

**Associations Between Prenatal Exposure to  
Air Pollutants and Autism Spectrum Disorder:  
A Population-Based Cohort Study in  
Metro Vancouver, Canada**

by  
**Lief Pagalan**

B.Sc., Simon Fraser University, 2016

B.F.A., OCAD University, 2009

Thesis Submitted in Partial Fulfillment of the  
Requirements for the Degree of  
Master of Science

in the  
Master of Science Program  
Faculty of Health Sciences

© Lief Pagalan 2018  
SIMON FRASER UNIVERSITY  
Summer 2018

# Approval

**Name:** Lief Pagalan

**Degree:** Master of Science (Health Sciences)

**Title:** Associations Between Prenatal Exposure to Air Pollutants and Autism Spectrum Disorder: A Population-Based Cohort Study in Metro Vancouver, Canada

**Examining Committee:**

**Chair:** Bohdan Nosyk  
Associate Professor

**Meghan Winters**  
Senior Supervisor  
Associate Professor

**Bruce Lanphear**  
Co-Supervisor  
Professor

**Michael Brauer**  
Supervisor  
Professor  
School of Population and Public Health  
Faculty of Medicine  
University of British Columbia

**Ryan Allen**  
External Examiner  
Associate Professor

**Date Defended/Approved:** August 1, 2018

## Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

- a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

- b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

- c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library  
Burnaby, British Columbia, Canada

Update Spring 2016

## Abstract

The etiology of autism spectrum disorder (ASD) is poorly understood, but emerging evidence suggests that environmental factors may have a larger contributing role to the risk of ASD than previously thought. Recent studies suggest an association between air pollution and ASD. We conducted a retrospective, population-based cohort study of nearly all births in Metro Vancouver, Canada, from 2004–2009 to evaluate prenatal exposures to PM<sub>2.5</sub>, NO, and NO<sub>2</sub> as modifiable environmental risk factors for ASD. Children were diagnosed with ASD using a standardized assessment of Autism Diagnostic Interview-Revised and Autism Diagnostic Observation Schedule. Using temporally adjusted land use regression models, we estimated monthly mean ambient air pollution exposures during pregnancy for each person at their home address. We observed positive associations between air pollutant exposures and increased risk of ASD, especially for males, in a population-based cohort living in an area with relatively low levels of air pollution.

**Keywords:** autism spectrum disorder; air pollution; fine particulate matter; nitric oxide; nitrogen dioxide

## **Acknowledgements**

I wish to express my gratitude to my supervisory committee, Dr. Meghan Winters, Dr. Bruce Lanphear, and Dr. Michael Brauer, for their guidance and support. Your mentorship has enabled me to succeed in my graduate studies and helped me become a better researcher.

I would also like to thank the team at the BC Children's Hospital Research Institute and the Sunny Hill Health Centre for Children, Celeste Bickford, Dr. Whitney Weikum, Dr. Gillian E. Hanley, Dr. Nancy Lanphear, and Dr. Tim F. Oberlander, for their expert knowledge on autism spectrum disorder and assistance navigating the intricacies of population linked data. And to Dr. Hind Sbihi, Dr. Lawrence McCandless, and Dr. Hui Xie for their methodological and biostatistical insights.

My work was facilitated by funding provided by the Canadian Institutes of Health Research, the Fraser Basin Council BC Clean Air Research Fund, the Sunny Hill Health Centre for Children, and the Djavad Mowafaghian Foundation.

# Table of Contents

Approval.....	ii
Ethics Statement.....	iii
Abstract.....	iv
Acknowledgements.....	v
Table of Contents.....	vi
List of Acronyms .....	viii
<b>Chapter 1. Introduction .....</b>	<b>1</b>
1.1. Autism Spectrum Disorder .....	1
1.1.1. ASD Prevalence .....	2
1.2. Environmental Risk Factors for ASD.....	3
1.2.1. Air Pollution and Neurodevelopment.....	4
1.2.2. Air Pollution and ASD .....	4
1.3. Research Context .....	5
1.3.1. Linked Population Health Data .....	6
1.3.2. ASD Diagnoses in British Columbia .....	6
ADOS and ADI-R .....	7
1.3.3. Air Pollution in Metro Vancouver .....	9
Metro Vancouver Air Quality Monitoring Network .....	9
Metro Vancouver Land Use Regression Models .....	10
Temporal Stability of Land Use Regression Models .....	10
1.4. Aims and Objectives .....	11
1.5. References.....	13
<b>Chapter 2. Prenatal Exposure to Air Pollution and ASD .....</b>	<b>19</b>
2.1. Introduction .....	19
2.2. Methods .....	20
2.2.1. Population and Study Design .....	20
2.2.2. Case Ascertainment.....	21
2.2.3. Exposure Assessment.....	22
2.2.4. Statistical Analyses .....	22
2.3. Results .....	23
2.4. Discussion.....	24
2.5. Conclusion .....	27
2.6. Tables and Figures .....	28
2.7. Disclaimer .....	31
2.8. References.....	32
<b>Chapter 3. Conclusion.....</b>	<b>37</b>
3.1. Summary.....	37
3.2. Limitations and Future Work .....	38
3.2.1. Area vs Individual-Level Air Pollution Exposures .....	38

3.2.2.	Timing of Exposure and Critical Windows of Vulnerability .....	39
3.2.3.	Case Definition for ASD .....	40
3.2.4.	Phenotypes and Etiological Heterogeneity.....	41
3.2.5.	Gene-Environment Interactions.....	42
3.3.	Conclusion and Policy Implications.....	42
3.4.	References.....	44
<b>Appendix A.</b>	<b>Chapter 2 Manuscript Abstract .....</b>	<b>47</b>
<b>Appendix B.</b>	<b>Chapter 2 Supplemental Tables and Figures .....</b>	<b>48</b>
<b>Appendix C.</b>	<b>Temporally Adjusted LUR Model Estimates: Protocols .....</b>	<b>57</b>
<b>Appendix D.</b>	<b>Temporally Adjusted LUR Model Estimates: R Scripts .....</b>	<b>67</b>
Data Sources .....		67
Land Use Regression Models .....		67
Metro Vancouver Air Quality Monitoring Network .....		67
Postal Code Conversion Files.....		68
GIS Boundary Files.....		69
Disclaimer .....		69
R Scripts .....		69
01_process_pccf.R .....		69
02_process_pccf_retired.R .....		72
03_subset_pc_within_van.R .....		74
04_get_pollutant_at_pc.R .....		76
05_get_pollutant_in_time.R .....		80
06_temporally_adjust_pollutant.R.....		88
07_trim_temporally_adjusted_models.R.....		104
08_prep_models_for_popdata.R.....		104
09_export_final_models.R .....		107
<b>Appendix D.</b>	<b>Temporally Adjusted LUR Model Estimates: Data File .....</b>	<b>108</b>

## List of Acronyms

ADI-R	Autism Diagnostic Interview-Revised
ADOS	Autism Diagnostic Observation Schedule
ASD	Autism spectrum disorder
BC	British Columbia
BCAAN	BC Autism Assessment Network
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
IQR	Interquartile range
LUR	Land use regression
MCFD	Ministry of Children and Family Development
NASS	National Autism Spectrum Disorder Surveillance System
NO	Nitric oxide
NO <sub>2</sub>	Nitrogen dioxide
NO <sub>x</sub>	Nitrogen oxides
OR	Odds ratio
PM <sub>10</sub>	Particulate matter with a diameter of less than 10 micrometers
PM <sub>2.5</sub>	Particulate matter with a diameter of less than 2.5 micrometers
ppb	Parts per billion
SES	Socioeconomic status
µg/m <sup>3</sup>	Micrograms per cubic meter

# Chapter 1.

## Introduction

The objective of this thesis is to investigate whether prenatal exposures to ambient air pollution—specifically particulate matter with a diameter smaller than 2.5 micrometres (PM<sub>2.5</sub>), nitric oxide (NO), and nitrogen dioxide (NO<sub>2</sub>)—are potential modifiable environmental risk factors for autism spectrum disorder (ASD) in Metro Vancouver, Canada. Emerging evidence suggests that environmental factors may have a greater contributing role to the risk of ASD than previously thought and identifying these modifiable risk factors can provide opportunities and strategies for prevention.

### 1.1. Autism Spectrum Disorder

ASD is a neurodevelopmental condition characterized by challenges with social interaction, repetitive behaviours, and communication. The Diagnostic and Statistical Manual of Mental Disorders, fifth edition, (DSM-5) defines the diagnostic criteria for ASD as persistent deficits in social communication and interaction and restricted, repetitive patterns of behaviour, interests, or activities<sup>1</sup>. These symptoms are present in early developmental periods, cause clinically significant impairment in social and occupational functioning, and are not explained by intellectual disability or global developmental delay (although ASD and intellectual disability may be frequent comorbidities)<sup>1</sup>. There are no biological tests for ASD, and diagnoses rely on behavioural assessments.

“Spectrum” refers to ASD as a group of neurodevelopmental disorders with a wide variation in the type and severity of symptoms. The etiology of ASD is poorly understood, but it is suspected to be multifactorial and believed to include genetic, environmental, and epigenetic risk factors<sup>2</sup>. Moreover, ASD subtypes are hypothesized to have different etiological pathways caused by different combinations of genetic and environmental factors, making it a challenging health outcome to study.

### 1.1.1. ASD Prevalence

In Canada, the National Autism Spectrum Disorder Surveillance System (NASS) estimated the 2015 prevalence of ASD at 15.2 per 1,000 children 5–17 years of age in 2015 (1 in 66) across six provinces and one territory (i.e. Newfoundland and Labrador, Nova Scotia, Prince Edward Island, New Brunswick, Quebec, British Columbia and the Yukon)<sup>3</sup>. (Data for the Northwest Territories, Nunavut, Ontario, Alberta, Saskatchewan, and Manitoba are still under review.) Males were four times more likely to be diagnosed with ASD than females<sup>3</sup>. Based on regions with reported historical data, prevalence rates in Canada have increased from 2003–2015 among 5–14-year-olds: from 6.0 per 1,000 to 19.6 per 1,000 in Newfoundland and Labrador; from 5.0 per 1,000 to 17.7 per 1,000 in Prince Edward Island; and, from 3.5 per 1,000 to 15.7 per 1,000 in Quebec<sup>3</sup>. The prevalence in BC, based on 99.4% of children ages 5–17 receiving social services, was estimated at 14.7 per 1,000, closest to the overall, national prevalence. Out of a total of 29,099 children with ASD, Quebec and BC had the largest proportion of cases in Canada (58.4% and 31.2%, respectively), but also the largest population of 5–17-year-olds. NASS estimates for the different geographic regions harmonize various education, health, and social service data from each province or territory. Additionally, the case definition used by NASS requires that the diagnosis of ASD be provided or confirmed by a licensed health care professional and that it meet criteria from the Diagnostic and Statistical Manual for Mental Disorders or International Classification for Diseases.

The U.S. Centers for Disease Control and Prevention has reported a similar increase in ASD prevalence from 6.7 per 1,000 children age eight years (1 in 150) in 2000 to 16.8 per 1,000 (1 in 59) in 2014<sup>4,5</sup>. Prevalence estimates between Canada and the U.S. are similar although each rely on different methodologies, such as differences in time periods, ages of population under surveillance, and data collection methods (e.g. U.S case ascertainment use a records-review method which can positively identify a child with ASD based on documented behaviours consistent with ASD without exclusively relying on an ASD diagnoses). The average prevalence of ASD across studies in North America, Europe, and Asia is estimated at 1-2%<sup>5</sup>.

This increase in prevalence has been attributed to various factors<sup>6</sup>, such as changes in diagnostic criteria, diagnostic accretion and substitution, greater awareness about ASD, and people having children at an older age (a risk factor for ASD). Studies

have attributed 25–60% of the observed increase in ASD prevalence to changes in ASD diagnoses practices<sup>7,8</sup>. However, improvement to case ascertainment and broadening of clinical definitions does not rule out a true rise of ASD prevalence over time nor the contribution of other etiological factors to this increase<sup>8–11</sup>. A real rise that is not fully explained by these factors presents an epidemiologic opportunity not only to investigate what could be contributing to the increase in prevalence but also to test suspected and undiscovered etiological risks that may be contributing to ASD incidence<sup>6,10</sup>.

## **1.2. Environmental Risk Factors for ASD**

ASD is highly heritable, but environmental factors also contribute to the risk of ASD. In addition to genetics, suspected environmental risk factors also include environmental chemicals, maternal factors, and pharmaceutical drugs<sup>2</sup>. Twin studies that attempt to estimate the relative contribution of genes to ASD provide evidence for a genetic component, but they also highlight non-inherited genetic risk factors, suggesting that the environment also has a contributing role in ASD etiology<sup>12–14</sup>. Overall, ASD is a complex condition, and its etiological pathways and risk factors remain largely uncertain.

Prior research into modifiable environmental risk factors for ASD and other neurodevelopmental disorders have focused on chemicals and mixtures that are suspected of causing developmental neurotoxicity, including metals (e.g. arsenic, lead, manganese, mercury), polychlorinated biphenyls, polybrominated diphenyl ethers, polycyclic aromatic hydrocarbons, organophosphates and organochlorine pesticides, and other air pollutants<sup>2,12,15,16</sup>. Environmental risk factors for ASD are suspected of disrupting fetal and early life neural development. Hypothesized pathophysiological mechanisms for environmental risk factors include alterations to the nervous system, endocrine or immune system disruption, epigenetic alterations in methylation, and mitochondrial dysregulation<sup>16</sup>. Precise periods of susceptibility to environmental factors are uncertain, but studies focus on prenatal and early postnatal periods because they are critical and sensitive developmental windows when environmental exposures can have causal influences on developmental disorders<sup>16</sup>.

### **1.2.1. Air Pollution and Neurodevelopment**

Evidence from human epidemiological and animal toxicological studies have accumulated over the past decade to suggest that the central nervous system may be a target of air pollution through oxidative stress and inflammation, potentially leading to neurotoxicity, and may be an etiological factor in neurodegenerative or neurodevelopmental disorders, such as ASD<sup>17–20</sup>. Oxidative stress, when an excess of reactive oxygen species overwhelms the body's antioxidant defences, can lead to inflammation, tissue damage, and cell death<sup>21</sup>. Recent studies on the effects of air pollution on children show that air pollution exposure is associated with slower brain maturation<sup>22</sup>, poorer neurobehavioral function<sup>23</sup>, altered brain structures of the cerebral cortex and impaired inhibitory control<sup>24</sup>, and lower verbal ability<sup>25</sup>, school readiness<sup>25</sup>, and IQ<sup>26</sup>.

### **1.2.2. Air Pollution and ASD**

To date, three systematic reviews have examined the association between air pollution and ASD: Lam et al.<sup>27</sup> examined studies on ASD and airborne pollutants broadly (e.g. occupational exposures, pesticides, metals, hazardous air pollutants, proximity to major roads, and general air pollutants, such as PM, NO<sub>2</sub>, and ozone); Flores-Pajot et al.<sup>28</sup>, PM and NO<sub>2</sub>; and Morales-Suárez-Varela et al.<sup>29</sup>, PM. The systematic review by Flores-Pajot et al. provides the latest meta-analyses, including trimester-specific analyses.

Lam et al.<sup>27</sup> concluded that there is some evidence of an association between early life exposure to air pollution as a whole and ASD, with the strongest evidence for PM. The systematic review by Morales-Suárez-Varela et al.<sup>29</sup> on PM reported an association between diesel PM and ASD; a strong association between PM<sub>2.5</sub>; and a very weak association with PM<sub>10</sub>. Associations between diesel PM and PM<sub>10</sub> with ASD appeared to be dependent on total concentrations, whereas no trend was attributable to total PM<sub>2.5</sub> exposure, suggesting that risk is present even at low levels<sup>29</sup>. The meta-analyses by Flores-Pajot et al.<sup>28</sup> reported odds ratios of 1.34 (95% CI:0.83, 2.17, *n* = 8 cohorts) per 10 µg/m<sup>3</sup> increase in exposure in PM<sub>2.5</sub> and 1.05 (95% CI:0.99, 1.11, *n* = 8 cohorts) per 10 ppb increase in NO<sub>2</sub>.

Studies have observed associations between ASD and both pre- and post-natal exposures to NO<sub>2</sub> and PM<sub>2.5</sub>, and stronger associations in the third trimester compared to the first and second<sup>28,30</sup>. However, an insufficient number of studies have examined exposure windows, so no definite conclusions can be made whether risk varies based on the perinatal timing of exposure (i.e. between trimesters or prenatal versus postnatal)<sup>28,29</sup>.

Study results appear to vary by location. More specifically, studies from the United States, Israel, and Taiwan have reported positive associations between PM, NO, and NO<sub>2</sub> exposures and increased risk for ASD<sup>30–36</sup>. However, conflicting results have been observed in Europe, where studies reported no associations with ASD<sup>37</sup> or autistic traits<sup>38,39</sup> and PM, NO<sub>2</sub>, and nitrogen oxides.

The systematic reviews highlighted limitations of current studies on confounding, exposure assessment, and case ascertainment. Adjustment for confounders varied across studies, and a standardized list of confounding variables could help identify bias and improve comparability across studies<sup>27,29</sup>. Exposure assessment could be improved by refining spatial accuracy, accounting for timing of exposure, and incorporating time-activity patterns<sup>27–29</sup>. Tracking residential history to reduce exposure misclassification is particularly important when testing for critical windows of development and risk differences across various perinatal periods<sup>27–29</sup>. Lastly, the systematic reviews reported that potential differences in detection rates, which may introduce bias, can be a result from the variety of methods used to assess ASD outcome (e.g. validated instruments, meeting diagnostic criteria, parental reports, etc.) and differences in case definition, diagnostic practices, and access to health services<sup>28,29</sup>.

### **1.3. Research Context**

To investigate the association between prenatal exposures to PM<sub>2.5</sub>, NO, and NO<sub>2</sub> with ASD, we used a large population-based cohort in Metro Vancouver, a metropolitan area with relatively low levels of air pollution. ASD diagnosis was made on strict, established clinically verified diagnoses of ASD provided by the BC Autism Assessment Network (BCAAN). For exposure assessment, we used spatiotemporal air pollution models to develop temporally and spatially resolved person-specific exposure estimates. These spatiotemporal air pollution models were developed by combining

spatial land use regression (LUR) models <sup>40</sup> with temporal data from Metro Vancouver's air quality monitoring network <sup>41</sup>. The cohort and the temporally adjusted land-use regression models were linked by six-digit postal codes.

### **1.3.1. Linked Population Health Data**

The birth cohort was developed as part of the larger Canadian Institutes of Health Research–funded study “Developmental Origins of Autism: A Population-Level Linked Data Study of Potentially Modifiable Risk Factors” (PI: Oberlander), which examines four areas: (a.) mental health, psychotropics, anticonvulsants; (b.) exposures during labour and delivery; (c.) preterm birth and intrauterine growth; and (d.) air pollution. This study was nested within the Oberlander et al. study.

The birth cohort linked province-wide person-specific health data across multiple administrative health databases through Population Data BC. Birth records of all births in British Columbia (BC) from 2000–2009 were linked from the British Columbia Perinatal Data Registry to datasets from BC Vital Statistics, BC Medical Service Plan, hospital discharge data, Statistics Canada, and BCAAN. The linked dataset contains: data abstracted from obstetrical and neonatal medical records on nearly all of the in-hospital births and home births in the province; demographic information; residential location; records of all medical visits paid for on a fee-for-service basis; data on births, deaths, and hospital discharges; neighbourhood level data on urbanicity and socioeconomic status; and person-specific ASD diagnoses based on clinical evaluations.

For this thesis, the BC-wide cohort was subset to all births from 2004–2009 in Metro Vancouver based on the spatial domain of the 2003 LUR models. Beginning the cohort in 2004 enabled us to develop up to 12 months of prenatal exposure estimates for the first births in 2004 and to rely on standardized ASD assessments implemented in 2004.

### **1.3.2. ASD Diagnoses in British Columbia**

In BC, diagnostic assessment for ASD was standardized in 2004 with the *Standards and Guidelines for the Assessment and Diagnosis of Young Children with Autism Spectrum Disorder in British Columbia* <sup>42</sup>. These standards and guidelines

provided minimum standards to diagnose children under the age of six with ASD, established eligibility for ASD intervention services, and created consistency for ASD diagnostic processes across the province<sup>42</sup>. The report also established guidelines for clinical pathways of diagnosis, surveillance and screening, primary care assessment and referrals, and diagnostic multidisciplinary assessment<sup>42</sup>. Children suspected of having ASD are referred by primary care practitioners or specialists to the BC Autism Assessment Network (BCAAN), where they are assessed by a pediatrician, psychiatrist, or psychologist who is a qualified ASD diagnostician. BCAAN is a provincial program that provides diagnostic assessments for children and youth up to the age of 19 who are suspected of having ASD. Assessment services are provided across the five provincial health authorities: Northern Health, Interior Health, Vancouver Island Health Authority, Vancouver Coastal Health, and Fraser Health. As of December 2017, the overall wait time for assessment in BC was approximately 41.7 weeks<sup>43</sup>.

Assessment through BCAAN, which is covered by the provincial Medical Services Plan, is available at no cost to patients. Free diagnostic services improve case ascertainment bias because low socioeconomic status (SES) is a barrier to paid diagnostic services. Free diagnostic services help to reduce differential case misclassification by SES, which is particularly important because low SES is also typically associated with higher air pollution exposures<sup>44</sup>.

Families of children who are younger than six years of age and diagnosed with ASD receive up to \$22,000 annually from the Ministry of Children and Family Development (MCFD) to help pay for eligible autism intervention services<sup>45</sup>. At age 6, the MCFD direct funding drops to \$6,000, with \$18,850 directed to the school district for in-school interventions<sup>45</sup>. Families may also choose to receive a private, non-BCAAN diagnosis, costing around \$2,000–3,000<sup>46</sup>. However, to be eligible for MCFD funding, private diagnostic assessments completed by private clinicians must meet the same standards as a BCAAN assessment<sup>47</sup>.

### ***ADOS and ADI-R***

There are no established biological tests or markers for diagnosing ASD. Instead, ASD diagnoses rely on observed behaviour and cognitive function. As part of BCAAN's clinical diagnostic assessment, diagnosticians utilize the Autism Diagnostic Observation Schedule (ADOS)<sup>48–50</sup> and the Autism Diagnostic Interview-Revised (ADI-R)<sup>51,52</sup> as

standardized instruments—alongside detailed clinical history, evaluation of developmental status, and reports from health practitioners—to arrive at a final diagnosis of ASD based on clinical judgement and criteria in the DSM and ICD <sup>42,46</sup>. ADOS and ADI-R are considered the ‘gold standard’ tools for research and clinical diagnoses of ASD <sup>53,54</sup> and meet the DSM-IV and the International Statistical Classification of Diseases and Related Health Problems 10th Revision criteria for ASD <sup>55,56</sup>.

ADOS is a semi-structured observation scale designed to help diagnosticians differentiate between ASD from other disorders and typically developing individuals and to characterize social and communicative behaviours associated with ASD <sup>55</sup>. ADOS consists of four 35–60 minute modules with similar formats and structures, but each module is based on an age and language level <sup>55</sup>. Modules 1 and 2 include play-based tasks for individuals with no to limited expressive language, and Modules 3 and 4 involve conversations for individuals with fluent speech <sup>55</sup>. Diagnosticians perform standardized behaviours (e.g. conversations with older children or adults, play-based activities with younger children) to observe the presence or absence of expected responses and particular social behaviours <sup>55</sup>. Diagnosticians rate an individuals’ frequency of behaviours and quality of interaction on an ordinal scale from 0 to 3 (with 0 indicating no evidence of abnormality related to ASD and 3 indicating definite evidence) across various communication, social interaction, play or imagination, stereotyped behaviours, and restricted interests domains <sup>55</sup>. A classification of ASD or non-ASD is made when separate and combined thresholds in social affect, and restricted and repetitive behaviour domains are exceeded <sup>55</sup>.

ADI-R is a two-to-three hour standardized, semi-structured interview administered by a trained diagnostician to parents or caregivers of individuals who may have ASD <sup>56</sup>. The interview gathers quantitative and qualitative information—including detailed descriptions of behaviours and a history of symptoms associated with ASD—and 93 items across four domains: (a.) reciprocal social interactions; (b.) language and communication; (c.) restricted, repetitive, and stereotyped behaviors and interests; and (d.) abnormality of development evident at or before 36 months <sup>56</sup>. These items are related to behaviours that are rare in individuals without ASD or intellectual disabilities <sup>56</sup>. Like ADOS, a classification of ASD requires meeting or exceeding thresholds in all four areas <sup>56</sup>.

ADI-R provides a history and description of individual's functioning from a broad range of contexts, but ADI-R alone cannot be used to make clinical diagnoses<sup>56</sup>. When ADOS and ADI-R are used together, they provide complementary and improved diagnostic accuracy compared to when each instrument is used alone<sup>54,56–58</sup>. Used together, ADOS and ADI-R result in 81–82% sensitivity and 86–87% specificity, compared to 59–72% specificity when each is used alone<sup>57</sup>.

ADOS and ADI-R are instruments used to diagnose ASD, but valid administration and interpretation of results rely on the clinical skill, training, and practice by the diagnostician to administer reliably<sup>55,56</sup>.

### **1.3.3. Air Pollution in Metro Vancouver**

#### ***Metro Vancouver Air Quality Monitoring Network***

In the Vancouver metropolitan area, Metro Vancouver operates 23 air quality monitoring stations, part of a total of 29 monitoring stations that make up the Lower Fraser Valley Air Quality Monitoring Network. This network monitors air pollutants, including ozone, carbon monoxide, sulphur dioxide, nitric oxide, nitrogen dioxide, and particulate matter, to provide an overview of air quality in the region and to monitor ambient regional air quality objectives. The air quality objectives for the region are an hourly average of 25 µg/m<sup>3</sup> and an annual average of 6 µg/m<sup>3</sup> for PM<sub>2.5</sub>, and 106 ppb and 21 ppb for NO<sub>2</sub>, respectively<sup>59</sup>. Air quality has improved with short-term peak and average concentrations declining since the mid-nineties for carbon monoxide, NO<sub>2</sub>, sulphur dioxide, and PM<sup>60</sup>. PM<sub>2.5</sub> emissions are predominantly from transportation, wood and natural gas heating, industrial sources, and wildfires<sup>60</sup>. During summer, a large portion of PM<sub>2.5</sub> is comprised of secondary PM<sub>2.5</sub> (formed by reactions of nitrogen oxides, sulphur dioxide, and ammonia in the air), and episodic smoke from forest fires can increase PM<sub>2.5</sub> concentrations<sup>60</sup>. NO<sub>2</sub> emissions are dominated by transportation sources, so higher concentrations of NO<sub>2</sub> are observed in more densely trafficked areas and near busy roads<sup>60</sup>. NO<sub>2</sub> is a secondary pollutant formed when NO from vehicle exhaust mixes with oxygen in the air, so NO<sub>2</sub> exhibits less spatially heterogeneity compared to NO, which is found in higher concentrations near roadways but rapidly decreases with distance (Figures B2 and B3 in Appendix B).

## ***Metro Vancouver Land Use Regression Models***

LUR models for Metro Vancouver were developed by Henderson et al. <sup>40</sup> with the objective of developing exposure assessment tools capable of generating reliable estimates of ambient air pollution concentrations for epidemiological applications. LUR models capture the statistical relationship between land use characteristics and pollutant measurements to predict exposure estimates at high spatial resolutions over a large area <sup>40</sup>.

Henderson et al. developed the NO and NO<sub>2</sub> LUR models based on measurements from 116 passive samplers deployed throughout Metro Vancouver over two 14-day sampling periods. The PM<sub>2.5</sub> LUR model was based on measurements at 25 locations over a 2-month period. Measurements were temporally adjusted to estimate annual average concentrations. Land use characteristic variables from measurement sites were used in linear regression models to develop smooth, 10 m<sup>2</sup> resolution spatial surfaces of predicted annual average concentrations for the study area.

Predictor variables for the NO model ( $r^2 = 0.62$ ) included the number of major roads within 100 m and 1,000 m radius circular buffers of the measurement sites, the number of secondary roads within a 100 m buffer, the population density within a 2,500 m radius, and elevation. The NO<sub>2</sub> model ( $r^2 = 0.56$ ) included the same predictor variables as well as commercial land use within 750 m. Variables for the PM<sub>2.5</sub> model ( $r^2 = 0.52$ ) were the amount of commercial and industrial land use within 300 m, the amount of residential land use within 750 m, and elevation.

## ***Temporal Stability of Land Use Regression Models***

The NO and NO<sub>2</sub> LUR models were previously evaluated by Wang et al. for temporal stability and shown to provide reliable estimates from 2003–2010 <sup>61</sup>. LUR models for 2010 were developed using the same protocol for 2003 LUR models. The 2010 LUR models were temporally extrapolated by hindcasting and the 2003 models by forecasting. Model stability was evaluated by comparing predicted values against observed measurements (i.e. temporally extrapolated 2003 LUR models were compared to 2010 measurements, and vice-versa for 2010 LUR models and 2003 measurements). Four temporal interpolation methods were used: (1) applying a temporal trend; (2)

updating values of predictor values; (3) applying both methods 1 and 2; and (4) calibrating coefficients using the same predictor variables.

Overall, the LUR models performed better through forecasting than hindcasting, and decreasing pollutant concentrations maintained or improved forecasting performance. Wang et al. showed that applying an annual temporal trend provided an easy and effective temporal interpolation approach without requiring additional measurements.

The LUR NO<sub>2</sub> model was also evaluated for the ability to predict seasonal variability. Air quality monitoring network data was used to derive temporal adjustment factors, and sub-annual predictions were compared to observed measurements. The NO<sub>2</sub> model showed a similar, but slightly lower predictive ability for sub-annual predictions compared to annual average predictions.

## **1.4. Aims and Objectives**

Given the need identified by previous literature and the opportunities provided by rich data, this thesis advanced previous research by using (a.) a population-based cohort, among the largest to date, that captures the entire population of interest and all children positively identified with ASD through the provincial health program to reduce ascertainment bias; (b.) strict and established clinically verified diagnoses of ASD to reduce outcome misclassification; and (c.) high-resolution, temporally-adjusted land use regression models to reduce exposure misclassification, especially for trimester-focused analyses.

This study aimed to test and characterize the associations between prenatal exposures to ambient PM<sub>2.5</sub>, NO, and NO<sub>2</sub> and developing ASD using a retrospective, population-based cohort study from 2004–2009 in Metro Vancouver. Our objectives were to:

- Characterize the spatiotemporal trends in air pollution in Metro Vancouver from 2003–2009 (with air pollution estimates beginning one year before the start of the cohort to account for the gestational period of the first births in 2004).

- Quantify the association of PM<sub>2.5</sub>, NO, and NO<sub>2</sub> as risk factors for the incidence of ASD.
- Evaluate sex–exposure interactions and differences by sex in the odds of developing ASD given prenatal exposures to air pollution.
- Test whether exposures during the third trimester, a potential window of increased vulnerability<sup>28,34,35</sup>, are associated with increased odds of ASD compared to first and second trimester exposures.

This work is presented as a manuscript-based thesis. Chapter 2 presents a population-based cohort study examining the associations between ASD and average prenatal exposures to air pollution during the full pregnancy and per trimester. Chapter 2 was prepared and submitted as an article for peer-review to *JAMA Pediatrics*. Chapter 3 delves into the strengths and limitations of the work and its implications for public health policy. The Appendix contains the original abstract of the manuscript submitted to *JAMA Pediatrics*, supplemental tables and figures for Chapter 2, and the protocols, R scripts, and data file of the temporally adjusted LUR model estimates.

