-1.13)	1.22 (1.04-1.44)*	1.06 (0.89-1.27)	0.94 (0.76-1.17)
Q3	1.19 (1.01-1.41)*	1.02 (0.86-1.22)	0.91 (0.74-1.13)	1.20 (1.02-1.43)*	1.13 (0.94-1.35)	1.01 (0.82-1.26)
Q4	1.21 (1.03-1.43)*	1.05 (0.88-1.25)	1.06 (0.86-1.31)	1.03 (0.87-1.22)	0.95 (0.80-1.14)	0.87 (0.71-1.08)
Q5 [least deprived]	1.00	1.00	1.00	1.00	1.00	1.00
City Analysis						
Q1 [most deprived]	1.66 (1.31-2.09)*	1.46 (1.13-1.89)*	1.38 (1.01-1.88)*	1.60 (1.26-2.03)*	1.44 (1.11-1.87)*	1.18 (0.86-1.61)
Q2	1.21 (0.96-1.54)	1.05 (0.81-1.35)	0.92 (0.67-1.25)	1.35 (1.06-1.72)*	1.21 (0.94-1.56)	1.08 (0.80-1.49)
Q3	1.10 (0.87-1.40)	1.00 (0.78-1.29)	1.00 (0.74-1.36)	1.16 (0.91-1.48)	1.09 (0.85-1.41)	1.01 (0.74-1.37)
Q4	0.92 (0.72-1.17)	0.83 (0.64-1.07)	0.85 (0.63-1.16)	0.88 (0.69-1.12)	0.83 (0.64-1.07)	0.73 (0.54-1.00)
Q5 [least deprived]	1.00	1.00	1.00	1.00	1.00	1.00

* $p \leq .05$

[a] Controls for age, sex, study province.

[b] Controls for age, sex, study province, education, and household income.

[c] Controls for age, sex, study province, education, household income and all other variables listed in Table 18.

Assessing Mediators between Long-Term Neighborhood SES and Lung Cancer Risk

To what degree is the association between long-term neighborhood SES and lung cancer risk mediated by health behaviours, and occupational and environmental exposures? Tables 21 and 22 summarize results for long-term neighborhood SES when several hypothesized mediating factors are controlled: smoking, health behaviors, and occupational and environmental exposures – for the national and urban analyses, respectively.

For Table 21 (the national analysis), model 1 shows results for a baseline model for long-term neighborhood SES adjusted for individual-level income and education and sociodemographics. Model 2 (including smoking variables) further attenuated the neighborhood effect by 53%, to 1.20 (95% CI: 0.97-1.48). Models 3 (other health-behaviours), 4 (occupational exposures) and 5 (environmental exposures) had small influence on the neighborhood SES and lung cancer association, even though most individual-health behaviours and occupational exposures were significantly associated with lung cancer incidence and corresponded to hypothesized effect directions. In model 6 (fully adjusted model), smoking pack years, years since cessation, residential second hand smoke exposure, meat consumption, and occupational exposure to industrial dusts and odors are significant.

In terms of the city analysis (Table 22), including income and education (model 1) attenuated the unadjusted neighborhood OR to 1.46 (95% CI: 1.13-1.89). Inclusion of smoking variables again attenuated the neighborhood effect observed in model 2 (although only by 19.5%), but the OR for the most versus least deprived neighborhood SES quintiles remained statistically significant (OR: 1.37; 95% CI: 1.01-1.87). Including the other individual health-behaviours (model 3), occupational exposures (model 4) and environmental exposures (model 5) had little influence on the neighborhood effect. In the fully adjusted model (model 6), all smoking variables and meat consumption were statistically significant.

Table 21. Odds ratios for lung cancer incidence in the national incremental models.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Neighborhood SES Index						
Q1 [most deprived]	1.36 (1.14-1.62)*	1.20 (0.97-1.48)	1.34 (1.12-1.60)*	1.33 (1.11-1.59)*	1.36 (1.14-1.63)*	1.18 (0.95-1.46)
Q2	0.94 (0.78-1.12)	0.90 (0.73-1.11)	0.92 (0.77-1.11)	0.93 (0.78-1.11)	0.95 (0.79-1.14)	0.91 (0.73-1.13)
Q3	1.02 (0.86-1.22)	0.91 (0.74-1.13)	1.00 (0.84-1.20)	0.98 (0.82-1.18)	1.04 (0.87-1.24)	0.91 (0.74-1.13)
Q4	1.05 (0.88-1.25)	1.06 (0.86-1.31)	1.06 (0.89-1.27)	1.03 (0.87-1.23)	1.07 (0.90-1.27)	1.06 (0.86-1.31)
Q5 [least deprived]	1	1	1	1	1	1
Smoking						
Smoking pack years		3.24 (2.74-3.84)*				3.15 (2.65-3.75)*
Residential SHS exposure		1.12 (1.02-1.24)*				1.12 (1.01-1.24)*
Yrs since quit smoking						
Never Smoker		1				1
>25 years		1.25 (0.92-1.69)				1.23 (0.91-1.67)
11-25 years		2.58 (1.99-3.34)*				2.62 (2.02-3.41)*
6-10 years		4.78 (3.54-6.45)*				4.82 (3.56-6.05)*
2-5 years		7.25 (5.24-10.0)*				7.32 (5.28-10.1)*
Current Smoker		6.68 (5.14-8.68)*				6.48 (4.98-8.44)*
Health-Behaviours						
Moderate PA (hrs/month)			0.82 (0.74-0.91)			0.98 (0.86-1.10)
Rigorous PA (hrs/month)			0.92 (0.89-0.95)			0.97 (0.93-1.01)
Vegetable servings/wk			0.89 (0.84-0.94)			0.95 (0.89-1.01)
Meat servings/week			1.20 (1.13-1.26)*			1.11 (1.04-1.18)*
Alcohol servings/wk			1.16 (1.11-1.20)*			1.03 (0.98-1.07)
Occupational Exposures						
Occupational SHS				1.28 (1.18-1.39)*		1.00 (0.91-1.10)
Years working with industrial dust/odors				1.14 (1.05-1.24)*		1.10 (1.00-1.22)*
Years working with hazardous substances				1.01 (1.00-1.03)*		1.01 (0.99-1.02)
Environmental Exposures						
Average HR Radon (Bq/m^3)					1.05 (0.94-1.16)	1.06 (0.95-1.19)
NO_2 (ppb)					1.10 (0.96-1.27)	1.13 (0.97-1.31)
Residing 100m of Major Rd					0.99 (0.91-1.07)	0.99 (0.91-1.07)

* $p<.05$. NOTE: All models control for age, sex, study province, education, and household income. For continuous variables, ORs are calculated for one IQR increase.

Table 22. Odds ratios for lung cancer incidence in the city incremental models.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Neighborhood SES Index						
Q1 [most deprived]	1.46 (1.13-1.89)*	1.37 (1.01-1.87)*	1.44 (1.11-1.88)*	1.45 (1.12-1.87)*	1.45 (1.12-1.88)*	1.38 (1.01-1.88)*
Q2	1.05 (0.81-1.35)	0.90 (0.67-1.23)	1.03 (0.80-1.34)	1.05 (0.81-1.36)	1.06 (0.82-1.36)	0.92 (0.67-1.25)
Q3	1.00 (0.78-1.29)	0.99 (0.73-1.34)	1.01 (0.78-1.31)	0.99 (0.77-1.28)	1.01 (0.78-1.30)	1.00 (0.74-1.36)
Q4	0.83 (0.64-1.07)	0.84 (0.62-1.13)	0.84 (0.65-1.09)	0.84 (0.65-1.08)	0.83 (0.65-1.07)	0.85 (0.63-1.16)
Q5 [least deprived]	1	1	1	1	1	1
Smoking						
Smoking pack years		2.72 (2.15-3.43)*				2.66 (2.09-3.38)*
Residential SHS exposure		1.10 (0.82-1.46)				1.09 (0.95-1.26)
Yrs since quit smoking						
Never Smoker		1				1
>25 years		1.13 (0.74-1.72)				1.19 (0.77-1.80)
11-25 years		2.80 (1.95-4.02)*				2.83 (1.97-4.07)*
6-10 years		5.26 (3.45-8.03)*				5.33 (3.48-8.16)*
2-5 years		7.55 (4.78-11.9)*				7.63 (4.81-12.1)*
Current Smoker		7.46 (5.17-10.7)*				7.18 (4.96-10.3)*
Health-Behaviours						
Moderate PA (hrs/month)			0.73 (0.62-0.85)			0.91 (0.75-1.09)
Rigorous PA (hrs/month)			0.92 (0.88-0.97)			0.98 (0.92-1.04)
Vegetable servings/wk			0.84 (0.77-0.92)			0.92 (0.82-1.03)
Meat servings/week			1.28 (1.18-1.40)*			1.17 (1.06-1.29)*
Alcohol servings/wk			1.17 (1.11-1.24)*			1.04 (0.97-1.11)
Occupational Exposures						
Occupational SHS				1.31 (1.16-1.49)*		1.00 (0.85-1.16)
Years working with industrial dust/odors				1.03 (0.91-1.16)		1.01 (0.87-1.17)
Years working with hazardous substances				1.00 (0.99-1.01)		1.00 (0.98-1.01)
Environmental Exposures						
Average HR Radon (Bq/m^3)					1.23 (0.98-1.55)	1.28 (0.96-1.58)
NO_2 (ppb)					1.02 (0.83-1.24)	1.09 (0.87-1.37)
Residing 100m of Major Rd					1.05 (0.88-1.26)	1.00 (0.80-1.24)

* $p<.05$. NOTE: All models control for age, sex, study province, education, and household income. For continuous variables, ORs are calculated for one IQR increase.

To further examine potential mediation, Table 23 models the individual mediating factors found to be statistically significant in the previous fully-adjusted models shown in Tables 21 and 22 with the long-term neighborhood SES exposures. In the national analysis, individuals in the most deprived neighborhood SES quintile had significantly higher smoking pack years (2.01 years) and were more likely to be current smokers or to have quit smoking more recently than individuals in least deprived neighborhoods after controlling for individual SES and other behavioural, occupational and environmental factors. Individuals in the more deprived neighborhood SES quintile were also more likely to work with industrial dusts and odors (1.65 additional years). Individual meat consumption and second hand smoke exposures were not statistically significant, suggesting they are not mediators of the neighborhood-lung cancer association. For the city analysis, individuals in the most deprived neighborhood SES quintile had higher, although not statistically significant, smoking pack years (1.17 years). Neighborhood SES was only significantly related to years since quit smoking, with similar associations as the national analysis.

Table 23. Predicting mediating variables based on individual's long-term neighborhood SES index.

	Smoking Pack Years[a]	Years Since Quit Smoking[b]	Residential SHS Exposure[a]	Meat Consumption[a]	Working with Industrial Dust/Odors[a]
National Analysis					
Long-term Neighborhood SES[a]					
Q1 [most deprived]	2.01 (0.82)*	1.40 (1.15-1.71)*	2.11 (1.70)	-0.08 (0.33)	1.65 (0.58)*
Q2	0.81 (0.83)	1.26 (1.01-1.59)*	1.82 (1.71)	-0.21 (0.33)	0.43 (0.58)
Q3	0.33 (0.83)	1.25 (1.01-1.56)*	2.16 (1.69)	-0.12 (0.33)	0.56 (0.58)
Q4	-0.67 (0.81)	1.09 (0.89-1.33)	2.80 (1.67)	0.16 (0.32)	0.78 (0.57)
Q5 [least deprived]	1	1	1	1	1
Urban Analysis					
Long-term Neighborhood SES[a]					
Q1 [most deprived]	1.17 (1.26)	1.37 (1.09-1.71)*	--	-0.56 (0.45)	--
Q2	0.84 (1.23)	1.24 (1.00-1.54)*	--	0.02 (0.43)	--
Q3	0.02 (1.21)	1.13 (0.91-1.40)	--	-0.31 (0.43)	--
Q4	-0.26 (1.20)	1.00 (0.80-1.23)	--	-0.68 (0.43)	--
Q5 [least deprived]	1	1	--	1	--

*P<.05; each model controls for age, sex, study province and all independent vars. reported in Tables 4 and 5.
[a] Linear regression model; estimates shown are variable coefficients.
[b] Ordinal logistic regression model; estimates shown are odds ratios for being current smoker.

6.4. Discussion

We used a large population-based lung cancer case-control study to: (1) investigate whether there is an association between long-term neighborhood SES, derived from 20 years of residential histories and five censuses, and lung cancer incidence; (2) compare long-term neighborhood estimates to common single-point-in-time neighborhood SES measures; and (3) examine the extent to which the association between long-term neighborhood SES and lung cancer is mediated by a range of behavioral, environmental and occupational factors.

For our first aim, we examined the association between long-term neighborhood SES and lung cancer incidence. Given that lung cancer incidence rates follow SES gradients (Sidorchuk et al., 2009), it was not surprising that we observed the odds of lung cancer elevated by sixty percent in the most deprived SES neighborhoods, suggesting that individual

SES inequalities are reflected in neighborhood environments. After controlling for all individual and geographic lung cancer risk factors, positive associations remained between long-term neighborhood SES and lung cancer incidence, with significant associations in the city analyses (OR: 1.38; 95% CI: 1.01-1.88). We hypothesize that the "left over" association may be due to chronic stressors associated with neighborhood socioeconomic deprivation; however, no measures of individual-level stress were available. In addition, we did not observe dose-response gradients and associations were primarily limited to the most deprived quintile of neighborhood SES, suggesting that the influence of neighborhood SES on lung cancer risk may be restricted to the most socioeconomically deprived neighborhoods. Our finding is similar to a study in Denmark that found lung cancer incidence was lower in areas with the lowest unemployment rate compared to the highest (HR: 0.88; 95% CI 0.84- 0.92) (Meijer et al., 2012a); however, other studies have not observed differences in lung cancer mortality and neighborhood SES after accounting for individual characteristics (Chaix et al., 2006; Martikainen et al., 2003). More research is therefore needed to examine whether the remaining association between neighborhood SES and cancer risk is due to chronic stress, unmeasured risk factors, or residual confounding from important etiological factors.

For our second aim, we examined the impact of using long-term neighborhood measures compared to common single-point-in-time measures of neighborhood SES in studying lung cancer risk. Despite high correlation between different measures, we found that a single point-in-time measure substantially attenuated effect estimates of neighborhood SES on lung cancer risk. For example, the effect sizes in the fully-adjusted models for study-entry versus long-term neighborhood SES measures were reduced by 72% and 53% in the national and city analyses. In addition, we found that the study-entry versus long-term neighborhood SES measure may be capturing different constructs. Including individual smoking, physical

activity, and diet and alcohol consumption eliminated most of the study-entry neighborhood SES association, while the long-term measure remained elevated and statistically significant in the city analysis. Similar results were found in a case-control study in Montreal, Canada, where the influence of neighborhood income (at time of study-entry) on lung cancer incidence was completely removed in the fully-adjusted model (Nkosi et al., 2012). Conceptually, long-term exposure to neighborhood SES would be required to influence lung cancer risk, due to the long latency period associated with lung cancer development. These finding highlight the importance of incorporating long-term measures of neighborhood context, especially in studies examining chronic diseases with long latency periods, and supports prior findings that single point-in-time neighborhood SES measures underestimate the impact of contextual conditions on health (Clarke et al., 2013; Lloyd et al., 2010; Murray et al., 2010; Wheaton & Clarke, 2003).

Our third aim was to examine mediation of the relationship between long-term neighborhood SES and lung cancer risk. A number of health behaviours, and occupational and environmental exposures were significantly associated with lung cancer incidence in this case-control study; however, the major mediating pathway identified was individual smoking behaviour, and to a smaller degree occupational exposures to industrial dusts/odors. The incremental models shown in Tables 21 and 22 highlighted the importance of smoking behaviour as a mediating factor and this finding is supported by the robust literature linking neighborhood SES to smoking, as well as smoking cessation, after accounting for individual factors (Pearce et al., 2012). In a review of the pathways linking neighborhood context and smoking, two general categories were identified: place-based practices (including social capital, social practices, contagion, and crime, disorder and related stressors) and place-based regulations (including smoking cessation policies, tobacco retailing and availability, tobacco

advertising, and urban renewal) (Pearce et al., 2012). For the other mediating factor indentified, very little research is available that examines whether occupational exposures may operate as a mediator of neighborhood SES; however, Martikainen et al. (2003) also demonstrated the important connection between neighborhood context, occupational composition (i.e. proportion of manual labor workers) and lung cancer risk. Overall, these findings suggest that individual smoking behaviour and occupational exposures to industrial dusts and odors are partial mediators of the long-term neighborhood SES effect on lung cancer risk.

Strengths and Limitations

Our study had a considerable number of strengths: a large population-based case-control study; inclusion of urban and rural residents; availability of long-term neighborhood measures; comprehensive individual-level information; and exposure estimates for occupational and environmental exposures. Nevertheless, the results of this study need to be interpreted with a number of limitations in mind.

First is the use of a broad indicator of neighborhood SES. Ideally, objective neighborhood measures other than SES alone should be included, such as the availability of tobacco retailers, physical activity and health food resources, perceived safety, etc., as well as individual perceptions of neighborhood SES conditions. Unfortunately, these types of variables were not available for this study. Second, we were only able to measure twenty years of neighborhood SES exposure and further life-course neighborhood measures (including those in early-life) should be evaluated. Third, we used CTs to represent neighborhoods in cities and CSDs to represent neighborhoods in non-city areas. The use of larger geographic sizes to represent neighborhoods in these areas may explain the weaker associations seen in the national analysis compared to the city analysis. The difficulty in defining neighborhoods in

rural areas is also reflected in the lower Cronbach's alpha for the five final census variables in the national analysis (0.62) compared to the city analysis (0.80). While much more work is needed to define and measure neighborhoods in rural areas, this population-based case-control study of neighborhood SES is one of the first studies to attempt to include both rural and urban locations and neighborhood measures. Fourth, this study has the potential for response bias in the recruitment of cases and population controls, even though the response rate for lung cancer cases (61.7%) and population controls (67%) was relatively high for a population-based case-control study. While a random sample of individuals is not a random sample of all places, we saw good coverage of Canadian neighborhoods in this study and z-scores for individual census variables corresponded to the census average, with small expected differences given the older age of the case-control study population compared to the general Canadian population. Fifth, the influence of residual confounding from important individual and geographic characteristics may exist as well as confounding by unmeasured individual-level factors. This is a particular concern for smoking, as shown when simple smoking status is used (Nkosi et al., 2012); however, we were able to control for a comprehensive set of individual variables and used multiple measures for smoking exposures (i.e. smoking pack-years, years since quite smoking, and residential and occupational second hand smoke exposure). Fifth and finally, while we modeled neighborhood SES over a twenty year period, our results may not necessarily be causal as residents may select into the neighborhoods in which they live on the basis of health-related behaviours, their occupation, and environmental characteristics. Nevertheless, we have taken an important step to furthering our understanding of how neighborhoods many influence lung cancer risk by including long-term neighborhood measures and examining mediation by health-behaviours and occupational and environmental exposures.

6.5. Conclusions

This study has highlighted three important findings that can be used to guide future research on the role of neighborhood context to lung cancer. First, estimating long-term neighborhood SES is important for evaluating the influence of neighborhood context on lung cancer risk, a result that is likely applicable to all chronic diseases. For such outcomes, study results may be severely attenuated if single point-in-time measures of neighborhood characteristics are used, as long-term and study-entry neighborhood SES measures, while correlated, may be capturing different exposure constructs. Our finding give further support for the need to include longitudinal and life-course measures of neighborhood context in health research to identify the true magnitude of neighborhood influences on health.

Second, using long-term neighborhood SES exposures, we found that lung cancer cases were more likely to reside in the most deprived socioeconomic neighborhoods of Canada and that individual smoking behaviour and potentially occupational exposure to industrial dusts and fumes are the main mediating pathways of the effect of neighborhood SES on lung cancer risk. While substantial research has examined potential intervention options for contextual influences on smoking, these have tended not to include cumulative life-course impacts of neighborhood context, which may operate through different pathways to influence smoking behaviour. Much more research is also needed to examine how neighborhood SES context may shape occupational opportunities or vice-versa how occupational opportunities may shape neighborhood SES.

Third, long-term neighborhood SES was also associated with lung cancer risk, even after accounting for individual-level smoking behaviour and a comprehensive set of other potential confounding factors. Hence, more research is needed to examine whether the

remaining association is due to chronic stress, unmeasured risk factors, or residual confounding from important etiological factors. This information could then be used to further examine specific, modifiable pathways by which neighborhood context may influence lung cancer risk.

Chapter 7

Conclusions of Dissertation

The objective of this dissertation was to examine whether three exposures associated with the physical and social residential environment – specifically, ambient air pollution, radon and neighborhood SES – are risk factors for the development of lung cancer in Canada. Each research chapter in this thesis was designed for independent publication and contains specific conclusions, contributions and limitations. Therefore, this final chapter serves to synthesize the overall findings and key contributions, highlight overall limitations of the research, and propose future research directions.

7.1. Summary of Research and Contributions

This dissertation contributed to the literature as well as to public and population health through: (1) advancing spatiotemporal exposure assessment methods; (2) creating new epidemiological knowledge for three environmental exposures; and (3) developing new knowledge and tools specific to Canada that can be used to reduce the population health burden associated with air pollution, radon and neighborhood SES exposures.

7.1.1. Methodological advances in spatiotemporal exposure assessment

Advances in spatiotemporal exposure assessments were made for all three exposures examined in this dissertation. In terms of air pollution, considerable effort was taken to create novel spatiotemporal air pollution exposure assessment methods. The national LUR method developed in Chapter 2 was one of the first applications of the LUR approach to fixed-site national monitoring data and the first to use satellite-based predictor variables. While this

approach was very successful in estimating current spatial concentrations of $PM_{2.5}$ and NO_2, it was not possible to estimate historical pollution concentrations using this approach (due to the detailed geographic data required). Therefore, Chapter 3 modified the national modeling strategy so it could be applied to estimate historical air pollution concentrations for residential histories within the NECSS lung cancer case-control study. This exposure assessment represents one of the most thorough retrospective exposures assessments conducted to date and was able to estimate exposures for participants in rural and urban areas of Canada, thereby allowing for one of the first national-level epidemiological analyses of air pollution. Chapter 3 also highlighted the importance of residential mobility to accurate exposure assessments and the potential biases that can be incorporated into air pollution epidemiological studies that do not incorporate such information.

The radon exposure assessment presented in Chapter 5 developed a method for estimating long-term ecological estimates for radon exposure using residential histories applied to mapped radon concentration and potential areas of Canada. This was the third study conducted to-date that has used ecological radon estimates with individual-level outcome and covariate data. A very large sample of residential radon measurements (n=14,000) was used to create the health region radon maps and significant variation was identified across Canada, which has important implications for prevention activities. In addition, this was the first study to include both radon concentration and potential maps in the same study, and important spatial similarities and differences were observed that may influence the use of such maps for educational and policy/risk assessment purposes.

This dissertation also contributed substantially to the methodological literature on neighborhood SES exposure assessment. Similar to the air pollution exposure assessment, Chapter 6 stressed the importance of estimating long-term spatiotemporal neighborhood SES

measures from residential histories and longitudinal census data. Important differences were observed between long-term and single point-in-time neighborhood SES measures, with the latter substantially attenuating effect estimates. In addition, long-term and study-entry neighborhood measures seemed to capture different exposure constructs. These findings have important implications for all neighborhood context research that uses single-point-in-time measures (typically neighborhoods at study entry) and strongly suggests that all neighborhood context research should use longitudinal measures. This was also one of the first studies to examine neighborhood influences on health in a population-based sample that included multiple cities and urban and rural locations.

Overall, this dissertation has stressed the need for comprehensive spatiotemporal exposure assessment approaches in epidemiology. While epidemiologists are now routinely capitalizing on the spatial dimensions of health data, there is a lack of examples (and methods) available that incorporate both space and time, especially for chronic diseases with long latency. This dissertation has attempted to address this gap by using twenty years of residential histories and applying spatiotemporal exposure assessment methods to estimate long-term air pollution, radon and neighborhood SES exposures and by evaluating what is gained by using such approaches compared to common spatial-only exposure assessments.

7.1.2. Epidemiological findings

New and important epidemiological findings were reported in this dissertation for lung cancer incidence and long-term air pollution, radon and neighborhood SES exposures. The air pollution and lung cancer findings reported in Chapter 4 support the growing body of evidence linking ambient air pollution to lung cancer risk. In summary, ORs for lung cancer incidence in the national analyses associated with a 10-unit increase in $PM_{2.5}$ ($\mu g/m^3$), NO_2 (ppb) and O_3

(ppb) were 1.29 (0.95-1.75), 1.11 (1.00-1.24), and 1.09 (0.85-1.16), respectively. The results of this analysis suggest that increased lung cancer risk is associated with ambient air pollution levels in Canada, with the strongest association for NO_2 and $PM_{2.5}$, but that more research is needed to examine the risks associated with O_3 exposure and proximity to traffic. This study represents one of largest air pollution and lung cancer studies conducted to date (and the first large study in Canada). This study is unique in that it: conducted risk estimates for three ambient air pollutants and proximity to traffic; stratified models by histological subtypes; included a population-based sample of both urban and rural populations; and adjusting for a comprehensive set of individual and geography information on potential confounders, such as second hand smoke exposure, occupational exposures and radon exposures. These epidemiological results highlight that lung cancer is associated with the comparability low ambient air pollution levels found in Canada and supports further reductions of ambient air pollution levels.

The radon epidemiological findings shown in Chapter 5 correspond to the existing literature that residential radon exposure is a risk factor for lung cancer. A 50 Bq/m^3 increase in average Health Region radon concentration was associated with a 7% (-6-21%) increase in the odds of lung cancer, and exposure quintiles demonstrated a dose-response relationship. This risk estimate corresponds closely to existing meta-analyses of residential radon and lung cancer risk. In addition, for every ten years that an individual lived in a high radon potential zone, the odds of lung cancer incidence increased by 11% (1-23%). To date, only one residential radon epidemiological study had been conducted in Canada and reported no associations between residential radon concentrations and lung cancer. The research presented in this dissertation demonstrates that radon is a significant risk factor for lung cancer in

Canada and that these risks are unevenly distributed across Canada – suggesting targeted prevention could be used to increase cost-effectiveness of prevention programs.

Chapter 6 presented the first epidemiological analysis of long-term exposure to neighborhood SES and lung cancer risk. The odds of lung cancer cases residing in the most versus least deprived long-term neighborhood SES quintiles were significantly elevated in the national (OR: 1.36; 95% CI: 1.14-1.62) and city (OR: 1.46; 95% CI: 1.13-1.89) analyses after adjustment for individual SES. This increase remained significant (OR: 1.38; 95% CI: 1.01-1.88) in the city analysis after adjusting for smoking and other known and suspected lung cancer risk factors. For both analyses there was no dose-response relationship, rather all associations were observed for the most deprived SES neighborhoods. Smoking behavior was identified as the predominant mediating pathway of long-term neighborhood SES, and to a smaller degree occupational exposure to industrial dusts and odors in the national analysis. The fact that long-term neighborhood SES was associated with lung cancer risk, even after accounting for individual-level smoking behavior and a comprehensive set of other potential confounding factors, suggests that more research is needed to examine whether the remaining association is due to chronic stress, unmeasured risk factors, or residual confounding from important etiological factors. This information could then be used to further examine specific, modifiable pathways by which neighborhood context may influence lung cancer risk.

7.1.3. Implications for reducing the burden of lung cancer in Canada

The epidemiological studies reported in this dissertation represent the first and/or largest studies conducted in Canada for air pollution, radon and neighborhood SES. Therefore, these studies contribute substantially to estimating the attributable lung cancer burden

associated with these risk factors in Canada. Such estimates are important for identifying population health impact and for developing population health policy.

All three environmental exposures examined in this dissertation were found to be associated with increased lung cancer risk (although not all were statistically significant). The magnitude of these associations is only one component contributing to their population health impact, and equally important is the prevalence of these risk factors within the Canadian population. The population attributable fraction (PAF) includes both the magnitude of risk and exposure prevalence and can be used to estimate the disease burden in a population that could be eliminated if the effects of certain risk factors were removed. From a population health perspective, PAFs are only useful if the association between a risk factor and disease is casual – the weight of evidence for the risk factors examined in this dissertation is strongest for radon, followed by ambient air pollution and then neighborhood SES.

PAFs are presented for $PM_{2.5,}$ $NO_{2,}$ radon and neighborhood SES (O_3 was not included as little evidence is available for an association with lung cancer). The WHO (2012) comparative risk assessment approach was used for $PM_{2.5,}$ NO_2 and neighborhood SES, while an existing radon PAF of 16% was used, which was calculated from the same residential radon measurement data used in Chapter 4 (Chen et al., 2012). Appendix 5 provides detailed PAF calculations. Briefly, $PM_{2.5}$ and NO_2 exposure prevalence estimates from the national exposure models reported in Chapter 2 of 8.4 (±1.4) $\mu g/m^3$ and 23.4 (±11.5) ppb are used along with the epidemiological findings that lung cancer increased by 29% and 11% per 10-unit increases. The counterfactual population exposure used for $PM_{2.5}$ and NO_2 were 2.5 (±1) $\mu g/m^3$ and 10 (±5) ppb, respectively. For neighborhood SES an increased risk of 36% for the most versus least deprived neighborhood SES quintile was used with a counterfactual of no neighborhood SES effects. The resulting PAFs are shown in Figure 15.

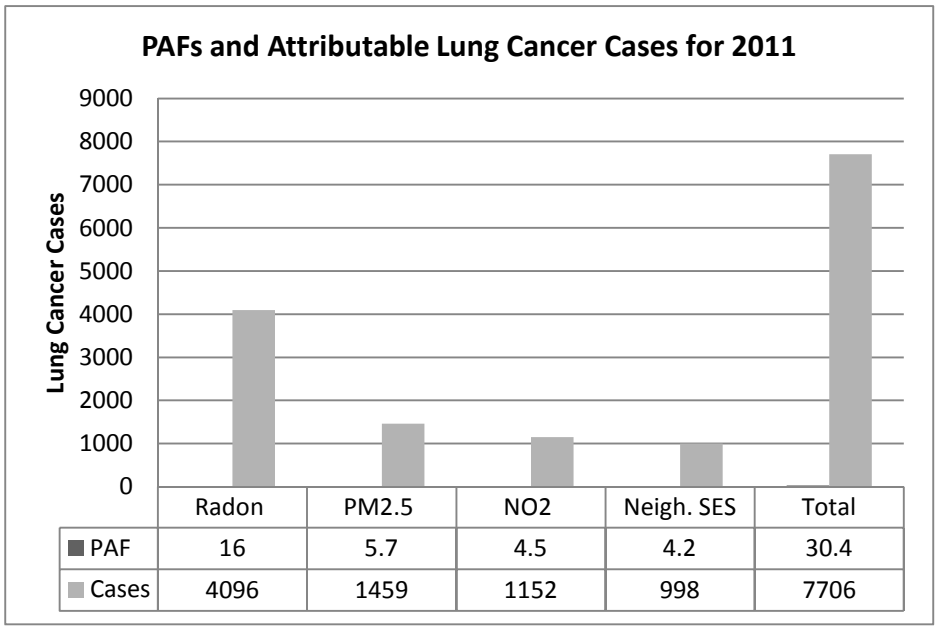

Figure 15. Summary of PAFs and attributable lung cancer cases during 2011 for the different environmental risk factors examined in this dissertation.