## 1.5. References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5). *Diagnostic Stat Man Ment Disord 4th Ed TR*. 2013. doi:10.1176/appi.books.9780890425596.744053
2. Dietert RR, Dietert JM, DeWitt JC. Environmental risk factors for autism. *Emerg Health Threats J*. 2011;4(1). doi:10.3402/ehhj.v4i0.7111
3. Ofner M, Coles A, Decou M Lou, et al. *Autism Spectrum Disorder Among Children and Youth in Canada 2018*. Ottawa; 2018. <https://www.canada.ca/en/public-health/services/publications/diseases-conditions/autism-spectrum-disorder-children-youth-canada-2018.html>.
4. Christensen DL, Baio J, Braun KVN, et al. Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *MMWR Surveill Summ*. 2016;65(3):1-23. doi:10.15585/mmwr.ss6503a1
5. Centers for Disease Control and Prevention. Autism Spectrum Disorder (ASD): Data & Statistics. <https://www.cdc.gov/ncbddd/autism/data.html>. Published 2016. Accessed March 1, 2018.
6. Weintraub K. The prevalence puzzle: Autism counts. *Nature*. 2011;479(7371):22-24. doi:10.1038/479022a
7. Hansen SN, Schendel DE, Parner ET. Explaining the Increase in the Prevalence of Autism Spectrum Disorders. *JAMA Pediatr*. 2015;169(1):56. doi:10.1001/jamapediatrics.2014.1893
8. King M, Bearman P. Diagnostic change and the increased prevalence of autism. *Int J Epidemiol*. 2009;38(5):1224-1234. doi:10.1093/ije/dyp261
9. Rutter M. Incidence of autism spectrum disorders: Changes over time and their meaning\*. *Acta Paediatr*. 2007;94(1):2-15. doi:10.1111/j.1651-2227.2005.tb01779.x
10. Newschaffer CJ, Falb MD, Gurney JG. National autism prevalence trends from United States special education data. *Pediatrics*. 2005;115(3):e277-82. doi:10.1542/peds.2004-1958
11. Ornoy A, Weinstein-Fudim L, Ergaz Z. Prenatal factors associated with autism spectrum disorder (ASD). *Reprod Toxicol*. 2015. doi:10.1016/j.reprotox.2015.05.007

12. Bakulski KM, Singer AB, Fallin MD. Genes and Environment in Autism Spectrum Disorders: An Integrated Perspective. In: Hu VW, ed. *Frontiers in Autism Research: New Horizons for Diagnosis and Treatment*. World Scientific; 2014:335-374. doi:10.1142/9789814602167\_0014
13. Hallmayer J, Cleveland S, Torres A, et al. Genetic Heritability and Shared Environmental Factors Among Twin Pairs With Autism. *Arch Gen Psychiatry*. 2011;68(11):1095. doi:10.1001/archgenpsychiatry.2011.76
14. Lichtenstein P, Carlström E, Råstam M, Gillberg C, Anckarsäter H. The Genetics of Autism Spectrum Disorders and Related Neuropsychiatric Disorders in Childhood. *Am J Psychiatry*. 2010;167(11):1357-1363. doi:10.1176/appi.ajp.2010.10020223
15. Landrigan PJ, Lambertini L, Birnbaum LS. A research strategy to discover the environmental causes of autism and neurodevelopmental disabilities. *Environ Health Perspect*. 2012;120(7):a258-a260. doi:10.1289/ehp.1104285
16. Kalkbrenner AE, Schmidt RJ, Penlesky AC. Environmental chemical exposures and autism spectrum disorders: a review of the epidemiological evidence. *Curr Probl Pediatr Adolesc Health Care*. 2014;44(10):277-318. doi:10.1016/j.cppeds.2014.06.001
17. Costa LG, Chang YC, Cole TB. Developmental Neurotoxicity of Traffic-Related Air Pollution: Focus on Autism. *Curr Environ Heal reports*. 2017;4(2). doi:10.1007/s40572-017-0135-2
18. Allen JL, Oberdorster G, Morris-Schaffer K, et al. Developmental neurotoxicity of inhaled ambient ultrafine particle air pollution: Parallels with neuropathological and behavioral features of autism and other neurodevelopmental disorders. *Neurotoxicology*. 2017;59. doi:10.1016/j.neuro.2015.12.014
19. Babadjouni RM, Hodis DM, Radwanski R, et al. Clinical effects of air pollution on the central nervous system; a review. *J Clin Neurosci*. 2017. doi:10.1016/j.jocn.2017.04.028
20. Brockmeyer S, D'Angiulli A. How air pollution alters brain development: The role of neuroinflammation. *Transl Neurosci*. 2016. doi:10.1515/tnsci-2016-0005
21. Burton GJ, Jauniaux E. Oxidative stress. *Best Pract Res Clin Obstet Gynaecol*. 2011;25(3):287-299. doi:10.1016/J.BPOBGYN.2010.10.016
22. Pujol J, Martínez-Vilavella G, Macià D, et al. Traffic pollution exposure is associated with altered brain connectivity in school children. 2016. doi:10.1016/j.neuroimage.2016.01.036

23. Wang S, Zhang J, Zeng X, Zeng Y, Wang S, Chen S. Association of Traffic-Related Air Pollution with Children's Neurobehavioral Functions in Quanzhou, China. *Environ Health Perspect.* 2009;117(10):1612-1618. doi:10.1289/ehp.0800023
24. Guxens M, Lubczy MJ, Muetzel RL, et al. Air Pollution Exposure During Fetal Life, Brain Morphology, and Cognitive Function in School-Age Children. 2018. doi:10.1016/j.biopsych.2018.01.016
25. Midouhas E, Kokosi T, Flouri E. Outdoor and indoor air quality and cognitive ability in young children. *Environ Res.* 2018. doi:10.1016/j.envres.2017.11.026
26. Perera FP, Li Z, Whyatt R, et al. Prenatal airborne polycyclic aromatic hydrocarbon exposure and child IQ at age 5 years. *Pediatrics.* 2009;124(2):e195-202. doi:10.1542/peds.2008-3506
27. Lam J, Sutton P, Kalkbrenner A, et al. A systematic review and meta-analysis of multiple airborne pollutants and autism spectrum disorder. Kesavachandran CN, ed. *PLoS One.* 2016;11(9):e0161851. doi:10.1371/journal.pone.0161851
28. Flores-Pajot MC, Ofner M, Do MT, Lavigne E, Villeneuve PJ. Childhood autism spectrum disorders and exposure to nitrogen dioxide, and particulate matter air pollution: A review and meta-analysis. *Environ Res.* 2016;151:763-776. doi:10.1016/j.envres.2016.07.030
29. Morales-Suárez-Varela M, Peraita-Costa I, Llopis- González A. Systematic review of the association between particulate matter exposure and autism spectrum disorders. *Environ Res.* 2017;153:150-160. doi:10.1016/j.envres.2016.11.022
30. Raz R, Levine H, Pinto O, Broday DM, Yuval, Weisskopf MG. Traffic-Related Air Pollution and Autism Spectrum Disorder: A Population-Based Nested Case-Control Study in Israel. *Am J Epidemiol.* August 2017. doi:10.1093/aje/kwx294
31. Becerra TA, Wilhelm M, Olsen J, Cockburn M, Ritz B. Ambient air pollution and autism in Los Angeles County, California. *Environ Health Perspect.* 2013;121(3):380-386. doi:10.1289/ehp.1205827
32. Jung C-R, Lin Y-T, Hwang B-F. Air Pollution and Newly Diagnostic Autism Spectrum Disorders: A Population-Based Cohort Study in Taiwan. Yoshikawa T, ed. *PLoS One.* 2013;8(9):e75510. doi:10.1371/journal.pone.0075510
33. Volk HE, Lurmann F, Penfold B, Hertz-Picciotto I, McConnell R. Traffic-Related Air Pollution, Particulate Matter, and Autism. *JAMA Psychiatry.* 2013;70(1):71. doi:10.1001/jamapsychiatry.2013.266

34. Raz R, Roberts AL, Lyall K, et al. Autism spectrum disorder and particulate matter air pollution before, during, and after pregnancy: A nested case-control analysis within the nurses' health study II cohort. *Environ Health Perspect.* 2015;123(3):264-270. doi:10.1289/ehp.1408133
35. Kalkbrenner AE, Windham GC, Serre ML, et al. Particulate matter exposure, prenatal and postnatal windows of susceptibility, and autism spectrum disorders. *Epidemiology.* 2015;26(1):30-42. doi:10.1097/EDE.0000000000000173
36. Talbott EO, Arena VC, Rager JR, et al. Fine particulate matter and the risk of autism spectrum disorder. *Environ Res.* 2015;140:414-420. doi:10.1016/j.envres.2015.04.021
37. Gong T, Dalman C, Wicks S, et al. Perinatal exposure to traffic-related air pollution and autism spectrum disorders. *Environ Health Perspect.* 2017;125(1):119-126. doi:10.1289/EHP118
38. Gong T, Almqvist C, Bölte S, et al. Exposure to air pollution from traffic and neurodevelopmental disorders in Swedish twins. *Twin Res Hum Genet.* 2014;17(6):553-562. doi:10.1017/thg.2014.58
39. Guxens M, Ghassabian A, Gong T, et al. Air pollution exposure during pregnancy and childhood autistic traits in four European population-based cohort studies: The ESCAPE project. *Environ Health Perspect.* 2016;124(1):133-140. doi:10.1289/ehp.1408483
40. Henderson SB, Beckerman B, Jerrett M, Brauer M. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ Sci Technol.* 2007;41(7):2422-2428. doi:10.1021/es0606780
41. Metro Vancouver. PM2.5, NO, and NO2 Daily Averages, 2003–2014. 2016.
42. Dua V. *Standards and Guidelines for the Assessment and Diagnosis of Young Children with Autism Spectrum Disorder in British Columbia: An Evidence-Based Report Prepared for The British Columbia Ministry of Health Planning.*; 2003. [http://www.phsa.ca/Documents/asd\\_standards\\_0318.pdf](http://www.phsa.ca/Documents/asd_standards_0318.pdf).
43. Provincial Health Services Authority. BC Autism Assessment Network. <http://www.phsa.ca/our-services/programs-services/bc-autism-assessment-network>. Accessed August 20, 2005.
44. Weisskopf MG, Kioumourtzoglou M-A, Roberts AL. Air Pollution and Autism Spectrum Disorders: Causal or Confounded? *Curr Environ Heal Reports.* 2015;2(4):430-439. doi:10.1007/s40572-015-0073-9

45. Government of British Columbia. Autism Funding. <https://www2.gov.bc.ca/gov/content/health/managing-your-health/healthy-women-children/child-behaviour-development/special-needs/autism-spectrum-disorder/autism-funding>. Accessed December 1, 2017.
46. Autism Community Training. Living & Working with Children with ASD in British Columbia. A Manual for Parents & Community Professionals. 2016. [www.actcommunity.ca/autism-manual](http://www.actcommunity.ca/autism-manual).
47. Ministry of Children and Family Development. *A Parent's Handbook: Your Guide to Autism Programs.*; 2017.
48. Lord C, Risi S, Lambrecht L, et al. The Autism Diagnostic Observation Schedule—Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. *J Autism Dev Disord.* 2000;30(3):205-223. doi:10.1023/A:1005592401947
49. Lord C, Rutter M, Goode S, et al. Autism diagnostic observation schedule: A standardized observation of communicative and social behavior. *J Autism Dev Disord.* 1989;19(2):185-212. doi:10.1007/BF02211841
50. Gotham K, Risi S, Pickles A, Lord C. The Autism Diagnostic Observation Schedule: Revised Algorithms for Improved Diagnostic Validity. *J Autism Dev Disord.* 2007;37(4):613-627. doi:10.1007/s10803-006-0280-1
51. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord.* 1994;24(5):659-685. doi:10.1007/BF02172145
52. Le Couteur A, Rutter M, Lord C, et al. Autism diagnostic interview: A standardized investigator-based instrument. *J Autism Dev Disord.* 1989;19(3):363-387. doi:10.1007/BF02212936
53. Filipek PA, Accardo PJ, Baranek GT, et al. The Screening and Diagnosis of Autistic Spectrum Disorders. *J Autism Dev Disord.* 1999;29(6):439-484. doi:10.1023/A:1021943802493
54. Falkmer T, Anderson K, Falkmer M, Horlin C. Diagnostic procedures in autism spectrum disorders: a systematic literature review. *Eur Child Adolesc Psychiatry.* 2013;22(6):329-340. doi:10.1007/s00787-013-0375-0
55. Carr T. Autism Diagnostic Observation Schedule. In: *Encyclopedia of Autism Spectrum Disorders*. New York, NY: Springer New York; 2013:349-356. doi:10.1007/978-1-4419-1698-3\_896
56. Kim SH, Hus V, Lord C. Autism Diagnostic Interview-Revised. In: *Encyclopedia of Autism Spectrum Disorders*. New York, NY: Springer New York; 2013:345-349. doi:10.1007/978-1-4419-1698-3\_894

57. Risi S, Lord C, Gotham K, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry*. 2006. doi:10.1097/01.chi.0000227880.42780.0e
58. Le Couteur A, Haden G, Hammal D, McConachie H. Diagnosing Autism Spectrum Disorders in Pre-school Children Using Two Standardised Assessment Instruments: The ADI-R and the ADOS. *J Autism Dev Disord*. 2008;38(2):362-372. doi:10.1007/s10803-007-0403-3
59. Metro Vancouver. *Ambient Air Quality Objectives*.; 2016.
60. Metro Vancouver. *2015 Lower Fraser Valley Air Quality Monitoring Report*. Burnaby, Canada; 2018.
61. Wang R, Henderson SB, Sbihi H, Allen RW, Brauer M. Temporal stability of land use regression models for traffic-related air pollution. *Atmos Environ*. 2013;64:312-319. doi:10.1016/J.ATMOSENV.2012.09.056

## Chapter 2.

# Prenatal Exposure to Air Pollution and ASD

Note: Modified and reproduced with permission from *JAMA Pediatrics*. 2019. 173(1): 86–92. doi:10.1001/jamapediatrics.2018.3101. Copyright © 2019 American Medical Association. All rights reserved.

### 2.1. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by varying degrees of difficulty in social interaction, verbal and non-verbal communication, and repetitive behaviours <sup>1</sup>. In the United States, the Centers for Disease Control and Prevention estimated a sharp increase in ASD prevalence from 6.7 per 1,000 children aged eight years (1 in 150) in 2000 to 14.6 per 1,000 (1 in 68) in 2012 <sup>2,3</sup>. Prevalence of ASD in Canada was 15.2 per 1,000 children 5–17 years of age in 2015 (1 in 66), and in British Columbia (BC) it was estimated to be 14.7 per 1,000 (1 in 68) <sup>4</sup>.

The etiology of ASD, which is suspected to be multifactorial, is poorly understood <sup>5</sup>. Prior research into modifiable environmental risk factors for ASD has targeted chemicals and mixtures, such as polychlorinated biphenyls, organophosphate pesticides, and automotive exhaust—due to concerns of suspected developmental neurotoxicity <sup>6</sup>. Recent systematic reviews have reported that exposure to air pollution is associated with increased risk for ASD <sup>7–9</sup>. In particular, studies from the United States, Israel, and Taiwan have reported positive associations between particulate matter (PM), nitric oxide (NO), and nitrogen dioxide (NO<sub>2</sub>) exposures and increased risk for ASD <sup>10–16</sup>. In contrast, three studies in Europe observed no associations with ASD <sup>17</sup> or autistic traits <sup>18,19</sup> and PM, NO<sub>2</sub>, and nitrogen oxides (NO<sub>x</sub>). The proportion of children with ASD from these studies ranged from 0.7% (out of 49,073) <sup>11</sup> to 53.2% (out of 524) <sup>12</sup>, with sample sizes ranging from 524 to 83,385 <sup>10,12</sup>. Outcome assessment in past studies relied on parent-reported questionnaires, diagnostic codes in various health registries, and in some studies, outcome validation of all or a partial subset of children positively identified with ASD.

Our study leveraged one of the largest population-based birth cohorts to date with strict diagnostic criteria for all children positively identified with ASD, and high-resolution exposure estimates to investigate the association between prenatal exposure to ambient air pollution in Metro Vancouver, Canada.

## **2.2. Methods**

### **2.2.1. Population and Study Design**

The cohort comprised births to mothers from 2004–2009 who were registered with the provincial health insurance plan for  $\geq 275$  days and resided in the Greater Vancouver area during the calendar year of their pregnancy. In British Columbia, the provincial public health insurance program covers nearly all residents<sup>20</sup>. This enabled us to link person-specific health data across multiple administrative health databases. We linked birth records in British Columbia from the British Columbia Perinatal Data Registry and BC Vital Statistics with BC Medical Service Plan and hospital discharge data, Statistics Canada data, and BC Autism Assessment Network (BCAAN) data through Population Data BC<sup>21–27</sup>. The British Columbia Perinatal Data Registry contains data abstracted from obstetrical and neonatal medical records on nearly 100% of births in the province of British Columbia from over 60 acute care facilities as well as births occurring at home attended by BC registered midwives, including women who had pregnancies ending in a live or stillbirth of at least 20 weeks gestation or 500 grams birth weight. The registry also collects data on maternal postpartum readmissions up to 42 days post-delivery and baby transfers and readmissions up to 28 days after birth. The estimated gestational age is calculated based on last menstrual period, first ultrasound ( $<20$  weeks), clinical estimate from newborn exam, and documentation from maternal chart<sup>28</sup>. Health services data from the BC Ministry of Health and BC Vital Statistics Agency includes demographic information, residential location, medical visits paid for on a fee-for-service basis, and data on births, deaths, and hospital discharges. Statistics Canada provides neighbourhood-level data on urbanicity and socioeconomic status (SES). SES was measured as income bands, 1,000 bands of equalized, disposable income for each postal code (1 lowest income, 1,000, highest). The Vancouver metropolitan area is highly urbanized, where a postal code may represent small geographical areas such as a high-rise building or one side of a city block, and on average include  $\sim 35$  individuals<sup>29</sup>.

The study area was based on the estimation domain of the land use regression (LUR) models used for exposure assessment<sup>30</sup>. We obtained a linked file of all singleton, twin, or triplet births to mothers ages 15–49 in British Columbia. Births with unknown gestational age or missing gender information, stillbirths, and babies who died at less than one year of age were not included in the data extract. Using residential postal codes, we excluded births outside of the study area.

To develop upwards of 12 months of exposure estimates for the first birth in 2004, we temporally adjusted LUR models beginning January 2003. The cohort was followed up until the end of 2014, establishing a 5-year minimum follow-up period for ASD diagnoses (for the last births in 2009) and a 10-year maximum (for the first births in 2004).

The study was approved by the University of British Columbia Children’s and Women’s Research Ethics Board (certificate no. H14-01161) and Simon Fraser University (certificate no. 2016s0146).

### **2.2.2. Case Ascertainment**

All primary care practitioners and specialists conduct general developmental surveillance, and if a child is suspected of having ASD or other brain-based disorders, they are referred for assessment. BCAAN data contains person-specific ASD diagnoses based on clinical evaluations made by pediatricians, psychiatrists, or psychologists who are qualified ASD diagnosticians. Diagnostic assessment for ASD has been standardized in British Columbia since 2004. In addition to detailed clinical history, evaluation of developmental status, and reports from health practitioners, clinical diagnostic assessment utilizes the Autism Diagnostic Interview-Revised (ADI-R) and the Autism Diagnostic Observation Schedule (ADOS) as standardized tools<sup>31</sup>. BCAAN provides clinical diagnostic assessment for children and youth up to 19 years of age. Assessment through BCAAN is available at no cost as it is covered by the provincial health plan.

### **2.2.3. Exposure Assessment**

We developed temporally adjusted, long-term average measures of air pollution by combining 2003 LUR models (10 m<sup>2</sup> spatial resolution) developed by Henderson et al.<sup>30</sup> with continuous monitoring data from Metro Vancouver's Air Quality Monitoring Network<sup>32</sup> (see Appendix C). Temporally adjusting the LUR models enabled us to incorporate annual and seasonal trends into the LUR models to estimate monthly mean exposures of ambient air pollution for each pregnant woman at their residential postal code for four time periods: their full pregnancy and each trimester. We used predictive mean matching to impute missing air monitoring station data.

First, we developed temporal factors as a ratio between the monthly mean of air pollutant concentrations across all air monitors for the month of interest over the annual mean of air pollutant concentrations across all air monitors for the LUR year<sup>33</sup>. Then, we derived minimum concentrations by calculating the mean from the lowest daily average in a month for all air monitors for each month of the LUR year. To avoid potentially estimating values of zero, LUR models were truncated before temporal adjustment (Figures B1–B3 in Appendix B). We extracted air pollutant concentrations from the truncated LUR models at all postal codes<sup>34–40</sup> within the study area and then multiplied them by the temporal factors to produce monthly air pollutant estimates across all postal codes. We then linked air pollutant estimates for each mother during each month of pregnancy using conception date, birth date, and residential postal code. Mothers in the cohort resided at 52,754 unique postal codes; out of these, we truncated 8.1% to 1.4 µg/m<sup>3</sup> for PM<sub>2.5</sub>, 23 postal codes to 1.3 ppb for NO, and 138 postal codes to 6.9 ppb for NO<sub>2</sub>. ArcGIS software version 10.3 and R versions 3.3–3.5, with *rgdal* and *rgeos* packages, were used for geoprocessing.

### **2.2.4. Statistical Analyses**

To examine the independence of pollutant concentrations, we calculated a Pearson correlation coefficient matrix of exposure estimates for all pollutant–time period pairs. We selected covariates a priori based on expert knowledge, data availability, past literature, and a list of confounders and effect modifiers identified by the Environmental Epidemiology of Autism Research Network<sup>7</sup>. We included: child sex (categorical), birth month (categorical), birth year (categorical), maternal age (continuous), mother's birth

country (categorical), and neighbourhood-level urbanicity (categorical) and income band (continuous). We did not adjust for gestational age or birth weight because they may be on the causal pathway between air pollution and ASD<sup>10,16</sup>.

We examined the association between air pollution exposure and the odds of ASD using single pollutant and single time period models. We calculated odds ratios (OR) and 95% confidence intervals (CI) using logistic regression to estimate increases in odds of ASD per interquartile range (IQR) change of PM<sub>2.5</sub>, NO, and NO<sub>2</sub>. Additionally, we conducted sex-stratified and sensitivity analyses by individually adding gestational age, small for gestational age, birth weight, multiple birth, parity, and maternal smoking to our original model; replaced birth month with conception month; replaced original exposure estimates with estimates using non-truncated, temporally adjusted LUR models; and restricted follow-up to a maximum of five years. All statistical analyses were performed in R versions 3.3–3.5.

## 2.3. Results

From 2004–2009, 132,256 children were born in the Vancouver metropolitan area; 1,307 (1.0%) were diagnosed with ASD by BCAAN. The median age at assessment for ASD was 4.2 years (IQR, 3.3–5.2). Children who were diagnosed with ASD tended to be male (at a ratio of approximately 1:5), were born to slightly older, multiparous mothers, and lived in lower income level neighbourhoods (Table 1).

Air pollution concentrations, which exhibited a seasonal pattern, decreased over the 2003–2009 study period, from an annual average of 5.7 to 5.0 µg/m<sup>3</sup> for PM<sub>2.5</sub>, 16.0 to 11.0 ppb for NO, and 16.0 to 13.5 ppb for NO<sub>2</sub> (Figure B4 in Appendix B). Concentrations of PM<sub>2.5</sub> were higher during summer due to secondary PM<sub>2.5</sub> formation and episodic smoke from forest fire activity<sup>41</sup>. In contrast, NO and NO<sub>2</sub> concentrations were higher in winter.

For mothers in the cohort, the median monthly average exposure of PM<sub>2.5</sub> during pregnancy was 3.5 µg/m<sup>3</sup> (IQR, 2.7–4.2); for NO, 18.3 ppb (IQR, 14.0–24.7); and for NO<sub>2</sub>, 14.3 ppb (IQR, 12.2–17.0). Values were similar across the full pregnancy and trimesters (Table B1 in Appendix B). Correlation across time periods (Figure B5 in

Appendix B) was high for both PM<sub>2.5</sub> ( $r = 0.8–1.0$ ) and NO<sub>2</sub> ( $r = 0.6–1.0$ ), suggesting a limited ability to differentiate between exposures across trimesters.

In unadjusted logistic regression models, we estimated an increase in odds of ASD with increased air pollution exposure for the full pregnancy (Table 2) and each trimester. Positive associations remained after adjusting for child sex, birth month, birth year, maternal age, maternal birthplace, urbanicity, and income band. Although some CIs cross one, point estimates trend in positive directions, suggesting increased odds of ASD with higher exposure to air pollutants (Table 2). We observed consistently higher effect estimates in males for all air pollutants (Table 2) and across all trimesters (Figure B6 in Appendix B), although interactions between sex and exposure for all air pollutants and time periods were not statistically significant ( $p = 0.30–0.77$ ).

We conducted sensitivity analyses to examine whether adjusting for gestational age, small for gestational age, birth weight, multiple birth, parity, and maternal smoking altered effect estimates. We also replaced birth month with conception month and compared temporally adjusted exposure estimates using truncated and non-truncated LUR models. Overall results remained similar after adjusting for the additional covariates added to the base models, with parity attenuating the association the most (Table B2 in Appendix B). Restricting follow-up for diagnosis to a maximum of five years reduced the number of children positively identified with ASD to 910 and resulted small changes to the ORs, an increase for PM<sub>2.5</sub> and NO<sub>2</sub>, and a decrease for NO (Table B2 in Appendix B).

We had limited ability to differentiate between trimester exposures due to their high correlation. Likelihood ratio tests on adjusted models with mutually adjusted trimesters (Table B3 in Appendix B) indicated no differences in odds of developing ASD between trimester-specific exposures for any of the pollutants.

## 2.4. Discussion

We observed increased odds for ASD associated with increased exposure to PM<sub>2.5</sub>, NO, and NO<sub>2</sub> in one of the largest population-based cohort studies of prenatal exposures to air pollution and the subsequent development of ASD. Effect estimates

were consistently larger for males. Notably, this study was conducted with a population-based sample and relied on strict, established clinically verified ASD diagnostic criteria.

We found positive associations between ASD and prenatal exposures to air pollutants. Figure 1 plots our results alongside two meta-analyses and all studies published to date examining the associations between ASD and prenatal exposures to PM<sub>2.5</sub>, NO, and NO<sub>2</sub>. Point estimates across all studies—except for one—show a positive trend for all pollutants. Consistent with the findings in the systematic review by Flores-Pajot et al.<sup>8</sup>, our study showed a positive association between ASD and PM<sub>2.5</sub> exposure during the full pregnancy, albeit with a slightly lower OR. Flores-Pajot et al.<sup>8</sup> also reported on studies with NO<sub>2</sub>, and their reported pooled risk ratio was similar to our observed OR. Although NO is not a regulated air pollutant, we included NO in the study as an indicator of traffic-related air pollution because our exposure models are capable of distinguishing between NO and NO<sub>2</sub> as primary and secondary traffic-related air pollutants.

We conducted sex-stratified analyses and, similar to two prior studies that considered associations between PM<sub>2.5</sub> or NO<sub>2</sub> with ASD and found positive associations<sup>13,16</sup>, we observed larger ORs among males than females in stratified analyses. Due to the fewer number of female children with ASD in our study, our sex-stratified models with female children had wider confidence intervals.

There is insufficient evidence to date of clear trends in increased odds by trimester. In a systematic review of PM and ASD, Morales-Suárez-Varela et al. reported no definitive conclusions between exposure periods, PM, and risk of ASD<sup>9</sup>. Flores-Pajot et al.<sup>8</sup> reported an increase in the strength of association between PM<sub>2.5</sub> and ASD from the first to the third trimester, from a risk ratio of 1.10 per 10 µg/m<sup>3</sup> to 1.33, and little difference in association across trimesters for NO<sub>2</sub>. Flores-Pajot et al.<sup>8</sup> caution, however, that these PM<sub>2.5</sub> trimester estimates are based on four studies, and NO<sub>2</sub>, on only two studies.

To date, positive associations between ASD and air pollution have been observed across seven studies in the United States<sup>10,12–15</sup>, Israel<sup>16</sup>, and Taiwan<sup>11</sup>, but not in three European studies<sup>17–19</sup>. Potential reasons for inconsistent findings across these studies could be due to differences in case definition, air pollution levels, or

composition of airborne pollutants<sup>16,17,19</sup>. Our study was in an area with relatively low air pollution and had similar levels of air pollution as other European cohorts; Guxens et al.<sup>19</sup> observed median air pollution levels during pregnancy ranging from 8.4–22.4  $\mu\text{g}/\text{m}^3$  of  $\text{PM}_{2.5}$  and 9.5–22.5 ppb of  $\text{NO}_2$  (converted from 17.9–42.2  $\mu\text{g}/\text{m}^3$ ) across cohorts in Sweden, the Netherlands, Italy, and Spain. As well, Guxens et al.<sup>19</sup> and Gong et al.<sup>18</sup> used measures of autistic traits, not diagnoses of ASD, and therefore may not be directly comparable to the other studies described above.

We did not adjust for birth weight, small for gestational age, and preterm birth in our primary analyses because these could be on the causal pathway between air pollution and ASD. Studies have found adverse associations between traffic-related air pollution and birth weight, small for gestational age, and preterm birth<sup>42,43</sup>, and some evidence that these may be associated with ASD<sup>44–46</sup>. Still, adjusting for these covariates in secondary analyses showed no appreciable differences in ORs.

This study has notable strengths and limitations. In our study, which is one of the largest population-based cohorts on ASD and air pollution, all children who were diagnosed with ASD received a clinical diagnosis by qualified diagnosticians who have undergone extensive training to provide standardized clinical assessments. This case definition, however, does not include children who received an ASD diagnosis from a private practitioner. With a linkage of our cohort to Ministry of Education data that contain ASD diagnoses from both BCAAN and private practitioners for children in the public-school system, 1,967 children were identified with ASD, although only 1,307 (66.4%) are identified by the strict diagnostic criteria through BCAAN. We did not see meaningful differences in air pollutant estimates associated with whether diagnoses were from BCAAN or private practitioners.

Our temporally adjusted LUR models had high spatial and temporal resolutions to help reduce measurement error. Spatially, exposure estimates were resolved to small geographic areas at a postal code level. Temporally adjusting the LUR models incorporated seasonal variability over time and allowed us to test for trimester-specific exposures. Wang et al.<sup>47</sup> demonstrated the ability of the NO and  $\text{NO}_2$  LUR models to reliably forecast annual and seasonal average estimates over seven years by temporally interpolating the LUR models to 2010 and comparing predictions against observed measurements. However, trimester-specific analyses were limited by the high correlation

between trimester exposures. Additionally, accounting for time-activity patterns during pregnancy could help improve personal exposure estimates<sup>48</sup>, although pregnant women tend to increasingly spend more time at home during the latter part of pregnancy<sup>49</sup>.

## **2.5. Conclusion**

Using strict person-specific diagnostic criteria and high-resolution exposure estimates applied to a large population-based birth cohort, we observed positive associations between exposures to air pollutants and odds of ASD in a metropolitan area with relatively low ambient air pollution levels. Our findings are consistent with emerging associations between exposure to air pollutants during fetal growth with elevated risks of ASD in other settings and suggest that reducing exposure to air pollutants for pregnant women could help reduce ASD incidence.