These PAF estimates reflect that a substantial burden of lung cancer is associated with the environmental exposures examined in this dissertation, with the largest risks coming from radon, followed by $PM_{2.5}$ and NO_2 air pollution and then neighborhood SES. While the sum of PAF's needs to be interpreted with caution, due to correlation between risk factors (primarily NO_2 and $PM_{2.5}$) the combined PAF for all risk factors is approximately 30%, translating into an estimated 7,706 lung cancer cases annually that could be eliminated if the effects of these risk factors were removed from the Canadian population. Importantly, these are theoretical estimates and more research is needed to establish causality, to incorporate mediating contributions (particularly for neighborhood SES), to account for correlated risk factors, and to determine counterfactual levels that can feasible be attained through public health and policy interventions. This estimate can be viewed as an upper bound estimate, due to exposure correlations. In addition, the use of a quintile exposure measure for neighborhood SES is not a

useful measure in the PAF framework, as there will always be a most deprived quintile (even though this may not have health implications once it reaches a certain threshold). Neighborhood context research is very much in its infancy and the body of evidence for specific causal associations has yet to be developed, although we were able to highlight the important role of smoking as a mediator of the neighborhood SES effect. Nevertheless, these estimates provide insights into the importance of these environmental exposures and how their risks compare with each other and with other risk factors.

The PAF is often misinterpreted in that a set of individual PAFs should add to 100%. In fact, PAFs will add to much more than 100% due to the multi-causal nature of lung cancer (and most chronic diseases). This is especially important for lung cancer, as PAFs for smoking of 80-90% (Parkin et al., 1994) are often interpreted as only 10-20% of the "causes" of lung cancer remain. Here we see that the PAFs for these four environmental risk factors combine to 30% and there remain many known lung cancer risk factors, such as occupational and second hand smoke exposures, which would also add significantly to this non-smoking PAF total. However, it is not necessarily beneficial to propose a hierarchy or compare one component cause against another – preventing or reducing cancer causing exposures wherever possible should be the goal of comprehensive cancer prevention programs, which is important to stress when interpreting and translating the finding of this dissertation.

In addition to providing information to estimate the lung cancer burden associated with air pollution, radon and neighborhood SES, this dissertation has created a number of spatial exposure assessment products that could be used to help reduce this burden. In terms of air pollution, the national LUR models created in Chapter 2 are now being used for population surveillance and risk assessments, to identify potential susceptible populations and neighborhoods in Canada (i.e. environmental justice research), and for several national-scale

epidemiological studies. The retrospective models developed in Chapter 3 are also being used to explore whether air pollution is a risk factor for other cancer sites in the NECSS case-control study (e.g. breast and prostate). The development of the radon concentration maps also address an important shortcoming to radon population health prevention in Canada – as there were no radon maps currently available. The maps presented in this dissertation are already being used to target population health and radon prevention programs. From a population health perspective, areas of Canada that have much greater radon risk could be the target of focused radon prevention programs, as geographical targeting could increase the impact and cost-effectiveness of such programs. Policies could also use the long-term neighborhood exposure methods developed in this dissertation to identify and address neighborhoods that have experienced consistent deprivation over the last 20-30 years. In addition, smoking cessation programs could be tailored and targeted to these persistently low SES neighborhoods of Canada.

7.2. Limitations

A number of limitations need to be highlighted to ensure that the results of this dissertation are properly interpreted and to stress further research needed in this field. Specific limitations were presented in each research Chapter and only overarching limitations associated with the NECSS case-control study design and the spatial exposure assessment methods conducted are discussed here.

7.2.1. Study Design

The NECSS case-control study has been used throughout this dissertation. The NECSS was designed specifically to examine environmental cancer risk factors, and therefore has a

number of unique characteristics compared to other studies. However, several limitations need to be taken into account when interpreting the epidemiological findings.

The provincially based population-based study design is both a strength and weakness of the NECSS study, particularly when exposures are assessed using GIS and spatial methods. Due to residential mobility, study participants were located in all areas of Canada (even though Quebec did not participate in the study), thus requiring national-scale exposure assessment methods. A large portion of the study population was also located in Southern Ontario, specifically in and around the city of Toronto and any response bias or exposure assessment error in this geographic area would have a large influence on the epidemiological results. Extensive sensitivity analyses were conducted and regional random effects were incorporated into all epidemiological models to account for any residual regional differences.

An inherent weakness of the case-control study design is response and recall bias. While the response rate for lung cancer cases (61.7%) and population controls (67%) was relatively high for a population-based case-control study, response bias cannot be completely ruled out. The fact that different provinces used different methods to recruit population controls (including random digit dialing, health insurance databases, and property assessment databases) is also a weakness, as these methods may have resulted in different response rates and spatial differences in the control population. Because exposures were derived using spatial exposure assessment methods this is an important limitation to highlight, and one that is rarely examined in epidemiological studies. To address this limitation, we conducted a thorough examination of residential histories, included study province and regional random effects in all models, and found consistent epidemiological results between provinces.

Recall bias is less of a concern in this study as we used objective exposure assessment methods derived in a GIS; however, these exposures were based on self-reported residential

histories. Recall bias was present for self-reported residential histories prior to 1975, with cases recalling older residences more often than controls. We therefore restricted the exposure period in our study to 1975-1994, where there was no significant difference (p=0.54) between the number of geocoded residential-years between cases and controls. Self-selection of residential location is also of concern specifically for the neighborhood analysis and our results may not necessarily be causal as residents may choose the neighborhoods in which they live on the basis of health-related characteristics; however, we took a step towards addressing this problem by examining 20 years of residential histories.

Although this was one of the largest studies of air pollution and lung cancer (in terms of the number of lung cancer cases), we were unable to examine risks for never-smokers due to the small percentage of cases (6%), which is similar to other air pollution and lung cancer studies (e.g. Lepeule et al., 2012). The strong influence of smoking on lung cancer limits study power for environmental exposure analyses, as seen in the relatively large confidence intervals in some epidemiological results presented in this dissertation; however, this limitation is inherent in all non-smoking lung cancer analyses.

Finally, the influence of residual confounding from important individual and geographic characteristics cannot be ruled out. The magnitude of any residual confounding is likely small, however, as we were able to control for a comprehensive set of individual variables and used multiple measures of smoking histories (i.e. smoking pack-years, years since quite smoking, and residential and occupational second hand smoke exposure).

7.2.2. Spatial exposure assessments

Comprehensive long-term spatiotemporal exposure assessments were conducted for air pollution, radon and neighborhood SES; nevertheless, potential exposure misclassification

remains a limitation of this research. Residential histories were coded using a standard geographic reference of 6-digit postal codes (due to privacy concerns) and postal codes can change over time and their spatial accuracy is limited in rural areas. The primary limitation of the air pollution exposure assessment was the lack of historical data available to assign air pollution concentrations and build (and evaluate) exposure models. The ability of models to capture fine-scale within-city pollution gradients were also limited by the lack of historical GIS data needed for LUR modeling. The primary limitation of the radon exposure assessment was the use of ecological and not household radon measures to estimate individual-level radon exposures and random error in exposure assessment is therefore present, which would have attenuated epidemiological results. The primary limitation of the neighborhood SES exposure assessment was the use of administrative derived boundaries and data to represent neighborhoods and the difference between neighborhoods estimated from CT and CSD boundaries. Measures for specific pathways linking neighborhood social context to lung cancer risk (e.g. tobacco availability, physical activity resources, social norms, community belonging, crime, or local policies and regulations) could also have been evaluated; however, little historical data was available to estimate such measures.

7.3. Recommendations for Further Research

In the preceding section, as well as in individual research chapters, I have highlighted specific limitations as they apply to the overall study design and to the air pollution, radon and neighborhood SES exposure assessments conducted in this dissertation. Here I expand from these specific limitations to examine broad future research directions that are required to enhance our knowledge of air pollution, radon and neighborhood SES, and more generally the

physical and social contextual environments (and their potential synergies) and lung cancer risk.

Foremost is the need for a multi-level life-course approach that integrates exposure to physical and social environments across an individual's life. While this dissertation examined exposures over a twenty-year period, it is likely that air pollution, radon and neighborhood SES have critical period, accumulative, latent, and trajectory impacts on lung cancer. The call for a life-course approach in epidemiology is not new (Ben-Shlomo and Kuh, 2002; Kuh et al., 2003; Lynch and Smith, 2005), yet integrative examples including both the physical and social environments are few. Throughout this dissertation I have conducted parallel studies of different risk factors, and while these have highlighted risks associated with specific exposures, they are unlikely to reveal the true risk profiles associated with these environmental exposures due to exposure assessment limitations, complex causal pathways, risk factor interactions, synergies with biological, behavioral, and psychosocial factors and individual susceptibility. For cancer in general, where established risk factors explain only a relative small proportion of cancers (Boffetta et al., 2008), a more integrated multilevel life-course research approach is required. To help meet this goal, further research is needed on conceptual frameworks and theory; enhancements in exposure assessment; as well as analytical methods.

First, inadequate attention has been paid to conceptual frameworks and theory for a multi-level life-course approach. A lack of theory on specific causal mechanisms linking social context to health outcomes has been referenced in many commentaries as the most pressing issue facing neighborhood context research (Cummins et al., 2007; Macintyre et al., 2002; O'Campo, 2003; Oakes, 2006). Very little theory and few conceptual frameworks exist that integrate both physical and social environmental exposures, and while there is a growing body of research that integrates ambient air pollution and psychosocial stress (Clougherty and

Kubzansky, 2009; Gee and Payne-Sturges, 2004; O'Neill et al., 2007, 2003), this research is limited to one physical and one social exposure. For example, Clougherty and Kubzansky (2009) created a framework for examining social stress and susceptibility to air pollution in respiratory health and recommended careful attention be paid to the relative temporalities of stress and pollution exposures, to non-linearities in their independent and combined effects, to physiologic pathways not elucidated by epidemiologic methods, and to the relative spatial distributions of social and physical exposures at multiple geographic scales.

In terms of broad conceptual frameworks, very few were identified. Schulz and Northridge (2004) provide a framework that incorporates a range of social and physical environmental characteristics and emphasizes the interplay of social processes with features of the physical environment, highlighting macro-level, community-level and inter-personal level pathways. Diez Roux (2003) also examined a broad range of physical and social residential environment and the specific pathways that may influence cardiovascular disease. Extending these comprehensive theoretical frameworks over the life-course is difficult, but an important challenge to future researchers in this area.

Testing multi-level life-course models will also require improved exposure assessments. In particular, there is a need to enhance the specificity of exposure assessments over the life-course, while at the same time reducing exposure misclassification. The specificity of exposure assessments is particularly relevant for neighborhood SES, which is a surrogate for a spectrum of potential pathways influencing health, compared to air pollution and radon epidemiology that examine specific pollutants and risk factors. However, even in air pollution epidemiology much more work is needed to examine the specific sources and components of air pollution that are responsible for lung cancer risk. As mentioned, while I estimated exposures over twenty years in this dissertation, such exposures assessments should

be extended over the entire life-course. These improvements in exposure assessments could come from new spatiotemporal exposure assessment methods or from other integrated exposure assessment approaches including spatial methods, questionnaires, biomarkers and linked administrative data. As an example, to further examine the chronic-stress hypothesis, studies could measure potentially stressful neighborhood characteristics (e.g. neighborhood crime) in childhood, adolescences and adulthood, as well as individual perceptions of neighborhood safety and individual biomarkers of allostatic load. This type of integrated exposure assessment approach would provide further evidence for distal and proximal influences of stress, and specific pathways and mechanisms linking stressful neighborhood environmental to health. Much more methodological research is needed to be able to estimate life-course exposures, which is a major barrier to understanding of the influence of the physical and social environments on health.

Finally, different analytical methods are needed to conduct multi-level life-course research. Advancements are needed to be able to test the different mechanism relevant to this framework (e.g. critical period, accumulative, latent, and trajectory), to account for clustered and correlated data structures, and to examine multiple interactions. Effects of place on health emerge from complex interdependent pathways and traditional epidemiologic study designs and statistical regression approaches have difficulties examining these types of dynamic processes (Auchincloss and Diez-Roux, 2008). In this dissertation, I examined air pollution, radon and neighborhood SES as distinct environmental risk factors, and future research could use network analysis and agent-based models (e.g. El-Sayed et al., 2012), cross-classified multi-level models (e.g. Chandola et al., 2005) or structural equation models (e.g. Cole and Maxwell, 2003) to further examine how these risk factors interact with each other and with individual characteristics.

7.4. Summary

The purpose of this dissertation was to examine whether three exposures associated with the physical and social residential environment – specifically, ambient air pollution, radon and neighborhood SES – are risk factors for lung cancer in Canada. Using a large population-based case-control study, I conducted long-term exposure assessments using twenty years of residential histories and spatiotemporal exposure models. Epidemiological results provided evidence that ambient air pollution, radon and neighborhood SES are all potential risk factors for lung cancer in Canada. Population attributable fraction estimates suggest that up to 7,706 lung cancer cases annually could be eliminated if the effects of these risk factors were removed from the Canadian population.

References

Adamkiewicz, G., Zota, A.R., Fabian, M.P., Chahine, T., Julien, R., Spengler, J.D., Levy, J.I., 2011. Moving environmental justice indoors: Understanding structural influences on residential exposure patterns in low-income communities. *American Journal of Public Health* 101, S238–S245.

Alberg, A.J., Ford, J.G., Samet, J.M., 2007. Epidemiology of lung cancer: ACCP evidence-based clinical practice guidelines (2nd Edition). *CHEST* 132, 29S–55S.

Andersen, C.E., Raaschou-Nielsen, O., Andersen, H.P., Lind, M., Gravesen, P., Thomsen, B.L., Ulbak, K., 2007. Prediction of 222Rn in Danish dwellings using geology and house construction information from central databases. *Radiat Prot Dosimetry* 123, 83–94.

Annear, M.J., Gidlow, B., Cushman, G., 2009. Neighbourhood deprivation and older adults' preferences for and perceptions of active leisure participation. *Annals of Leisure Research* 12, 96–128.

Auchincloss, A.H., Roux, A.V.D., 2008. A new tool for epidemiology: The usefulness of dynamic-agent models in understanding place effects on health. *Am. J. Epidemiol*. 168, 1–8.

Azzam, E.I., Little, J.B., 2004. The radiation-induced bystander effect: evidence and significance. *Hum Exp Toxicol* 23, 61–65.

Band, P.R., Jiang, H., Zielinski, J.M., 2011. Analysis of lung cancer incidence relating to air pollution levels adjusting for cigarette smoking: a case-control study. *Air Pollution XIX*, 445–453.

Barbone, F., Bovenzi, M., Cavallieri, F., Stanta, G., 1995. Air pollution and lung cancer in Trieste, Italy. *American Journal of Epidemiology* 141, 1161 –1169.

Beelen, R., Hoek, G., Brandt, P.A. van den, Goldbohm, R.A., Fischer, P., Schouten, L.J., Armstrong, B., Brunekreef, B., 2008. Long-term exposure to traffic-related air pollution and lung cancer risk. *Epidemiology* 19, 702.

Beelen, R., Hoek, G., Fischer, P., Brandt, P.A., Brunekreef, B., 2007. Estimated long-term outdoor air pollution concentrations in a cohort study. *Atmospheric Environment* 41, 1343–1358.

Beeson, W.L., Abbey, D.E., Knutsen, S.F., 1998. Long-term concentrations of ambient air pollutants and incident lung cancer in California adults: results from the AHSMOG study. *Environ Health Perspect* 106, 813–823.

Ben-Shlomo, Y., Kuh, D., 2002. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology* 31, 285 –293.

Bird, C.E., Seeman, T., Escarce, J.J., Basurto-Dávila, R., Finch, B.K., Dubowitz, T., Heron, M., Hale, L., Merkin, S.S., Weden, M., Lurie, N., 2010. Neighbourhood socioeconomic status and biological "wear and Tear" in a nationally representative sample of US adults. *J Epidemiol Community Health* 64, 860–865.

Black, J.L., Carpiano, R.M., Fleming, S., Lauster, N., 2011. Exploring the distribution of food stores in British Columbia: Associations with neighbourhood socio-demographic factors and urban form. *Health & Place* 17, 961–970.

Bochicchio, F., 2011. The newest international trend about regulation of indoor radon. *Radiat Prot Dosimetry* 146, 2–5.

Boffetta, P., Tubiana, M., Hill, C., Boniol, M., Aurengo, A., Masse, R., Valleron, A.J., Monier, R., 2008. The causes of cancer in France. *Annals of Oncology* 20(3), 550-555.

Boice J.D., 1990. Studies of atomic bomb survivors understanding radiation effects. *JAMA* 264, 622–623.

Boone-Heinonen, J., Diez Roux, A.V., Kiefe, C.I., Lewis, C.E., Guilkey, D.K., Gordon-Larsen, P., 2011. Neighborhood socioeconomic status predictors of physical activity through young to middle adulthood: The CARDIA study. *Social Science & Medicine* 72, 641–649.

Bräuner, E.V., Andersen, C.E., Sørensen, M., Jovanovic Andersen, Z., Gravesen, P., Ulbak, K., Hertel, O., Pedersen, C., Overvad, K., Tjønneland, A., Raaschou-Nielsen, O., 2012. Residential radon and lung cancer incidence in a Danish cohort. *Environmental Research* 118, 130-136.

Braveman P.A., C.C., 2005. Socioeconomic status in health research: One size does not fit all. *JAMA* 294, 2879–2888.

Brunekreef, B., Holgate, S.T., 2002. Air pollution and health. *The Lancet* 360, 1233–1242.

Buzzelli, M., Jerrett, M., 2004. Racial gradients of ambient air pollution exposure in Hamilton, Canada. *Environment and Planning* 36, 1855–1876.

Carpiano, R.M., 2007. Neighborhood social capital and adult health: An empirical test of a Bourdieu-based model. *Health and Place* 13, 639–655.

Carpiano, R.M., 2008. Actual or potential neighborhood resources and access to them: Testing hypotheses of social capital for the health of female caregivers. *Social Science & Medicine* 67, 568–582.

Carson, A.P., Rose, K.M., Catellier, D.J., Kaufman, J.S., Wyatt, S.B., Diez–Roux, A.V., Heiss, G., 2007. Cumulative socioeconomic status across the life course and subclinical atherosclerosis. *Annals of Epidemiology* 17, 296–303.

CCS, 2012. Canadian Cancer Statistic 2012. http://www.cancer.ca/canada-wide/about%20cancer/cancer%20statistics/stats%20at%20a%20glance/lung%20cancer.aspx?sc_lang=en (accessed 6.15.12).

Celik, I., Gallicchio, L., Boyd, K., Lam, T.K., Matanoski, G., Xuguang, T., Shiels, M., Hammond, E., Liwei, C., Robinson, K.A., Caulfield, L.E., Herman, J.G., Guallar, E., Alberg, A.J. 2008. Arsenic in drinking water and lung cancer: A systematic review. *Environmental Research* 108, 48-55.

Chaix, B., Rosvall, M., Lynch, J., Merlo, J., 2006. Disentangling contextual effects on cause-specific mortality in a longitudinal 23-year follow-up study: impact of population density or socioeconomic environment? *International Journal of Epidemiology* 35, 633–643.

Chandola, T., Clarke, P., Wiggins, R., Bartley, M., 2005. Who you live with and where you live: setting the context for health using multiple membership multilevel models. *J Epidemiol Community Health* 59, 170–175.

Chen, F., Jackson, H., Bina, W.F., 2009. Lung adenocarcinoma incidence rates and their relation to motor vehicle density. *Cancer Epidemiology Biomarkers & Prevention* 18, 760 – 764.

Chen, H., Goldberg, M.S., Villeneuve, P.J., 2008. A systematic review of the relation between long-term exposure to ambient air pollution and chronic diseases. *Rev Environ Health* 23, 243–297.

Chen, J., Moir, D., Whyte, J., 2012. Canadian population risk of radon induced lung cancer: A re-assessment based on the recent cross-Canada radon survey. *Radiat Prot Dosimetry* 152, 9-13.

Christopher J Paciorek, J.D.Y., 2009. Practical large-scale spatio-temporal modeling of particulate matter concentrations. *Annals of applied Statistics* 3(1), 370-397.

Clarke, P., Morenoff, J., Debbink, M., Golberstein, E., Elliott, M.R., Lantz, P.M., 2013. Cumulative exposure to neighborhood context: Consequences for health transitions over the adult life course. *Research on Aging*. doi: 10.1177/0164027512470702.

Clougherty, J.E., Kubzansky, L.D., 2009. A framework for examining social stress and susceptibility to air pollution in respiratory health. *Environmental Health Perspectives* 117(9), 1351-1358.

Cohen, A.J., Ross Anderson, H., Ostro, B., Pandey, K.D., Krzyzanowski, M., Künzli, N., Gutschmidt, K., Pope, A., Romieu, I., Samet, J.M., Smith, K., 2005. The global burden of disease due to outdoor air pollution. *Journal of Toxicology and Environmental Health* 68, 1301-1307.

Cole, D.A., Maxwell, S.E., 2003. Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology* 112, 558–577.

Crouse, D.L., Peters, P.A., Van Donkelaar, A., Goldberg, M.S., Villeneuve, P.J., Brion, O., Khan, S., Atari, D.O., Jerrett, M., Pope, C.A., Brauer, M., Brook, J.R., Martin, R.V., Stieb, D., Burnett, R.T., 2012. 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. *Environmental Health Perspectives* 120, 708–714.

Cummins, S., Curtis, S., Diez-Roux, A.V., Macintyre, S., 2007. Understanding and representing "place" in health research: a relational approach. *Social Science & Medicine* 65, 1825–1838.

Curry, S.J., Wagner, E.H., Cheadle, A., Diehr, P., Koepsell, T., Psaty, B., McBride, C., 1993. Assessment of community-level influences on individuals' attitudes about cigarette smoking, alcohol use, and consumption of dietary fat. *Am J Prev Med* 9, 78–84.

Curtis, S., Southall, H., Congdon, P., Dodgeon, B., 2004. Area effects on health variation over the life-course: analysis of the longitudinal study sample in England using new data on area of residence in childhood. *Social Science & Medicine* 58, 57–74.

Danaei, G., Vander Hoorn, S., Lopez, A.D., Murray, C.J., Ezzati, M., 2005. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *The Lancet* 366, 1784–1793.

Darby, S., 2005. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *BMJ* 330, 223–0.

Diez-Roux, A.V., Mair, C., 2010. Neighborhoods and health. *Annals of the New York Academy of Sciences* 1186, 125–145.

Diez-Roux, A., 2003. Residential environments and cardiovascular risk. *J Urban Health* 80, 569–589.

Diez-Roux, A., Merkin, S.S., Arnett, D., Chambless, L., Massing, M., Nieto, F.J., Sorlie, P., Szklo, M., Tyroler, H.A., Watson, R.L., 2001. Neighborhood of residence and incidence of coronary heart sisease. *New England Journal of Medicine* 345, 99–106.

Dockery, D.W., Pope, C.A., 3rd, Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G., Jr, Speizer, F.E., 1993. An association between air pollution and mortality in six U.S. cities. *N. Engl. J. Med.* 329, 1753–1759.

Doll, R., Peto, R., 1981. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *Journal of the National Cancer Institute* 66, 1191–1308.

Donaldson, K., Stone, V., 2003. Current hypotheses on the mechanisms of toxicity of ultrafine particles. *Ann lis Super Sanita* 39, 405–410.

Donaldson, M.S., 2004. Nutrition and cancer: A review of the evidence for an anti-cancer diet. *Nutr J* 3, 1–21.

Donkelaar, A. van, Martin, R.V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., Villeneuve, P.J., 2010. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environ Health Perspect* 118, 847-855.

Doubeni, C.A., Major, J.M., Laiyemo, A.O., Schootman, M., Zauber, A.G., Hollenbeck, A.R., Sinha, R., Allison, J., 2012. Contribution of behavioral risk factors and obesity to socioeconomic differences in colorectal cancer incidence. *JNCI J Natl Cancer Inst* 104, 1353–1362.

Duncan, C., Jones, K., Moon, G., 1999. Smoking and deprivation: are there neighbourhood effects? *Social Science & Medicine* 48, 497–505.

Earle, C.C., Venditti, L.N., Neumann, P.J., Gelber, R.D., Weinstein, M.C., Potosky, A.L., Weeks, J.C., 2000. Who gets chemotherapy for metastatic lung cancer? *Chest* 117, 1239–1246.

Ellaway, A., Macintyre, S., 2009. Are perceived neighbourhood problems associated with the likelihood of smoking? *J Epidemiol Community Health* 63, 78–80.

Elliott, M., 2000. The stress process in neighborhood context. *Health & Place* 6, 287–299.

El-Sayed, A.M., Scarborough, P., Seemann, L., Galea, S., 2012. Social network analysis and agent-based modeling in social epidemiology. *Epidemiologic Perspectives & Innovations* 9, 1-9.

Environment Canada, 2011. Canadian Regional and Hemispheric O3 and NOx System (CHRONOS). http://www.msc-smc.ec.gc.ca/aq_smog/chronos_e.cfm (accessed 02.06.2011).

EPA, 2012. EPA Assessment of risks from radon in homes. nepis.epa.gov/Adobe/PDF/00000ANV.PDF (accessed 10.11.2012).

Evans, G.W., Kantrowitz, E., 2002. Socioeconomic status and health: The potential role of environmental risk exposure. *Annual Review of Public Health* 23, 303–331.

Finch, B.K., Phuong Do, D., Heron, M., Bird, C., Seeman, T., Lurie, N., 2010. Neighborhood effects on health: Concentrated advantage and disadvantage. *Health & Place* 16, 1058–1060.

Freedman, V.A., Grafova, I.B., Rogowski, J., 2011. Neighborhoods and chronic disease onset in later life. *Am J Public Health* 101, 79–86.

Frohlich, K.L., Potvin, L., Chabot, P., Corin, E., 2002. A theoretical and empirical analysis of context: neighbourhoods, smoking and youth. *Social Science & Medicine* 54, 1401–1417.

Gee, G.C., Payne-Sturges, D.C., 2004. Environmental health disparities: a framework integrating psychosocial and environmental concepts. *Environmental Health Perspectives* 112, 1645.

Ghio, A.J., Kim, C., Devlin, R.B., 2000. Concentrated Ambient Air Particles Induce Mild Pulmonary Inflammation in Healthy Human Volunteers. *Am. J. Respir. Crit. Care Med.* 162, 981–988.

Giskes, K., Van Lenthe, F.J., Turrell, G., Brug, J., Mackenbach, J.P., 2006. Smokers living in deprived areas are less likely to quit: a longitudinal follow-up. *Tob Control* 15, 485–488.

Gray, A., Read, S., McGale, P., Darby, S., 2009. Lung cancer deaths from indoor radon and the cost effectiveness and potential of policies to reduce them. *BMJ* 338:a3110.

Gunier, R.B., Hertz, A., Behren, J.V., Reynolds, P., 2003. Traffic density in California: socioeconomic and ethnic differences among potentially exposed children. *Exposure Science and Environmental Epidemiology* 13, 240–246.

Hart, J.E., Garshick, E., Dockery, D.W., Smith, T.J., Ryan, L., Laden, F., 2011. Long-term ambient multipollutant exposures and mortality. *Am. J. Respir. Crit. Care Med.* 183, 73–78.

Hart, J.E., Yanosky, J.D., Puett, R.C., Ryan, L., Dockery, D.W., Smith, T.J., Garshick, E., Laden, F., 2009. Spatial modeling of PM10 and NO2 in the continental United States, 1985–2000. *Environmental Health Perspectives* 117, 1690-1696.

Hauri, D.D., Huss, A., Zimmermann, F., Kuehni, C.E., Röösli, M., 2012. A prediction model for assessing residential radon concentration in Switzerland. *Journal of Environmental Radioactivity* 112, 83–89.

Havard, S., Deguen, S., Zmirou-Navier, D., Schillinger, C., Bard, D., 2009. Traffic-related Air pollution and socioeconomic status: A spatial autocorrelation study to assess environmental equity on a Small-Area Scale. *Epidemiology* 20, 223.

Health Canada, 2012. Radon - Environmental and Workplace Health. http://www.hc-sc.gc.ca/ewh-semt/radiation/radon/index-eng.php (accessed 10.8.12).

Health Canada, 2012 Cross-Canada Survey of Radon Concentrations in Homes - Final Report http://www.hc-sc.gc.ca/ewh-semt/radiation/radon/survey-sondage-eng.php (accessed 11.18.12).

Health Effects Institute, 2010. Traffic-related air pollution: A critical review of the literature on emissions, exposure, and health effects. The Health Effects Institute: Boston, MA USA.

Heid, I.M., Küchenhoff, H., Miles, J., Kreienbrock, L., Wichmann, H.E., 2004. Two dimensions of measurement error: Classical and Berkson error in residential radon exposure assessment. *Journal of Exposure Science and Environmental Epidemiology* 14, 365–377.

Hurley, S.E., Reynolds, P., Goldberg, D.E., Hertz, A., Anton-Culver, H., Bernstein, L., Deapen, D., Peel, D., Pinder, R., Ross, R.K., 2005. Residential mobility in the California Teachers Study: implications for geographic differences in disease rates. *Social Science & Medicine* 60, 1547–1555.

Hystad, P., Carpiano, R.M., 2012. Sense of community-belonging and health-behaviour change in Canada. *J Epidemiol Community Health* 66, 277–283.

Hystad, P., Demers, P.A., Johnson, K.C., Brook, J., Donkelaar, A. van, Lamsal, L., Martin, R., Brauer, M., 2012. Spatiotemporal air pollution exposure assessment for a Canadian population-based lung cancer case-control study. *Environmental Health* 11, 22.

IARC, 2004. IARC Monographsy on the evaluation of carcinogenic risks to humans: Tobacco smoke and Involuntary Smoking, Vol 83. Lyon, France. http://monographs.iarc.fr/

IARC, 2012a. IARC Monographsy on the evaluation of carcinogenic risks to humans: Diesel and Gasoline Engine Exhausts and Some Nitroarenes, Vol 105. Lyon, France. http://monographs.iarc.fr/

IARC, 2012b. IARC Monographsy on the evaluation of carcinogenic risks to humans: Ambient Air Pollution.Upcoming Vol 109. http://monographs.iarc.fr/ENG/Meetings/vol109-callexperts.php

IARC, 2012c. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Manmade Mineral Fibres and Radon, Vol 43. Lyon, France. URL http://monographs.iarc.fr/

Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., Forman, D., 2011. Global cancer statistics. *Cancer J Clin* 61, 69-90.

Jerrett, M., Burnett, R.T., Kanaroglou, P., Eyles, J., Finkelstein, N., Giovis, C., Brook, J.R., 2001. A GIS-environmental justice analysis of particulate air pollution in Hamilton, Canada. *Environment and Planning* 33, 955–974.

Jerrett, M., Burnett, R.T., Ma, R., 3rd, C.A.P., Krewski, D., Newbold, K.B., Thurston, G., Shi, Y., Finkelstein, N., Calle, E.E., Thun, M.J., 2005. Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* 16, 727–736.

Johnson, K.C., Mao, Y., Argo, J., Dubois, S., Semenciw, R., Lava, J., 1998. The national enhanced cancer surveillance system: A case-control approach to environment-related cancer surveillance in Canada. *Environmetrics* 9, 495–504.

Kan, K., 2007. Residential mobility and social capital. *Journal of Urban Economics* 61, 436–457.

Karasek, D., Ahern, J., Galea, S., 2012. Social norms, collective efficacy, and smoking cessation in urban neighborhoods. *Am J Public Health* 102, 343–351.

Katanoda, K., Sobue, T., Satoh, H., Tajima, K., Suzuki, T., Nakatsuka, H., Takezaki, T., Nakayama, T., Nitta, H., Tanabe, K., 2011. An association between long-term exposure to ambient air pollution and mortality from lung cancer and respiratory diseases in Japan. *Epidemiology* 21, 132–143.

Katsouyanni, K., Trichopoulos, D., Kalandidi, A., Tomos, P., Riboli, E., 1991. A case-control study of air pollution and tobacco smoking in lung cancer among women in Athens. *Preventive Medicine* 20, 271–278.