## 2.6. Tables and Figures

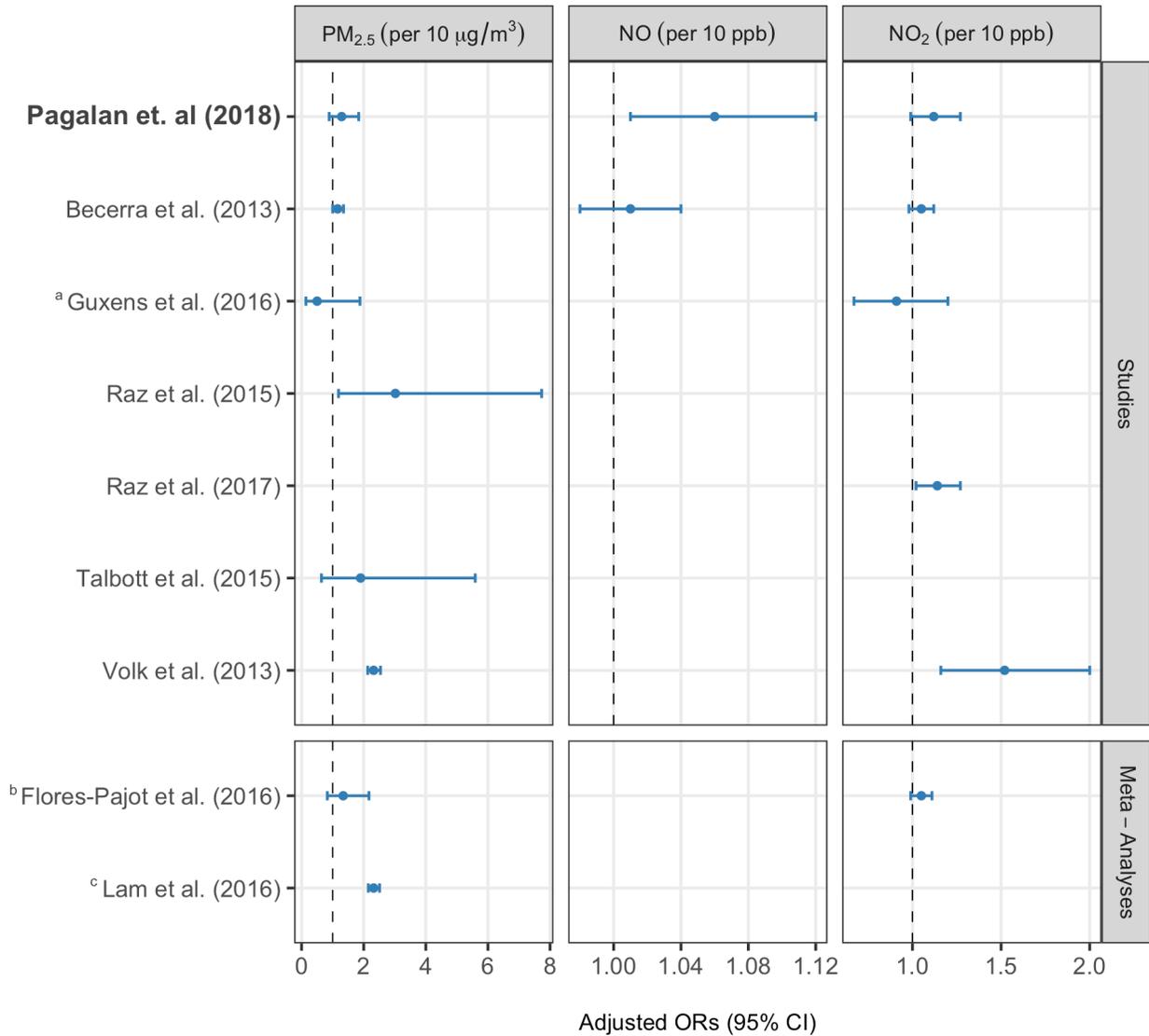
**Table 1. Characteristics of children with and without ASD.**

Characteristic	Children with ASD (n = 1,307)		Children without ASD (n = 130,949)	
Sex (%)				
Female	216	(16.5)	63,989	(48.9)
Male	1,091	(83.5)	66,960	(51.1)
Gestational Age at Birth (Weeks) (median, IQR)	39	37–40	39	38–40
Maternal Age at Delivery (Years) (median, IQR)	32.2	28.7–35.9	31.9	28.2–35.4
Multiple Birth (%)				
Singleton	1,250	(95.6)	126,942	(96.9)
Twins and Triplets	57	(4.4)	4,007	(3.1)
Parity (%)				
Nulliparous	541	(41.4)	67,834	(51.8)
Multiparous	766	(58.6)	63,114	(48.2)
Mother's Birthplace (%)				
Canada	563	(43.1)	66,284	(50.7)
India	83	(6.4)	12,275	(9.4)
China	128	(9.8)	11,004	(8.4)
Philippines	127	(9.7)	6,606	(5.1)
United Kingdom	24	(1.8)	2,512	(1.9)
United States	21	(1.6)	1,986	(1.5)
Vietnam	51	(3.9)	2,912	(2.2)
Hong Kong	28	(2.1)	2,045	(1.6)
Other Countries	280	(21.5)	25,102	(19.2)
Income Band (median, IQR)	390	164–679	449	189–738
Urbanicity (%)				
Rural	6	(0.5)	456	(0.4)
Semi-Urban	11	(0.9)	729	(0.6)
Urban	1,277	(98.7)	129,035	(99.1)
Month of Birth (%)				
January	119	(9.1)	10,423	(8.0)
February	88	(6.7)	9,717	(7.4)
March	116	(8.9)	10,762	(8.2)
April	109	(8.3)	10,679	(8.2)
May	111	(8.5)	11,258	(8.6)
June	114	(8.7)	11,096	(8.5)
July	92	(7.0)	11,379	(8.7)
August	109	(8.3)	11,556	(8.8)
September	109	(8.3)	11,591	(8.9)
October	109	(8.3)	11,267	(8.6)
November	129	(9.9)	10,719	(8.2)
December	102	(7.8)	10,502	(8.0)
Year of Birth (%)				
2004	235	(18.0)	20,700	(15.8)
2005	206	(15.8)	20,759	(15.9)
2006	200	(15.3)	21,465	(16.4)
2007	210	(16.1)	22,302	(17.0)
2008	232	(17.8)	22,516	(17.2)
2009	224	(17.1)	23,207	(17.7)

**Table 2. Odds of ASD for prenatal exposure to air pollutants.**

Pollutant	Units	Model <sup>a</sup>	Sample Size	Children with ASD	OR (95% CI)
PM <sub>2.5</sub>	Per IQR (1.5 µg/m <sup>3</sup> )	Unadjusted full pregnancy model	131,440	1,300	1.08 (1.03, 1.14)
		Adjusted full pregnancy model	129,439	1,276	1.04 (0.98, 1.10)
		Adjusted full pregnancy model: males	66,612	1,064	1.04 (0.98, 1.10)
		Adjusted full pregnancy model: females	62,827	212	1.03 (0.90, 1.18)
NO	Per IQR (10.7 ppb)	Unadjusted full pregnancy model	131,437	1,300	1.09 (1.03, 1.14)
		Adjusted full pregnancy model	129,436	1,276	1.07 (1.01, 1.13)
		Adjusted full pregnancy model: males	66,611	1,064	1.09 (1.02, 1.15)
		Adjusted full pregnancy model: females	62,825	212	0.98 (0.83, 1.13)
NO <sub>2</sub>	Per IQR (4.8 ppb)	Unadjusted full pregnancy model	131,437	1,300	1.11 (1.05, 1.17)
		Adjusted full pregnancy model	129,436	1,276	1.06 (0.99, 1.12)
		Adjusted full pregnancy model: males	66,611	1,064	1.07 (1.00, 1.13)
		Adjusted full pregnancy model: females	62,825	212	1.00 (0.86, 1.16)

<sup>a</sup> Models adjusted for child sex, birth month, birth year, maternal age, maternal birthplace, and neighborhood-level urbanicity and income band.



**Figure 1. Forest plot of ORs examining the association of ASD and prenatal exposure to air pollutants during full pregnancy.**

If multiple models were presented in a study, the chosen models were those that used temporally adjusted exposure estimates to account for temporal variation, and models that used full pregnancy exposure estimates without mutually adjusting for other time periods. For comparability, effect estimates were standardized to 10 µg/m<sup>3</sup> unit change for PM<sub>2.5</sub> and 10 ppb unit change for NO and NO<sub>2</sub>. <sup>a</sup> NO<sub>2</sub> conversion factor: 10 µg/m<sup>3</sup> = 5.32 ppb based on ambient pressure of 1 atm and a temperature of 25 °C. <sup>b</sup> Flores-Pajot et al. (2016) meta-analysis summary risk ratio for PM<sub>2.5</sub> comprised studies: Guxens et al. (2016), Becerra et al. (2013), Raz et al. (2015), Talbott et al. (2015), and Volk et al. (2013); and summary risk ratio for NO<sub>2</sub>: Guxens et al. (2016), Becerra et al. (2013), and Volk et al. (2013). <sup>b</sup> Lam et al. (2016) meta-analysis summary odds ratio for PM<sub>2.5</sub> comprised studies: Raz et al. (2015), Volk et al. (2013), and Becerra et al. (2013).

## **2.7. Disclaimer**

All inferences, opinions, and conclusions drawn in this study are those of the authors, and do not reflect the opinions or policies of the Data Stewards. The information provided in these data is intended for educational and informational purposes only. These data are not intended to endorse or recommend any particular product, material or service provider nor is it intended as a substitute for engineering, legal or other professional advice. Such advice should be sought from qualified professionals.

## 2.8. References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5). *Diagnostic Stat Man Ment Disord 4th Ed TR*. 2013. doi:10.1176/appi.books.9780890425596.744053
2. Christensen DL, Baio J, Braun KVN, et al. Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *MMWR Surveill Summ*. 2016;65(3):1-23. doi:10.15585/mmwr.ss6503a1
3. Centers for Disease Control and Prevention. Autism Spectrum Disorder (ASD): Data & Statistics. <https://www.cdc.gov/ncbddd/autism/data.html>. Published 2016. Accessed March 1, 2018.
4. Ofner M, Coles A, Decou M Lou, et al. *Autism Spectrum Disorder Among Children and Youth in Canada 2018*. Ottawa; 2018. <https://www.canada.ca/en/public-health/services/publications/diseases-conditions/autism-spectrum-disorder-children-youth-canada-2018.html>.
5. Dietert RR, Dietert JM, DeWitt JC. Environmental risk factors for autism. *Emerg Health Threats J*. 2011;4(1). doi:10.3402/ehth.v4i0.7111
6. Landrigan PJ, Lambertini L, Birnbaum LS. A research strategy to discover the environmental causes of autism and neurodevelopmental disabilities. *Environ Health Perspect*. 2012;120(7):a258-a260. doi:10.1289/ehp.1104285
7. Lam J, Sutton P, Kalkbrenner A, et al. A systematic review and meta-analysis of multiple airborne pollutants and autism spectrum disorder. Kesavachandran CN, ed. *PLoS One*. 2016;11(9):e0161851. doi:10.1371/journal.pone.0161851
8. Flores-Pajot MC, Ofner M, Do MT, Lavigne E, Villeneuve PJ. Childhood autism spectrum disorders and exposure to nitrogen dioxide, and particulate matter air pollution: A review and meta-analysis. *Environ Res*. 2016;151:763-776. doi:10.1016/j.envres.2016.07.030
9. Morales-Suárez-Varela M, Peraita-Costa I, Llopis- González A. Systematic review of the association between particulate matter exposure and autism spectrum disorders. *Environ Res*. 2017;153:150-160. doi:10.1016/j.envres.2016.11.022
10. Becerra TA, Wilhelm M, Olsen J, Cockburn M, Ritz B. Ambient air pollution and autism in Los Angeles County, California. *Environ Health Perspect*. 2013;121(3):380-386. doi:10.1289/ehp.1205827
11. Jung C-R, Lin Y-T, Hwang B-F. Air Pollution and Newly Diagnostic Autism Spectrum Disorders: A Population-Based Cohort Study in Taiwan. Yoshikawa T, ed. *PLoS One*. 2013;8(9):e75510. doi:10.1371/journal.pone.0075510

12. Volk HE, Lurmann F, Penfold B, Hertz-Picciotto I, McConnell R. Traffic-Related Air Pollution, Particulate Matter, and Autism. *JAMA Psychiatry*. 2013;70(1):71. doi:10.1001/jamapsychiatry.2013.266
13. Raz R, Roberts AL, Lyall K, et al. Autism spectrum disorder and particulate matter air pollution before, during, and after pregnancy: A nested case-control analysis within the nurses' health study II cohort. *Environ Health Perspect*. 2015;123(3):264-270. doi:10.1289/ehp.1408133
14. Kalkbrenner AE, Windham GC, Serre ML, et al. Particulate matter exposure, prenatal and postnatal windows of susceptibility, and autism spectrum disorders. *Epidemiology*. 2015;26(1):30-42. doi:10.1097/EDE.0000000000000173
15. Talbott EO, Arena VC, Rager JR, et al. Fine particulate matter and the risk of autism spectrum disorder. *Environ Res*. 2015;140:414-420. doi:10.1016/j.envres.2015.04.021
16. Raz R, Levine H, Pinto O, Broday DM, Yuval, Weisskopf MG. Traffic-Related Air Pollution and Autism Spectrum Disorder: A Population-Based Nested Case-Control Study in Israel. *Am J Epidemiol*. August 2017. doi:10.1093/aje/kwx294
17. Gong T, Dalman C, Wicks S, et al. Perinatal exposure to traffic-related air pollution and autism spectrum disorders. *Environ Health Perspect*. 2017;125(1):119-126. doi:10.1289/EHP118
18. Gong T, Almqvist C, Bölte S, et al. Exposure to air pollution from traffic and neurodevelopmental disorders in swedish twins. *Twin Res Hum Genet*. 2014;17(6):553-562. doi:10.1017/thg.2014.58
19. Guxens M, Ghassabian A, Gong T, et al. Air pollution exposure during pregnancy and childhood autistic traits in four European population-based cohort studies: The ESCAPE project. *Environ Health Perspect*. 2016;124(1):133-140. doi:10.1289/ehp.1408483
20. Chamberlayne R, Green B, Barer ML, Hertzman C, Lawrence WJ, Sheps SB. Creating a Population-based Linked Health Database: A New Resource for Health Services Research. *Can J Public Heal / Rev Can Sante'e Publique*. 1998;89(4):270-273.
21. Canadian Institute for Health Information [creator] (2016): Discharge Abstract Database (Hospital Separations). Population Data BC [publisher]. Data Extract. MOH (2015). <http://www.popdata.bc.ca/data>.
22. British Columbia Ministry of Health [creator] (2015): Consolidation File (MSP Registration & Premium Billing). Population Data BC [publisher]. Data Extract. MOH (2015). <http://www.popdata.bc.ca/data>.

23. British Columbia Ministry of Health [creator] (2016): Medical Services Plan (MSP) Payment Information File. Population Data BC [publisher]. Data Extract. MOH (2015). <http://www.popdata.bc.ca/data>.
24. Perinatal Services BC [creator] (2011): British Columbia Perinatal Data Registry. Population Data BC [publisher]. Data Extract. PSBC (2015). <http://www.perinataleservicesbc.ca/health-professionals/data-surveillance/perinatal-data-registry>.
25. BC Vital Statistics Agency [creator] (2015): Vital Statistics Births. Population Data BC [publisher]. Data Extract. BC Vital Statistics Agency (2015). 2015. <http://www.popdata.bc.ca/data>.
26. BC Vital Statistics Agency [creator] (2015): Vital Statistics Deaths. Population Data BC [publisher]. Data Extract BC Vital Statistics Agency (2015). <http://www.popdata.bc.ca/data>.
27. Statistics Canada [creator] (2009): Statistics Canada Income Band Data. Catalogue Number: 13C0016. Population Data BC [publisher]. Data Extract. Population Data BC (2016). <http://www.popdata.bc.ca/data>.
28. Perinatal Services BC. Data Dictionary for Use by Health Authority Analysts. <http://www.perinataleservicesbc.ca/health-professionals/data-surveillance/perinatal-data-registry/resources-for-analysts>. Published 2015. Accessed January 1, 2018.
29. Clark C, Sbihi H, Tamburic L, Brauer M, Frank LD, Davies HW. Association of long-term exposure to transportation noise and traffic-related air pollution with the incidence of diabetes: A prospective cohort study. *Environ Health Perspect*. 2017;125(8). doi:10.1289/EHP1279
30. Henderson SB, Beckerman B, Jerrett M, Brauer M. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ Sci Technol*. 2007;41(7):2422-2428. doi:10.1021/es0606780
31. Dua V. *Standards and Guidelines for the Assessment and Diagnosis of Young Children with Autism Spectrum Disorder in British Columbia: An Evidence-Based Report Prepared for The British Columbia Ministry of Health Planning.*; 2003. [http://www.phsa.ca/Documents/asd\\_standards\\_0318.pdf](http://www.phsa.ca/Documents/asd_standards_0318.pdf).
32. Metro Vancouver. PM2.5, NO, and NO2 Daily Averages, 2003–2014. 2016.
33. Sbihi H, Allen RW, Becker A, et al. Perinatal exposure to traffic-related air pollution and atopy at 1 year of age in a multi-center Canadian birth cohort study. *Environ Health Perspect*. 2015;123(9):902-908. doi:10.1289/ehp.1408700

34. 2009-11-19, "Census of Canada. Postal Code Conversion File, January 2003 Postal Codes, 2001", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/TF2D9 V3> [Version].
35. 2009-11-19, "Census of Canada. Postal Code Conversion File, July 2004 Postal Codes, 2001", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/SRHXX V2> [Version].
36. 2009-11-18, "Census of Canada. Postal Code Conversion File, October 2005 Postal Codes, 2001", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/92QKP V3> [Version].
37. 2009-11-19, "Census of Canada. Postal Code Conversion File, September 2006 Postal Codes, 2001", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/HQIOQ V2> [Version].
38. 2009-11-19, "Census of Canada. Postal Code Conversion File, September 2007 Postal Codes, 2006 [2007]", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/PVYYH V2> [Version].
39. 2009-11-18, "Census of Canada. Postal Code Conversion File, March 2008 Postal Codes, 2006 [2008]", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/LNEEC V3> [Version].
40. 2010-01-28, "Census of Canada. Postal Code Conversion File, July 2009 Postal Codes, 2006 [2009]", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/NTC6K V3> [Version].
41. Metro Vancouver. *2014 Lower Fraser Valley Air Quality Monitoring Report*. Burnaby, Canada; 2015.
42. Brauer M, Lencar C, Tamburic L, Koehoorn M, Demers P, Karr C. A Cohort Study of Traffic-Related Air Pollution Impacts on Birth Outcomes. *Environ Health Perspect*. 2008;116(5):680-686. doi:10.1289/ehp.10952
43. Smith RB, Fecht D, Gulliver J, et al. Impact of London's road traffic air and noise pollution on birth weight: retrospective population based cohort study. *BMJ*. 2017;359:j5299. doi:10.1136/bmj.j5299
44. Lampi KM, Lehtonen L, Tran PL, et al. Risk of autism spectrum disorders in low birth weight and small for gestational age infants. *J Pediatr*. 2012;161(5):830-836. doi:10.1016/j.jpeds.2012.04.058
45. Moore GS, Kneitel AW, Walker CK, Gilbert WM, Xing G. Autism risk in small- and large-for-gestational-age infants. In: *American Journal of Obstetrics and Gynecology*. Vol 206. Elsevier; 2012:314.e1-314.e9. doi:10.1016/j.ajog.2012.01.044

46. Leavey A, Zwaigenbaum L, Heavner K, Burstyn I. Gestational age at birth and risk of autism spectrum disorders in Alberta, Canada. *J Pediatr*. 2013;162(2):361-368. doi:10.1016/j.jpeds.2012.07.040
47. Wang R, Henderson SB, Sbihi H, Allen RW, Brauer M. Temporal stability of land use regression models for traffic-related air pollution. *Atmos Environ*. 2013;64:312-319. doi:10.1016/J.ATMOSENV.2012.09.056
48. Nethery E, Leckie SE, Teschke K, Brauer M. From measures to models: An evaluation of air pollution exposure assessment for epidemiological studies of pregnant women. *Occup Environ Med*. 2008;65(9):579-586. doi:10.1136/oem.2007.035337
49. Nethery E, Brauer M, Janssen P. Time-activity patterns of pregnant women and changes during the course of pregnancy. *J Expo Sci Environ Epidemiol*. 2008;19(3):317-324. doi:10.1038/jes.2008.24

## Chapter 3.

### Conclusion

#### 3.1. Summary

The objective of this thesis was to investigate air pollution as a potential risk factor for autism spectrum disorder (ASD). We assessed associations between prenatal exposures to ambient PM<sub>2.5</sub>, NO, and NO<sub>2</sub> and development of ASD.

In Chapter 2, we reported the results of a retrospective, population-based cohort study of all births from 2004–2009 in Metro Vancouver on the relationship between prenatal air pollution exposures and the odds of developing ASD using logistic regression models. All children positively identified with ASD through the provincial health system were evaluated by the BC Autism Assessment Network (BCAAN) using standardized diagnostic protocols and reliance on the Autism Diagnostic Observation Schedule and the Autism Diagnostic Interview-Revised diagnostic tools. Exposure estimates for each pregnant woman were calculated using temporally adjusted, high-resolution land use regression (LUR) models. The original LUR models were developed for epidemiological applications <sup>1</sup> and have been shown to predict air pollution concentrations reliably from 2003–2010 <sup>2</sup>. Population Data BC provided us with the capability to link all administrative health data into a large dataset for our cohort.

Our study, which is comprised of a cohort of 132,256 births with 1,307 (1.0%) children diagnosed with ASD, is among the largest studies to date on air pollution and ASD. We developed spatially and temporally resolved monthly exposure estimates at the residential address of each pregnant woman from conception to birth and calculated average exposures for the entire pregnancy and each trimester. In our analysis, we observed correlations among air pollutant exposures and high correlation across trimesters exposures, limiting our ability to test whether the odds of ASD were greater in the third trimester. Our results showed positive associations for all three pollutants and ASD, with stronger associations for males. We observed no differences in odds across trimesters.

Our findings were shared at three conferences: the Semiahmoo Symposium on Environmental, Occupational, and Public Health (January 2017), the 29th Annual Scientific Conference of the International Society for Environmental Epidemiology (September 2017), and the Cascadia Annual Symposium on Environmental, Occupational and Population Health (January 2018). A version of Chapter 2 was submitted to *JAMA Pediatrics* for peer-review.

## **3.2. Limitations and Future Work**

The strengths of our study include capturing nearly the entire population of interest, using strict diagnostic criteria, and developing high-resolution spatiotemporal models for exposure assessment. As well, our study is among the few that attempted to examine windows of vulnerability during prenatal development. Despite these study strengths, we acknowledge some important limitations, which we explore in the sections below.

### **3.2.1. Area vs Individual-Level Air Pollution Exposures**

Instead of using personal air monitoring devices to measure individual-level exposure, we used LUR models to estimate area-level measurements at residential locations to infer personal exposures. In a study comparing short-term, personal exposures of pregnant women to air pollution versus ambient air pollution estimates from the LUR models developed by Henderson et al. <sup>1</sup>, Nethery et al. <sup>3</sup> observed moderate correlations for NO and NO<sub>2</sub> and increased correlation when using both home and work locations. The alternative, using personal air monitoring devices to measure individual exposures for all pregnant women in Metro Vancouver over the course of their pregnancy, would be cost prohibitive. Additionally, sampling pregnant women and prospectively measuring exposures would be challenging because ASD is a rare outcome.

Dispersion and LUR models were the most common exposure assessment methods used by comparable air pollution and ASD studies <sup>4</sup>. Dispersion models simulate how pollutants are transported and dispersed in the atmosphere but are not feasible at a high spatial resolution for large areas <sup>1</sup>. Spatially interpolating air pollution concentrations using air quality monitoring stations can identify regional patterns but

cannot capture small-scale, localized variation <sup>1</sup>. The advantage of a LUR model over other exposure assessment methods is its ability to provide air pollution estimates at high spatial resolutions over large areas. The LUR models used in our study had 10 m<sup>2</sup> spatial resolution <sup>1</sup>, and we extracted air pollution concentrations at postal codes that represent a small geographic area, approximately one side of a city block <sup>5</sup>.

### **3.2.2. Timing of Exposure and Critical Windows of Vulnerability**

One of the key study questions that our study was unable to test adequately was whether exposure during the third trimester, compared to the first and second, had increased odds of ASD from air pollution exposure compared to other trimesters or the full duration of pregnancy. Identifying increased windows of vulnerability is important because it would provide clues into ASD etiology by identifying potential developmental processes that may be altered by air pollution. We found no difference in odds between trimesters, but high correlation of pollutant concentrations across trimesters limited our analyses. Moreover, attempting to simultaneously control for trimester exposures using mutually adjusted trimester models produced very wide confidence intervals. This problem could be minimized by incorporating residential history during pregnancy into our exposure assessment or by utilizing a natural experiment study design that compares typically low versus unusually high air pollution periods, such as during forest fires.

Incorporating residential history during pregnancy into exposure estimates could potentially increase exposure variation and reduce correlation of trimester exposures if pregnant women move between areas with noticeably different air pollution levels. Residential history of mothers who moved during pregnancy could increase exposure variability because air pollution variation is generally greater across space compared to time. In a study examining traffic-related air pollution impacts on birth outcomes in the same study area, Brauer et al. <sup>6</sup> observed that approximately 40% of mothers had at least two addresses during their pregnancy. Incorporating residential history would also have the added benefit of reducing exposure assessment bias. Studies that have reported higher associations in later trimesters may be observing bias as a result of utilizing residential address at birth <sup>4,7,8</sup>. This address better reflects a pregnant woman's residential location during the trimesters closest to birth and may not represent the location—and therefore the exposure—of earlier stages of pregnancy.

Additionally, although our study focused on trimester exposures, child development continues beyond birth. The associations between air pollution and ASD may be different between prenatal and postnatal periods<sup>8–12</sup>. Extending exposure estimates from prenatal to perinatal periods would allow us to test whether these associations differ and if they are stronger before or after birth.

Another way to answer the critical windows of vulnerability question is to rely on a natural experiment with a period of atypical air pollution levels, similar to the Rich et al.<sup>13</sup> study in China that compared the effect of decreased air pollution exposure for pregnant women during the 2008 Beijing Olympics on child birth weight. We could, for example, design a cross-sectional study of women at various stages of pregnancy and monitor for increased exposures of PM<sub>2.5</sub> during forest fire activity. The challenge for a study like this would be having a large enough sample to power the analyses mainly because ASD is a rare outcome.

### **3.2.3. Case Definition for ASD**

Outcome assessment was based on BCAAN diagnoses, but an alternative case definition was available through Ministry of Education data. Although BCAAN data have diagnoses that meet strict criteria, Ministry of Education data contain both BCAAN and diagnoses by private practitioners. In BC, ASD diagnostic services are available for free through the provincial health insurance system, but parents can also access private diagnostic services, costing approximately \$2,000–3,000<sup>14</sup>. Diagnoses made by private diagnosticians are not captured in the BCAAN dataset, but they are available in Ministry of Education data because school districts receive funding from the Ministry of Children and Family Development for in-school interventions and services for children who are enrolled in the public-school system and have been diagnosed with ASD.

Using both BCAAN and Ministry of Education datasets, we can identify 1,967 children with positive ASD diagnoses, compared to 1,307 children with ASD using BCAAN data alone. The current study case definition may be misclassifying 660 children who are positively identified with ASD in Ministry of Education data but not in BCAAN data. The BCAAN case definition is more conservative and has stricter diagnostic criteria, and it is unknown whether positive diagnoses rates differ between BCAAN and private diagnosticians. We compared average exposures for the full pregnancy period

among children with positive diagnoses of ASD in both BCAAN and Ministry of Education data (group A,  $n_a = 1,178$ ), only in BCAAN data (group B,  $n_b = 129$ ), and only in Ministry of Education data (group C,  $n_c = 660$ ). Mean exposures of  $PM_{2.5}$  were 3.7, 3.6, and 3.5  $\mu\text{g}/\text{m}^3$  for groups A, B, and C, respectively; NO, 21.8, 21.7, and 20.5 ppb; and  $NO_2$ , 15.8, 15.8, and 15.1 ppb. Across all pollutants, children in group C had lower average exposures for the full pregnancy period compared to children in groups A and B: absolute differences in exposure, using group C as a reference, ranged from 0.24–0.15  $\mu\text{g}/\text{m}^3$  for  $PM_{2.5}$ , 1.15–1.25 ppb of NO, and 0.64–0.66 ppb for  $NO_2$ . These differences were not meaningful, and we do not suspect differential misclassification by exposure.

This limitation is important because socioeconomic status may be associated with both exposure and outcome: low socioeconomic status is typically associated with higher air pollution exposure<sup>15,16</sup>, and higher socioeconomic status families are more likely to pay for a diagnosis by a private practitioner. Socioeconomic status is, therefore, a potential confounder. Our models control for socioeconomic status using an area-level income variable and maternal age at birth, but some residual confounding may remain that could contribute to the uncertainty of an alternative Ministry of Education case definition. Performing a sensitivity analysis that compares results between both case definitions would strengthen our study.

### **3.2.4. Phenotypes and Etiological Heterogeneity**

Individuals diagnosed with ASD present with a wide range of symptoms and considerable clinical heterogeneity, making it a challenge for health studies that compare individuals with and without ASD<sup>17</sup>. ASD is a collection of sign and symptoms with no validated biomarkers for diagnostic screening, so various ASD phenotypes could have various etiological pathways. Dividing individuals with ASD along sub-groups based on differences in behavioural symptoms, comorbidities, or metabolic and immune abnormalities can help distinguish the various potential physiological mechanisms for ASD<sup>17</sup>. Sub-group analyses of ASD subtypes by comorbidity, symptoms and severity, or ADOS and ADI-R score modules may highlight if air pollution has different impacts across different etiological pathways.

### **3.2.5. Gene-Environment Interactions**

ASD is a complex disorder with both genetic and environmental risk factors, but few studies have examined both risk factors together. Four potential gene-environment interaction mechanisms are described by Bakulski et al.<sup>18</sup>: effect of exposure modified by genetics, effect of genetics modified by environment, joint effects of both gene and environment, and effect of environment mediated by genetics. To date, only one air pollution study has explored gene-environment interactions: Volk et al. reported elevated risk of ASD in children with the MET rs1858830 CC genotype and high air pollution exposure<sup>19</sup>. Adopting a gene-environment paradigm is novel and promises to advance research on ASD and its etiology.

### **3.3. Conclusion and Policy Implications**

Our study, which found that ASD is associated with prenatal air pollution exposures, suggests that reducing exposure to air pollutants for pregnant women could help reduce ASD incidence. We observed a small effect size, but this small risk can result in a large burden of disease because exposure to air pollution is ubiquitous.

Individual-level interventions to reduce exposure, such as recommending that pregnant women avoid high pollution areas or install air filtration systems at home, are one approach. However, population-based approaches may be longer-lasting and more effective at improving population health, especially for exposures that are widespread. Strategies to reduce air pollution and meet air quality targets have been established by the Canada-wide Air Quality Management System and regionally by Metro Vancouver. More research, however, can be done to examine the health effects at levels below current air quality standards to help evaluate acceptable levels of air pollution. Additionally, motor vehicle traffic is a significant source of nitrogen oxides, so changes to the built environment that promote activate transportation<sup>20</sup> and set minimum distances between sources and individuals<sup>21</sup>, and policies that establish low-emission zones<sup>21</sup>, incentivize alternative fuel vehicles, and ban gas and diesel cars<sup>22</sup> could help reduce air pollution further.

We hope our findings encourage researchers to continue investigating the role of air pollution in neurodevelopmental disorders, and policy-makers to advance environmental policies that reduce the effects of air pollution on human health.