Kawachi, I., Subramanian, S.V., 2007. Neighbourhood influences on health. *Journal of Epidemiology and Community Health* 61, 3-4.

Kawakami, N., Li, X., Sundquist, K., 2011. Health-promoting and health-damaging neighbourhood resources and coronary heart disease: a follow-up study of 2,165,000 people. *Journal of Epidemiology and Community Health* 10, 866-872.

Kim, D., Masyn, K.E., Kawachi, I., Laden, F., Colditz, G.A., 2010. Neighborhood socioeconomic status and behavioral pathways to risks of colon and rectal cancer in women. *Cancer* 116, 4187–4196.

Knaapen, A.M., Borm, P.J.A., Albrecht, C., Schins, R.P.F., 2004. Inhaled particles and lung cancer. Part A: Mechanisms. *International Journal of Cancer* 109, 799–809.

Krewski, D., Lubin, J.H., Zielinski, J.M., Alavanja, M., Catalan, V.S., Field, R.W., Klotz, J.B., Létourneau, E.G., Lynch, C.F., Lyon, J.I., 2005. Residential radon and risk of lung cancer: A combined analysis of 7 North American case-control studies. *Epidemiology* 16, 137.

Kuh, D., Ben-Shlomo, Y., Lynch, J., Hallqvist, J., Power, C., 2003. Life course epidemiology. *J Epidemiol Community Health* 57, 778–783.

Laden, F., Schwartz, J., Speizer, F.E., Dockery, D.W., 2006. Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard six cities study. *Am. J. Respir. Crit. Care Med*. 173, 667-672.

Lall, R., Kendall, M., Ito, K., Thurston, G.D., 2004. Estimation of historical annual $PM_{2.5}$ exposures for health effects assessment. *Atmospheric Environment* 38, 5217–5226.

Lamsal, L.N., Martin, R.V., Donkelaar, A. van, Steinbacher, M., Celarier, E.A., Bucsela, E., Dunlea, E.J., Pinto, J.P., 2008. Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument. *J.Geophys.Res* 113, D16308.

Lantz, P.M., Mendez, D., Philbert, M.A., 2013. Radon, smoking, and lung cancer: The need to refocus radon control policy. *American Journal of Public Health* 113, 443-447.

Lepeule, J., Laden, F., Dockery, D., Schwartz, J., 2012. Chronic exposure to fine particles and mortality: An extended follow-up of the Harvard Six Cities study from 1974 to 2009. *Environ. Health Perspect*. 120, 965–970.

Létourneau, E.G., Krewski, D., Choi, N.W., Goddard, M.J., Mcgregor, R.G., Zielinski, J.M., Du, J., 1994. Case-control study of residential radon and lung cancer in Winnipeg, Manitoba, Canada. *Am. J. Epidemiol*. 140, 310–322.

Leyland, A.H., Næss, Ø., 2009. The effect of area of residence over the life course on subsequent mortality. *Journal of the Royal Statistical Society* 172, 555–578.

Liaw, Y.P., Ting, T.F., Ho, C., Chiou, Z., 2010. Cell type specificity of lung cancer associated with nitric oxide. *Science of the Total Environment* 408, 4931–4934.

Lim, S.S., Vos, T., Flaxman, A.D., Danaei, G., Shibuya, K., et al., 2012. 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. *The Lancet* 380, 2224–2260.

Lim, W.-Y., Seow, A., 2012. Biomass fuels and lung cancer. *Respirology* 17, 20–31.

Lloyd, J.E.V., Li, L., Hertzman, C., 2010. Early experiences matter: Lasting effect of concentrated disadvantage on children's language and cognitive outcomes. *Health & Place* 16, 371–380.

López-Cima, M.F., García-Pérez, J., Pérez-Gómez, B., Aragonés, N., López-Abente, G., Tardón, A., Pollán, M., 2011. Lung cancer risk and pollution in an industrial region of Northern Spain: a hospital-based case-control study. *Int J Health Geogr* 10, 1-10.

Lubin, J.H., Boice, J.D., Edling, C., Hornung, R.W., Howe, G.R., Kunz, E., Kusiak, R.A., Morrison, H.I., Radford, E.P., Samet, J.M., Tirmarche, M., Woodward, A., Yao, S.X., Pierce, D.A., 1995. Lung cancer in radon-exposed miners and estimation of risk from indoor exposure. *JNCI J Natl Cancer Inst* 87, 817–827.

Lubin, J.H., Wang, Z.Y., Boice, J.D., Jr, Xu, Z.Y., Blot, W.J., De Wang, L., Kleinerman, R.A., 2004. Risk of lung cancer and residential radon in China: pooled results of two studies. *Int. J. Cancer* 109, 132–137.

Lynch, J., Smith, G.D., 2005. A life course approach to chronic disease epidemiology. *Annual Review of Public Health* 26, 1–35.

Macintyre, S., Ellaway, A., Cummins, S., 2002. Place effects on health: how can we conceptualise, operationalise and measure them? *Social Science & Medicine* 55, 125–139.

Major, J.M., Doubeni, C.A., Freedman, N.D., Park, Y., Lian, M., Hollenbeck, A.R., Schatzkin, A., Graubard, B.I., Sinha, R., 2010. Neighborhood socioeconomic deprivation and mortality: NIH-AARP Diet and Health Study. *PLoS ONE* 5, e15538.

Mao, Y., Hu, J., Ugnat, A.M., Semenciw, R., Fincham, S., 2001. Socioeconomic status and lung cancer risk in Canada. *International Journal of Epidemiology* 30, 809–817.

Mao, Y., Pan, S., Wen, S.W., Johnson, K.C., 2003. Physical activity and the risk of lung cancer in Canada. *Am. J. Epidemiol*. 158, 564–575.

Martikainen, P., Kauppinen, T., Valkonen, T., 2003. Effects of the characteristics of neighbourhoods and the characteristics of people on cause specific mortality: a register based follow up study of 252 000 men. *J Epidemiol Community Health* 57, 210–217.

Matheson, F.I., Moineddin, R., Dunn, J.R., Creatore, M.I., Gozdyra, P., Glazier, R.H., 2006. Urban neighborhoods, chronic stress, gender and depression. *Social Science & Medicine* 63, 2604–2616.

Matukala Nkosi, T., Parent, M.É., Siemiatycki, J., Rousseau, M.C., 2012. Socioeconomic Position and Lung Cancer Risk. *Epidemiology* 23, 377–385.

Meijer, M., Bloomfield, K., Engholm, G., 2012a. Neighbourhoods matter too: the association between neighbourhood socioeconomic position, population density and breast, prostate and lung cancer incidence in Denmark between 2004 and 2008. *J Epidemiol Community Health* 67, 6-13.

Meijer, M., Röhl, J., Bloomfield, K., Grittner, U., 2012b. Do neighborhoods affect individual mortality? A systematic review and meta-analysis of multilevel studies. *Social Science & Medicine* 74, 1204–1212.

Menvielle, G., Boshuizen, H., Kunst, A.E., Vineis, P., Dalton, S.O., Bergmann, M.M., Hermann, S., Veglia, F., Ferrari, P., Overvad, K., Raaschou-Nielsen, O., Tjønneland, A., Kaaks, R., Linseisen, J., Palli, D., Krogh, V., Tumino, R., Rodriguez, L., Agudo, A., Sánchez, M.-J., Arozena, J.M.A., Cirera, L., Ardanaz, E., Bingham, S., Khaw, K.-T., Boffetta, P., Duell, E., Slimani, N., Gallo, V., Riboli, E., Bueno-de-Mesquita, H.B., 2010. Occupational exposures contribute to educational inequalities in lung cancer incidence among men: Evidence from the EPIC prospective cohort study. *International Journal of Cancer* 126, 1928–1935.

Miles, R., 2006. Neighborhood disorder and smoking: Findings of a European urban survey. *Social Science & Medicine* 63, 2464–2475.

Mohai, P., Pellow, D., Roberts, J.T., 2009. Environmental Justice. *Annual Reviews of Environment* 34, 405-430.

Morello-Frosch, R., Jesdale, B.M., 2006. Separate and unequal: residential segregation and estimated cancer risks associated with ambient air toxics in US metropolitan areas. *Environmental Health Perspectives* 114, 386.

Murray, E.T., Diez Roux, A.V., Carnethon, M., Lutsey, P.L., Ni, H., O'Meara, E.S., 2010. Trajectories of neighborhood poverty and associations with subclinical atherosclerosis and associated risk factors. *American Journal of Epidemiology* 171, 1099-1108.

Naess, O., Claussen, B., Smith, G.D., Leyland, A.H., 2008. Life course influence of residential area on cause-specific mortality. *J Epidemiol Community Health* 62, 29-34.

Nafstad, P., Håheim, L.L., Oftedal, B., Gram, F., Holme, I., Hjermann, I., Leren, P., 2003. Lung cancer and air pollution: a 27 year follow up of 16 209 Norwegian men. *Thorax* 58, 1071-1076.

Nagy, E., Zeisig, M., Kawamura, K., Hisamatsu, Y., Sugeta, A., Adachi, S., Möller, L., 2005. DNA adduct and tumor formations in rats after intratracheal administration of the urban air pollutant 3-nitrobenzanthrone. *Carcinogenesis* 26, 1821–1828.

Narayanan, P.K., Goodwin, E.H., Lehnert, B.E., 1997. α particles initiate biological production of superoxide anions and hydrogen peroxide in human cells. *Cancer Res* 57, 3963–3971.

Neumann, C.M., Forman, D.L., Rothlein, J.E., 1998. Hazard screening of chemical releases and environmental equity analysis of populations proximate to toxic release inventory facilities in Oregon. *Environmental Health Perspectives* 106, 217.

Nyberg, F., Gustavsson, P., Jarup, L., Bellander, T., Berglind, N., Jakobsson, R., Pershagen, G., 2000. Urban air pollution and lung cancer in Stockholm. *Epidemiology*, 487–495.

O'Campo, P., 2003. Invited commentary: advancing theory and methods for multilevel models of residential neighborhoods and health. *American Journal of Epidemiology* 157, 1-9.

O'Neill, M.S., Jerrett, M., Kawachi, I., Levy, J.I., Cohen, A.J., Gouveia, N., Wilkinson, P., Fletcher, T., Cifuentes, L., Schwartz, J., 2003. Health, wealth, and air pollution: advancing theory and methods. *Environmental Health Perspectives* 111, 1861.

O'Neill, M.S., McMichael, A.J., Schwartz, J., Wartenberg, D., 2007. Poverty, environment, and health: the role of environmental epidemiology and environmental epidemiologists. *Epidemiology* 18, 664.

Oakes, J.M., 2006. Commentary: Advancing neighbourhood-effects research-selection, inferential support, and structural confounding. *International Journal of Epidemiology* 35, 643–647.

Ohlsson, H., Merlo, J., 2011. Place effects for areas defined by administrative boundaries: A life course analysis of mortality and cause specific morbidity in Scania, Sweden. *Social Science &* Medicine 73, 1145–1151.

Parkin, D.M., Bray, F., Ferlay, J., Pisani, P., 2005. Global cancer statistics, 2002. *Cancer J Clin* 55, 74–108.

Parkin, D.M., Pisani, P., Lopez, A.D., Masuyer, E., 1994. At least one in seven cases of cancer is caused by smoking. Global estimates for 1985. International Journal of Cancer 59, 494–504.

Pearce, J., Barnett, R., Moon, G., 2012. Sociospatial inequalities in health-related behaviours Pathways linking place and smoking. *Prog Hum Geogr* 36, 3–24.

Pearce, J.R., Dorling, D., 2010. The influence of selective migration patterns among smokers and nonsmokers on geographical inequalities in health. *Annals of the Association of American Geographers* 100, 393–408.

Perlin, S.A., Wong, D., Sexton, K., 2001. Residential proximity to industrial sources of air pollution: interrelationships among race, poverty, and age. *Journal of the Air & Waste Management Association* 51, 406–421.

Pesch, B., Kendzia, B., Gustavsson, P., Jöckel, K., Johnen, G., Pohlabeln, H., Olsson, A., Ahrens, W., Gross, I.M., Brüske, I., Wichmann, H., Merletti, F., Richiardi, L., Simonato, L., Fortes, C., Siemiatycki, J., Parent, M., Consonni, D., Landi, M.T., Caporaso, N., Zaridze, D., Cassidy, A., Szeszenia-Dabrowska, N., Rudnai, P., Lissowska, J., Stücker, I., Fabianova, E., Dumitru, R.S., Bencko, V., Foretova, L., Janout, V., Rudin, C.M., Brennan, P., Boffetta, P., Straif, K., Brüning, T., 2012. Cigarette smoking and lung cancer—relative risk estimates for the major histological types from a pooled analysis of case–control studies. *International Journal of Cancer* 131m 1210-1219,

Pickett, K.E., Pearl, M., 2001. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J Epidemiol Community Health* 55, 111–122.

Pope, C.A., 3rd, Burnett, R.T., Turner, M.C., Cohen, A., Krewski, D., Jerrett, M., Gapstur, S.M., Thun, M.J., 2011. Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: shape of the exposure-response relationships. *Environ. Health Perspect*. 119, 1616–1621.

Pope, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., Thurston, G.D., 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* 287, 1132–1141.

Raaschou-Nielsen, O., Andersen, Z., Hvidberg, M., Jensen, S.S., Ketzel, M., Sorensen, M., Loft, S., Overvad, K., Tjonneland, A., 2011. Lung cancer incidence and long-term exposure to aAir Pollution from Traffic. *Environmental Health Perspectives* 119, 860-865.

Raaschou-Nielsen, O., Bak, H., Sørensen, M., Jensen, S.S., Ketzel, M., Hvidberg, M., Schnohr, P., Tjønneland, A., Overvad, K., Loft, S., 2010. Air Pollution from Traffic and Risk for Lung Cancer in Three Danish Cohorts. Cancer Epidemiology Biomarkers & Prevention 19, 1284–1291.

Radon Environmental. Radon Potential Map of Canada. http://www.radoncorp.com/mapping.html. (accessed 22.01.13).

Rahman, I., MacNee, W., 2000. Oxidative stress and regulation of glutathione in lung inflammation. *European Respiratory Journal* 16, 534–554.

Reiche, E.M.V., Nunes, S.O.V., Morimoto, H.K., 2004. Stress, depression, the immune system, and cancer. *The Lancet Oncology* 5, 617–625.

Reitzel, L.R., Nguyen, N., Zafereo, M.E., Li, G., Wei, Q., Sturgis, E.M., 2012. Neighborhood deprivation and clinical outcomes among head and neck cancer patients. *Health & Place* 18, 861–868.

Robert, S.A., Strombom, I., Trentham-Dietz, A., Hampton, J.M., McElroy, J.A., Newcomb, P.A., Remington, P.L., 2004. Socioeconomic risk factors for breast cancer: Distinguishing individual-and community-level effects. *Epidemiology* 15, 442.

Ross, C.E., Mirowsky, J., 2008. Neighborhood socioeconomic status and health: Context or composition? *City & Community* 7, 163–179.

Samet, J.M., Avila-Tang, E., Boffetta, P., Hannan, L.M., Olivo-Marston, S., Thun, M.J., Rudin, C.M., 2009. Lung cancer in never smokers: Clinical epidemiology and environmental risk factors. *Clin Cancer Res* 15, 5626–5645.

Sampson, R.J., Morenoff, J.D., Gannon-Rowley, T., 2002. Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology* 28, 443–478.

Sampson, R.J., Sharkey, P., Raudenbush, S.W., 2008. Durable effects of concentrated disadvantage on verbal ability among African-American children. *PNAS* 105, 845–852.

Sanderson, M., Coker, A.L., Perez, A., Du, X.L., Peltz, G., Fadden, M., 2006. A multilevel analysis of socioeconomic status and prostate cancer risk. *Annals of Epidemiology* 16, 901–907.

Sandhu, M.S., White, I.R., McPherson, K., 2001. Systematic review of the prospective cohort studies on meat consumption and colorectal cancer risk: A meta-analytical approach. *Cancer Epidemiol Biomarkers Prev* 10, 439–446.

Schulz, A., Northridge, M.E., 2004. Social determinants of health: Implications for environmental health promotion. *Health Education & Behavior* 31, 455–471.

Sethi, T.K., El-Ghamry, M.N., Kloecker, G.H., 2012. Radon and lung cancer. *Clin Adv Hematol Oncol* 10, 157–164.

Sidorchuk, A., Agardh, E.E., Aremu, O., Hallqvist, J., Allebeck, P., Moradi, T., 2009. Socioeconomic differences in lung cancer incidence: a systematic review and meta-analysis. *Cancer Causes and Control* 20, 459–471.

Siemiatycki, J., Richardson, L., Straif, K., Latreille, B., Lakhani, R., Campbell, S., Rousseau, M.-C., Boffetta, P., 2004. Listing occupational carcinogens. *Environmental Health Perspectives 112*, 1447–1459.

Sisti, J., Boffetta, P., 2012. What proportion of lung cancer in never-smokers can be attributed to known risk factors? *International Journal of Cancer* 131, 265–275.

Statistics Canada, 2010. Households and the environment survey, knowledge of radon and testing, Canada and provinces http://www5.statcan.gc.ca/cansim/a05?lang=eng&id=1530098#F3 (accessed 6.19.12).

Statistics Canada, 2012a. Census Dictionary - Census Tracts http://www12.statcan.ca/census-recensement/2011/ref/dict/geo013-eng.cfm (accessed 9.25.12).

Statistics Canada, 2012b. Postal Code Conversion File Plus. http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?lang=eng&catno=82F0086X (accessed 10.4.12).

Stidley, C.A., Samet, J.M., 1994. Assessment of ecologic regression in the study of lung cancer and indoor radon. *Am. J. Epidemiol.* 139, 312–322.

Sundquist, J., Li, X., Sundquist, K., 2012. Neighborhood deprivation and mortality in individuals with cancer. *European Journal of Cancer Prevention* 21, 387–394.

Tollefsen, T., Gruber, V., Bossew, P., Cort, M.D., 2011. Status of the European indoor radon map. *Radiat Prot Dosimetry* 145, 110–116.

Turner, M.C., Krewski, D., Chen, Y., Pope, C.A., Gapstur, S., Thun, M.J., 2011a. Radon and lung cancer in the American Cancer Society Cohort. *Cancer Epidemiol Biomarkers Prev* 20, 438–448.

Turner, M.C., Krewski, D., Pope, C.A., Chen, Y., Gapstur, S.M., Thun, M.J., 2011b. Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers. *American Journal of Respiratory and Critical Care Medicine* 184, 1374–1381.

Turrell, G., Hewitt, B.A., Miller, S.A., 2012. The influence of neighbourhood disadvantage on smoking cessation and its contribution to inequalities in smoking status. *Drug and Alcohol Review* 31, 645–652.

Tyczynski, J.E., Bray, F., Parkin, D.M., 2003. Lung cancer in Europe in 2000: Epidemiology, prevention, and early detection. *Lancet Oncol.* 4, 45–55.

USGS 2012. United States Geological Survey; 995. http://energy.cr.usgs.gov/radon/rnus.html. (accessed 11.15.12).

Villeneuve, P.J., Goldberg, M.S., Burnett, R.T., Donkelaar, A. van, Chen, H., Martin, R.V., 2011a. Associations between cigarette smoking, obesity, sociodemographic characteristics and remote-sensing-derived estimates of ambient $PM_{2.5}$: Results from a Canadian population-based survey. *Occup Environ Med* 68, 920–927.

Villeneuve, P.J., Parent, M.-É., Sahni, V., Johnson, K.C., 2011b. Occupational exposure to diesel and gasoline emissions and lung cancer in Canadian men. *Environmental Research* 111, 727–735.

Vineis, P., Hoek, G., Krzyzanowski, M., Vigna-Taglianti, F., Veglia, F., Airoldi, L., Autrup, H., Dunning, A., Garte, S., Hainaut, P., Malaveille, C., Matullo, G., Overvad, K., Raaschou-Nielsen, O., Clavel-Chapelon, F., Linseisen, J., Boeing, H., Trichopoulou, A., Palli, D., Peluso, M., Krogh, V., Tumino, R., Panico, S., Bueno-De-Mesquita, H.B., Peeters, P.H., Lund, E.E., Gonzalez, C.A., Martinez, C., Dorronsoro, M., Barricarte, A., Cirera, L., Quiros, J.R., Berglund, G., Forsberg, B., Day, N.E., Key, T.J., Saracci, R., Kaaks, R., Riboli, E., 2006. Air pollution and risk of lung cancer in a prospective study in Europe. *International Journal of Cancer* 119, 169–174.

Waitzman, N.J., Smith, K.R., 1998. Phantom of the area: poverty-area residence and mortality in the United States. *Am J Public Health* 88, 973–976.

Webster, T.F., Hoffman, K., Weinberg, J., Vieira, V., Aschengrau, A., 2008. Community-and individual-level socioeconomic status and breast cancer risk: Multilevel modeling on Cape Cod, Massachusetts. *Environmental Health Perspectives* 116, 1125–1129.

Weden, M.M., Carpiano, R.M., Robert, S.A., 2008. Subjective and objective neighborhood characteristics and adult health. *Social Science & Medicine* 66, 1256–1270.

Wheaton, B., Clarke, P., 2003. Space Meets Time: Integrating temporal and contextual influences on mental health in early adulthood. *American Sociological Review* 68, 680–706.

Wight, R.G., Cummings, J.R., Karlamangla, A.S., Aneshensel, C.S., 2009. Urban neighborhood context and change in depressive symptoms in late life. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 64, 247–251.

Wilcox, H.B., Al-Zoughool, M., Garner, M.J., Jiang, H., Klotz, J.B., Krewski, D., Nicholson, W.J., Schoenberg, J.B., Villeneuve, P.J., Zielinski, J.M., 2008. Case-control study of radon and lung cancer in New Jersey. *Radiat Prot Dosimetry* 128, 169–179.

Yanosky, J.D., Paciorek, C.J., Schwartz, J., Laden, F., Puett, R., Suh, H.H., 2008. Spatio-temporal modeling of chronic PM_{10} exposure for the Nurses' Health Study. *Atmospheric Environment* 42, 4047–4062.

Yen, I.H., Michael, Y.L., Perdue, L., 2009. Neighborhood environment in studies of health of older adults: A Systematic review. *American Journal of Preventive Medicine* 37, 455–463.

Yorifuji, T., Kashima, S., Tsuda, T., Takao, S., Suzuki, E., Doi, H., Sugiyama, M., Ishikawa-Takata, K., Ohta, T., 2010. Long-term exposure to traffic-related air pollution and mortality in Shizuoka, Japan. *Occupational and Environmental Medicine* 67, 111 –117.

Appendix 1

Table 1. Summary statistics of monitoring data collected from previous city-specific LUR monitoring in Canada and used to evaluate national NO_2 and benzene models.

Substance[a]	Year	N	Mean +/-SD	Min	Max
$NO_2(\mu g/m^3)$					
Edmonton[b]	2008	50	28.90 (5.88)	17.09	42.96
Montreal[c]	05/06	135	22.00 (5.77)	7.44	35.81
Sarnia[d]	2005	35	19.66 (6.28)	2.39	31.41
Toronto[d,e]	04/06	196	24.20 (8.46)	8.98	52.18
Victoria[b]	2006	40	9.49 (4.87)	0.75	19.18
Vancouver[b]	2003	114	30.08 (7.76)	14.49	52.64
Winnipeg[b]	2008	49	16.06 (5.43)	4.25	32.98
Benzene($\mu g/m^3$)					
Montreal[c]	05/06	135	1.05 (0.44)	0.39	3.35
Sarnia[d]	2005	37	0.93 (0.56)	0.28	3.36
Toronto[d]	2006	45	0.75 (0.54	0.40	4.10
Winnipeg[b]	2008	94	0.44 (0.26)	0.08	1.12

[a]Each city conducted monitoring for 2 week periods.
[b]Fixed site monitoring was used to adjust to yearly average.
[c]Average of 3 seasons of monitoring (Dec., May., Aug.) used to capture yearly average.
[d]No yearly adjustment conducted (Sarnia data collected in October and Toronto in September).
[e]Combined 2004 and 2006 data for evaluation.

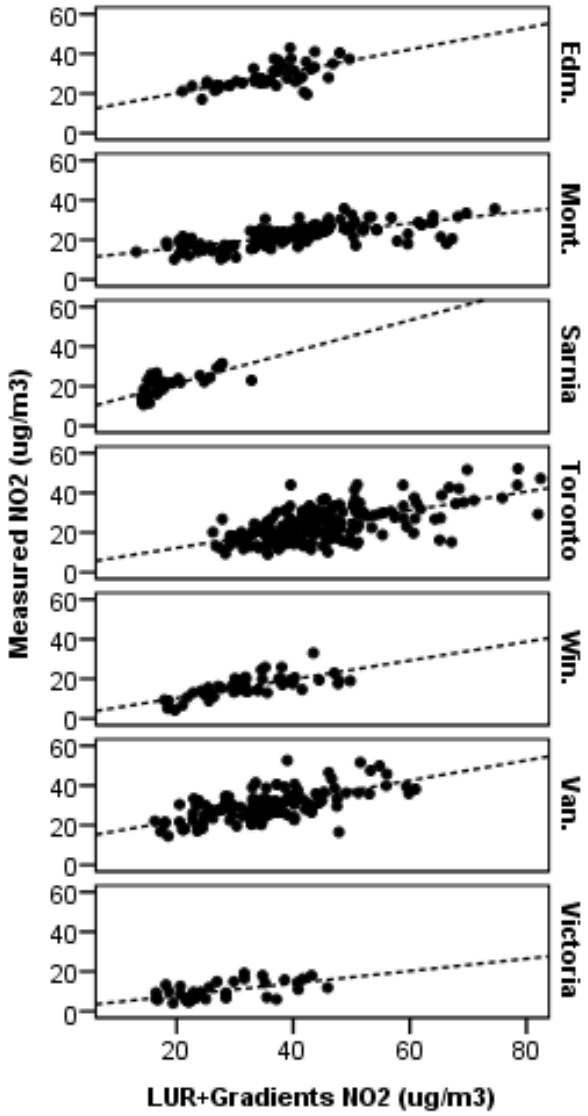

Figure 1. Evaluation of national NO_2 model, incorporating satellite data, geographic landuse variables and deterministic gradients, with independent within-city measurements.

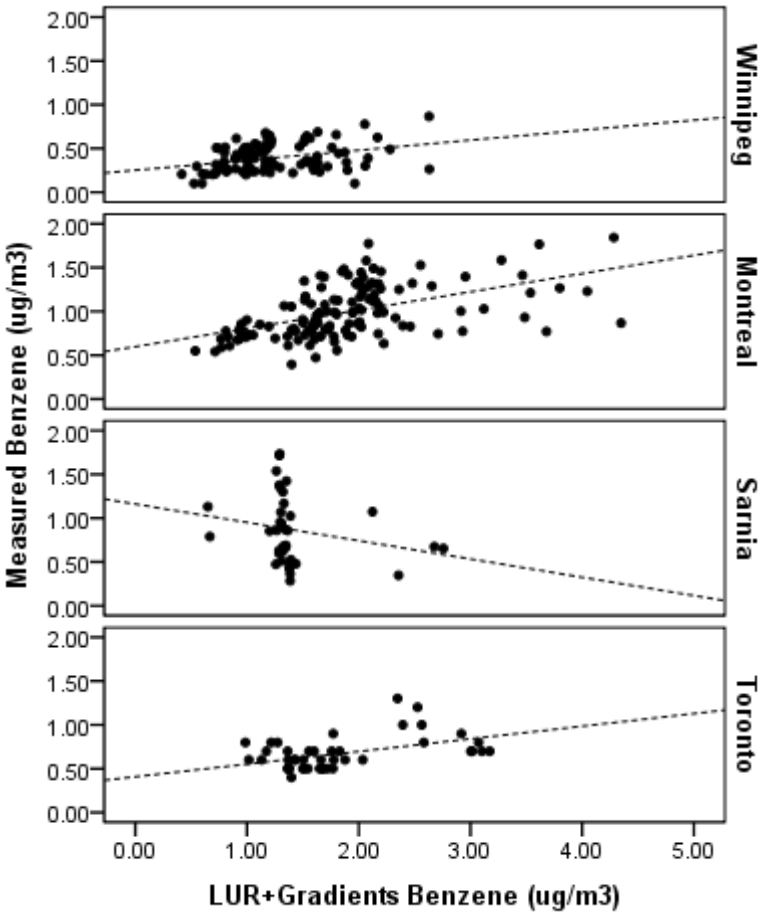

Figure 2. Evaluation of national Benzene model, incorporating geographic landuse variables and deterministic gradients, with independent within-city measurements.

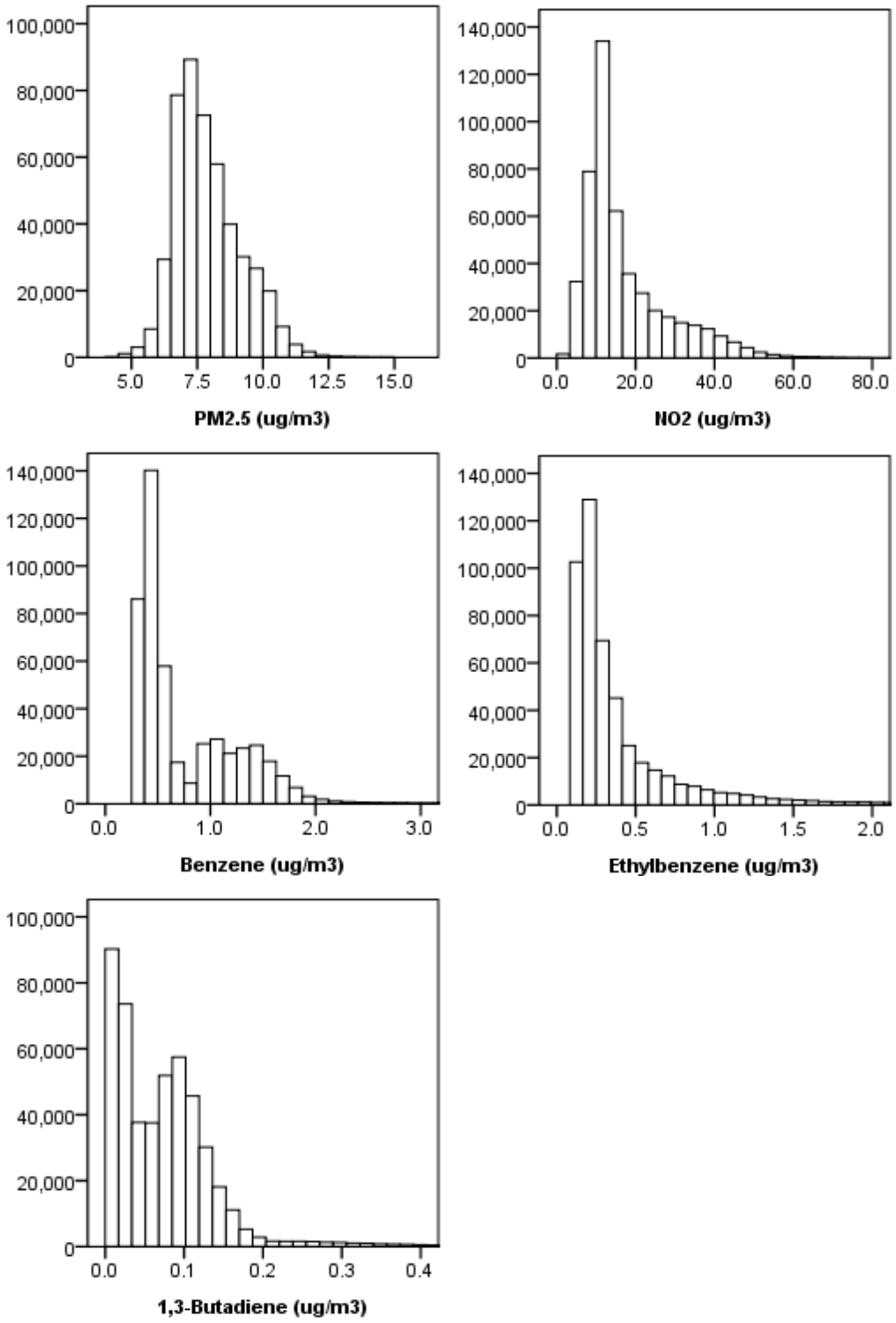

Figure 3. Annual 2006 Canadian population exposure estimates from national LUR plus gradient models (Frequencies represent street block points, each containing approximately 89 individuals).