### 3.4. References

1. Henderson SB, Beckerman B, Jerrett M, Brauer M. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ Sci Technol*. 2007;41(7):2422-2428. doi:10.1021/es0606780
2. Wang R, Henderson SB, Sbihi H, Allen RW, Brauer M. Temporal stability of land use regression models for traffic-related air pollution. *Atmos Environ*. 2013;64:312-319. doi:10.1016/J.ATMOSENV.2012.09.056
3. Nethery E, Leckie SE, Teschke K, Brauer M. From measures to models: An evaluation of air pollution exposure assessment for epidemiological studies of pregnant women. *Occup Environ Med*. 2008;65(9):579-586. doi:10.1136/oem.2007.035337
4. Flores-Pajot MC, Ofner M, Do MT, Lavigne E, Villeneuve PJ. Childhood autism spectrum disorders and exposure to nitrogen dioxide, and particulate matter air pollution: A review and meta-analysis. *Environ Res*. 2016;151:763-776. doi:10.1016/j.envres.2016.07.030
5. Clark C, Sbihi H, Tamburic L, Brauer M, Frank LD, Davies HW. Association of long-term exposure to transportation noise and traffic-related air pollution with the incidence of diabetes: A prospective cohort study. *Environ Health Perspect*. 2017;125(8). doi:10.1289/EHP1279
6. Brauer M, Lencar C, Tamburic L, Koehoorn M, Demers P, Karr C. A Cohort Study of Traffic-Related Air Pollution Impacts on Birth Outcomes. *Environ Health Perspect*. 2008;116(5):680-686. doi:10.1289/ehp.10952
7. Lam J, Sutton P, Kalkbrenner A, et al. A systematic review and meta-analysis of multiple airborne pollutants and autism spectrum disorder. Kesavachandran CN, ed. *PLoS One*. 2016;11(9):e0161851. doi:10.1371/journal.pone.0161851
8. Kalkbrenner AE, Windham GC, Serre ML, et al. Particulate matter exposure, prenatal and postnatal windows of susceptibility, and autism spectrum disorders. *Epidemiology*. 2015;26(1):30-42. doi:10.1097/EDE.0000000000000173
9. Talbott EO, Arena VC, Rager JR, et al. Fine particulate matter and the risk of autism spectrum disorder. *Environ Res*. 2015;140:414-420. doi:10.1016/j.envres.2015.04.021
10. Volk HE, Lurmann F, Penfold B, Hertz-Picciotto I, McConnell R. Traffic-Related Air Pollution, Particulate Matter, and Autism. *JAMA Psychiatry*. 2013;70(1):71. doi:10.1001/jamapsychiatry.2013.266

11. Raz R, Levine H, Pinto O, Broday DM, Yuval, Weisskopf MG. Traffic-Related Air Pollution and Autism Spectrum Disorder: A Population-Based Nested Case-Control Study in Israel. *Am J Epidemiol*. August 2017. doi:10.1093/aje/kwx294
12. Raz R, Roberts AL, Lyall K, et al. Autism spectrum disorder and particulate matter air pollution before, during, and after pregnancy: A nested case-control analysis within the nurses' health study II cohort. *Environ Health Perspect*. 2015;123(3):264-270. doi:10.1289/ehp.1408133
13. Rich DQ, Liu K, Zhang J, et al. Differences in Birth Weight Associated with the 2008 Beijing Olympic Air Pollution Reduction: Results from a Natural Experiment. *Environ Health Perspect*. April 2015. doi:10.1289/ehp.1408795
14. Autism Community Training. Living & Working with Children with ASD in British Columbia. A Manual for Parents & Community Professionals. 2016. [www.actcommunity.ca/autism-manual](http://www.actcommunity.ca/autism-manual).
15. Becerra TA, Wilhelm M, Olsen J, Cockburn M, Ritz B. Ambient air pollution and autism in Los Angeles County, California. *Environ Health Perspect*. 2013;121(3):380-386. doi:10.1289/ehp.1205827
16. Weisskopf MG, Kioumourtzoglou M-A, Roberts AL. Air Pollution and Autism Spectrum Disorders: Causal or Confounded? *Curr Environ Heal Reports*. 2015;2(4):430-439. doi:10.1007/s40572-015-0073-9
17. Hu VW. Phenotype Definition: A Cornerstone of Autism Research, Diagnosis and Therapy. In: *Frontiers in Autism Research: New Horizons for Diagnosis and Treatment*. ; 2014:3-22. doi:10.1142/9789814602167\_0001
18. Bakulski KM, Singer AB, Fallin MD. Genes and Environment in Autism Spectrum Disorders: An Integrated Perspective. In: Hu VW, ed. *Frontiers in Autism Research: New Horizons for Diagnosis and Treatment*. World Scientific; 2014:335-374. doi:10.1142/9789814602167\_0014
19. Volk HE, Kerin T, Lurmann F, Hertz-Picciotto I, McConnell R, Campbell DB. Autism Spectrum Disorder: Interaction of Air Pollution with the MET Receptor Tyrosine Kinase Gene. *Epidemiology*. 2014;25(1):44-47. doi:10.1097/EDE.0000000000000030
20. Rabl A, de Nazelle A. Benefits of shift from car to active transport. *Transp Policy*. 2012;19(1):121-131. doi:10.1016/j.tranpol.2011.09.008
21. Giles L V., Barn P, Künzli N, et al. From Good Intentions to Proven Interventions: Effectiveness of Actions to Reduce the Health Impacts of Air Pollution. *Environ Health Perspect*. 2010;119(1):29-36. doi:10.1289/ehp.1002246

22. Castle S. Britain to Ban New Diesel and Gas Cars by 2040 - The New York Times. The New York Times.  
<https://www.nytimes.com/2017/07/26/world/europe/uk-diesel-petrol-emissions.html>. Published 2017. Accessed May 27, 2018.

## Appendix A.

### Chapter 2 Manuscript Abstract

**Importance:** The etiology of autism spectrum disorder is poorly understood, but prior studies suggest associations with environmental exposures.

**Objective:** To evaluate the association between prenatal exposures to airborne pollutants and autism spectrum disorder in a large population-based cohort.

**Design:** We developed a cohort encompassing nearly all births in Metro Vancouver, Canada, from 2004–2009, with follow-up through 2014. Children were diagnosed with autism spectrum disorder using a standardized assessment of Autism Diagnostic Interview-Revised and Autism Diagnostic Observation Schedule. We used temporally adjusted, high-resolution land use regression models to estimate monthly mean exposures to particulate matter with a diameter of less than 2.5 micrometers (PM<sub>2.5</sub>), nitric oxide (NO), and nitrogen dioxide (NO<sub>2</sub>) at the maternal residence during pregnancy. Logistic regression models were used to evaluate the relationship between prenatal air pollution exposures and the odds of developing autism spectrum disorder, adjusting for child sex, birth month, birth year, maternal age, maternal birthplace, and neighborhood-level urbanicity and income band.

**Setting:** Population-based cohort in Vancouver, British Columbia, Canada.

**Participants:** A cohort of 132,256 births, with 1,307 (1.0%) children diagnosed with autism spectrum disorder by five years of age.

**Results:** Adjusted odds ratios (OR) for ASD per interquartile range (IQR) (1.5 µg/m<sup>3</sup>) of PM<sub>2.5</sub> during pregnancy was 1.04 (95% CI, 0.98–1.10), 1.07 (95% CI, 1.01–1.13) per IQR (10.7 ppb) of NO, and 1.06 (95% CI, 0.99–1.12) per IQR (4.8 ppb) of NO<sub>2</sub>. Effect estimates were larger for males than females for all pollutants, with percentage differences ranging from 1.0–10.6%.

**Conclusions:** We observed associations between exposure to air pollutants and incidence of autism spectrum disorder in a population-based birth cohort, especially among males.

## Appendix B.

### Chapter 2 Supplemental Tables and Figures

**Table B1. Median monthly average prenatal exposure estimates among children born in Metro Vancouver, Canada, between 2004–2009.**

Pollutant	Units	Time Period	Sample Size	Median Monthly Average Exposure Estimates (IQR) <sup>a</sup>	
PM <sub>2.5</sub>	µg/m <sup>3</sup>	Full Pregnancy	131,440	3.5	(2.7, 4.2)
		Trimester 1	131,447	3.4	(2.6, 4.3)
		Trimester 2	131,640	3.4	(2.6, 4.3)
		Trimester 3	131,206	3.4	(2.5, 4.3)
NO	ppb	Full Pregnancy	131,437	18.3	(14.0, 24.7)
		Trimester 1	131,444	17.4	(10.0, 28.3)
		Trimester 2	131,637	16.6	(9.5, 27.8)
		Trimester 3	131,203	15.3	(9.1, 26.5)
NO <sub>2</sub>	ppb	Full Pregnancy	131,437	14.3	(12.2, 17.0)
		Trimester 1	131,444	14.5	(12.1, 17.6)
		Trimester 2	131,637	14.3	(11.9, 17.5)
		Trimester 3	131,203	14.0	(11.6, 17.2)

<sup>a</sup> Monthly average exposure estimates based on complete monthly exposures for the period of interest. Incomplete prenatal monthly exposures, approximately 0.5–0.8% of the cohort ( $n = 132,256$  children), are a result of postal codes retiring during the period of interest, preventing spatial referencing.

**Table B2. Sensitivity analyses of odds of ASD for prenatal exposure to air pollutants.**

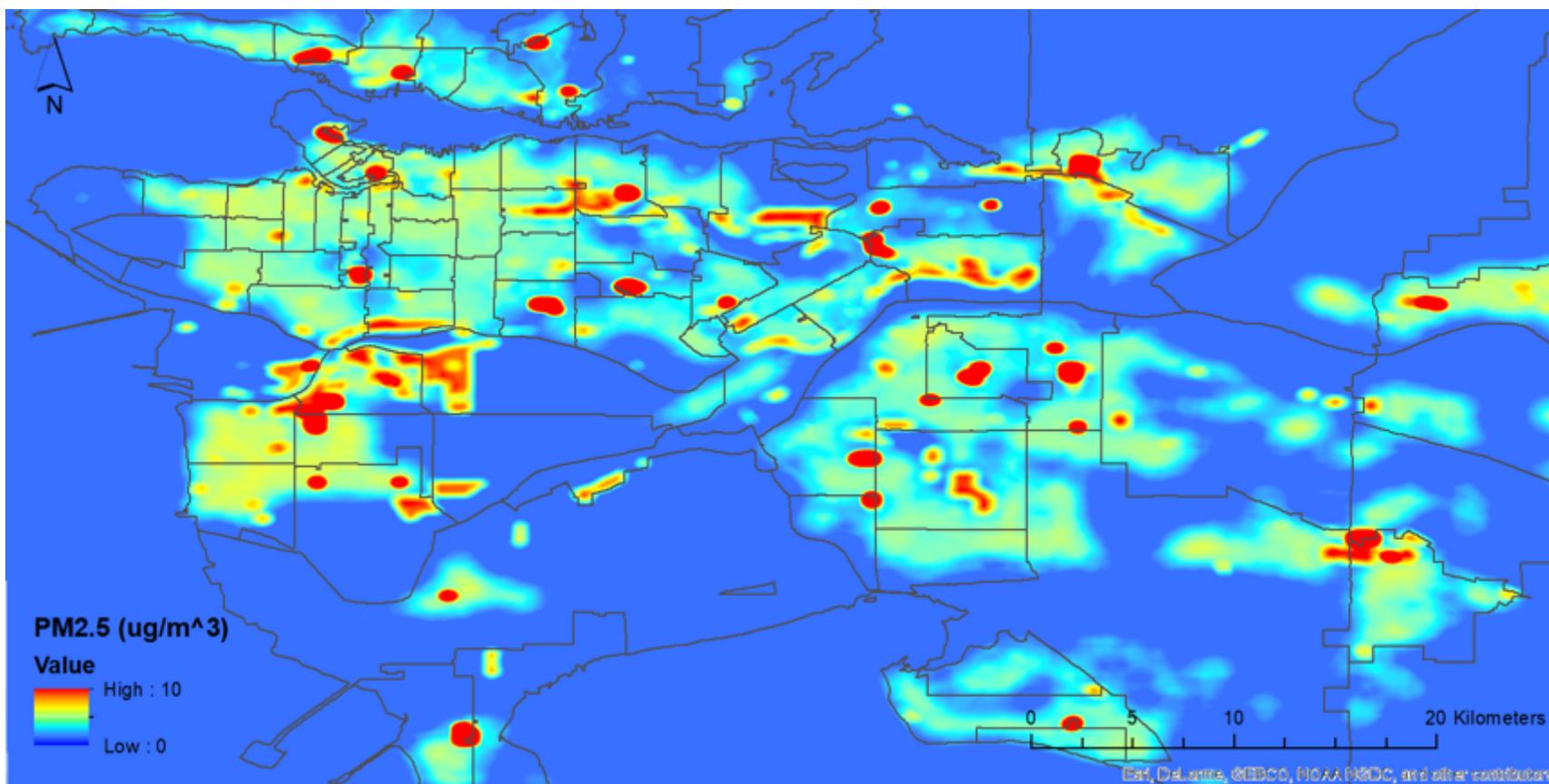
Pollutant	Units	Model <sup>a</sup>	Sample Size	Children with ASD	OR (95% CI)
PM <sub>2.5</sub>	Per IQR (1.5 µg/m <sup>3</sup> )	Unadjusted full pregnancy model	131,440	1,300	1.08 (1.03, 1.14)
		Adjusted full pregnancy model	129,439	1,276	1.04 (0.98, 1.10)
		Adjusted full pregnancy model + gestational age	129,439	1,276	1.04 (0.99, 1.10)
		Adjusted full pregnancy model + small for gestational age	129,423	1,276	1.04 (0.98, 1.10)
		Adjusted full pregnancy model + birth weight	129,423	1,276	1.04 (0.98, 1.10)
		Adjusted full pregnancy model + multiple birth	129,439	1,276	1.04 (0.98, 1.10)
		Adjusted full pregnancy model + parity	129,438	1,276	1.03 (0.97, 1.08)
		Adjusted full pregnancy model + smoking	129,439	1,276	1.04 (0.98, 1.10)
		Adjusted full pregnancy model replacing birth month with conception month	129,439	1,276	1.04 (0.98, 1.10)
		Adjusted full pregnancy model using non-truncated, temporally adjusted LUR	129,439	1,276	1.04 (0.98, 1.09)
		Adjusted full pregnancy model restricting follow-up to a maximum of five years	129,439	891	1.06 (1.00, 1.13)
NO	Per IQR (10.7 ppb)	Unadjusted full pregnancy model	131,437	1,300	1.09 (1.03, 1.14)
		Adjusted full pregnancy model	129,436	1,276	1.07 (1.01, 1.13)
		Adjusted full pregnancy model + gestational age	129,436	1,276	1.07 (1.01, 1.13)
		Adjusted full pregnancy model + small for gestational age	129,420	1,276	1.07 (1.01, 1.13)
		Adjusted full pregnancy model + birth weight	129,420	1,276	1.07 (1.01, 1.13)
		Adjusted full pregnancy model + multiple birth	129,436	1,276	1.07 (1.01, 1.13)
		Adjusted full pregnancy model + parity	129,435	1,276	1.05 (0.99, 1.11)
		Adjusted full pregnancy model + smoking	129,436	1,276	1.07 (1.01, 1.13)
		Adjusted full pregnancy model replacing birth month with conception month	129,436	1,276	1.07 (1.01, 1.13)
		Adjusted full pregnancy model using non-truncated, temporally adjusted LUR	129,436	1,276	1.07 (1.01, 1.13)
		Adjusted full pregnancy model restricting follow-up to a maximum of five years	129,436	891	1.06 (0.98, 1.13)
NO <sub>2</sub>	Per IQR (4.8 ppb)	Unadjusted full pregnancy model	131,437	1,300	1.11 (1.05, 1.17)
		Adjusted full pregnancy model	129,436	1,276	1.06 (0.99, 1.12)
		Adjusted full pregnancy model + gestational age	129,436	1,276	1.06 (1.00, 1.12)
		Adjusted full pregnancy model + small for gestational age	129,420	1,276	1.05 (0.99, 1.12)
		Adjusted full pregnancy model + birth weight	129,420	1,276	1.06 (0.99, 1.12)
		Adjusted full pregnancy model + multiple birth	129,436	1,276	1.06 (0.99, 1.12)
		Adjusted full pregnancy model + parity	129,435	1,276	1.03 (0.97, 1.09)
		Adjusted full pregnancy model + smoking	129,436	1,276	1.05 (0.99, 1.12)
		Adjusted full pregnancy model replacing birth month with conception month	129,436	1,276	1.06 (0.99, 1.12)
		Adjusted full pregnancy model using non-truncated, temporally adjusted LUR	129,436	1,276	1.06 (0.99, 1.12)
		Adjusted full pregnancy model restricting follow-up to a maximum of five years	129,436	891	1.07 (1.00, 1.15)

<sup>a</sup> Models adjusted for child sex, birth month, birth year, maternal age, maternal birthplace, and neighborhood-level urbanicity and income band.

**Table B3. Comparison of adjusted odds ratios between individual trimester models versus a mutually adjusted trimester model.**

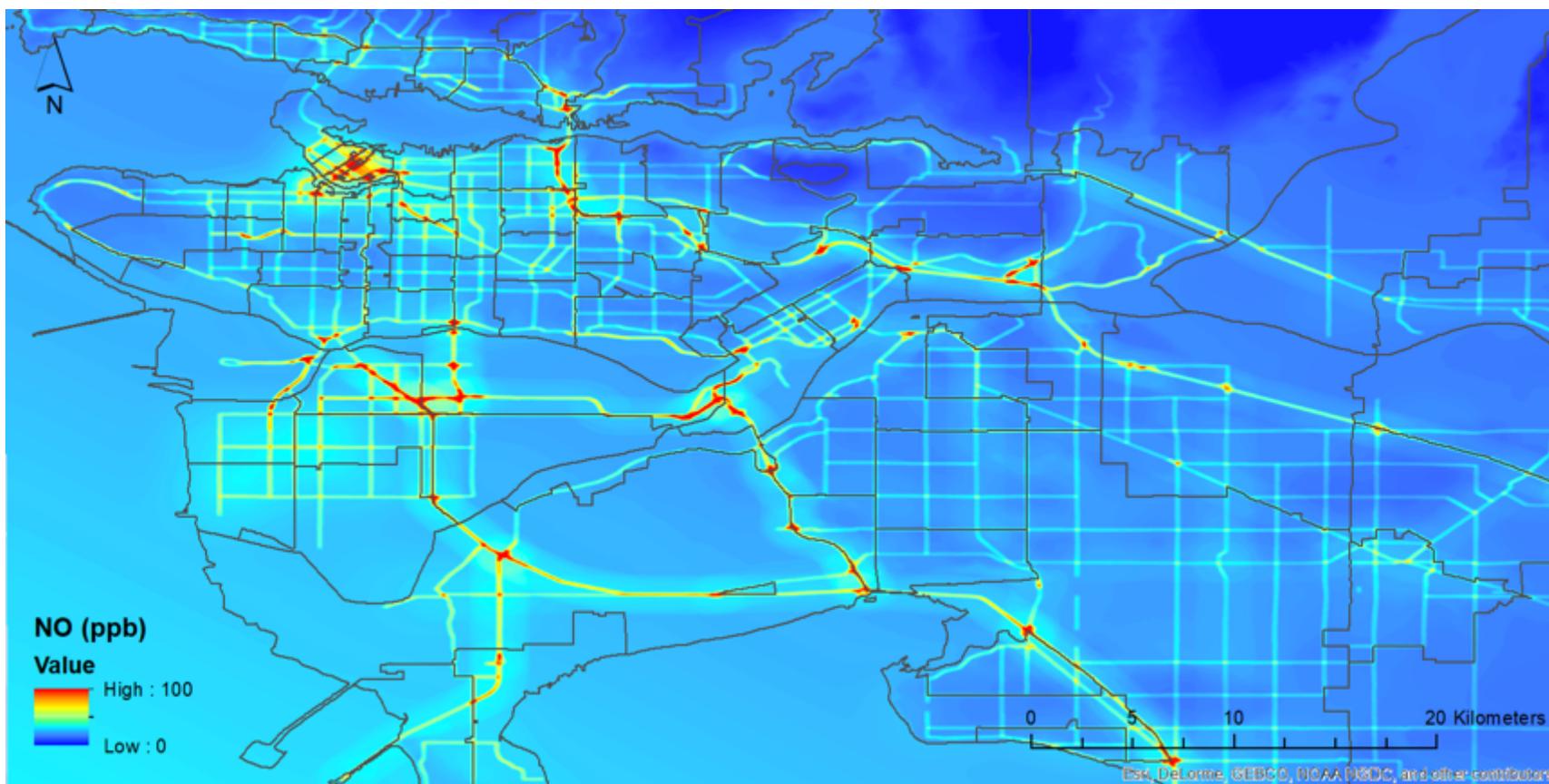
Pollutant	Units	Model <sup>a</sup>	Trimester	OR (95 %)
PM <sub>2.5</sub>	Per IQR (1.5 µg/m <sup>3</sup> )	Adjusted trimester 1 model	Trimester 1	1.03 (0.98, 1.09)
		Adjusted trimester 2 model	Trimester 2	1.04 (0.98, 1.09)
		Adjusted trimester 3 model	Trimester 3	1.04 (0.99, 1.10)
		Adjusted model with mutually adjusted trimesters	Trimester 1	0.92 (0.79, 1.08)
			Trimester 2	1.06 (0.90, 1.24)
			Trimester 3	1.06 (0.93, 1.22)
		NO	Per IQR (10.7 ppb)	Adjusted trimester 1 model
Adjusted trimester 2 model	Trimester 2			1.04 (0.99, 1.08)
Adjusted trimester 3 model	Trimester 3			1.04 (0.98, 1.09)
Adjusted model with mutually adjusted trimesters	Trimester 1			1.06 (1.00, 1.13)
	Trimester 2			0.98 (0.91, 1.05)
	Trimester 3			1.02 (0.96, 1.09)
NO <sub>2</sub>	Per IQR (4.8 ppb)	Adjusted trimester 1 model	Trimester 1	1.06 (1.00, 1.12)
		Adjusted trimester 2 model	Trimester 2	1.04 (0.98, 1.10)
		Adjusted trimester 3 model	Trimester 3	1.05 (0.99, 1.11)
		Adjusted model with mutually adjusted trimesters	Trimester 1	1.09 (0.93, 1.28)
			Trimester 2	0.90 (0.75, 1.09)
			Trimester 3	1.07 (0.90, 1.26)

<sup>a</sup> Models adjusted for child sex, birth month, birth year, maternal age, maternal birthplace, and neighborhood-level urbanicity and income band.



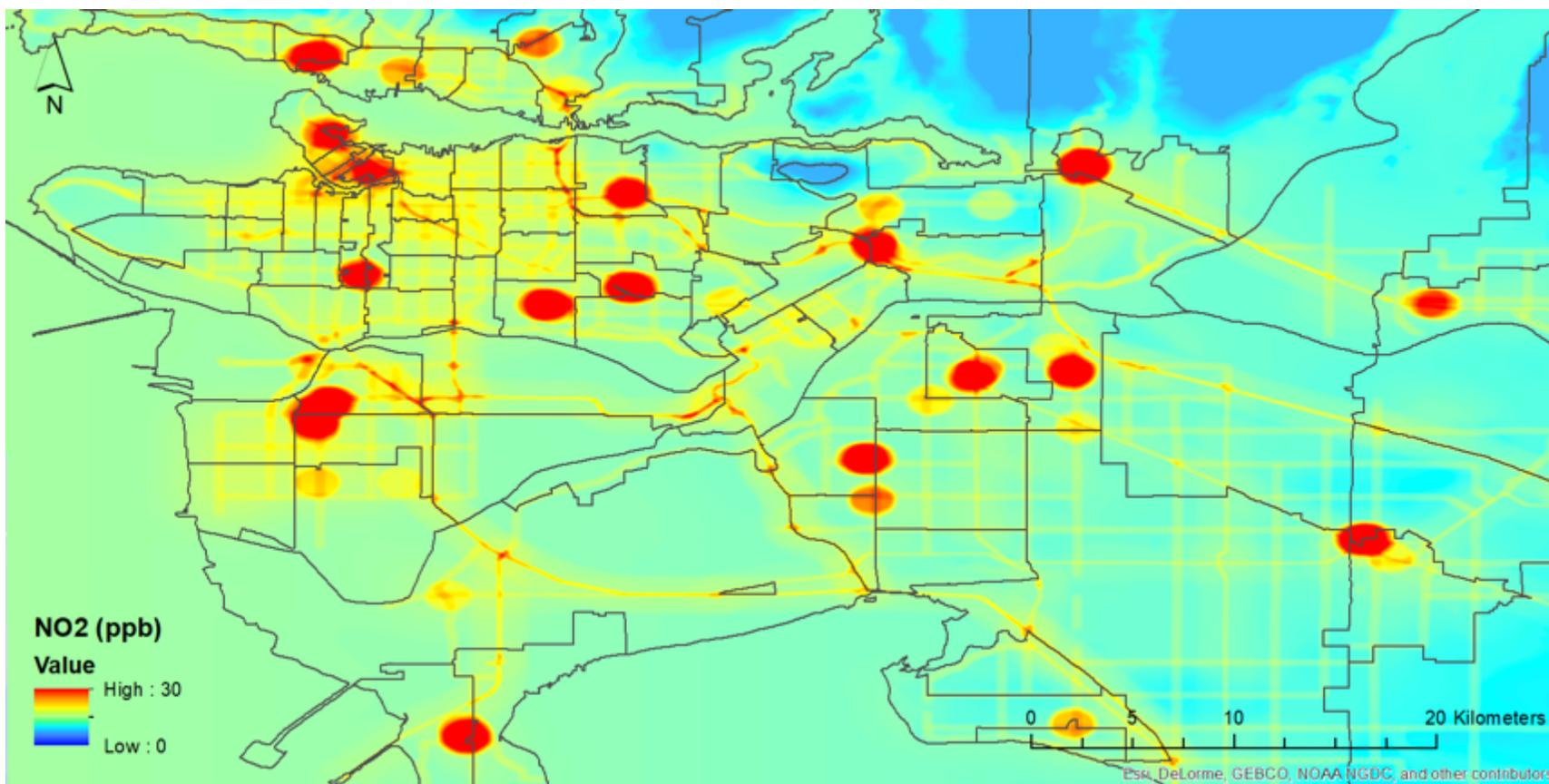
**Figure B1. Truncated land use regression model for PM<sub>2.5</sub> (2003) in Metro Vancouver, Canada.**

The original LUR models for Metro Vancouver were developed by Henderson et al. (2007) as exposure assessment tools for epidemiological applications. The original PM<sub>2.5</sub> LUR model was truncated to 1.4  $\mu\text{g}/\text{m}^3$  prior to temporal adjustment to prevent estimating values of zero.



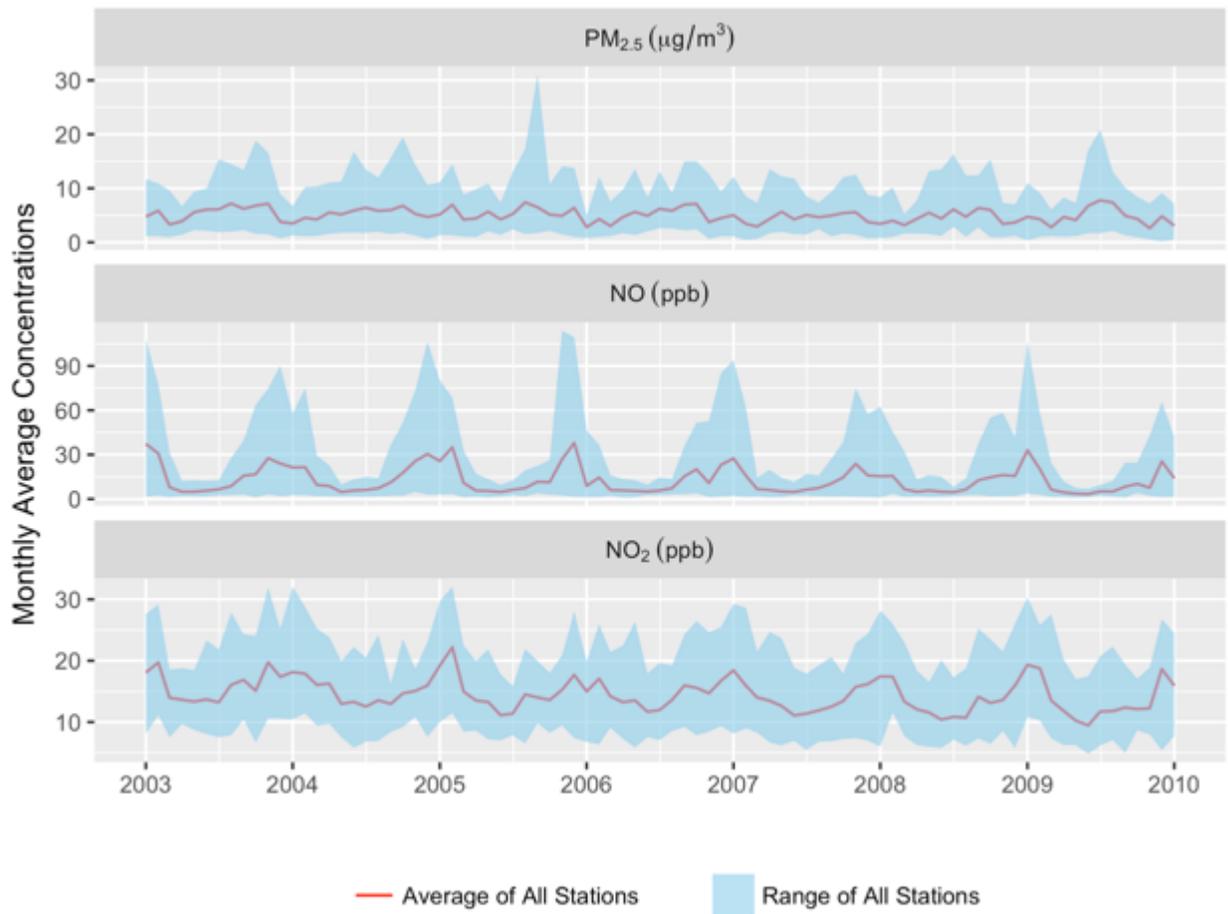
**Figure B2. Truncated land use regression model for NO (2003) in Metro Vancouver, Canada.**

The original LUR models for Metro Vancouver were developed by Henderson et al. (2007) as exposure assessment tools for epidemiological applications. The original NO LUR model was truncated to 1.3 ppb prior to temporal adjustment to prevent estimating values of zero.



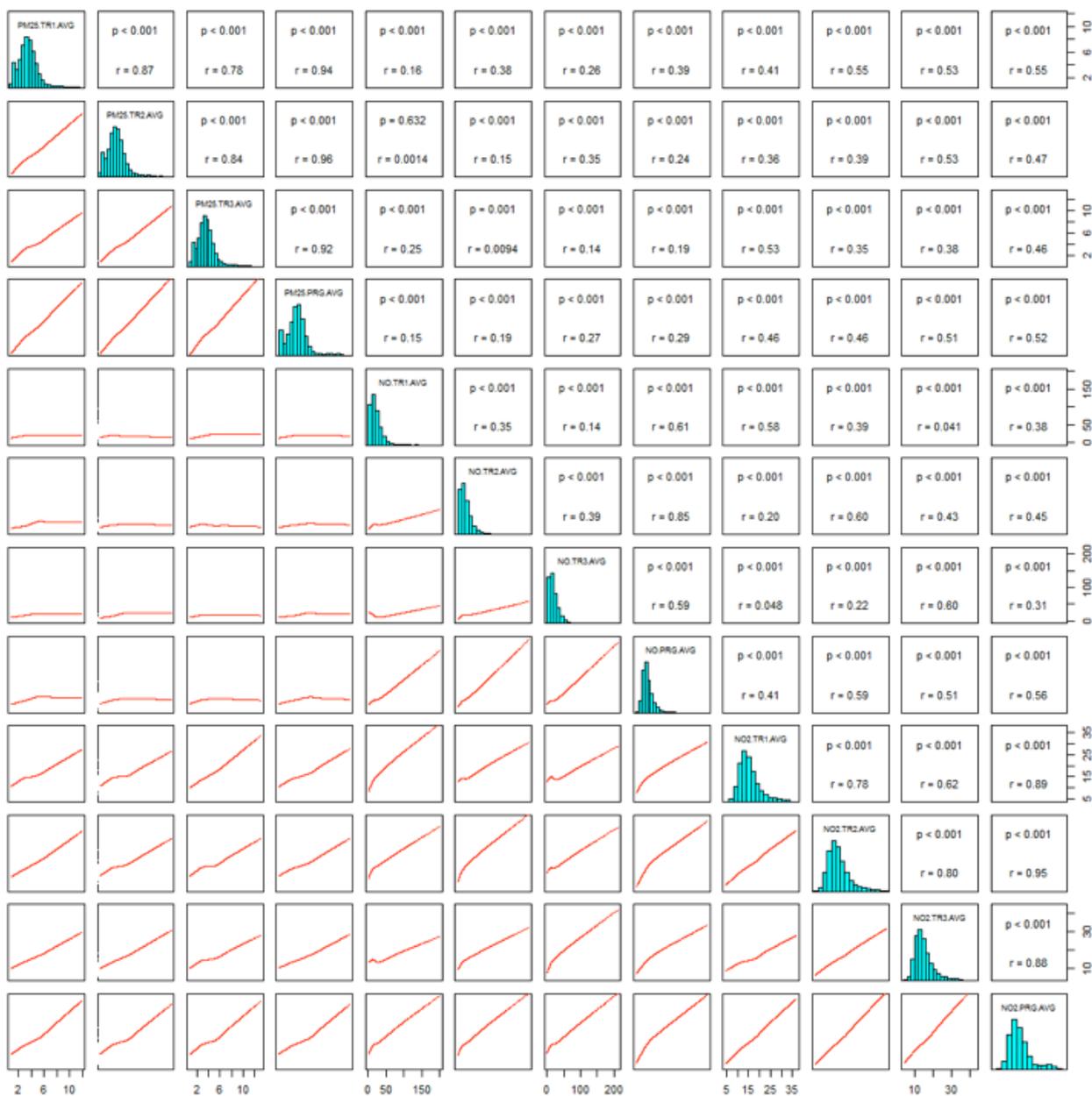
**Figure B3. Truncated land use regression model for NO<sub>2</sub> (2003) in Metro Vancouver, Canada.**

The original LUR models for Metro Vancouver were developed by Henderson et al. (2007) as exposure assessment tools for epidemiological applications. The original NO<sub>2</sub> LUR model was truncated to 6.9 ppb prior to temporal adjustment to prevent estimating values of zero.



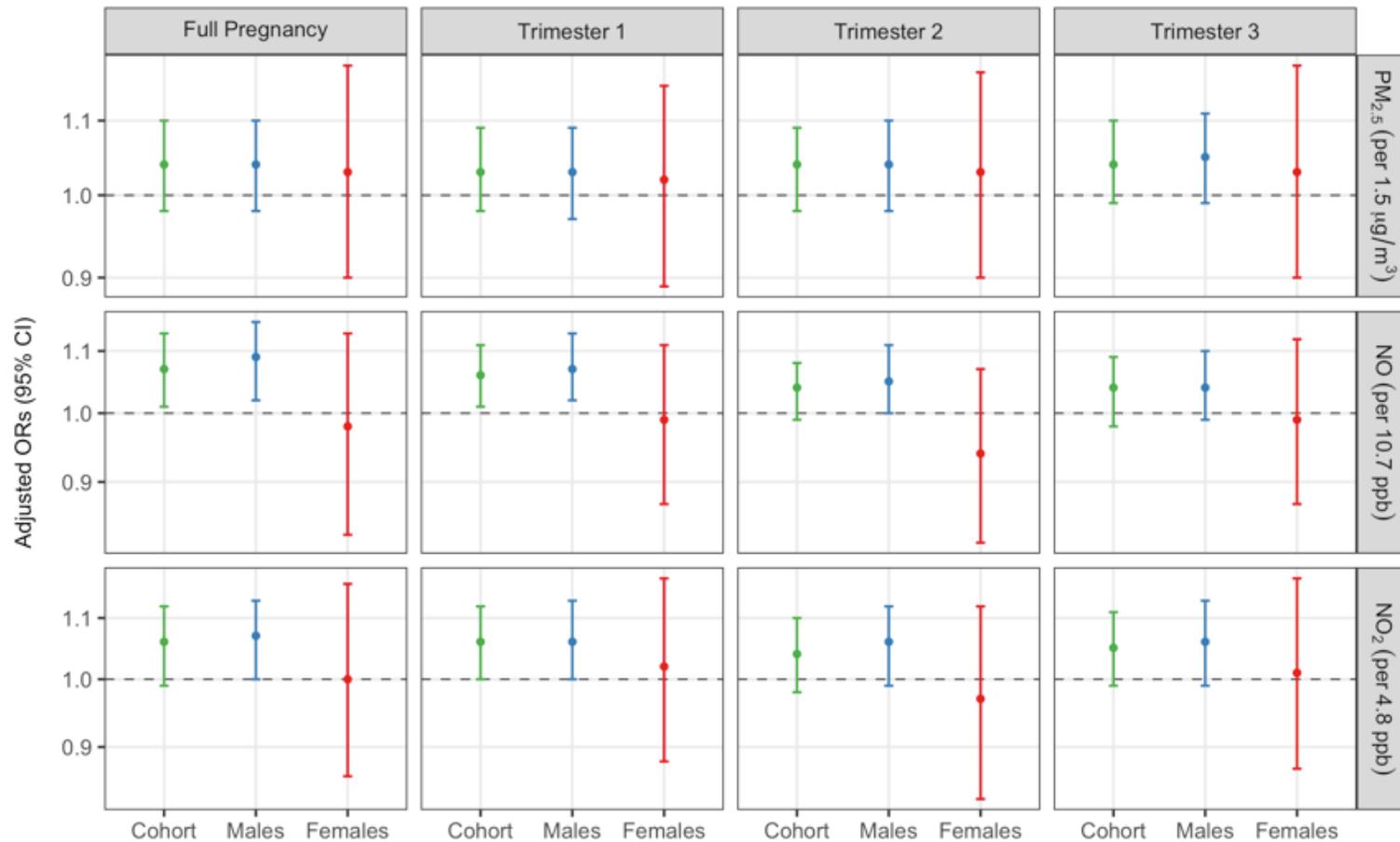
**Figure B4. Monthly average concentrations of PM<sub>2.5</sub>, NO, and NO<sub>2</sub> from air quality monitoring stations in Metro Vancouver, 2003–2009.**

Temporal trends in monthly average concentrations PM<sub>2.5</sub>, NO, and NO<sub>2</sub> from air quality monitoring stations in Metro Vancouver. Annual average concentrations have decreased slightly: from an annual average of 5.7 to 5.0 µg/m<sup>3</sup> for PM<sub>2.5</sub>, 16.0 to 11.0 ppb for NO, and 16.0 to 13.5 ppb for NO<sub>2</sub>. Concentrations of PM<sub>2.5</sub> tended to be higher in summer, whereas NO and NO<sub>2</sub> concentrations were higher in winter.



**Figure B5. Correlation matrix of air pollutant exposures across pregnancy and trimesters.**

The correlation matrix shows the distribution of exposure estimates for each pollutant across four time periods (i.e. full pregnancy and trimesters 1–3) and the correlation coefficient among all pollutant and time periods pairs. NO<sub>2</sub> showed moderate positive correlations with both PM<sub>2.5</sub> and NO. Correlations across time periods within pollutants was high for PM<sub>2.5</sub> and NO<sub>2</sub>. Abbreviations: PRG, full pregnancy; TR1, trimester 1; TR2, trimester 2; and TR3, trimester 3.



**Figure B6. Overall and sex-stratified odds of ASD for prenatal exposure to air pollutants during pregnancy and trimesters among children born in Metro Vancouver, Canada, between 2004–2009.**

Single pollutant and single time period models adjusted for child sex, birth month, birth year, maternal age, maternal birthplace, and neighborhood-level urbanicity and income band.

## Appendix C.

### Temporally Adjusted LUR Model Estimates: Protocols

Land use regression (LUR) models were temporally adjusted by extracting LUR values at each postal code, developing temporal factors using air monitoring data, and then multiplying LUR values by a temporal factor. We tested various methods to develop temporal factors and to truncate temporally adjusted minimum concentration values.

First, postal Code Conversion Files were used to associate all 6-digit postal codes within LUR model domains to a longitude and latitude. In a GIS, these 6-digit postal codes were mapped and overlaid on the LUR models to extract air pollutant values at each postal code location (Figure C1). Postal codes coordinates shift year to year (Figure C2), so this process was repeated for each study year with a respective Postal Code Conversion File.

Prior to developing temporal factors, missing air monitoring station data was imputed using Predictive Mean Matching (PMM), a semi-parametric imputation approach, using all available air monitoring data from 2003–2014. Missing air pollution data from one monitor were imputed using a regression of the other air monitors, and each missing value was assigned a random value among five observed values that were closest to the regression-predicted value. PMM produces imputed values using observed values, so the distribution between the imputed and observed data are similar, and the range of imputed values are always within the range of observed values. The probability density function between imputed and observed data were compared and inspected visually (Figures C3–C5).

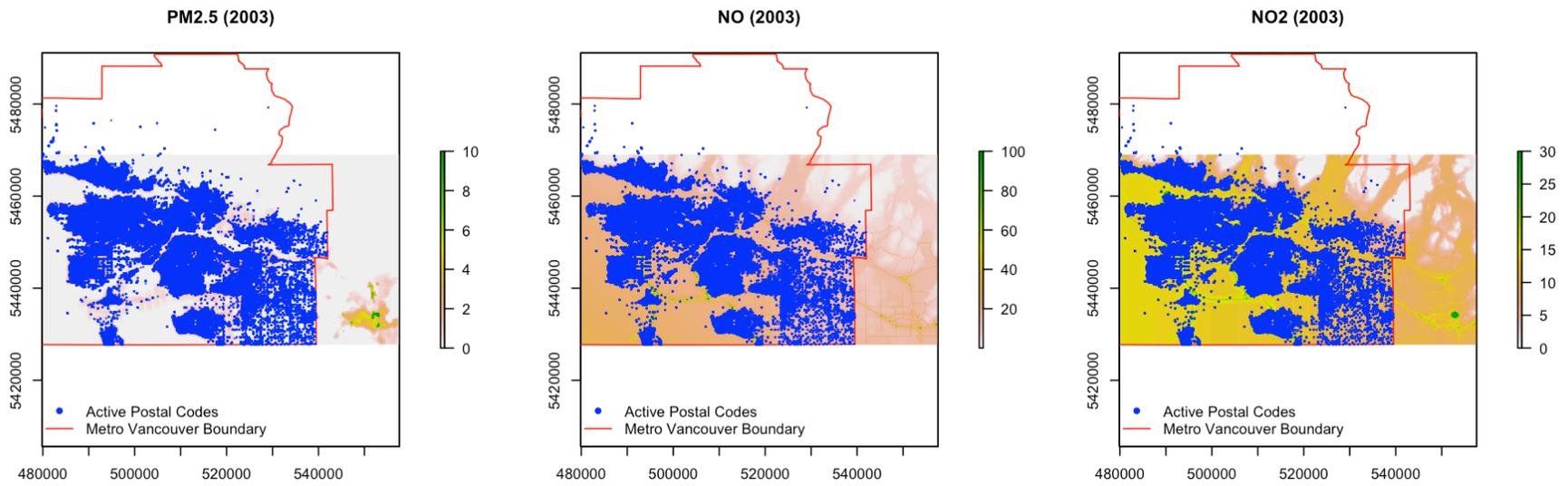
Approximately 11% of PM<sub>2.5</sub> air pollution monitoring data was missing across 9 air monitoring stations (excluding 5 air monitoring stations that have 100% missing data prior to becoming operational in 2010) and 2,557 days from 2003–2009, inclusive; between 1–47% missing data among air monitoring stations and between 3–36% per year. For NO and NO<sub>2</sub>, 3% of air pollution monitoring data was missing across 16 air monitoring stations (excluding 1 air monitoring station that has 100% missing data prior

to becoming operational in 2010) and 2,557 days from 2003–2009, inclusive; between 1–12% missing data among air monitoring stations and between 1–6% per year.

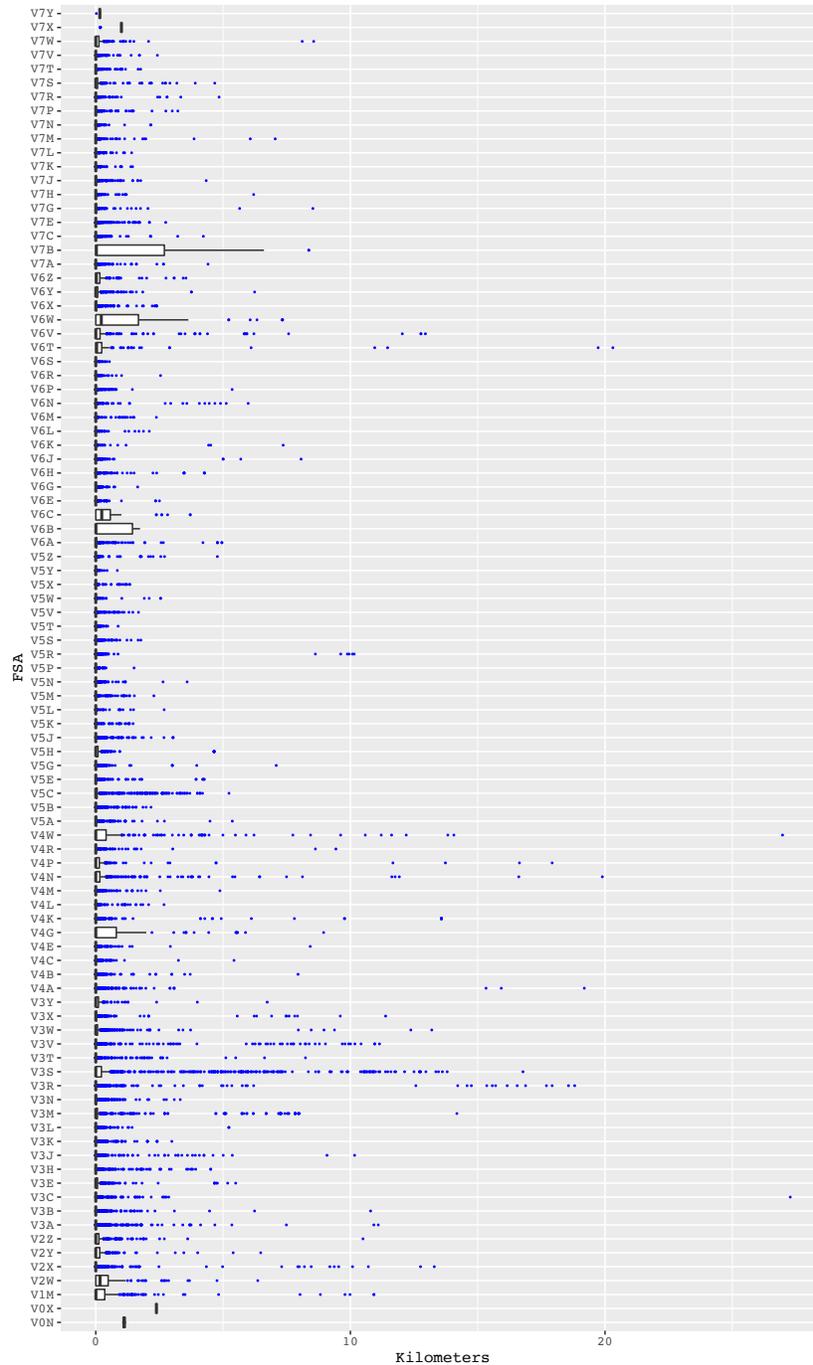
We then temporally adjusted the spatially resolved air pollution values using a temporal factor for each month of each study year using air monitoring data after imputation. The temporal factor was derived using air pollution monitoring data: a ratio between the average monthly air pollutant levels across all air monitors for the month of interest over the average yearly air pollutant levels across all air monitors for the LUR year. Temporal factors derived using median values produced larger temporally adjusted values, especially for NO.

Lastly, to prevent zero values, we truncated all zeros to a minimum value. Minimum concentrations were derived by calculating the mean from the lowest daily average in a month for all air monitors for each month of the LUR year. An alternative truncation method involved temporally adjusting the original spatially resolved values and then truncating minimum concentrations, derived by calculating the mean from the lowest daily average in a month for all air monitors for the predicted month. There were no meaningful differences between truncation methods. However, truncating LUR values before temporal interpolation modified fewer postal code values.

Dot and box plots in Figures C6–C8 show the temporally adjusted LUR model values for all postal codes across 2003–2009, comparing temporal adjustment results by: (a.) deriving temporal factors using mean versus median; (b.) deriving temporal factors using original versus imputed air monitoring data; and (c.) not truncating zeros, truncating minimum LUR-extracted values before temporal interpolation, or truncating minimum temporally adjusted LUR values.

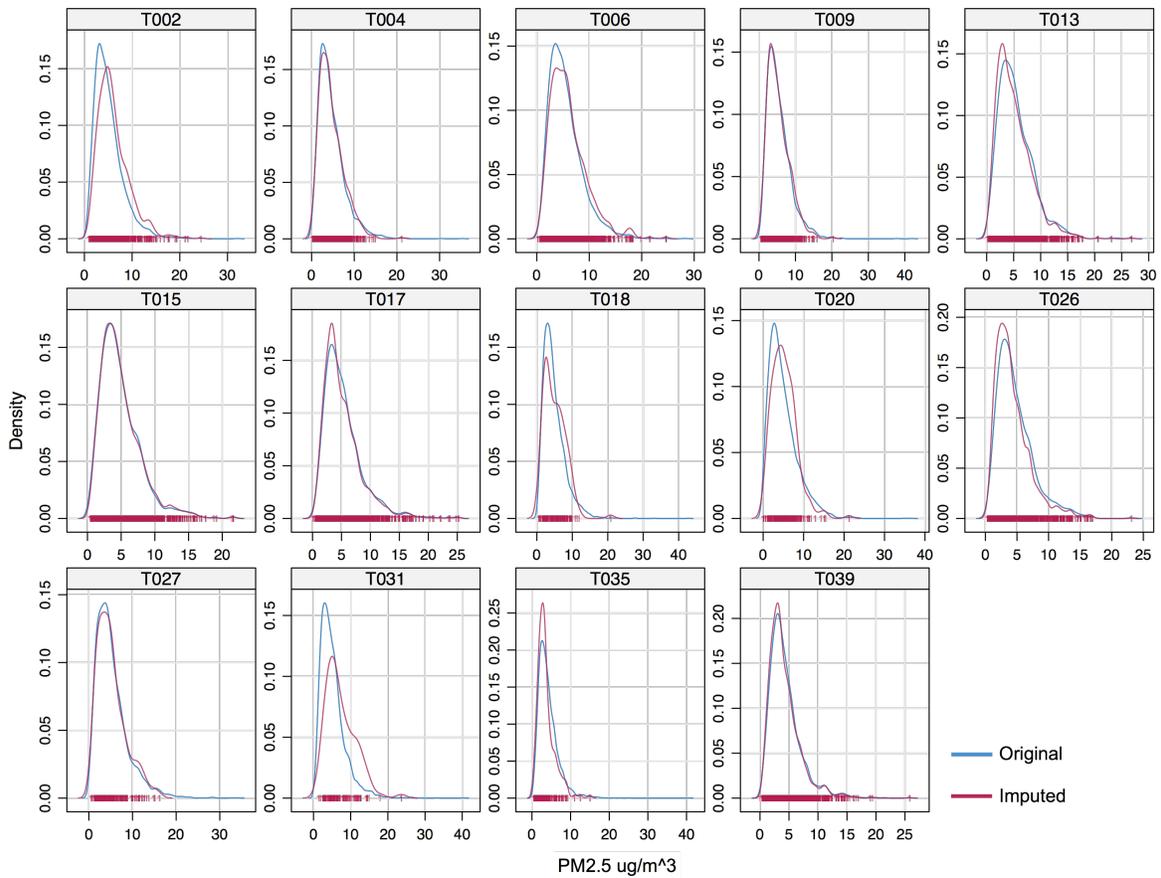


**Figure C1. Map of LUR surfaces overlaid with all active postal codes in Metro Vancouver from 2003–2009.**



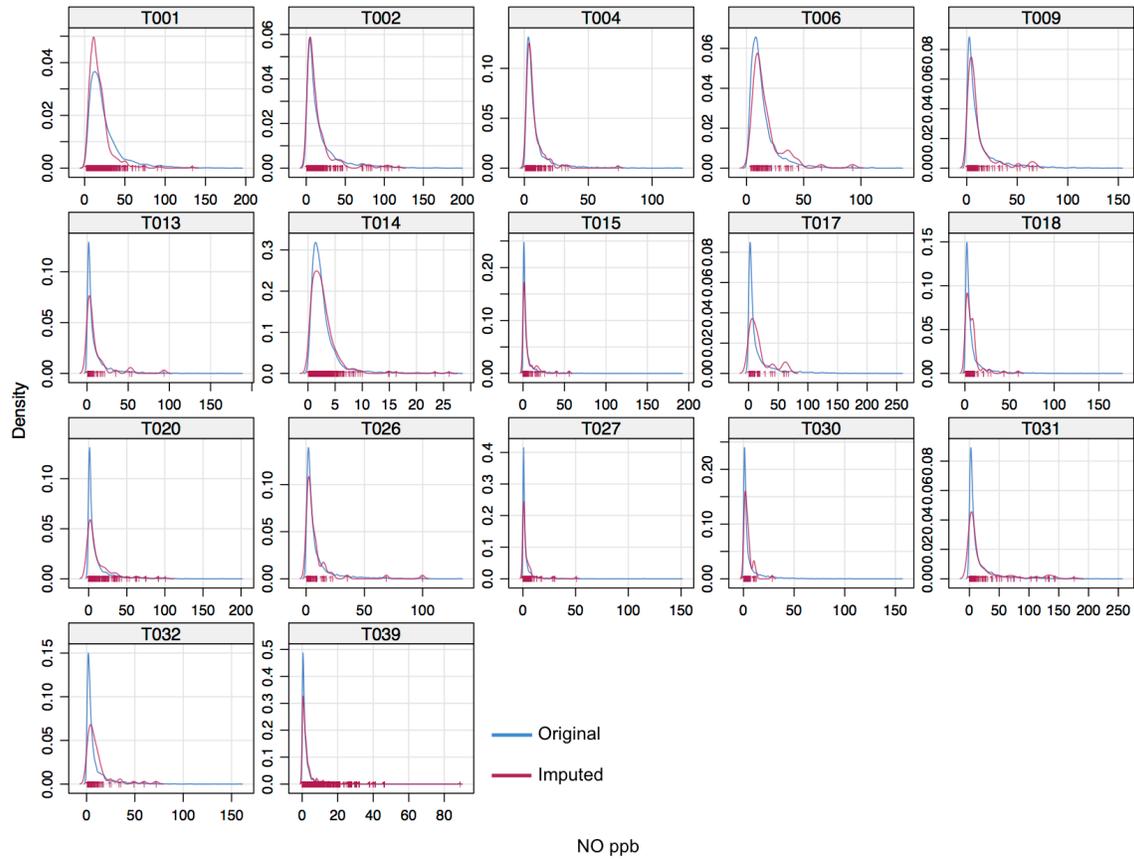
**Figure C2. Distance between postal code coordinate locations in 2003 versus 2009.**

Box-and-whisker plot showing the movement of postal codes, grouped by forward sortation areas, from 2003 to 2009. The point coordinates of postal codes shift from year to year, and this was our rationale for geocoding a postal code to its corresponding coordinate per year. A postal code covers a geographic area, but it is assigned a point coordinate in postal code conversion files by a single link indicator, which is meant to identify a geographic area with the majority of dwellings assigned to that particular postal code.



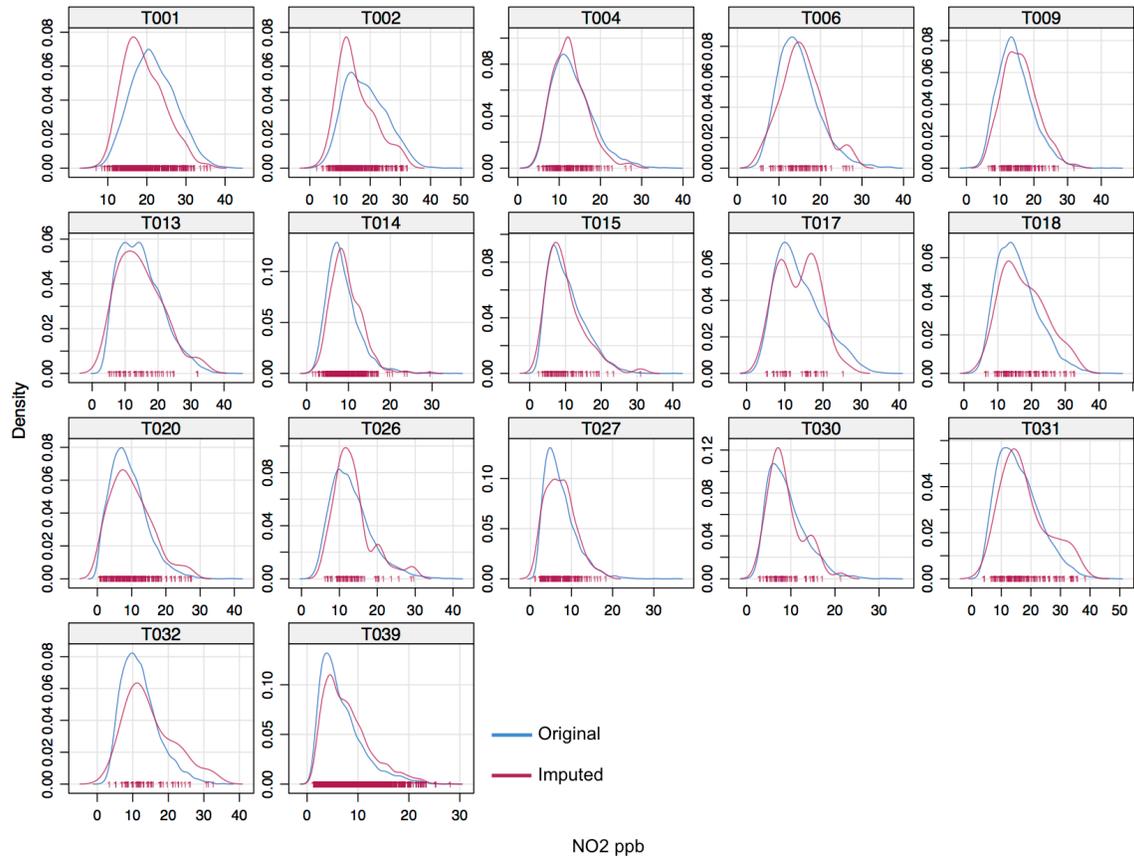
**Figure C3. Density plots of original versus imputed air monitoring data for PM<sub>2.5</sub> using predictive mean matching.**

Density plots, one for each air quality monitoring station, were developed to visually compare the distribution between original air monitoring data, containing missing values, versus complete, imputed values.



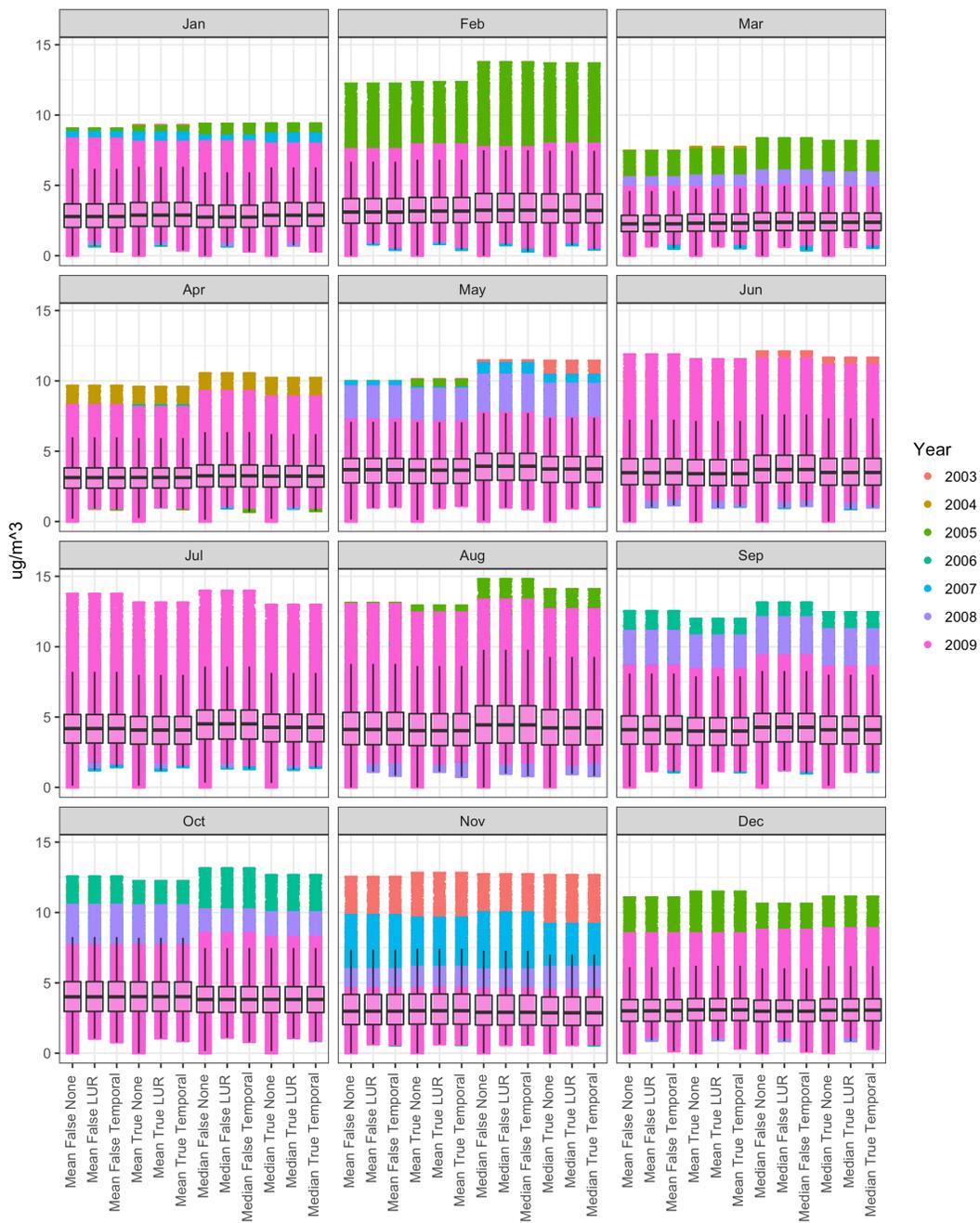
**Figure C4. Density plots of original versus imputed air monitoring data for NO using predictive mean matching.**

Density plots, one for each air quality monitoring station, were developed to visually compare the distribution between original air monitoring data, containing missing values, versus complete, imputed values.



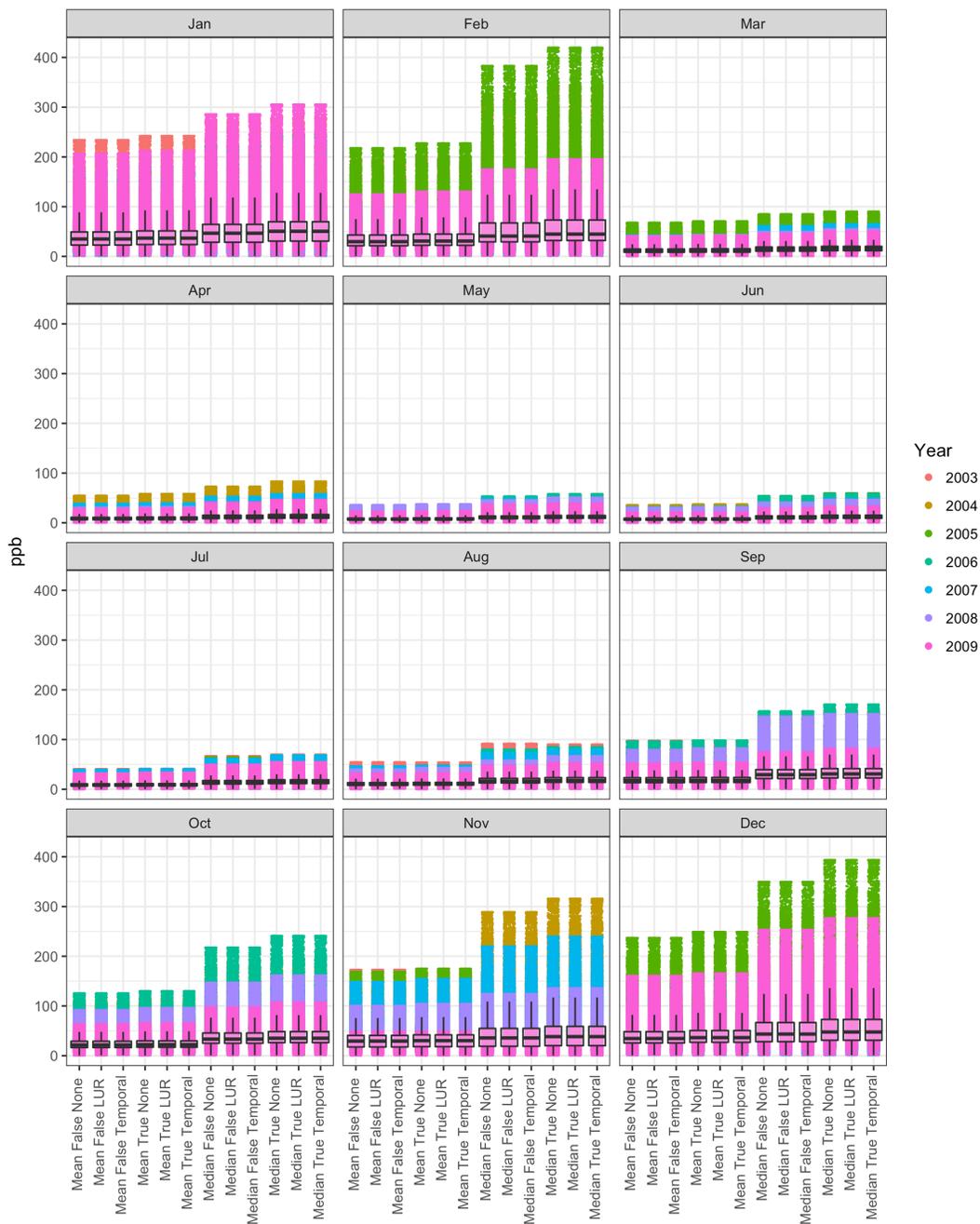
**Figure C5. Density plots of original versus imputed air monitoring data for NO<sub>2</sub> using predictive mean matching.**

Density plots, one for each air quality monitoring station, were developed to visually compare the distribution between original air monitoring data, containing missing values, versus complete, imputed values.