Appendix 2

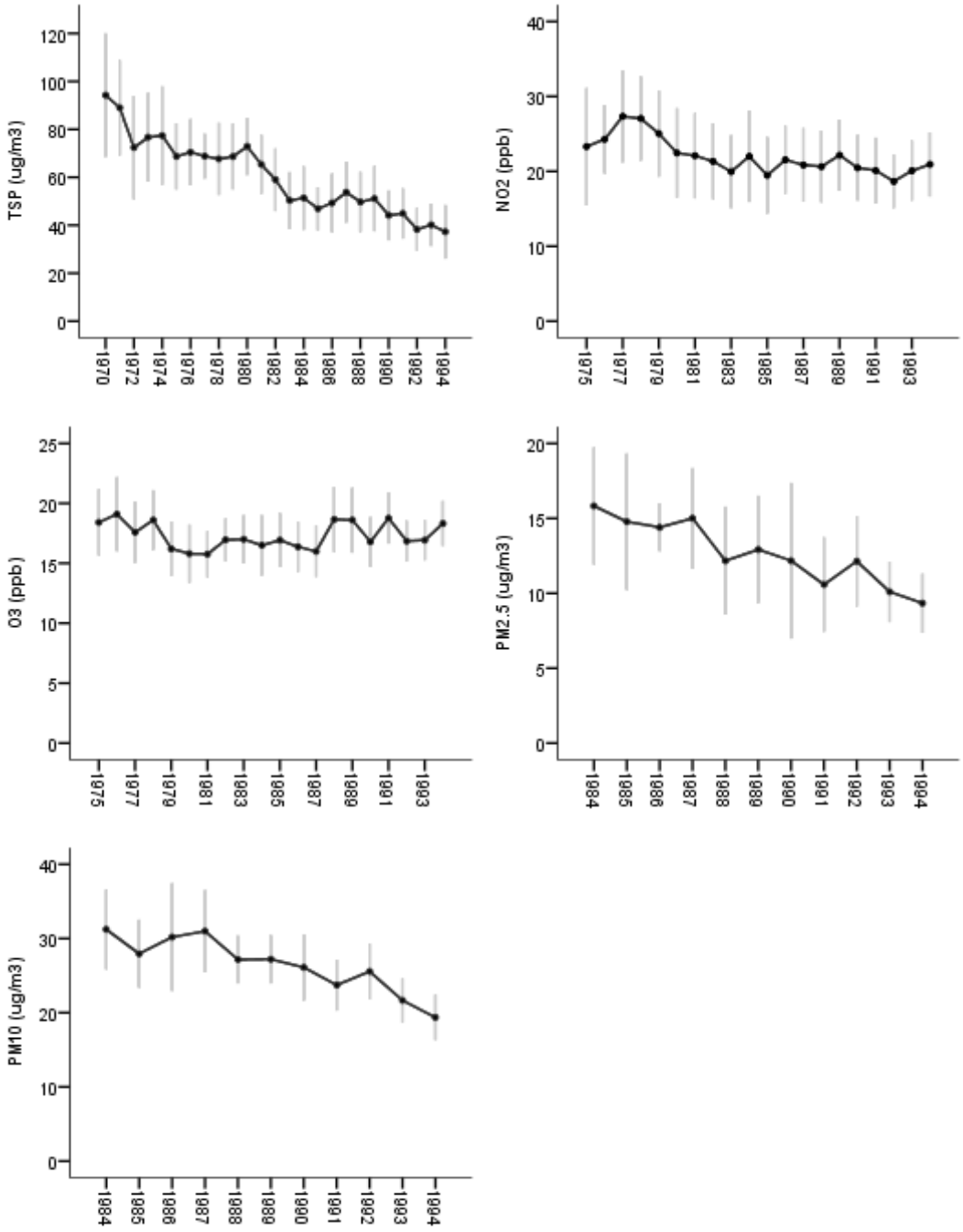

Figure 1. Annual average (SD) pollutant concentrations from all valid historical NAPS monitoring stations that were operating for the entire study period.

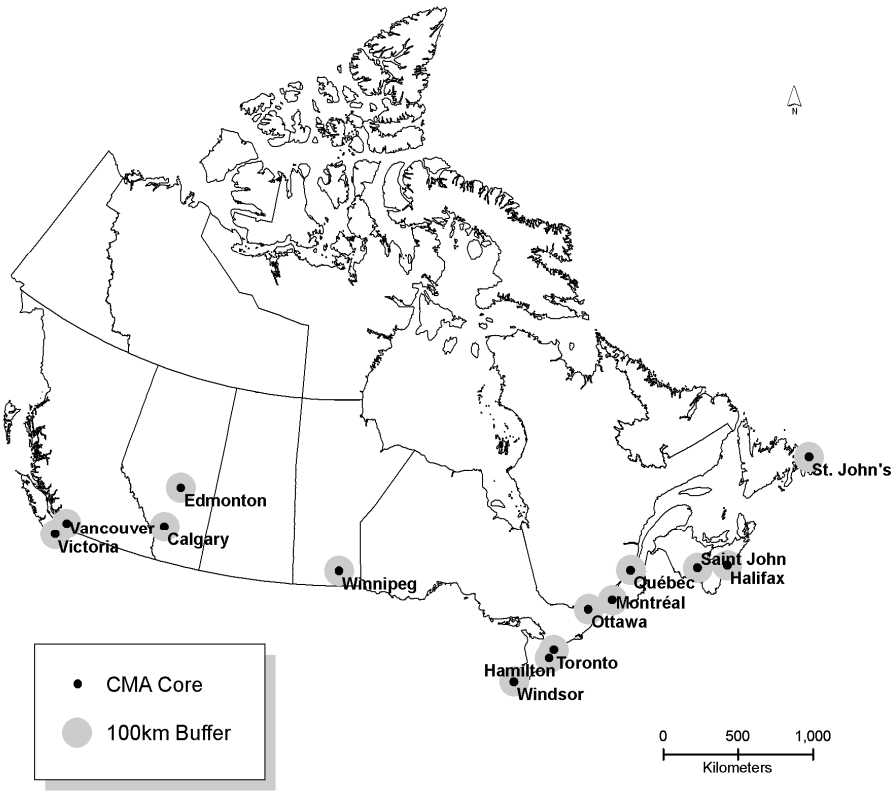

Figure 2. Census Metropolitan Areas (CMA's) in Canada with $PM_{2.5}$ and TSP measurements used to create predictive models of historical $PM_{2.5}$ concentrations.

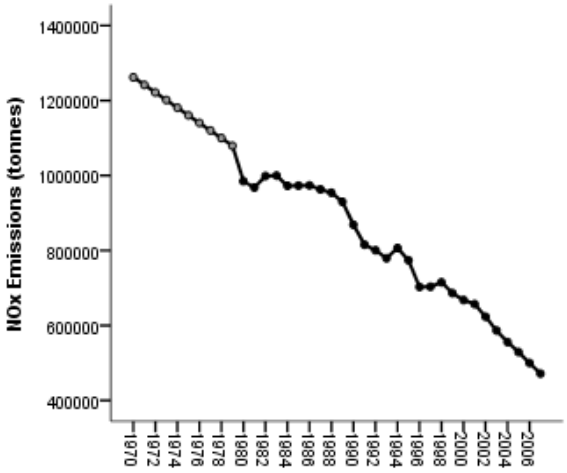

Figure 3. Yearly NO_x on-road mobile emissions in Canada from 1980 to 2007 and extrapolated levels to 1970.

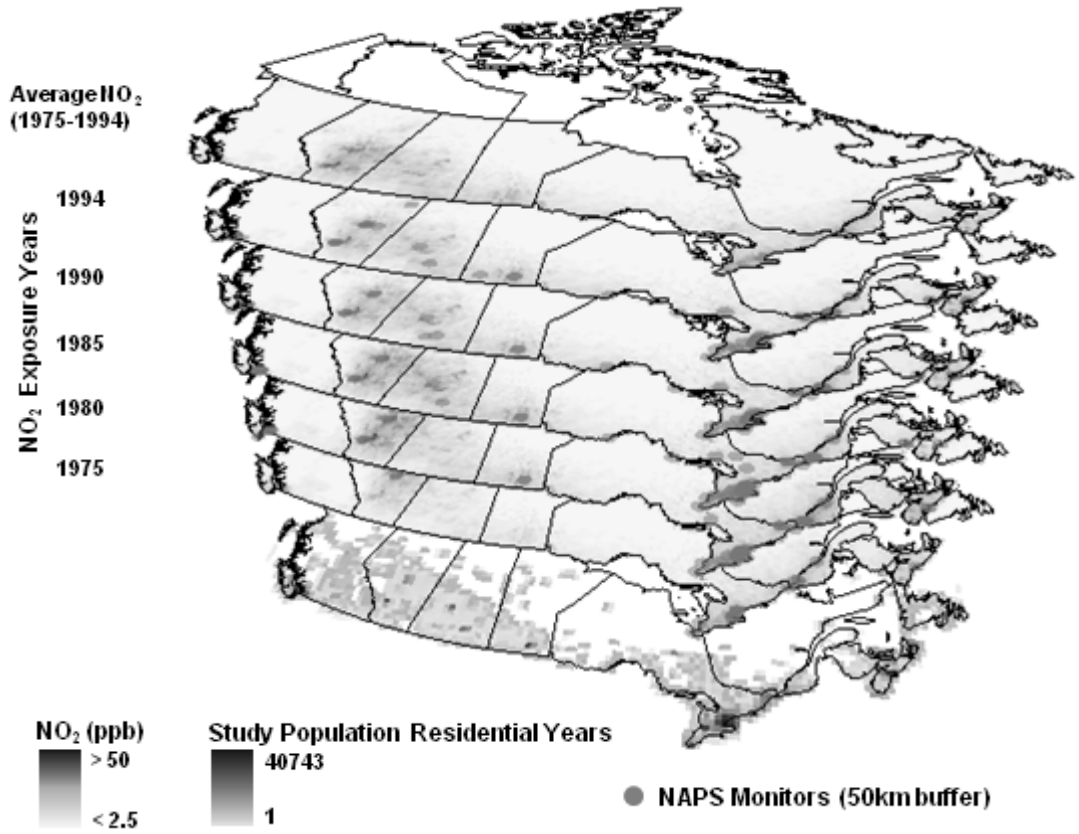

Figure 4. NO_2 exposure surfaces (note: 20 annual surfaces were created but only 5 are shown here) and locations of NAPS monitors with 50km buffers. The study population residential years represents all residential locations between 1975 and 1994 summed within a 50km grid.

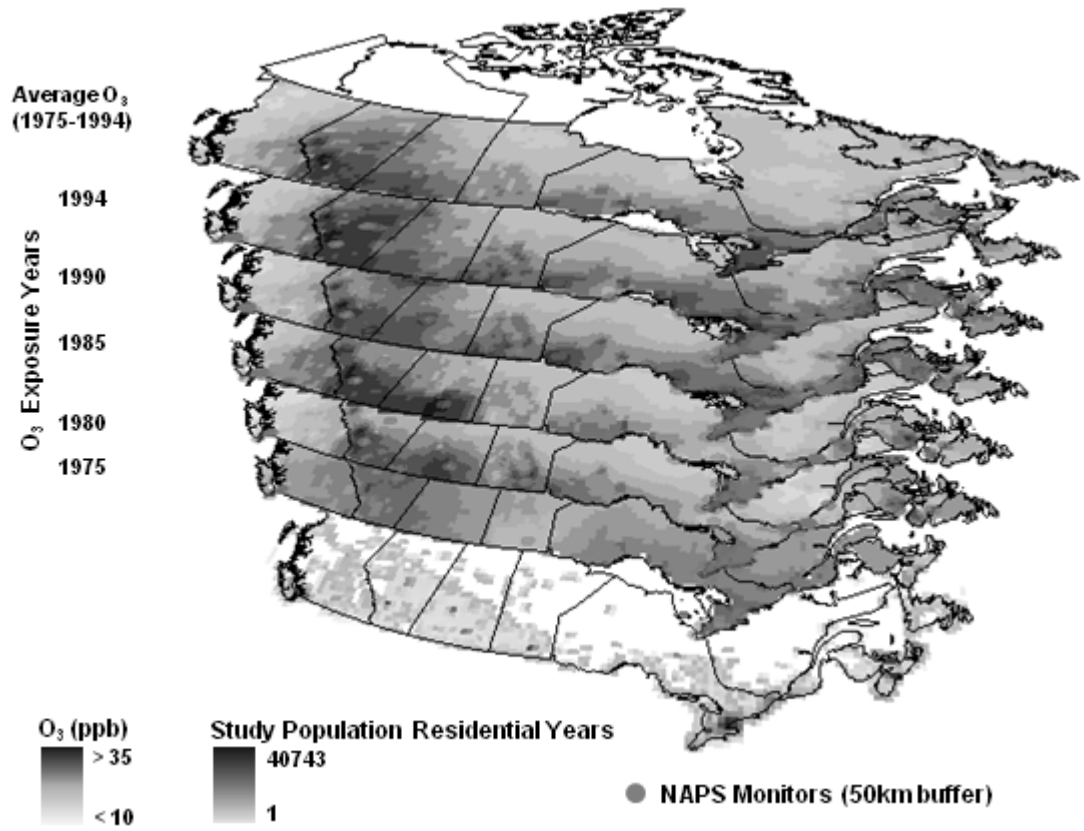

Figure 5. O_3 exposure surfaces (note: 20 annual surfaces were created but only 5 are shown here) and locations of NAPS monitors with 50km buffers. The study population residential years represents all residential locations between 1975 and 1994 summed within a 50km grid.