**Figure C6. Comparison between various temporal adjustment and truncation methods for temporally adjusting PM<sub>2.5</sub> land use regression estimates.**

Dot and box-and-whisker plots comparing temporally adjusted land use regression estimates at all postal codes in Metro Vancouver derived using (a.) temporal factors based on mean versus median of temporal air monitoring data; (b.) imputed [True] versus original air monitoring data [False]; and (c.) no truncation [None], truncation of LUR before temporal adjustment [LUR], or truncation after temporal adjustment of LUR values [Temporal].



**Figure C7. Comparison between various temporal adjustment and truncation methods for temporally adjusting NO land use regression estimates.**

Dot and box-and-whisker plots comparing temporally adjusted land use regression estimates at all postal codes in Metro Vancouver derived using (a.) temporal factors based on mean versus median of temporal air monitoring data; (b.) imputed [True] versus original air monitoring data [False]; and (c.) no truncation [None], truncation of LUR before temporal adjustment [LUR], or truncation after temporal adjustment of LUR values [Temporal].



**Figure C8. Comparison between various temporal adjustment and truncation methods for temporally adjusting NO<sub>2</sub> land use regression estimates.**

Dot and box-and-whisker plots comparing temporally adjusted land use regression estimates at all postal codes in Metro Vancouver derived using (a.) temporal factors based on mean versus median of temporal air monitoring data; (b.) imputed [True] versus original air monitoring data [False]; and (c.) no truncation [None], truncation of LUR before temporal adjustment [LUR], or truncation after temporal adjustment of LUR values [Temporal].

## Appendix D.

### Temporally Adjusted LUR Model Estimates: R Scripts

This code was used to develop temporally adjusted average measures of air pollution in Metro Vancouver from 2003–2009 by combining land use regression (LUR) models developed by Henderson et al. with continuous air monitoring data from Metro Vancouver’s Air Quality Monitoring Network. Temporally adjusting the LUR models enabled us to incorporate annual and seasonal trends to estimate monthly mean exposures of ambient air pollution at all 6-digit postal codes within Metro Vancouver. We used predictive mean matching to impute missing air monitoring station data.

We developed temporal factors as a ratio between the monthly mean of air pollutant concentrations across all air monitors for the month of interest over the annual mean of air pollutant concentrations across all air monitors for the LUR year. Then, we derived minimum concentrations by calculating the mean from the lowest daily average in a month for all air monitors for each month of the LUR year. To avoid potentially estimating values of zero, LUR models were truncated before temporal adjustment. We extracted air pollutant concentrations from the truncated LUR models at all postal codes within the study area, and then multiplied them by the temporal factors to produce monthly air pollutant estimates for all postal codes.

### Data Sources

#### Land Use Regression Models

Henderson SB, Beckerman B, Jerrett M, Brauer M. Application of land use regression to estimate ambient concentrations of traffic-related NO<sub>x</sub> and fine particulate matter. *Environmental Science and Technology*. 2007; 41 (7):2422-2428.

Wang R, Henderson SB, Sbihi H, Allen RW, Brauer M. Temporal Stability of Land Use Regression Models for Traffic-Related Air Pollution. *Atmospheric Environment*, 2012. 64: 312-319.

#### Metro Vancouver Air Quality Monitoring Network

Metro Vancouver. PM<sub>2.5</sub>, NO, and NO<sub>2</sub> Daily Averages, 2003–2014. 2016.

## Postal Code Conversion Files

- 2003: 2009-11-19, "Census of Canada. Postal Code Conversion File, January 2003 Postal Codes, 2001", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/TF2D9> V3 [Version]
- 2004: 2009-11-19, "Census of Canada. Postal Code Conversion File, July 2004 Postal Codes, 2001", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/SRHXK> V2 [Version]
- 2005: 2009-11-18, "Census of Canada. Postal Code Conversion File, October 2005 Postal Codes, 2001", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/92QKP> V3 [Version]
- 2006: 2009-11-19, "Census of Canada. Postal Code Conversion File, September 2006 Postal Codes, 2001", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/HQIOQ> V2 [Version]
- 2007: 2009-11-19, "Census of Canada. Postal Code Conversion File, September 2007 Postal Codes, 2006 [2007]", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/PVYH> V2 [Version]
- 2008: 2009-11-18, "Census of Canada. Postal Code Conversion File, March 2008 Postal Codes, 2006 [2008]", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/LNEEC> V3 [Version]
- 2009: 2010-01-28, "Census of Canada. Postal Code Conversion File, July 2009 Postal Codes, 2006 [2009]", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/NTC6K> V3 [Version]
- 2010: 2011-02-07, "Census of Canada. Postal Code Conversion File, October 2010 Postal Codes, 2006 [2011]", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/KO7Q0> V3 [Version]
- 2011: 2011-07-20, "Census of Canada. Postal Code Conversion File, May 2011 Postal Codes, 2006 [2011]", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/DCQV4> V3 [Version]
- 2012: 2013-09-11, "Platinum Postal Suite, v2012.3, [2012]", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/XUMWO> V2 [Version]
- 2013: 2013, "Census of Canada. Postal Code Conversion File, June 2013 Postal Codes, 2011 [2013]", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/10029> Statistics Canada [Distributor] V4 [Version]
- 2014: 2014-04, "Postal Code Conversion File, November 2014 Postal Codes, 2015", <http://hdl.handle.net.proxy.lib.sfu.ca/11272/10193> Statistics Canada [Distributor] V3 [Version]



```

        "Retired")

# Fixed width variables for 2004-2006
} else if(year >= 2004 && year <= 2006) {
  fixedWidth <- c(6,3,8,2,9,11,1,2,4,3,70,3,3,3,1,7,2,4,5,4,1,1,1,
                 30,1,1,8,8,5)
  colIndex <- c(1,2,5,6,7,8,27,28)
  colNames <- c("PostalCode",
                "FSA",
                "Lat",
                "Long",
                "SLI",
                "PR",
                "Birth",
                "Retired")

# Fixed width variables for 2007-2010
} else if(year >= 2007 && year <= 2010) {
  fixedWidth <- c(6,3,2,4,7,70,3,3,3,1,7,2,4,5,4,1,8,2,1,11,13,1,1,
                 30,1,1,8,8,1,3,1)
  colIndex <- c(1,2,3,20,21,22,27,28)
  colNames <- c("PostalCode",
                "FSA",
                "PR",
                "Lat",
                "Long",
                "SLI",
                "Birth",
                "Retired")

# Fixed width variables for 2011-2015, except 2012, which has headers
} else if(year >= 2011 && year <= 2015 && year != 2012) {
  fixedWidth <- c(6,3,2,4,7,70,3,3,3,1,7,2,4,5,4,1,8,2,1,11,13,1,1,
                 30,1,1,8,8,1,3,1,1)
  colIndex <- c(1,2,3,20,21,22,27,28)
  colNames <- c("PostalCode",
                "FSA",
                "PR",
                "Lat",
                "Long",
                "SLI",
                "Birth",
                "Retired")
}

# Import PCCF files and parse by fixed width fields
print(paste("Importing",year,"postal code conversion file"))

if(year != 2012 ) {
  postalCodes <- read.fwf(path, widths = fixedWidth, fileEncoding = "latin1")
} else if(year == 2012) {
  postalCodes <- read.csv(path, header = T)
}

# Select, rename, and add province and FSA columns
print(paste("Selecting",year,"postal code variables"))

if(year != 2012) {
  postalCodes <- postalCodes[ , colIndex]
  names(postalCodes) <- colNames
} else if(year == 2012) {
  postalCodes <- select(postalCodes, SLI,
                       PostalCode = POSTALCODE,

```

```

        PR      = PROV,
        Birth   = BIRTH_DATE,
        Retired = RET_DATE,
        Long    = LONGITUDE,
        Lat     = LATITUDE) %>%
  mutate(PR = 59, FSA = substr(PostalCode, 1,3))
}

# Add columns for source year, and string split retired and birth years

postalCodes$Year <- year
postalCodes$Retired <- as.numeric(substr(postalCodes$Retired, 1,4))
postalCodes$Birth <- as.numeric(substr(postalCodes$Birth, 1,4))

# Subset data by province, SLI, active for current year

print(paste("Filtering",year,"data by province and retired date"))

postalCodes <- filter(postalCodes, PR == 59 & SLI == 1,
  Retired == 1900 | Retired == year)

# Reorder columns, sort by Year and Postal Code, and remove duplicates

postalCodes <- select(postalCodes,
  Year,
  PostalCode,
  FSA,
  PR,
  Lat,
  Long,
  Birth,
  Retired,
  SLI) %>%
  arrange(Year, PostalCode) %>%
  distinct(Year, PostalCode, .keep_all = T)

print(paste(year, "PCCF succesfully processed"))
print("---")

return(postalCodes)
}

#' ---
#' Build data structure for PCCF parsing function. Create dataframe containing
#' path to PCCF and PCCF year.
#' ---

# Path prefix to original data folder

path.original <- "Data/Original/Postal Code Conversion Files/"

# Vector of file paths to parse

pccf.files <- c(
  paste(path.original,"2003/Data/pccf_59.txt", sep=""),
  paste(path.original,"2004/Data/pccf59_JUL04_fccp59.txt", sep=""),
  paste(path.original,"2005/Data/pccf59_oct05.txt", sep=""),
  paste(path.original,"2006/Data/pccf59_SEPT06_fccp59.txt", sep=""),
  paste(path.original,"2007/Data/pccf59_SEPT07_fccp59.txt", sep=""),
  paste(path.original,"2008/Data/pccf59_MAR08_fccp59.txt", sep=""),
  paste(path.original,"2009/Data/pccf59_JUL09_fccp59.txt", sep=""),
  paste(path.original,"2010/Data/pccf59_OCT10_fccp59.txt", sep=""),
  paste(path.original,"2011/Data/pccf59_MAY11_fccp59.txt", sep=""),
  paste(path.original,"2012/Data/CANmep_BC_merged.csv", sep=""),
  paste(path.original,"2013/Data/pccfNat_JUN13_fccpNat.txt", sep=""),
  paste(path.original,"2014/Data/pccfNat_NOV14_fccpNat.txt", sep=""),
  paste(path.original,"2015/Data/pccfNat_AUG15_fccpNat.txt", sep=""))

```

```

# Column bind year and path vectors into a dataframe

pccf.files <- data.frame(cbind(pccf.files, c(2003:2015)))
colnames(pccf.files) <- c("Path", "Year")
pccf.files$Path <- as.character(pccf.files$Path)
pccf.files$Year <- as.numeric(levels(pccf.files$Year))[pccf.files$Year]

rm(path.original)

#' ---
#' Process all PCCF files from 2003-2015 in a loop by passing pccf.files
#' dataframe in pccf.parse function. Append results for each iteration.
#' ---

# Process first year: 2003

pccf.data.bc <- pccf.parse(pccf.files$Path[1], pccf.files$Year[1])

# Process remaining years and bind to pccf.data.bc

for(i in 2:nrow(pccf.files)) {
  pccf.data.bc <- rbind(pccf.data.bc,
                       pccf.parse(pccf.files$Path[i], pccf.files$Year[i]))
}; rm(i)

# Save files

write_csv(pccf.data.bc, "Data/Cleaned/pccf_data_bc.csv")

rm(list = ls())

```

## 02\_process\_pccf\_retired.R

```

#' ---
#' Import 2015 PCCF and retired postal code files, merge them, process them.
#' 2015 retired file contains all expired postal codes from previous years
#' ---

library(readr)
library(dplyr)

#' ---
#' Function loads 2015 PCCF and retired files. Then parses relevant variables
#' and subset data to all retired postal codes in BC
#'
#' path : char string of file location
#' year : num of source PCCF year
#' ---

pccf.parse.retired <- function(path, year) {

  # Fixed width variables for 2015

  if(year == 2015) {
    fixedWidth <- c(6,3,2,4,7,70,3,3,3,1,7,2,4,5,4,1,8,2,1,11,13,1,1,
                   30,1,1,8,8,1,3,1,1)
    colNames <- c("PostalCode", "FSA",
                  "PR", "CDuid",
                  "CSDuid", "CSDname",
                  "CSDtype", "CCScode",
                  "SAC", "SACtype",
                  "CTname", "ER",
                  "DPL", "FED13uid",
                  "POP_CNTR_RA", "POP_CNTR_RA_type",
                  "DAuid", "Dissemination block",
                  "Rep_Pt_Type", "Lat",

```

```

        "Long",          "SLI",
        "PCtype",       "Comm_Name",
        "DMT",          "H_DMT",
        "Birth",        "Retired",
        "PO",           "QI",
        "Source",       "POP_CNTR_RA_SIZE_CLASS")
  colIndex <- c(1,2,3,20,21,22,27,28)
}

# Import PCCF files, parse by fixed width fields, and name columns

print(paste("Importing",year,"postal code conversion file"))
postalCodes <- read.fwf(path, widths = fixedWidth,
                        fileEncoding = "latin1",
                        fill = T,
                        col.names = colNames)

# Select, rename, and add province and FSA columns

print(paste("Selecting",year,"postal code variables"))
postalCodes <- postalCodes[ , colIndex]

# Add columns for source year, and string split retired and birth years

postalCodes$Year <- year
postalCodes$Retired <- as.numeric(substr(postalCodes$Retired, 1,4))
postalCodes$Birth <- as.numeric(substr(postalCodes$Birth, 1,4))

# Subset data by province, SLI, and keep only retired postal codes

print(paste("Filtering",year,"data by province and retired date"))
postalCodes <- filter(postalCodes, PR == 59 & SLI == 1, Retired >= 2003)

# Reorder columns, sort by Year and Postal Code, and remove duplicates

postalCodes <- select(postalCodes,
                      Year,
                      PostalCode,
                      FSA,
                      PR,
                      Lat,
                      Long,
                      Birth,
                      Retired,
                      SLI) %>%
  arrange(Year, PostalCode) %>%
  distinct(PostalCode, Retired, .keep_all = T)

print(paste(year, "Retired PCCF succesfully processed"))
print("---")

return(postalCodes)
}

#' ---
#' Build data structure for retired PCCF parsing function. Create dataframe
#' containing path to both files.
#' ---

# Path prefix to original data folder

path.original <- "Data/Original/Postal Code Conversion Files/"

# Vector of file paths to parse

pccf.files <- c(
  paste(path.original,"2015/Data/pccfNat_AUG15_fccpNat.txt", sep=""),

```

```

paste(path.original, "2015/Data/R2010.txt",          sep=""))
# Column bind year and path vectors into a dataframe
pccf.files <- data.frame(cbind(pccf.files, c(2015, 2015)))
colnames(pccf.files) <- c("Path", "Year")
pccf.files$Path <- as.character(pccf.files$Path)
pccf.files$Year <- as.numeric(levels(pccf.files$Year))[pccf.files$Year]
rm(path.original)

#' ---
#' Process retired PCCF files from 2015 in a loop by passing pccf.files
#' dataframe in pccf.parse function. Append results for each iteration.
#' ---

# Process 2015 PCCF, containing expired postal codes after 2010
pccf.data.ret.bc <- pccf.parse.retired(pccf.files$Path[1], pccf.files$Year[1])

# Process pre-2010 retired file and bind to 2015 retired postal codes
pccf.data.ret.bc <- rbind(pccf.data.ret.bc,
                          pccf.parse.retired(pccf.files$Path[2],
                                              pccf.files$Year[2]))

# Save files
write_csv(pccf.data.ret.bc, "Data/Cleaned/pccf_data_ret_bc.csv")
rm(list = ls())

```

### 03\_subset\_pc\_within\_van.R

```

#' ---
#' Import cleaned postal code data for all of BC, and select only
#' postal codes within Metro Vancouver geographic extent.
#' ---

library(sp)
library(rgdal)
library(rgeos)
library(dplyr)
library(readr)

# Load BC postal code data
pccf.data.bc <- read_csv("Data/Cleaned/pccf_data_bc.csv")
pccf.data.ret.bc <- read_csv("Data/Cleaned/pccf_data_ret_bc.csv")

# Load shape file of Metro Vancouver boundary to serve as geographic extent
extent.van <- readOGR("Data/Original/Metro Vancouver Open Data",
                     "AdminBoundaryDissolved")

# Load shape file of BC boundary to visually test geographic transformations
extent.bc <- readOGR("Data/Original/Stats Can Provincial Boundaries",
                    "bc_boundary")

# Duplicate original GEOGCS NAD83 lat long vars to preserve post-geocoding
pccf.data.bc <- mutate(pccf.data.bc,
                       Long_GCSNAD83 = Long,
                       Lat_GCSNAD83 = Lat)

```

```

pccf.data.ret.bc <- mutate(pccf.data.ret.bc,
                          Long_GCSNAD83 = Long,
                          Lat_GCSNAD83  = Lat)

# Turn BC postal code data into spatial point object using lat long variables

coordinates(pccf.data.bc)      <- c("Long", "Lat")
coordinates(pccf.data.ret.bc) <- c("Long", "Lat")

# Add original NAD83 geographic coordinate system to spatial objects

proj4string(pccf.data.bc)      <- CRS("+init=EPSG:4269")
proj4string(pccf.data.ret.bc) <- CRS("+init=EPSG:4269")

# Transform geographic coordinate system to match extent projection

pccf.data.bc      <- spTransform(pccf.data.bc,      proj4string(extent.van))
pccf.data.ret.bc <- spTransform(pccf.data.ret.bc, proj4string(extent.van))

# Map BC extent and postal codes to verify correct transformation

plot(extent.bc)
plot(pccf.data.bc,      add = T, pch = 1, cex = 0.25, col = "red")
plot(pccf.data.ret.bc, add = T, pch = 1, cex = 0.25, col = "blue")

# Subset BC postal codes within Metro Vancouver extent

pccf.data.van <- pccf.data.bc[extent.van, ]
pccf.data.ret.van <- pccf.data.ret.bc[extent.van, ]

# Map BC extent and Van postal codes to verify clipping at provincial scale

plot(extent.bc)
plot(pccf.data.van,      add = T, pch = 1, cex = 0.25, col = "red")
plot(pccf.data.ret.van, add = T, pch = 1, cex = 0.25, col = "blue")

# Map Van extent and Van postal codes to verify clipping at regional scale

plot(extent.van)
plot(pccf.data.van,      add = T, pch = 1, cex = 0.25, col = "red")
plot(pccf.data.ret.van, add = T, pch = 1, cex = 0.25, col = "blue")

# Convert spatial objects back to data frames

pccf.data.van <- as.data.frame(pccf.data.van)
pccf.data.ret.van <- as.data.frame(pccf.data.ret.van)

# Reorder columns and sort rows

pccf.data.van <- select(pccf.data.van,
                      Year,
                      PostalCode,
                      FSA,
                      PR,
                      Lat_GCSNAD83,
                      Long_GCSNAD83,
                      Lat_UTM10 = Lat,
                      Long_UTM10 = Long,
                      Birth,
                      Retired,
                      SLI) %>%
  rename(Lat = Lat_GCSNAD83,
         Long = Long_GCSNAD83) %>%
  arrange(Year, PostalCode)

pccf.data.ret.van <- select(pccf.data.ret.van,
                          Year,
                          PostalCode,

```

```

        FSA,
        PR,
        Lat_GCSNAD83,
        Long_GCSNAD83,
        Lat_UTM10 = Lat,
        Long_UTM10 = Long,
        Birth,
        Retired,
        SLI) %>%
  rename(Lat = Lat_GCSNAD83,
         Long = Long_GCSNAD83) %>%
  arrange(Year, PostalCode)

# Save spatial subsets

write_csv(pccf.data.van, "Data/Cleaned/pccf_data_van.csv")
write_csv(pccf.data.ret.van, "Data/Cleaned/pccf_data_ret_van.csv")

rm(list = ls())

```

## 04\_get\_pollutant\_at\_pc.R

```

#' ---
#' Import and geocode cleaned postal code data for Metro Vancouver, overlay it
#' on land use regression layer, and get pollutant value at each postal code.
#' ---

library(raster)
library(rgdal)
library(rgeos)
library(dplyr)
library(readr)
library(ggplot2)

# Load Metro Vancouver postal codes

pccf.data.van <- read_csv("Data/Cleaned/pccf_data_van.csv")

# Load LUR as rasters

lur.2003.pm25 <- raster("Data/Original/LUR Surfaces/PM25_2003/pm25_2003.tif")
lur.2003.no <- raster("Data/Original/LUR Surfaces/NO_2003/no_2003.tif")
lur.2010.no <- raster("Data/Original/LUR Surfaces/NO_2010/no_2010.tif")
lur.2003.no2 <- raster("Data/Original/LUR Surfaces/NO2_2003/no2_2003.tif")
lur.2010.no2 <- raster("Data/Original/LUR Surfaces/NO2_2010/no2_2010.tif")

extent.van <- readOGR("Data/Original/Metro Vancouver Open Data",
                    "AdminBoundaryDissolved")

#' ---
#' Analyze LUR air pollution values. Extract values from LUR and plot
#' their distribution. Look for zero and NA values.
#' ---

# Extract cell values from rasters

val.lur.2003.pm25 <- data.frame(Value = getValues(lur.2003.pm25))
val.lur.2003.no <- data.frame(Value = getValues(lur.2003.no))
val.lur.2010.no <- data.frame(Value = getValues(lur.2010.no))
val.lur.2003.no2 <- data.frame(Value = getValues(lur.2003.no2))
val.lur.2010.no2 <- data.frame(Value = getValues(lur.2010.no2))

# Trim precision

trim.precision <- function(df) {
  var <- "Value"

```

```

  digits <- 4
  df[, var] <- round(df[, var], digits)
  return(df)
}

val.lur.2003.pm25 <- trim.precision(val.lur.2003.pm25)
val.lur.2003.no <- trim.precision(val.lur.2003.no)
val.lur.2010.no <- trim.precision(val.lur.2010.no)
val.lur.2003.no2 <- trim.precision(val.lur.2003.no2)
val.lur.2010.no2 <- trim.precision(val.lur.2010.no2)

# Summarise LUR values

summary(val.lur.2003.pm25)
summary(val.lur.2003.no)
summary(val.lur.2010.no)
summary(val.lur.2003.no2)
summary(val.lur.2010.no2)

# Count zeros

filter(val.lur.2003.pm25, Value == 0) %>% summarise(N.Zeros = n())
filter(val.lur.2003.no, Value == 0) %>% summarise(N.Zeros = n())
filter(val.lur.2010.no, Value == 0) %>% summarise(N.Zeros = n())
filter(val.lur.2003.no2, Value == 0) %>% summarise(N.Zeros = n())
filter(val.lur.2010.no2, Value == 0) %>% summarise(N.Zeros = n())

# Look at smallest values after filtering out for zero

filter(val.lur.2003.pm25, Value > 0) %>% arrange(Value) %>% head()
filter(val.lur.2003.no, Value > 0) %>% arrange(Value) %>% head()
filter(val.lur.2010.no, Value > 0) %>% arrange(Value) %>% head()
filter(val.lur.2003.no2, Value > 0) %>% arrange(Value) %>% head()
filter(val.lur.2010.no2, Value > 0) %>% arrange(Value) %>% head()

# Plot density distribution of pollutants

plot.lur <- function(val.lur, pollutant, year, note, binwidth) {
  n <- base::nrow(val.lur)
  plot <- ggplot(val.lur, aes(Value)) +
    geom_histogram(binwidth = binwidth) +
    xlab(paste0(pollutant, " (", year, "). ",
               "N-Cells = ", prettyNum(n, big.mark = ","), ". ", note)) +
    ylab("Count")
  return(plot)
}

note <- ""

plot.lur(val.lur.2003.pm25, "PM2.5", "2003", note, 1)
plot.lur(val.lur.2003.no, "NO", "2003", note, 10)
plot.lur(val.lur.2010.no, "NO", "2010", note, 10)
plot.lur(val.lur.2003.no2, "NO2", "2003", note, 1)
plot.lur(val.lur.2010.no2, "NO2", "2010", note, 1)

# For 2010 NO2, change NA to zero for plotting purposes

val.lur.2010.no2.noNA <- val.lur.2010.no2
val.lur.2010.no2.noNA$Value[is.na(val.lur.2010.no2.noNA$Value)] <- 0

note <- "NAs set as zero"

plot.lur(val.lur.2010.no2.noNA, "NO2", "2010", note, 1)

# Plot with zeros and NA removed

note <- "Zeros & NAs Excluded."

```

```

val.lur.2003.pm25.trim <- val.lur.2003.pm25      %>% filter(Value > 0)
val.lur.2003.no.trim  <- val.lur.2003.no       %>% filter(Value > 0)
val.lur.2010.no.trim  <- val.lur.2010.no       %>% filter(Value > 0)
val.lur.2003.no2.trim <- val.lur.2003.no2     %>% filter(Value > 0)
val.lur.2010.no2.trim <- val.lur.2010.no2.noNA %>% filter(Value > 0)

plot.lur(val.lur.2003.pm25.trim, "PM2.5", "2003", note, 1)
plot.lur(val.lur.2003.no.trim, "NO", "2003", note, 10)
plot.lur(val.lur.2010.no.trim, "NO", "2010", note, 10)
plot.lur(val.lur.2003.no2.trim, "NO2", "2003", note, 1)
plot.lur(val.lur.2010.no2.trim, "NO2", "2010", note, 1)

rm(list = c(ls(pattern='^val'), "note", "plot.lur"))

#' ---
#' Function performs GIS functions to calculate pollutant values at a given
#' postal code. It transforms list of PCs into geographic object,
#' matches projection between PCs and LUR, overlays the PCs unto LUR, and then
#' selects LUR value at given PC location.
#'
#' pccf.data.van : data frame with postal codes
#' lur           : raster layer
#' map.label     : string
#' ---

get.pollutant.at.cord <- function(df.pc, lur, map.label) {

  # Give each observation a unique ID, used later to check merged data frames

  df.pc <- mutate(df.pc, ID = as.numeric(rownames(df.pc)))

  # Duplicate original GEOGCS NAD83 lat long vars to preserve post-geocoding

  df.pc.spatial <- mutate(df.pc, Lat_GCSNAD83 = Lat, Long_GCSNAD83 = Long) %>%
    mutate(Lat = Lat_UTM10, Long = Long_UTM10)

  # Transform PCs to spatial object using UTM10 coordinate system for LUR

  coordinates(df.pc.spatial) <- c("Long_UTM10", "Lat_UTM10")
  proj4string(df.pc.spatial) <- lur@crs

  # Map postal codes on top of LUR to verify transformation

  plot(lur, main = map.label)
  plot(extent.van, add = T, border = "red")
  plot(df.pc.spatial, add = T, pch = 1, cex = 0.1, col = "blue")
  legend("bottomleft",
        legend = c("Active Postal Codes", "Metro Vancouver Boundary"),
        col = c("blue", "red"),
        lty = c(NA, 1),
        pch = c(16, NA),
        bty = "n")

  # Extract LUR value at postal code coordinates

  pc.lur.values <- data.frame(df.pc.spatial$Year,
                             df.pc.spatial$PostalCode,
                             coordinates(df.pc.spatial),
                             df.pc.spatial$ID,
                             extract(lur, df.pc.spatial))

  names(pc.lur.values) <- c("Year",
                          "PostalCode",
                          "Long_UTM10",
                          "Lat_UTM10",
                          "ID",
                          "Value")

```

```

# Merge postal code and lur value data frames based
df.pollutant <- merge(df.pc,
                      pc.lur.values,
                      by = c("Year",
                              "PostalCode",
                              "Long_UTM10",
                              "Lat_UTM10",
                              "ID"),
                      all = T)

# Clean variables of interest, rename columns, and sort data frame
df.pollutant <- dplyr::select_(df.pollutant,
                               "Year",
                               "PC" = "PostalCode",
                               "FSA",
                               "PR",
                               "Lat_NAD83" = "Lat",
                               "Lon_NAD83" = "Long",
                               "Lat_UTM10",
                               "Lon_UTM10" = "Long_UTM10",
                               "Birth",
                               "Retired",
                               "SLI",
                               "Value") %>%
  arrange(Year, PC)
return(df.pollutant)
}

# Get LUR value at postal code coordinate
png("Figures/lur_postalcode_map.png", width = 1800, height = 1200, res = 150)
par(mfrow = c(2, 3),
    mar = c(4.1, 4.1, 4.1, 6.1))

lur.pc.val.pm25 <- get.pollutant.at.cord(
  pccf.data.van,
  lur.2003.pm25,
  "PM2.5 (2003)")

lur.pc.val.no.2003 <- get.pollutant.at.cord(
  filter(pccf.data.van, Year < 2010),
  lur.2003.no,
  "NO (2003)")

lur.pc.val.no.2010 <- get.pollutant.at.cord(
  filter(pccf.data.van, Year >= 2010),
  lur.2010.no,
  "NO (2010)")

lur.pc.val.no2.2003 <- get.pollutant.at.cord(
  filter(pccf.data.van, Year < 2010),
  lur.2003.no2,
  "NO2 (2003)")

lur.pc.val.no2.2010 <- get.pollutant.at.cord(
  filter(pccf.data.van, Year >= 2010),
  lur.2010.no2,
  "NO2 (2010)")

dev.off()

# Merge NO and NO2 data frames, respectively
lur.pc.val.no <- rbind(lur.pc.val.no.2003, lur.pc.val.no.2010)
lur.pc.val.no2 <- rbind(lur.pc.val.no2.2003, lur.pc.val.no2.2010)

```

```

rm(lur.pc.val.no.2003, lur.pc.val.no2.2003,
   lur.pc.val.no.2010, lur.pc.val.no2.2010)

# Trim observations

lur.pc.val.pm25 <- filter(lur.pc.val.pm25, Year >= 2003 & Year <= 2014)
lur.pc.val.no <- filter(lur.pc.val.no, Year >= 2003 & Year <= 2014)
lur.pc.val.no2 <- filter(lur.pc.val.no2, Year >= 2003 & Year <= 2014)

# Check values

summary(lur.pc.val.pm25)
summary(lur.pc.val.no)
summary(lur.pc.val.no2)

filter(lur.pc.val.pm25, Value == 0) %>% summarise(N.Zeros = n())
filter(lur.pc.val.no, Value == 0) %>% summarise(N.Zeros = n())
filter(lur.pc.val.no2, Value == 0) %>% summarise(N.Zeros = n())

# Save files

save.files <- function(df, filename) {
  write_csv(df, paste0("Data/Cleaned/", filename, ".csv"))
}

save.files(lur.pc.val.pm25, "pm25_lur_val_at_pc")
save.files(lur.pc.val.no, "no_lur_val_at_pc")
save.files(lur.pc.val.no2, "no2_lur_val_at_pc")

# Plot density distribution of pollutant values at pc

lur.pc.val.pm25$Year <- factor(lur.pc.val.pm25$Year)
lur.pc.val.no$Year <- factor(lur.pc.val.no$Year)
lur.pc.val.no2$Year <- factor(lur.pc.val.no2$Year)

ggplot(lur.pc.val.pm25, aes(x = Value, colour = Year)) + geom_density()
ggplot(lur.pc.val.no, aes(x = Value, colour = Year)) + geom_density()
ggplot(lur.pc.val.no2, aes(x = Value, colour = Year)) + geom_density()

# Clear global environment

rm(list = ls())

```

## 05\_get\_pollutant\_in\_time.R

```

#' ---
#' Load temporal air monitoring data, and calculate monthly-spatial and
#' baseline-spatial averages to create a spatially homogenous temporal factor
#' used to model LUR values across time. Temporal factor is based on air
#' monitoring station data, which has missing at random data and will first
#' require multiple imputation. Baseline-spatial average is based on
#' LUR year.
#' ---

library(dplyr)
library(readr)
library(reshape2)
library(mice)
library(lattice)
library(VIM)

ptm <- proc.time()

#' ---
#' Load data and analyze missing data in preparation multiple imputation