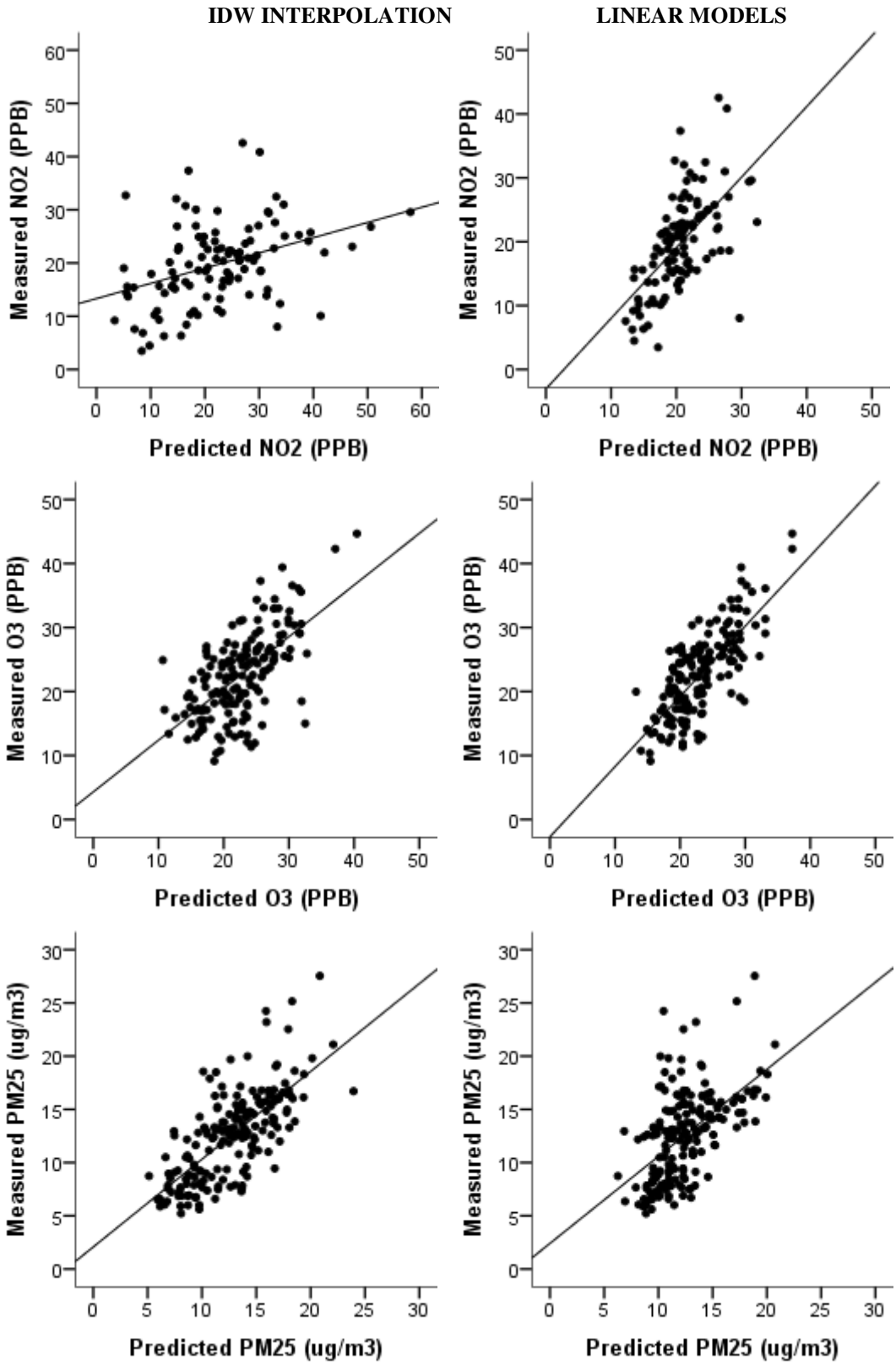

Figure 6. Scatter plots of measured versus predicted $PM_{2.5}$, NO_2 and O_3 for IDW interpolation and linear regression models.

Appendix 3

Table 1. Descriptive statistics of individual-level variables for study participants with 20 years of complete residential histories between 1975 and 1994.

Variable	Cases (n=2390)	Controls (n=3507)
Individual-Level Variables		
Age (Mean +/- SD)	63.5 (8.2)	59.0 (12.6)
Sex		
Female	1152 (48%)	1719 (49%)
Male	1238 (52%)	1788 (51%)
Education		
< High school	1379 (58%)	1514 (43%)
High school	406 (17%)	607 (17%)
> High school	590 (25%)	1373 (39%)
Smoking Status		
Never Smoking	130 (6%)	1337 (38%)
Former Smoker	969 (41%)	1446 (41%)
Current Smoker	1288 (54%)	718 (2%)
Age Started Smoking		
Non Smoker	130 (6%)	1337 (38%)
<15	894 (38%)	765 (22%)
15-19	958 (40%)	925 (26%)
20-24	293 (12%)	312 (9%)
≥25	112 (5%)	162 (5%)
Smoking Pack Years		
Non Smoker	130 (6%)	1337 (38%)
1 - 19	319 (14%)	1169 (34%)
20 -29	467 (20%)	392 (11%)
30 - 39	519 (22%)	247 (7%)
40 - 49	446 (19%)	149 (4%)
50 - 59	205 (9%)	69 (2%)
≥ 60	235 (10%)	79 (2%)
Years Since Cessation of Smoking		
Non Smoker	130 (6%)	1337 (38%)
>35	29 (1%)	177 (5%)
26-35	70 (3%)	312 (9%)
16-25	158 (7%)	383 (11%)
11-15	168 (7%)	223 (6%)
6-10	268 (11%)	208 (6%)
2-5	276 (12%)	143 (4%)

Current Smoker	1273 (54%)	715 (20%)
Median Household Income		
> $100,000	47 (2%)	137 (4%)
$50,000 - $99,999	283 (12%)	630 (18%)
$30,000 - 49,000	474 (20%)	840 (24%)
$20,000 - 29,999	398 (17%)	548 (16%)
$10,000 - 19,999	366 (15%)	363 (10%)
< $10,000	133 (6%)	100 (3%)
Prefer not to Report	689 (29%)	889 (25%)
Alcohol (servings/week)		
0	898 (38%)	1311 (37%)
1 - 2	426 (18%)	814 (23%)
3 - 5	223 (9%)	378 (11%)
5 - 10	342 (14%)	520 (15%)
> 10	501 (21%)	484 (14%)
Meat (servings/week)		
≤ 2	161 (7%)	391 (11%
3 - 5	462 (19%)	808 (23%)
6 - 10	925 (39%)	1310 (37%)
10 - 15	456 (19%)	595 (17%)
> 15	386 (16%)	403 (12%)
Residential Second Hand Smoke Exposure (exposure-years)[a]		
0	356 (15%)	978 (28%)
1 - 24	289 (12%)	664 (19%)
25 - 49	523 (22%)	781 (22%)
50 - 74	527 (22%)	602 (17%)
> 74	695 (29%)	482 (14%)
Occupational Second Hand Smoke Exposure (exposure-years)[b]		
0	750 (32%)	1244 (36%)
1 - 29	306 (13%)	620 (18%)
30 - 69	343 (14%)	577 (17%)
70 - 139	451 (19%)	506 (15%)
> 140	532 (22%)	549 (16%)
Years Working with Occupational Dust and Odors[c]		
0	1144 (48%)	1842 (53%)
1 - 9	287 (12%)	506 (15%)
10 - 19	205 (9%)	308 (9%)
20 - 29	237 (10%)	283 (8%)
> 29	514 (22%)	559 (16%)
Years Working with Hazardous Substances[d]		
0	1557 (65%)	2453 (70%)

1 - 9	200 (8%)	295 (8%)
10 - 19	132 (6%)	206 (6%)
20 - 39	234 (10%)	259 (7%)
> 40	267 (11%)	294 (8%)

* Summaries for individuals with 20 years of complete residential histories between 1975 and 1994.
[a] Person-years defined by the number of smokers in home multiplied by number of residential years.
[b] Person-years defined by the number of smokers in the immediate work environment multiplied by number of residential years.
[c] Self-reported daily or weekly exposure to industrial dusts and odors at work.
[d] Hazards include: arsenic, asbestos, asphalt, benzene, benzedine, cadmium salts, chromium salts, isopropyl oil, vinyl chloride, mustard gas, radiation sources, welding, wood dust.

Table 2. Descriptive statistics of geographic-level variables for study participants with 20 years of complete residential histories between 1975 and 1994.

Variable	**Cases (n=2390)**	**Controls (n=3507)**
Study Province		
Newfoundland	101 (4%)	213 (6%)
Prince Edward Island	60 (3%)	174 (5%)
Nova Scotia	314 (13%)	458 (13%)
Ontario	838 (35%)	1296 (37%)
Manitoba	157 (7%)	236 (7%
Saskatchewan	141 (6%)	195 (6%)
Alberta	325 (14%)	411 (12%)
British Columbia	454 (19%)	524 (15%)
Urban Size Category[a]		
>500,000	783 (33%)	1139 (33%)
100,000-499,999	376 (16%)	516 (15%)
30,000-99,999	217 (9%)	316 (9%)
1,000-29,999	441 (19%)	688 (20%)
<1,000	573 (24%)	848 (24%)
Average Health Region Indoor Radon Measurements (bq/m^3) (Mean +/- SD)	81.3 (41)	78.6 (39.5)
Neighborhood Characteristics (Years in most deprived relative Quintile)[b]		
Neighborhood Household Median Income		
0	1584 (67%)	2467 (71%)
>0-2	165 (7%)	240 (7%)
>2 - 5	144 (6%)	233 (7%)
>5 - 10	201 (9%)	259 (7%)
>10	279 (12%)	287 (8%)
% No high school diploma		
0	1555 (66%)	2408 (69%)
>0-2	189 (8%)	257 (7%)

>2 - 5	158 (6%)	214 (6%)
>5 - 10	186 (8%)	277 (8%)
>10	285 (12%)	330 (10%)
% Old Dwellings		
0	1565 (66%)	2245 (64%)
>0-5	277 (12%)	413 (12%)
>5 - 10	164 (7%)	309 (9%)
>10 - 15	173 (7%)	235 (7%)
>15	194 (8%)	284 (8%)

* Summaries for individuals with 20 years of complete residential histories between 1975 and 1994.
[a] Not included in the final multivariate models due to the high correlation with NO_2.

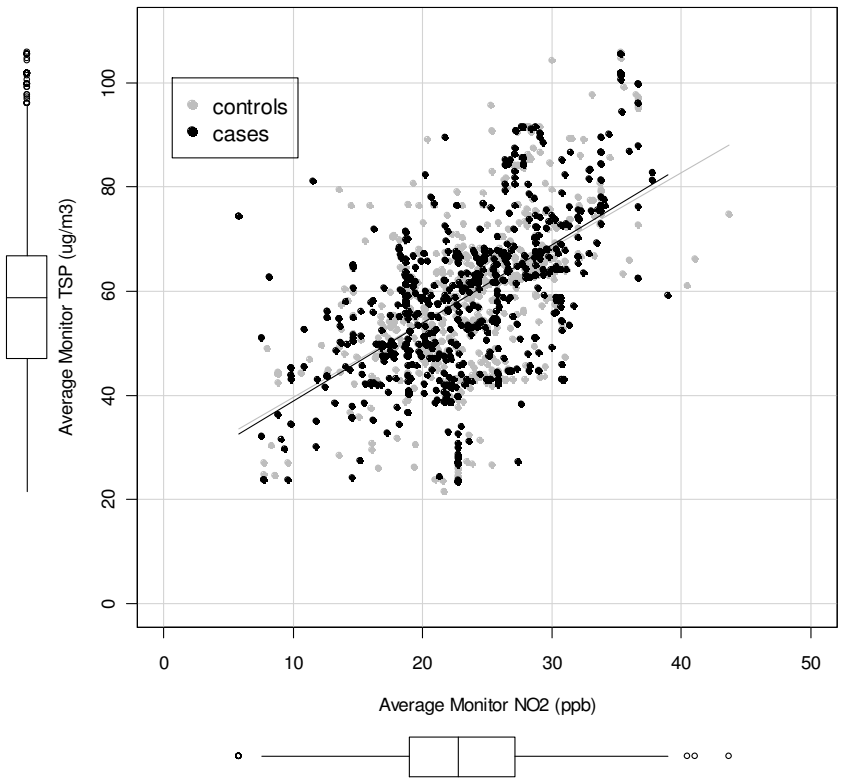

Figure 1. Relationship between monitor based NO_2 and TSP concentrations in the urban monitor bases analysis ($PM_{2.5}$ measurements were only available starting in 1984 and had limited coverage).

Appendix 4

PAFs were calculated for $PM_{2.5,}$ NO_2 and neighborhood SES to compare their impact on lung cancer in Canada. An existing PAF estimate was used that has already been developed from the residential radon measurement data mapped in Chapter 4 (Chen et al., 2012). O_3 estimates were not included as more research is needed to determine whether the association with lung cancer is causal. To estimate the PAF for continuous exposures ($PM_{2.5}$ and NO_2 air pollution), Equation 1 was used, while for discreet exposures (neighborhood SES) Equation 2 was used.

Equation 1:
$$\mathrm{PAF\,(cont.)} = \frac{\int_{x=0}^{m} \mathrm{RR}(x)\mathrm{P}(x)dx - \int_{x=0}^{m} \mathrm{RR}(x)\mathrm{P}'(x)dx}{\int_{x=0}^{m} \mathrm{RR}(x)\mathrm{P}(x)dx}$$

where RR(x) is the relative risk at exposure level x; P(x) is the population distribution of exposure; P'(x) is the counterfactual distribution of exposure; m is the maximum exposure level.

Equation 2:
$$\mathrm{PAF\,(dich.)} = \frac{P(RR-1)}{1+P(RR-1)}$$

where P is the prevalence of the dichotomous exposure; and RR is the relative risk.

The parameters used for each model are shown in Table 1, including risk estimates derived in this dissertation as well as those from existing meta-analyses (for comparison). The current and counterfactual exposure prevalence estimates are also shown, with the current

population exposure estimates to $PM_{2.5}$ and NO_2 estimated from the national LUR models presented in Chapter 2 and the counterfactual concentrations determined using the lower bound of annual NAPS monitoring data. The current radon exposure estimates and counterfactual, as well as the resulting PAF shown in Figure 1, are from (Chen et al., 2012). Figure 1 illustrates the resulting PAFs for lung cancer. The combined PAF is 30.4% translating to an estimated 7,706 lung cancer cases annually that could be eliminated if the effects of these risk factors were removed from the Canadian population.

Table 1. Parameters used in equations 1 and 2 to estimate population attributable fractions.

Risk Factor	**Estimated Risk**		**Current Exposures (mean ±SD)**	**Counterfactual Exposures (mean ±SD)**
	This Study	**Meta-Analyses**		
$PM_{2.5}$ (per 10ug/m^3)	29%	21%[b]	8.4 (1.4)[c]	2.5 (1.0)[d]
NO_2 (per 10ug/m^3)	11%	11%[b]	23.4 (11.5)[c]	10 (5.0)[d]
Radon (per 100 Bq/m^3)	13%	18%[e]	41.9 (2.8)[f]	0 (0.0)
Neighborhood SES (Q5 vs Q1)[a]	36%	n.a	36%	0%

[a] National analysis estimate adjusted for age, sex, study province, education and household income.
[b] Pooled estimate from meta-analyses (Chen et al. 2008).
[c] Estimated using national LUR exposure models presented in Chapter 2.
[d] Correspond to lower-bound NAPS annual measurements.
[e] Meta-analyses from Drewski et al., (2005).
[f] Population weighted geometric mean and standard deviation.

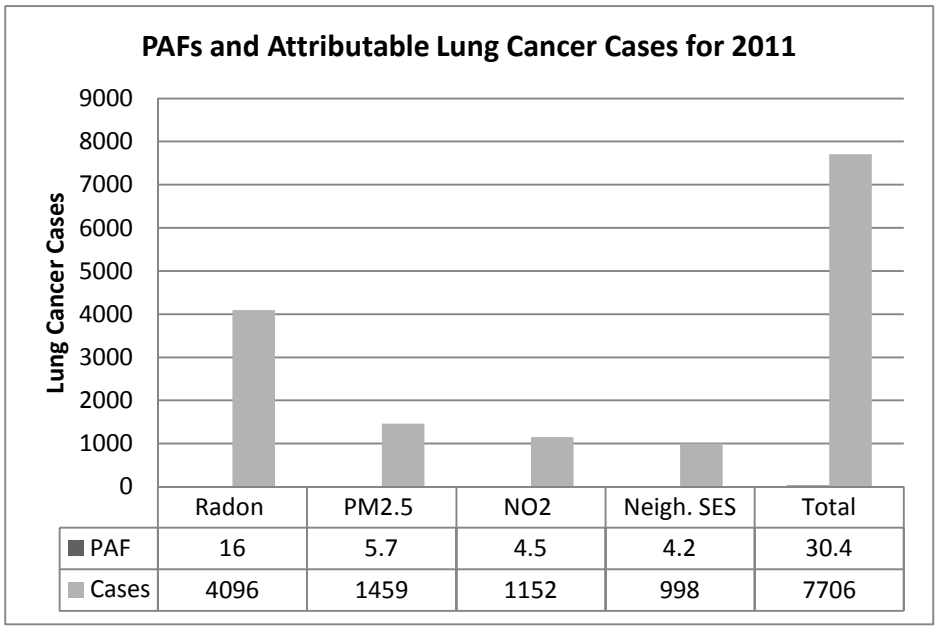

Figure 1. Summary of PAFs and attributable lung cancer cases during 2011 for the different environmental risk factors examined in this dissertation.