```

```

#' ---
# Load Metro Vancouver air monitoring data

load.data <- function(filename) {
  filepath <- "Data/Original/Air Quality Monitoring Network/"
  df <- read.csv(paste0(filepath, filename),
                header = T,
                stringsAsFactors = F,
                na.strings = c("<Samp", "InVld", "NoData", "Down",
                              "Purge", "RS232", "FailPwr"))

  colnames(df)[1] <- "Time"
  df <- mutate(df,
              Year = as.numeric(substr(Time, 7, 10)),
              Month = as.numeric(substr(Time, 4, 5)),
              Day = as.numeric(substr(Time, 1, 2))) %>%
    select(-Time)
  return(df)
}

pm25.A <- load.data("PM25_A.csv")
pm25.B <- load.data("PM25_B.csv")

temporal.pm25 <- rbind(pm25.A, pm25.B)
temporal.no <- load.data("NO.csv")
temporal.no2 <- load.data("NO2.csv")

rm(pm25.A, pm25.B, load.data)

# Make all columns numeric

make.cols.numeric <- function(df) {
  for(i in 1:length(df)) { df[, i] <- as.numeric(df[, i]) }
  return(df)
}

temporal.pm25 <- make.cols.numeric(temporal.pm25)
temporal.no <- make.cols.numeric(temporal.no)
temporal.no2 <- make.cols.numeric(temporal.no2)

rm(make.cols.numeric)

# Reshape date from wide to long

reshape.long <- function(df) {
  df <- melt(df, id.vars = c("Year", "Month", "Day"),
            na.rm = F,
            variable.name = "Monitor",
            value.name = "Value")
  return(df)
}

long.pm25 <- reshape.long(temporal.pm25)
long.no <- reshape.long(temporal.no)
long.no2 <- reshape.long(temporal.no2)

# Calculate number of missing values

count.missing <- function(df) {

  # Filter all NAs and count observations

  df.missing <- filter(df, is.na(df["Value"])) %>%
    group_by(Year, Month, Monitor) %>%
    summarise(Missing = n())

  # Filter all observed and count observations

```

```

df.observed <- filter(df, !is.na(df["Value"])) %>%
  group_by(Year, Month, Monitor) %>%
  summarise(Observed = n())

df <- merge(df.missing, df.observed,
            by = c("Year", "Month", "Monitor"), all = T)

df[is.na(df)] <- 0

# Calculate percentage of missing observations

df <- mutate(df, Total = Missing + Observed,
             Pct.Missing = signif(Missing / Total * 100, 0))

df[df == 0] <- NA

return(df)
}

missing.pm25 <- count.missing(long.pm25)
missing.no <- count.missing(long.no)
missing.no2 <- count.missing(long.no2)

# Summary of missing data

Pollutant <- c("PM25", "NO", "NO2")

Missing <- c(sum(missing.pm25$Missing, na.rm = T),
            sum(missing.no$Missing, na.rm = T),
            sum(missing.no2$Missing, na.rm = T))

Observed <- c(sum(missing.pm25$Observed, na.rm = T),
             sum(missing.no$Observed, na.rm = T),
             sum(missing.no2$Observed, na.rm = T))

Total <- c(sum(missing.pm25$Total, na.rm = T),
          sum(missing.no$Total, na.rm = T),
          sum(missing.no2$Total, na.rm = T))

missing.data <- data.frame(Pollutant, Missing, Observed, Total) %>%
  mutate(Percent = Missing / Total * 100)

rm(Pollutant, Missing, Observed, Total, missing.data)

# Reshape data

reshape.wide <- function(df, value.var) {
  df <- dcast(df, Year + Month ~ Monitor, value.var = value.var)
  return(df)
}

missing.wide.pm25 <- reshape.wide(missing.pm25, "Missing")
missing.wide.no <- reshape.wide(missing.no, "Missing")
missing.wide.no2 <- reshape.wide(missing.no2, "Missing")

# Export summary of missing values to be analyzed in Excel

# write_csv(missing.wide.pm25, path="Results/pm25_missing_summary.csv", na="")
# write_csv(missing.wide.no, path="Results/no_missing_summary.csv", na="")
# write_csv(missing.wide.no2, path="Results/no2_missing_summary.csv", na="")

rm(list = c(ls(pattern = "^long"), ls(pattern = "^missing"),
            "count.missing", "reshape.wide"))

# Plot missing data

plot.missing.data <- function(df) {
  plot <- agr(select(df, -(Year:Day)),

```

```

        col      = "red",
        numbers  = TRUE,
        labels   = names(select(df, -(Year:Day))),
        cex.axis = 0.8,
        gap      = 3,
        ylab     = c("Histogram of Missing Data",
                    "Patterns"),
        digits   = 2)
    return(plot)
}

plot.missing.pm25 <- plot.missing.data(temporal.pm25)
plot.missing.no  <- plot.missing.data(temporal.no)
plot.missing.no2 <- plot.missing.data(temporal.no2)

rm(list = ls(pattern = "^plot.missing"))

#' ---
#' Impute missing data using single imputation
#' ---

impute.na <- function(df, num.imp) {
  exclude.var = c("Year", "Month", "Day")
  imp.result <- mice(df,
                    m   = num.imp,
                    seed = 170703,
                    pred = quickpred(df, exclude = exclude.var))
  return(imp.result)
}

imp.pm25 <- impute.na(temporal.pm25, 1)
imp.no   <- impute.na(temporal.no,   1)
imp.no2  <- impute.na(temporal.no2,  1)

summary(imp.pm25)
summary(imp.no)
summary(imp.no2)

# Make density plot observed and imputed values: red-imputed, blue-observed
pdf("Figures/imputation_density_plot.pdf", width = 11, height = 8.5)

densityplot(imp.pm25,
            grid      = T,
            thicker   = 1,
            plot.points = T,
            xlab      = "PM2.5")

densityplot(imp.no,
            grid      = T,
            thicker   = 1,
            plot.points = T,
            xlab      = "NO")

densityplot(imp.no2,
            grid      = T,
            thicker   = 1,
            plot.points = T,
            xlab      = "NO2")

dev.off()

# Fill in original NA data with imputation results

temporal.imp.pm25 <- complete(imp.pm25, 1)
temporal.imp.no   <- complete(imp.no,   1)
temporal.imp.no2  <- complete(imp.no2,  1)

# Verify that no NAs remain

```

```

verify.na <- function(df) {
  results <- sapply(df, function(x) sum(is.na(x))) %>%
    data.frame()
  colnames(results) <- "NA.Count"
  return(results)
}

na.pm25 <- verify.na(temporal.imp.pm25)
na.no <- verify.na(temporal.imp.no)
na.no2 <- verify.na(temporal.imp.no2)

rm(list = c(ls(pattern = "^na"), ls(pattern = "^imp"), "verify.na"))

#' ---
#' Develop temporal factors for LUR adjustments over time
#' ---

# Convert data to long, rbind, and save

temporal.pm25 <- reshape.long(temporal.pm25) %>% mutate(Imputed = F)
temporal.no <- reshape.long(temporal.no) %>% mutate(Imputed = F)
temporal.no2 <- reshape.long(temporal.no2) %>% mutate(Imputed = F)

temporal.imp.pm25 <- reshape.long(temporal.imp.pm25) %>% mutate(Imputed = T)
temporal.imp.no <- reshape.long(temporal.imp.no) %>% mutate(Imputed = T)
temporal.imp.no2 <- reshape.long(temporal.imp.no2) %>% mutate(Imputed = T)

daily.pm25 <- rbind(temporal.pm25, temporal.imp.pm25)
daily.no <- rbind(temporal.no, temporal.imp.no)
daily.no2 <- rbind(temporal.no2, temporal.imp.no2)

write_csv(daily.pm25, "Data/Cleaned/pm25_daily.csv")
write_csv(daily.no, "Data/Cleaned/no_daily.csv")
write_csv(daily.no2, "Data/Cleaned/no2_daily.csv")

rm(list = c(ls(pattern = "^temporal"), "reshape.long"))

#' ---
#' Calculate yearly averages and medians with and without imputation
#' ---

calc.mean.year <- function(df, is.imputed) {
  if(is.imputed) { df <- filter(df, Imputed == T)
  } else { df <- filter(df, Imputed == F) }
  mean <- group_by(df, Year) %>%
    summarise(Value = mean(Value, na.rm = T)) %>%
    mutate(Measure = "Mean", Imputed = is.imputed)
  return(mean)
}

calc.median.year <- function(df, is.imputed) {
  if(is.imputed) { df <- filter(df, Imputed == T)
  } else { df <- filter(df, Imputed == F) }
  median <- group_by(df, Year) %>%
    summarise(Value = median(Value, na.rm = T)) %>%
    mutate(Measure = "Median", Imputed = is.imputed)
  return(median)
}

# Calculate yearly averages with no imputation

yearly.mean.pm25 <- calc.mean.year(daily.pm25, F)
yearly.mean.no <- calc.mean.year(daily.no, F)
yearly.mean.no2 <- calc.mean.year(daily.no2, F)

# Calculate yearly averages with imputation

```

```

yearly.mean.imp.pm25 <- calc.mean.year(daily.pm25, T)
yearly.mean.imp.no <- calc.mean.year(daily.no, T)
yearly.mean.imp.no2 <- calc.mean.year(daily.no2, T)

# Calculate yearly medians with no imputation

yearly.median.pm25 <- calc.median.year(daily.pm25, F)
yearly.median.no <- calc.median.year(daily.no, F)
yearly.median.no2 <- calc.median.year(daily.no2, F)

# Calculate yearly medians with imputation

yearly.median.imp.pm25 <- calc.median.year(daily.pm25, T)
yearly.median.imp.no <- calc.median.year(daily.no, T)
yearly.median.imp.no2 <- calc.median.year(daily.no2, T)

# Bind data frames into one

yearly.pm25 <- bind_rows(yearly.mean.pm25,
                        yearly.mean.imp.pm25,
                        yearly.median.pm25,
                        yearly.median.imp.pm25)

yearly.no <- bind_rows(yearly.mean.no,
                      yearly.mean.imp.no,
                      yearly.median.no,
                      yearly.median.imp.no)

yearly.no2 <- bind_rows(yearly.mean.no2,
                       yearly.mean.imp.no2,
                       yearly.median.no2,
                       yearly.median.imp.no2)

rm(list = c(ls(pattern = "^yearly.mean"), ls(pattern = "^yearly.median")))

#' ---
#' Calculate monthly averages and medians with and without imputation
#' ---

calc.mean.month <- function(df, is.imputed) {
  if(is.imputed) { df <- filter(df, Imputed == T)
  } else { df <- filter(df, Imputed == F) }
  mean <- group_by(df, Year, Month) %>%
    summarise(Value = mean(Value, na.rm = T)) %>%
    mutate(Measure = "Mean", Imputed = is.imputed)
  return(mean)
}

calc.median.month <- function(df, is.imputed) {
  if(is.imputed) { df <- filter(df, Imputed == T)
  } else { df <- filter(df, Imputed == F) }
  median <- group_by(df, Year, Month) %>%
    summarise(Value = median(Value, na.rm = T)) %>%
    mutate(Measure = "Median", Imputed = is.imputed)
  return(median)
}

# Calculate monthly averages with no imputation

monthly.mean.pm25 <- calc.mean.month(daily.pm25, F)
monthly.mean.no <- calc.mean.month(daily.no, F)
monthly.mean.no2 <- calc.mean.month(daily.no2, F)

# Calculate monthly averages with imputation

monthly.mean.imp.pm25 <- calc.mean.month(daily.pm25, T)
monthly.mean.imp.no <- calc.mean.month(daily.no, T)
monthly.mean.imp.no2 <- calc.mean.month(daily.no2, T)

```

```

# Calculate monthly medians with no imputation

monthly.median.pm25 <- calc.median.month(daily.pm25, F)
monthly.median.no   <- calc.median.month(daily.no,   F)
monthly.median.no2  <- calc.median.month(daily.no2,  F)

# Calculate monthly medians with imputation

monthly.median.imp.pm25 <- calc.median.month(daily.pm25, T)
monthly.median.imp.no   <- calc.median.month(daily.no,   T)
monthly.median.imp.no2  <- calc.median.month(daily.no2,  T)

# Bind data frames into one

monthly.pm25 <- bind_rows(monthly.mean.pm25,
                          monthly.mean.imp.pm25,
                          monthly.median.pm25,
                          monthly.median.imp.pm25)

monthly.no <- bind_rows(monthly.mean.no,
                       monthly.mean.imp.no,
                       monthly.median.no,
                       monthly.median.imp.no)

monthly.no2 <- bind_rows(monthly.mean.no2,
                        monthly.mean.imp.no2,
                        monthly.median.no2,
                        monthly.median.imp.no2)

rm(list = c(ls(pattern = "^monthly.mean"), ls(pattern = "^monthly.median")))

#' ---
#' Calc monthly minimum and max values
#' ---

calc.min.month <- function(df, is.imputed) {
  df <- na.omit(df, cols = "Value") %>%
    filter(Imputed == is.imputed)
  min <- group_by(df, Year, Month, Monitor) %>%
    summarise(Value = min(Value, na.rm = T)) %>%
    mutate(Measure = "Min", Imputed = is.imputed)
  return(min)
}

calc.max.month <- function(df, is.imputed) {
  df <- na.omit(df, cols = "Value") %>%
    filter(Imputed == is.imputed)
  max <- group_by(df, Year, Month, Monitor) %>%
    summarise(Value = max(Value, na.rm = T)) %>%
    mutate(Measure = "Max", Imputed = is.imputed)
  return(max)
}

# Get observed monthly min with no imputation

monthly.min.pm25 <- calc.min.month(daily.pm25, F)
monthly.min.no   <- calc.min.month(daily.no,   F)
monthly.min.no2  <- calc.min.month(daily.no2,  F)

monthly.min.imp.pm25 <- calc.min.month(daily.pm25, T)
monthly.min.imp.no   <- calc.min.month(daily.no,   T)
monthly.min.imp.no2  <- calc.min.month(daily.no2,  T)

# Get observed monthly max with no imputation

monthly.max.pm25 <- calc.max.month(daily.pm25, F)
monthly.max.no   <- calc.max.month(daily.no,   F)

```

```

monthly.max.no2 <- calc.max.month(daily.no2, F)

monthly.max.imp.pm25 <- calc.max.month(daily.pm25, T)
monthly.max.imp.no <- calc.max.month(daily.no, T)
monthly.max.imp.no2 <- calc.max.month(daily.no2, T)

# Bind data frames into one

monthly.bounds.pm25 <- bind_rows(list(monthly.min.pm25,
                                     monthly.max.pm25,
                                     monthly.min.imp.pm25,
                                     monthly.max.imp.pm25))

monthly.bounds.no <- bind_rows(list(monthly.min.no,
                                    monthly.max.no,
                                    monthly.min.imp.no,
                                    monthly.max.imp.no))

monthly.bounds.no2 <- bind_rows(list(monthly.min.no2,
                                     monthly.max.no2,
                                     monthly.min.imp.no2,
                                     monthly.max.imp.no2))

rm(list = c(ls(pattern = "^monthly.min"), ls(pattern = "^monthly.max")))

# Save min and max bounds

write_csv(monthly.bounds.pm25, "Data/Cleaned/pm25_bounds.csv")
write_csv(monthly.bounds.no, "Data/Cleaned/no_bounds.csv")
write_csv(monthly.bounds.no2, "Data/Cleaned/no2_bounds.csv")

rm(list = ls(pattern = "^monthly.bounds"))

#' ---
#' Calc temporal factor based on monthly and baseline avg
#' ---

# Establish baseline years based on LUR creation dates

baseline.pm25 <- filter(yearly.pm25, Year == 2003) %>%
  rename(Baseline = Value, LUR = Year)

baseline.no <- filter(yearly.no, Year == 2003 | Year == 2010) %>%
  rename(Baseline = Value, LUR = Year)

baseline.no2 <- filter(yearly.no2, Year == 2003 | Year == 2010) %>%
  rename(Baseline = Value, LUR = Year)

# Prepare temporal factor data frame with baseline year key

temp.factor.pm25 <- mutate(monthly.pm25,
                          LUR = 2003)

temp.factor.no <- mutate(monthly.no,
                        LUR = ifelse(Year < 2010, 2003, 2010))

temp.factor.no2 <- mutate(monthly.no2,
                        LUR = ifelse(Year < 2010, 2003, 2010))

# Merge monthly average dataframe to yearly baseline dataframe

merge.factor.baseline <- function(df.x, df.y) {
  df <- merge(df.x, df.y, all.x = T) %>%
    select(LUR, Year, Month, Measure, Imputed, Value, Baseline)
  return(df)
}

temp.factor.pm25 <- merge.factor.baseline(temp.factor.pm25, baseline.pm25)

```

```

temp.factor.no <- merge.factor.baseline(temp.factor.no, baseline.no)
temp.factor.no2 <- merge.factor.baseline(temp.factor.no2, baseline.no2)

# Calculate temporal factor

calc.temp.factor <- function(df) {
  return(mutate(df, Factor = Value / Baseline))
}

temp.factor.pm25 <- calc.temp.factor(temp.factor.pm25)
temp.factor.no <- calc.temp.factor(temp.factor.no)
temp.factor.no2 <- calc.temp.factor(temp.factor.no2)

rm(list = c(ls(pattern = "^calc."), ls(pattern = "^merge.")))

# Save output

write_csv(temp.factor.pm25, "Data/Cleaned/pm25_temporal_factor.csv")
write_csv(temp.factor.no, "Data/Cleaned/no_temporal_factor.csv")
write_csv(temp.factor.no2, "Data/Cleaned/no2_temporal_factor.csv")

proc.time() - ptm

rm(list = ls())

```

## 06\_temporally\_adjust\_pollutant.R

```

# Description -----

# Create spatiotemporal model of pollutants for Metro Van from 2003–2014.
# Take LUR values at a given postal code and multiple it by temporal factor
# of a given month and year.

# Load libraries and prepare data -----

library(tidyverse)

# Functions -----

dir <- "Data/Cleaned/"
dir.temp <- "Data/Temp/"

save.files <- function(df, dir, filename) {
  write_csv(df, paste0(dir, filename, ".csv"))
}

remove.files <- function(dir, filename) {
  file.remove(paste0(dir, filename, ".csv"))
}

if(!dir.exists(dir.temp)) { dir.create(dir.temp) }

# Load cleaned spatial air pollution data

lur.val.pm25 <- read_csv("Data/Cleaned/pm25_lur_val_at_pc.csv")
lur.val.no <- read_csv("Data/Cleaned/no_lur_val_at_pc.csv")
lur.val.no2 <- read_csv("Data/Cleaned/no2_lur_val_at_pc.csv")

# Load temporal factors

temp.factor.pm25 <- read_csv("Data/Cleaned/pm25_temporal_factor.csv")
temp.factor.no <- read_csv("Data/Cleaned/no_temporal_factor.csv")
temp.factor.no2 <- read_csv("Data/Cleaned/no2_temporal_factor.csv")

# Split temporal factor data frames

```

```

pm25.mean      <- filter(temp.factor.pm25, Measure == "Mean",  Imputed == F)
pm25.mean.imp  <- filter(temp.factor.pm25, Measure == "Mean",  Imputed == T)
pm25.median    <- filter(temp.factor.pm25, Measure == "Median", Imputed == F)
pm25.median.imp <- filter(temp.factor.pm25, Measure == "Median", Imputed == T)

no.mean        <- filter(temp.factor.no,   Measure == "Mean",  Imputed == F)
no.mean.imp    <- filter(temp.factor.no,   Measure == "Mean",  Imputed == T)
no.median      <- filter(temp.factor.no,   Measure == "Median", Imputed == F)
no.median.imp  <- filter(temp.factor.no,   Measure == "Median", Imputed == T)

no2.mean       <- filter(temp.factor.no2,  Measure == "Mean",  Imputed == F)
no2.mean.imp   <- filter(temp.factor.no2,  Measure == "Mean",  Imputed == T)
no2.median     <- filter(temp.factor.no2,  Measure == "Median", Imputed == F)
no2.median.imp <- filter(temp.factor.no2,  Measure == "Median", Imputed == T)

# Prepare PC data frame -----

# Prepare data frame with PC-extracted LUR values for vectorized operations.
# First create monthly columns, and then reshape data frame to long format.

add.months <- function(df) {

  # Add month columns

  df.months <- df
  for(month in 1:12) {
    df.months[, paste0("M", as.character(month))] = month
  }

  # Gather month columns to long

  df.long <- gather(df.months, Column, Month, M1:M12) %>%
    select(Year, Month, PC, Value) %>%
    arrange(Year, PC, Month)

  return(df.long)
}

lur.val.pm25 <- add.months(lur.val.pm25)
lur.val.no   <- add.months(lur.val.no)
lur.val.no2  <- add.months(lur.val.no2)

summary.pm25 <- group_by(lur.val.pm25, Year, PC) %>%
  summarise(SD = sd(Value), N = n())

summary.no <- group_by(lur.val.no, Year, PC) %>%
  summarise(SD = sd(Value), N = n())

summary.no2 <- group_by(lur.val.no2, Year, PC) %>%
  summarise(SD = sd(Value), N = n())

all(summary.pm25$SD == 0, na.rm = T)
all(summary.no$SD   == 0, na.rm = T)
all(summary.no2$SD  == 0, na.rm = T)

all(summary.pm25$N == 12, na.rm = T)
all(summary.no$N   == 12, na.rm = T)
all(summary.no2$N  == 12, na.rm = T)

rm(list = ls(pattern = "^summary."))

# Temporally adjusted estimates: no truncation -----

# Temporally adjust pollutants to create spatiotemporal models. Develop
# temporally adjusted estimate by multiplying LUR value at PC by temporal
# factor for each month for each year.

model.pollutant <- function(df.lur.val, df.factor) {

```

```

# Merge temporal factor to data frame with LUR values at PC
df <- merge(df.lur.val,
            select(df.factor, Year, Month, Measure, Imputed, Factor),
            by = c("Year", "Month"),
            all.x = T) %>%

# Select variables of interest

select(Measure, Imputed, Year, Month, PC, Value, Factor) %>%

# Sort columns

arrange(Measure, Imputed, Year, PC, Month) %>%

# Calculate temporally adjusted estimate

mutate(Modelled = Value * Factor)
return(df)
}

model.pm25.mean      <- model.pollutant(lur.val.pm25, pm25.mean)
model.pm25.mean.imp  <- model.pollutant(lur.val.pm25, pm25.mean.imp)
model.pm25.median    <- model.pollutant(lur.val.pm25, pm25.median)
model.pm25.median.imp <- model.pollutant(lur.val.pm25, pm25.median.imp)

model.no.mean        <- model.pollutant(lur.val.no, no.mean)
model.no.mean.imp    <- model.pollutant(lur.val.no, no.mean.imp)
model.no.median      <- model.pollutant(lur.val.no, no.median)
model.no.median.imp  <- model.pollutant(lur.val.no, no.median.imp)

model.no2.mean       <- model.pollutant(lur.val.no2, no2.mean)
model.no2.mean.imp   <- model.pollutant(lur.val.no2, no2.mean.imp)
model.no2.median     <- model.pollutant(lur.val.no2, no2.median)
model.no2.median.imp <- model.pollutant(lur.val.no2, no2.median.imp)

save.files(model.pm25.mean,      dir.temp, "model.pm25.mean")
save.files(model.pm25.mean.imp,  dir.temp, "model.pm25.mean.imp")
save.files(model.pm25.median,    dir.temp, "model.pm25.median")
save.files(model.pm25.median.imp, dir.temp, "model.pm25.median.imp")

save.files(model.no.mean,        dir.temp, "model.no.mean")
save.files(model.no.mean.imp,    dir.temp, "model.no.mean.imp")
save.files(model.no.median,      dir.temp, "model.no.median")
save.files(model.no.median.imp,  dir.temp, "model.no.median.imp")

save.files(model.no2.mean,       dir.temp, "model.no2.mean")
save.files(model.no2.mean.imp,   dir.temp, "model.no2.mean.imp")
save.files(model.no2.median,     dir.temp, "model.no2.median")
save.files(model.no2.median.imp, dir.temp, "model.no2.median.imp")

rm(list = c(ls(pattern = "^model.pm"),
            ls(pattern = "^model.no")))

# Temporally adjusted estimates: LUR truncation -----

# Temporally adjust pollutants, but first truncate LUR value floor based on
# observed air monitoring station values.

# Load data with min and max observed values

monthly.bounds.pm25 <- read_csv("Data/Cleaned/pm25_bounds.csv")
monthly.bounds.no  <- read_csv("Data/Cleaned/no_bounds.csv")
monthly.bounds.no2 <- read_csv("Data/Cleaned/no2_bounds.csv")

# Subset data to only minimum observations for LUR year

```

```

get.lur.min <- function(df, year, is.imputed) {
  df.min <- filter(df,
    Year == year,
    Measure == "Min",
    Imputed == is.imputed) %>%
    select(Measure, Imputed, Monitor, Year, Month, Value) %>%
    arrange(Measure, Imputed, Monitor, Year, Month)
  return(df.min)
}

lur.min.pm25 <- get.lur.min(monthly.bounds.pm25, 2003, F)
lur.min.no.03 <- get.lur.min(monthly.bounds.no, 2003, F)
lur.min.no.10 <- get.lur.min(monthly.bounds.no, 2010, F)
lur.min.no2.03 <- get.lur.min(monthly.bounds.no2, 2003, F)
lur.min.no2.10 <- get.lur.min(monthly.bounds.no2, 2010, F)

lur.min.pm25.imp <- get.lur.min(monthly.bounds.pm25, 2003, T)
lur.min.no.03.imp <- get.lur.min(monthly.bounds.no, 2003, T)
lur.min.no.10.imp <- get.lur.min(monthly.bounds.no, 2010, T)
lur.min.no2.03.imp <- get.lur.min(monthly.bounds.no2, 2003, T)
lur.min.no2.10.imp <- get.lur.min(monthly.bounds.no2, 2010, T)

# Calculate floor based on median-median & imp of min values for LUR year

annual.mean.min.pm25 <- mean(lur.min.pm25$Value, na.rm = T)
annual.mean.min.no.03 <- mean(lur.min.no.03$Value, na.rm = T)
annual.mean.min.no.10 <- mean(lur.min.no.10$Value, na.rm = T)
annual.mean.min.no2.03 <- mean(lur.min.no2.03$Value, na.rm = T)
annual.mean.min.no2.10 <- mean(lur.min.no2.10$Value, na.rm = T)

annual.mean.min.pm25.imp <- mean(lur.min.pm25.imp$Value, na.rm = T)
annual.mean.min.no.03.imp <- mean(lur.min.no.03.imp$Value, na.rm = T)
annual.mean.min.no.10.imp <- mean(lur.min.no.10.imp$Value, na.rm = T)
annual.mean.min.no2.03.imp <- mean(lur.min.no2.03.imp$Value, na.rm = T)
annual.mean.min.no2.10.imp <- mean(lur.min.no2.10.imp$Value, na.rm = T)

annual.median.min.pm25 <- median(lur.min.pm25$Value, na.rm = T)
annual.median.min.no.03 <- median(lur.min.no.03$Value, na.rm = T)
annual.median.min.no.10 <- median(lur.min.no.10$Value, na.rm = T)
annual.median.min.no2.03 <- median(lur.min.no2.03$Value, na.rm = T)
annual.median.min.no2.10 <- median(lur.min.no2.10$Value, na.rm = T)

annual.median.min.pm25.imp <- median(lur.min.pm25.imp$Value, na.rm = T)
annual.median.min.no.03.imp <- median(lur.min.no.03.imp$Value, na.rm = T)
annual.median.min.no.10.imp <- median(lur.min.no.10.imp$Value, na.rm = T)
annual.median.min.no2.03.imp <- median(lur.min.no2.03.imp$Value, na.rm = T)
annual.median.min.no2.10.imp <- median(lur.min.no2.10.imp$Value, na.rm = T)

# Save LUR floors

floor.1 <- c(annual.mean.min.pm25, "Mean", FALSE, "PM25", 2003)
floor.2 <- c(annual.mean.min.no.03, "Mean", FALSE, "NO", 2003)
floor.3 <- c(annual.mean.min.no.10, "Mean", FALSE, "NO", 2010)
floor.4 <- c(annual.mean.min.no2.03, "Mean", FALSE, "NO2", 2003)
floor.5 <- c(annual.mean.min.no2.10, "Mean", FALSE, "NO2", 2010)

floor.6 <- c(annual.mean.min.pm25.imp, "Mean", TRUE, "PM25", 2003)
floor.7 <- c(annual.mean.min.no.03.imp, "Mean", TRUE, "NO", 2003)
floor.8 <- c(annual.mean.min.no.10.imp, "Mean", TRUE, "NO", 2010)
floor.9 <- c(annual.mean.min.no2.03.imp, "Mean", TRUE, "NO2", 2003)
floor.10 <- c(annual.mean.min.no2.10.imp, "Mean", TRUE, "NO2", 2010)

floor.11 <- c(annual.median.min.pm25, "Median", FALSE, "PM25", 2003)
floor.12 <- c(annual.median.min.no.03, "Median", FALSE, "NO", 2003)
floor.13 <- c(annual.median.min.no.10, "Median", FALSE, "NO", 2010)
floor.14 <- c(annual.median.min.no2.03, "Median", FALSE, "NO2", 2003)
floor.15 <- c(annual.median.min.no2.10, "Median", FALSE, "NO2", 2010)

```

```

floor.16 <- c(annual.median.min.pm25.imp, "Median", TRUE, "PM25", 2003)
floor.17 <- c(annual.median.min.no.03.imp, "Median", TRUE, "NO", 2003)
floor.18 <- c(annual.median.min.no.10.imp, "Median", TRUE, "NO", 2010)
floor.19 <- c(annual.median.min.no2.03.imp, "Median", TRUE, "NO2", 2003)
floor.20 <- c(annual.median.min.no2.10.imp, "Median", TRUE, "NO2", 2010)

lur.floors <- as.data.frame(cbind(floor.1, floor.2,
                                floor.3, floor.4,
                                floor.5, floor.6,
                                floor.7, floor.8,
                                floor.9, floor.10,
                                floor.11, floor.12,
                                floor.13, floor.14,
                                floor.15, floor.16,
                                floor.17, floor.18,
                                floor.19, floor.20))

lur.floors <- as.data.frame(t(lur.floors))

names(lur.floors) <- c("Floor", "Measure", "Imputed", "Pollutant", "Year")

rownames(lur.floors) <- 1:nrow(lur.floors)

lur.floors <- mutate(lur.floors,
                    Floor = as.numeric(as.character(Floor)),
                    Measure = as.character(Measure),
                    Imputed = as.logical(as.character(Imputed)),
                    Pollutant = as.character(Pollutant),
                    Year = as.numeric(as.character(Year)))

save.files(lur.floors, dir, "lur_floors")

rm(list = c(ls(pattern = "^lur.min."),
            ls(pattern = "^annual.mean."),
            ls(pattern = "^annual.median."),
            ls(pattern = "^floor.")))

# Truncate LUR value floors

ltrunc.floor <- function(df, df.floor, measure, imputed, pollutant, year) {
  df.floor <- filter(df.floor,
                    Measure == measure,
                    Imputed == imputed,
                    Pollutant == pollutant,
                    Year == year)$Floor[1]
  df <- mutate(df, Value = ifelse(Value < df.floor, df.floor, Value))
  return(df)
}

lur.val.no.03 <- filter(lur.val.no, Year < 2010)
lur.val.no.10 <- filter(lur.val.no, Year >= 2010)

lur.val.no2.03 <- filter(lur.val.no2, Year < 2010)
lur.val.no2.10 <- filter(lur.val.no2, Year >= 2010)

# PM2.5

lur.trunc.pm25.mean <- ltrunc.floor(lur.val.pm25, lur.floors,
                                   "Mean", F,
                                   "PM25", 2003)

lur.trunc.pm25.mean.imp <- ltrunc.floor(lur.val.pm25, lur.floors,
                                       "Mean", T,
                                       "PM25", 2003)

lur.trunc.pm25.median <- ltrunc.floor(lur.val.pm25, lur.floors,
                                       "Median", F,
                                       "PM25", 2003)

```

```

lur.trunc.pm25.median.imp <- ltrunc.floor(lur.val.pm25, lur.floors,
                                         "Median", T,
                                         "PM25", 2003)

filter(lur.floors, Pollutant == "PM25", Year == 2003)

summary(lur.trunc.pm25.mean$Value)
summary(lur.trunc.pm25.mean.imp$Value)
summary(lur.trunc.pm25.median$Value)
summary(lur.trunc.pm25.median.imp$Value)

# NO - 2003

lur.trunc.no.03.mean      <- ltrunc.floor(lur.val.no.03, lur.floors,
                                         "Mean", F,
                                         "NO", 2003)

lur.trunc.no.03.mean.imp <- ltrunc.floor(lur.val.no.03, lur.floors,
                                         "Mean", T,
                                         "NO", 2003)

lur.trunc.no.03.median   <- ltrunc.floor(lur.val.no.03, lur.floors,
                                         "Median", F,
                                         "NO", 2003)

lur.trunc.no.03.median.imp <- ltrunc.floor(lur.val.no.03, lur.floors,
                                         "Median", T,
                                         "NO", 2003)

filter(lur.floors, Pollutant == "NO", Year == 2003)

summary(lur.trunc.no.03.mean$Value)
summary(lur.trunc.no.03.mean.imp$Value)
summary(lur.trunc.no.03.median$Value)
summary(lur.trunc.no.03.median.imp$Value)

# NO - 2010

lur.trunc.no.10.mean     <- ltrunc.floor(lur.val.no.10, lur.floors,
                                         "Mean", F,
                                         "NO", 2010)

lur.trunc.no.10.mean.imp <- ltrunc.floor(lur.val.no.10, lur.floors,
                                         "Mean", T,
                                         "NO", 2010)

lur.trunc.no.10.median   <- ltrunc.floor(lur.val.no.10, lur.floors,
                                         "Median", F,
                                         "NO", 2010)

lur.trunc.no.10.median.imp <- ltrunc.floor(lur.val.no.10, lur.floors,
                                         "Median", T,
                                         "NO", 2010)

filter(lur.floors, Pollutant == "NO", Year == 2010)

summary(lur.trunc.no.10.mean$Value)
summary(lur.trunc.no.10.mean.imp$Value)
summary(lur.trunc.no.10.median$Value)
summary(lur.trunc.no.10.median.imp$Value)

# NO2 - 2003

lur.trunc.no2.03.mean    <- ltrunc.floor(lur.val.no2.03, lur.floors,
                                         "Mean", F,
                                         "NO2", 2003)

```

```

lur.trunc.no2.03.mean.imp <- ltrunc.floor(lur.val.no2.03, lur.floors,
                                         "Mean", T,
                                         "NO2", 2003)

lur.trunc.no2.03.median <- ltrunc.floor(lur.val.no2.03, lur.floors,
                                         "Median", F,
                                         "NO2", 2003)

lur.trunc.no2.03.median.imp <- ltrunc.floor(lur.val.no2.03, lur.floors,
                                         "Median", T,
                                         "NO2", 2003)

filter(lur.floors, Pollutant == "NO2", Year == 2003)

summary(lur.trunc.no2.03.mean$Value)
summary(lur.trunc.no2.03.mean.imp$Value)
summary(lur.trunc.no2.03.median$Value)
summary(lur.trunc.no2.03.median.imp$Value)

# NO2 - 2010

lur.trunc.no2.10.mean <- ltrunc.floor(lur.val.no2.10, lur.floors,
                                       "Mean", F,
                                       "NO2", 2010)

lur.trunc.no2.10.mean.imp <- ltrunc.floor(lur.val.no2.10, lur.floors,
                                       "Mean", T,
                                       "NO2", 2010)

lur.trunc.no2.10.median <- ltrunc.floor(lur.val.no2.10, lur.floors,
                                       "Median", F,
                                       "NO2", 2010)

lur.trunc.no2.10.median.imp <- ltrunc.floor(lur.val.no2.10, lur.floors,
                                       "Median", T,
                                       "NO2", 2010)

filter(lur.floors, Pollutant == "NO2", Year == 2010)

summary(lur.trunc.no2.10.mean$Value)
summary(lur.trunc.no2.10.mean.imp$Value)
summary(lur.trunc.no2.10.median$Value)
summary(lur.trunc.no2.10.median.imp$Value)

rm(list = ls(pattern = "^lur.val"))

# Bind respective NO and NO2 columns

lur.trunc.no.mean <- rbind(lur.trunc.no.03.mean,
                           lur.trunc.no.10.mean)

lur.trunc.no.mean.imp <- rbind(lur.trunc.no.03.mean.imp,
                               lur.trunc.no.10.mean.imp)

lur.trunc.no.median <- rbind(lur.trunc.no.03.median,
                              lur.trunc.no.10.median)

lur.trunc.no.median.imp <- rbind(lur.trunc.no.03.median.imp,
                                  lur.trunc.no.10.median.imp)

rm(list = c(ls(pattern = "^lur.trunc.no.03"),
            ls(pattern = "^lur.trunc.no.10")))

lur.trunc.no2.mean <- rbind(lur.trunc.no2.03.mean,
                             lur.trunc.no2.10.mean)

lur.trunc.no2.mean.imp <- rbind(lur.trunc.no2.03.mean.imp,
                                 lur.trunc.no2.10.mean.imp)

```

```

lur.trunc.no2.median <- rbind(lur.trunc.no2.03.median,
                             lur.trunc.no2.10.median)

lur.trunc.no2.median.imp <- rbind(lur.trunc.no2.03.median.imp,
                                  lur.trunc.no2.10.median.imp)

rm(list = c(ls(pattern = "^lur.trunc.no2.03"),
            ls(pattern = "^lur.trunc.no2.10")))

# Check LUR truncation

check.min <- function(df, floor, pollutant, measure, imputed) {
  min <- filter(floor,
                Pollutant == pollutant,
                Measure == measure,
                Imputed == imputed);
  print(min)
  drop_na(df, Value) %>%
  group_by(Year) %>%
  summarise(Min = min(Value))
}

check.min(lur.trunc.pm25.mean,      lur.floors, "PM25", "Mean",  F)
check.min(lur.trunc.pm25.mean.imp,  lur.floors, "PM25", "Mean",  T)
check.min(lur.trunc.pm25.median,    lur.floors, "PM25", "Median", F)
check.min(lur.trunc.pm25.median.imp, lur.floors, "PM25", "Median", T)

check.min(lur.trunc.no.mean,        lur.floors, "NO",   "Mean",  F)
check.min(lur.trunc.no.mean.imp,    lur.floors, "NO",   "Mean",  T)
check.min(lur.trunc.no.median,      lur.floors, "NO",   "Median", F)
check.min(lur.trunc.no.median.imp,  lur.floors, "NO",   "Median", T)

check.min(lur.trunc.no2.mean,       lur.floors, "NO2",  "Mean",  F)
check.min(lur.trunc.no2.mean.imp,   lur.floors, "NO2",  "Mean",  T)
check.min(lur.trunc.no2.median,     lur.floors, "NO2",  "Median", F)
check.min(lur.trunc.no2.median.imp, lur.floors, "NO2",  "Median", T)

# Check number of postal codes that were LUR floor truncated

lur.floors.check <- filter(lur.floors,
                          Year == 2003,
                          Measure == "Mean",
                          Imputed == TRUE) %>%
  select(Floor, Pollutant)

check.num.pc.lur.trunc <- function(df, pollutant) {
  floor <- lur.floors.check$Floor[lur.floors.check$Pollutant == pollutant]
  total <- filter(df, Year == 2003) %>%
    select(Value)
  filtered <- filter(df, Year == 2003,
                    Value == floor) %>%
    select(Value) %>%
    na.omit()
  print(paste0("Total PC's filtered: ", nrow(filtered),
              " out of ", nrow(total),
              " (" , round(nrow(filtered) / nrow(total) * 100, 1), "%)")
)
}

check.num.pc.lur.trunc(lur.trunc.pm25.mean.imp, "PM25")
check.num.pc.lur.trunc(lur.trunc.no.mean.imp,   "NO")
check.num.pc.lur.trunc(lur.trunc.no2.mean.imp,  "NO2")

rm(lur.floors.check, check.num.pc.lur.trunc)

# Temporally adjust LUR truncated values

model.pm25.mean.ltrc <- model.pollutant(lur.trunc.pm25.mean,

```

```

                                pm25.mean)
model.pm25.mean.imp.ltrc <- model.pollutant(lur.trunc.pm25.mean.imp,
                                pm25.mean.imp)
model.pm25.median.ltrc <- model.pollutant(lur.trunc.pm25.median,
                                pm25.median)
model.pm25.median.imp.ltrc <- model.pollutant(lur.trunc.pm25.median.imp,
                                pm25.median.imp)
model.no.mean.ltrc <- model.pollutant(lur.trunc.no.mean,
                                no.mean)
model.no.mean.imp.ltrc <- model.pollutant(lur.trunc.no.mean.imp,
                                no.mean.imp)
model.no.median.ltrc <- model.pollutant(lur.trunc.no.median,
                                no.median)
model.no.median.imp.ltrc <- model.pollutant(lur.trunc.no.median.imp,
                                no.median.imp)
model.no2.mean.ltrc <- model.pollutant(lur.trunc.no2.mean,
                                no2.mean)
model.no2.mean.imp.ltrc <- model.pollutant(lur.trunc.no2.mean.imp,
                                no2.mean.imp)
model.no2.median.ltrc <- model.pollutant(lur.trunc.no2.median,
                                no2.median)
model.no2.median.imp.ltrc <- model.pollutant(lur.trunc.no2.median.imp,
                                no2.median.imp)

rm(list = ls(pattern = "^lur.trunc."))

save.files(model.pm25.mean.ltrc, dir.temp, "model.pm25.mean.ltrc")
save.files(model.pm25.mean.imp.ltrc, dir.temp, "model.pm25.mean.imp.ltrc")
save.files(model.pm25.median.ltrc, dir.temp, "model.pm25.median.ltrc")
save.files(model.pm25.median.imp.ltrc, dir.temp, "model.pm25.median.imp.ltrc")

save.files(model.no.mean.ltrc, dir.temp, "model.no.mean.ltrc")
save.files(model.no.mean.imp.ltrc, dir.temp, "model.no.mean.imp.ltrc")
save.files(model.no.median.ltrc, dir.temp, "model.no.median.ltrc")
save.files(model.no.median.imp.ltrc, dir.temp, "model.no.median.imp.ltrc")

save.files(model.no2.mean.ltrc, dir.temp, "model.no2.mean.ltrc")
save.files(model.no2.mean.imp.ltrc, dir.temp, "model.no2.mean.imp.ltrc")
save.files(model.no2.median.ltrc, dir.temp, "model.no2.median.ltrc")
save.files(model.no2.median.imp.ltrc, dir.temp, "model.no2.median.imp.ltrc")

rm(list = c(ls(pattern = "^model.pm"),
            ls(pattern = "^model.no")))

# Temporally adjusted estimates: temporal truncation -----

# Temporally adjust pollutants then truncate temporally adjusted floor based
# on observed air monitoring station values.

# Calculate monthly floors based on central tendency and imputation

get.month.min <- function(df) {
  df <- filter(df, Measure == "Min") %>%
    drop_na(Value)
  df.min <- group_by(df, Imputed, Year, Month) %>%
    summarise(Mean = mean(Value),
              Median = median(Value),

```

```

      N.Monitors = n())
df.min <- gather(df.min, Measure, Value, c(Mean, Median)) %>%
  arrange(Imputed, Measure, Year, Month) %>%
  rename(Floor = Value)
return(df.min)
}

tem.month.floors.pm25 <- get.month.min(monthly.bounds.pm25)
tem.month.floors.no <- get.month.min(monthly.bounds.no)
tem.month.floors.no2 <- get.month.min(monthly.bounds.no2)

save.files(tem.month.floors.pm25, dir, "pm25_tem_floors_monthly")
save.files(tem.month.floors.no, dir, "no_tem_floors_monthly")
save.files(tem.month.floors.no2, dir, "no2_tem_floors_monthly")

floor.pm25.mean <- filter(tem.month.floors.pm25,
  Measure == "Mean",
  Imputed == F)

floor.pm25.mean.imp <- filter(tem.month.floors.pm25,
  Measure == "Mean",
  Imputed == T)

floor.pm25.median <- filter(tem.month.floors.pm25,
  Measure == "Median",
  Imputed == F)

floor.pm25.median.imp <- filter(tem.month.floors.pm25,
  Measure == "Median",
  Imputed == T)

floor.no.mean <- filter(tem.month.floors.no,
  Measure == "Mean",
  Imputed == F)

floor.no.mean.imp <- filter(tem.month.floors.no,
  Measure == "Mean",
  Imputed == T)

floor.no.median <- filter(tem.month.floors.no,
  Measure == "Median",
  Imputed == F)

floor.no.median.imp <- filter(tem.month.floors.no,
  Measure == "Median",
  Imputed == T)

floor.no2.mean <- filter(tem.month.floors.no2,
  Measure == "Mean",
  Imputed == F)

floor.no2.mean.imp <- filter(tem.month.floors.no2,
  Measure == "Mean",
  Imputed == T)

floor.no2.median <- filter(tem.month.floors.no2,
  Measure == "Median",
  Imputed == F)

floor.no2.median.imp <- filter(tem.month.floors.no2,
  Measure == "Median",
  Imputed == T)

# Trucante floor of temporally adjusted models

ttrunc.floor <- function(df, floor) {
  df.ttrunc <- merge(df, floor, by = c("Year",
    "Month",

```

```

                                "Measure",
                                "Imputed")) %>%
mutate(Truncated = ifelse(Modelled < Floor, Floor, Modelled)) %>%
select(Measure, Imputed,
       Year,      Month,
       PC,       Value,
       Factor,   Truncated) %>%
arrange(Measure, Imputed, Year, PC, Month) %>%
rename(Modelled = Truncated)
return(df.ttrunc)
}

# Truncate floor of temporally adjusted models: PM2.5

model.pm25.mean      <- read_csv("Data/Temp/model.pm25.mean.csv")
model.pm25.mean.imp  <- read_csv("Data/Temp/model.pm25.mean.imp.csv")
model.pm25.median    <- read_csv("Data/Temp/model.pm25.median.csv")
model.pm25.median.imp <- read_csv("Data/Temp/model.pm25.median.imp.csv")

model.pm25.mean.ttrc <- ttrunc.floor(model.pm25.mean,
                                     floor.pm25.mean)

model.pm25.mean.imp.ttrc <- ttrunc.floor(model.pm25.mean.imp,
                                         floor.pm25.mean.imp)

model.pm25.median.ttrc <- ttrunc.floor(model.pm25.median,
                                       floor.pm25.median)

model.pm25.median.imp.ttrc <- ttrunc.floor(model.pm25.median.imp,
                                           floor.pm25.median.imp)

save.files(model.pm25.mean.ttrc,      dir.temp, "model.pm25.mean.ttrc")
save.files(model.pm25.mean.imp.ttrc,  dir.temp, "model.pm25.mean.imp.ttrc")
save.files(model.pm25.median.ttrc,    dir.temp, "model.pm25.median.ttrc")
save.files(model.pm25.median.imp.ttrc, dir.temp, "model.pm25.median.imp.ttrc")

rm(list = ls(pattern = "^model.pm"))

# Truncate floor of temporally adjusted models: NO

model.no.mean      <- read_csv("Data/Temp/model.no.mean.csv")
model.no.mean.imp  <- read_csv("Data/Temp/model.no.mean.imp.csv")
model.no.median    <- read_csv("Data/Temp/model.no.median.csv")
model.no.median.imp <- read_csv("Data/Temp/model.no.median.imp.csv")

model.no.mean.ttrc <- ttrunc.floor(model.no.mean,
                                     floor.no.mean)

model.no.mean.imp.ttrc <- ttrunc.floor(model.no.mean.imp,
                                         floor.no.mean.imp)

model.no.median.ttrc <- ttrunc.floor(model.no.median,
                                       floor.no.median)

model.no.median.imp.ttrc <- ttrunc.floor(model.no.median.imp,
                                           floor.no.median.imp)

save.files(model.no.mean.ttrc,      dir.temp, "model.no.mean.ttrc")
save.files(model.no.mean.imp.ttrc,  dir.temp, "model.no.mean.imp.ttrc")
save.files(model.no.median.ttrc,    dir.temp, "model.no.median.ttrc")
save.files(model.no.median.imp.ttrc, dir.temp, "model.no.median.imp.ttrc")

rm(list = ls(pattern = "^model.no"))

# Truncate floor of temporally adjusted models: NO2

model.no2.mean      <- read_csv("Data/Temp/model.no2.mean.csv")
model.no2.mean.imp  <- read_csv("Data/Temp/model.no2.mean.imp.csv")

```

```

model.no2.median <- read_csv("Data/Temp/model.no2.median.csv")
model.no2.median.imp <- read_csv("Data/Temp/model.no2.median.imp.csv")

model.no2.mean.ttrc <- ttrunc.floor(model.no2.mean,
                                   floor.no2.mean)

model.no2.mean.imp.ttrc <- ttrunc.floor(model.no2.mean.imp,
                                       floor.no2.mean.imp)

model.no2.median.ttrc <- ttrunc.floor(model.no2.median,
                                     floor.no2.median)

model.no2.median.imp.ttrc <- ttrunc.floor(model.no2.median.imp,
                                         floor.no2.median.imp)

save.files(model.no2.mean.ttrc, dir.temp, "model.no2.mean.ttrc")
save.files(model.no2.mean.imp.ttrc, dir.temp, "model.no2.mean.imp.ttrc")
save.files(model.no2.median.ttrc, dir.temp, "model.no2.median.ttrc")
save.files(model.no2.median.imp.ttrc, dir.temp, "model.no2.median.imp.ttrc")

rm(list = ls(pattern = "^model.no"))

# Bind pollutant: PM25 -----

model.pm25.mean <- read_csv("Data/Temp/model.pm25.mean.csv")
model.pm25.mean.imp <- read_csv("Data/Temp/model.pm25.mean.imp.csv")
model.pm25.median <- read_csv("Data/Temp/model.pm25.median.csv")
model.pm25.median.imp <- read_csv("Data/Temp/model.pm25.median.imp.csv")
model.pm25.mean.ltrc <- read_csv("Data/Temp/model.pm25.mean.ltrc.csv")
model.pm25.mean.imp.ltrc <- read_csv("Data/Temp/model.pm25.mean.imp.ltrc.csv")
model.pm25.median.ltrc <- read_csv("Data/Temp/model.pm25.median.ltrc.csv")
model.pm25.median.imp.ltrc <- read_csv("Data/Temp/model.pm25.median.imp.ltrc.csv")
model.pm25.mean.ttrc <- read_csv("Data/Temp/model.pm25.mean.ttrc.csv")
model.pm25.mean.imp.ttrc <- read_csv("Data/Temp/model.pm25.mean.imp.ttrc.csv")
model.pm25.median.ttrc <- read_csv("Data/Temp/model.pm25.median.ttrc.csv")
model.pm25.median.imp.ttrc <- read_csv("Data/Temp/model.pm25.median.imp.ttrc.csv")

# Compare all models: PM2.5

summary(model.pm25.mean)
summary(model.pm25.mean.imp)
summary(model.pm25.median)
summary(model.pm25.median.imp)

summary(model.pm25.mean.ltrc)
summary(model.pm25.mean.imp.ltrc)
summary(model.pm25.median.ltrc)
summary(model.pm25.median.imp.ltrc)

summary(model.pm25.mean.ttrc)
summary(model.pm25.mean.imp.ttrc)
summary(model.pm25.median.ttrc)
summary(model.pm25.median.imp.ttrc)

# Add a column for imputation method

model.pm25.mean$Truncation = "None"
model.pm25.mean.imp$Truncation = "None"
model.pm25.median$Truncation = "None"
model.pm25.median.imp$Truncation = "None"

model.pm25.mean.ltrc$Truncation = "LUR"
model.pm25.mean.imp.ltrc$Truncation = "LUR"
model.pm25.median.ltrc$Truncation = "LUR"
model.pm25.median.imp.ltrc$Truncation = "LUR"

model.pm25.mean.ttrc$Truncation = "Temporal"
model.pm25.mean.imp.ttrc$Truncation = "Temporal"

```

```

model.pm25.median.ttrc$Truncation      = "Temporal"
model.pm25.median.imp.ttrc$Truncation = "Temporal"

# Bind models together

model.pm25 <- bind_rows(list(model.pm25.mean,
                             model.pm25.mean.imp,
                             model.pm25.median,
                             model.pm25.median.imp,
                             model.pm25.mean.ltrc,
                             model.pm25.mean.imp.ltrc,
                             model.pm25.median.ltrc,
                             model.pm25.median.imp.ltrc,
                             model.pm25.mean.ttrc,
                             model.pm25.mean.imp.ttrc,
                             model.pm25.median.ttrc,
                             model.pm25.median.imp.ttrc))

# Save file

save.files(model.pm25, dir, "pm25_model")

rm(list = ls(pattern = "^model.pm25"))

# Bind pollutant: NO -----

model.no.mean          <- read_csv("Data/Temp/model.no.mean.csv")
model.no.mean.imp      <- read_csv("Data/Temp/model.no.mean.imp.csv")
model.no.median        <- read_csv("Data/Temp/model.no.median.csv")
model.no.median.imp    <- read_csv("Data/Temp/model.no.median.imp.csv")
model.no.mean.ltrc     <- read_csv("Data/Temp/model.no.mean.ltrc.csv")
model.no.mean.imp.ltrc <- read_csv("Data/Temp/model.no.mean.imp.ltrc.csv")
model.no.median.ltrc   <- read_csv("Data/Temp/model.no.median.ltrc.csv")
model.no.median.imp.ltrc <- read_csv("Data/Temp/model.no.median.imp.ltrc.csv")
model.no.mean.ttrc     <- read_csv("Data/Temp/model.no.mean.ttrc.csv")
model.no.mean.imp.ttrc <- read_csv("Data/Temp/model.no.mean.imp.ttrc.csv")
model.no.median.ttrc   <- read_csv("Data/Temp/model.no.median.ttrc.csv")
model.no.median.imp.ttrc <- read_csv("Data/Temp/model.no.median.imp.ttrc.csv")

# Compare all models: NO

summary(model.no.mean)
summary(model.no.mean.imp)
summary(model.no.median)
summary(model.no.median.imp)

summary(model.no.mean.ltrc)
summary(model.no.mean.imp.ltrc)
summary(model.no.median.ltrc)
summary(model.no.median.imp.ltrc)

summary(model.no.mean.ttrc)
summary(model.no.mean.imp.ttrc)
summary(model.no.median.ttrc)
summary(model.no.median.imp.ttrc)

# Add a column for imputation method

model.no.mean$Truncation      = "None"
model.no.mean.imp$Truncation  = "None"
model.no.median$Truncation    = "None"
model.no.median.imp$Truncation = "None"

model.no.mean.ltrc$Truncation = "LUR"
model.no.mean.imp.ltrc$Truncation = "LUR"
model.no.median.ltrc$Truncation = "LUR"
model.no.median.imp.ltrc$Truncation = "LUR"

```

```

model.no.mean.ttrc$Truncation      = "Temporal"
model.no.mean.imp.ttrc$Truncation  = "Temporal"
model.no.median.ttrc$Truncation    = "Temporal"
model.no.median.imp.ttrc$Truncation = "Temporal"

# Bind models together

model.no <- bind_rows(list(model.no.mean,
                           model.no.mean.imp,
                           model.no.median,
                           model.no.median.imp,
                           model.no.mean.ltrc,
                           model.no.mean.imp.ltrc,
                           model.no.median.ltrc,
                           model.no.median.imp.ltrc,
                           model.no.mean.ttrc,
                           model.no.mean.imp.ttrc,
                           model.no.median.ttrc,
                           model.no.median.imp.ttrc))

# Save file

save.files(model.no, dir, "no_model")

rm(list = ls(pattern = "^model.no"))

# Bind pollutant: NO2 -----
model.no2.mean      <- read_csv("Data/Temp/model.no2.mean.csv")
model.no2.mean.imp  <- read_csv("Data/Temp/model.no2.mean.imp.csv")
model.no2.median    <- read_csv("Data/Temp/model.no2.median.csv")
model.no2.median.imp <- read_csv("Data/Temp/model.no2.median.imp.csv")
model.no2.mean.ltrc <- read_csv("Data/Temp/model.no2.mean.ltrc.csv")
model.no2.mean.imp.ltrc <- read_csv("Data/Temp/model.no2.mean.imp.ltrc.csv")
model.no2.median.ltrc <- read_csv("Data/Temp/model.no2.median.ltrc.csv")
model.no2.median.imp.ltrc <- read_csv("Data/Temp/model.no2.median.imp.ltrc.csv")
model.no2.mean.ttrc  <- read_csv("Data/Temp/model.no2.mean.ttrc.csv")
model.no2.mean.imp.ttrc <- read_csv("Data/Temp/model.no2.mean.imp.ttrc.csv")
model.no2.median.ttrc <- read_csv("Data/Temp/model.no2.median.ttrc.csv")
model.no2.median.imp.ttrc <- read_csv("Data/Temp/model.no2.median.imp.ttrc.csv")

# Compare all models: NO2

summary(model.no2.mean)
summary(model.no2.mean.imp)
summary(model.no2.median)
summary(model.no2.median.imp)

summary(model.no2.mean.ltrc)
summary(model.no2.mean.imp.ltrc)
summary(model.no2.median.ltrc)
summary(model.no2.median.imp.ltrc)

summary(model.no2.mean.ttrc)
summary(model.no2.mean.imp.ttrc)
summary(model.no2.median.ttrc)
summary(model.no2.median.imp.ttrc)

# Add a column for imputation method

model.no2.mean$Truncation      = "None"
model.no2.mean.imp$Truncation  = "None"
model.no2.median$Truncation    = "None"
model.no2.median.imp$Truncation = "None"

model.no2.mean.ltrc$Truncation  = "LUR"
model.no2.mean.imp.ltrc$Truncation = "LUR"
model.no2.median.ltrc$Truncation = "LUR"

```

```

model.no2.median.imp.ltrc$Truncation = "LUR"

model.no2.mean.ttrc$Truncation      = "Temporal"
model.no2.mean.imp.ttrc$Truncation  = "Temporal"
model.no2.median.ttrc$Truncation    = "Temporal"
model.no2.median.imp.ttrc$Truncation = "Temporal"

# Bind models together

model.no2 <- bind_rows(list(model.no2.mean,
                             model.no2.mean.imp,
                             model.no2.median,
                             model.no2.median.imp,
                             model.no2.mean.ltrc,
                             model.no2.mean.imp.ltrc,
                             model.no2.median.ltrc,
                             model.no2.median.imp.ltrc,
                             model.no2.mean.ttrc,
                             model.no2.mean.imp.ttrc,
                             model.no2.median.ttrc,
                             model.no2.median.imp.ttrc))

# Save file

save.files(model.no2, dir, "no2_model")

rm(list = ls(pattern = "^model.no2"))

rm(list = ls())

# Verify rbind and count truncated -----
count.truncated <- function(model) {

  filtered.lur <- mutate(model,
                         Measure = as.factor(Measure),
                         Year = as.factor(Year),
                         Month = as.factor(Month),
                         Truncation = as.factor(Truncation)) %>%
    filter(Measure == "Mean",
           Imputed == T,
           Truncation == "LUR")

  filtered.non <- mutate(model,
                         Measure = as.factor(Measure),
                         Year = as.factor(Year),
                         Month = as.factor(Month),
                         Truncation = as.factor(Truncation)) %>%
    filter(Measure == "Mean",
           Imputed == T,
           Truncation == "None")

  filtered.join <- full_join(filtered.lur, filtered.non, by = c("Measure",
                                                             "Imputed",
                                                             "Year",
                                                             "Month",
                                                             "PC",
                                                             "Factor"))

  filtered.join <- mutate(filtered.join,
                         LUR.Val.Diff = Value.x - Value.y,
                         Truncated = ifelse(LUR.Val.Diff != 0, T, F))

  summary <- group_by(filtered.join, Year, Truncated) %>%
    summarise(N = n()) %>%
    spread(Truncated, N) %>%
    select(-`<NA>`) %>%
    mutate(Total = `TRUE` + `FALSE`,
           Percent = `TRUE` / Total * 100)
}

```

```

    return(summary)
  }

  calc.percent.trunc <- function(summary) {
    summary.total <- c("Total",
                      sum(summary$`FALSE`, na.rm = T),
                      sum(summary$`TRUE`, na.rm = T),
                      sum(summary$`TRUE`, na.rm = T) +
                      sum(summary$`FALSE`, na.rm = T),
                      sum(summary$`TRUE`, na.rm = T) /
                      (sum(summary$`TRUE`, na.rm = T) +
                      sum(summary$`FALSE`, na.rm = T)) * 100)

    names(summary.total) <- c("Year", "FALSE", "TRUE", "Total", "Percent")
    summary.total <- as.data.frame(t(summary.total))
    return(summary.total)
  }

  # PM25

  model.pm25 <- read_csv("Data/Cleaned/pm25_model.csv")

  group_by(model.pm25, Measure, Imputed, Truncation) %>%
    summarise(Max = max(Modelled, na.rm = T),
              Mean = mean(Modelled, na.rm = T),
              Median = median(Modelled, na.rm = T),
              Min = min(Modelled, na.rm = T),
              N = n())

  trunc.count.pm25 <- count.truncated(model.pm25)
  trunc.summary.pm25 <- calc.percent.trunc(trunc.count.pm25)

  rm(model.pm25)

  # NO

  model.no <- read_csv("Data/Cleaned/no_model.csv")

  group_by(model.no, Measure, Imputed, Truncation) %>%
    summarise(Max = max(Modelled, na.rm = T),
              Mean = mean(Modelled, na.rm = T),
              Median = median(Modelled, na.rm = T),
              Min = min(Modelled, na.rm = T),
              N = n())

  trunc.count.no <- count.truncated(model.no)
  trunc.summary.no <- calc.percent.trunc(trunc.count.no)

  rm(model.no)

  # NO2

  model.no2 <- read_csv("Data/Cleaned/no2_model.csv")

  group_by(model.no2, Measure, Imputed, Truncation) %>%
    summarise(Max = max(Modelled, na.rm = T),
              Mean = mean(Modelled, na.rm = T),
              Median = median(Modelled, na.rm = T),
              Min = min(Modelled, na.rm = T),
              N = n())

  trunc.count.no2 <- count.truncated(model.no2)
  trunc.summary.no2 <- calc.percent.trunc(trunc.count.no2)

  rm(model.no2)

  rm(list = ls())

```

## 07\_trim\_temporally\_adjusted\_models.R

```
#' ---
#' Trim and subset data to prepare for visualization and analyses. Drop
#' "Value" and "Factor" as unnecessary variables, and subset data to only
#' Years 2003-2009.
#' ---

library(tidyverse)

save.files <- function(df, dir, filename) {
  write_csv(df, paste0(dir, filename, ".csv"))
}

dir <- "Data/Cleaned/"

prep.data <- function(df.path) {
  min.year <- 2003
  max.year <- 2009
  df <- read_csv(df.path)
  df.trim <- select(df, -c(Value, Factor)) %>%
    filter(Year >= min.year & Year <= max.year) %>%
    select(Measure, Imputed, Truncation, Year, Month, PC, Modelled) %>%
    mutate(Measure = factor(Measure),
           Imputed = factor(Imputed),
           Truncation = factor(Truncation),
           Year = factor(Year),
           Month = factor(Month),
           PC = factor(PC)) %>%
    arrange(Measure, Imputed, Truncation, Year, PC, Month)
}

# Trim and save data: PM2.5

model.pm25 <- prep.data("Data/Cleaned/pm25_model.csv")
save.files(model.pm25, dir, "pm25_model_trimmed")
rm(model.pm25); gc()

# Trim and save data: NO

model.no <- prep.data("Data/Cleaned/no_model.csv")
save.files(model.no, dir, "no_model_trimmed")
rm(model.no); gc()

# Trim and save data: NO2

model.no2 <- prep.data("Data/Cleaned/no2_model.csv")
save.files(model.no2, dir, "no2_model_trimmed")
rm(model.no2); gc()

# Clear global environment

rm(list = ls())
```

## 08\_prep\_models\_for\_popdata.R

```
# Libraries -----
library(tidyverse)

# Export Models -----

file.export <- function(pollutant) {
  df <- read_csv(paste0("Data/Cleaned/", pollutant, "_model_trimmed.csv"))
```

```

# Make data wide and rename columns based on model variant

df.model <- filter(df, Imputed == T) %>%
  rename(YEAR = Year,
         MONTH = Month,
         MODELLED = Modelled) %>%
  mutate(MODEL = paste0(toupper(Measure), ".",
                       toupper(as.character(Imputed)), ".",
                       toupper(substr(Truncation, 1, 3)))) %>%
  select(YEAR, MONTH, PC, MODEL, MODELLED) %>%
  spread(MODEL, MODELLED) %>%
  select(YEAR, MONTH, PC, MEAN.TRUE.NON,
         MEAN.TRUE.LUR,
         MEAN.TRUE.TEM,
         MEDIAN.TRUE.NON,
         MEDIAN.TRUE.LUR,
         MEDIAN.TRUE.TEM) %>%
  arrange(YEAR, PC, MONTH)

write_csv(df.model, paste0("Data/Cleaned/",
                          pollutant,
                          "_model_trimmed_long.csv"))

print(summary(df.model))

# Create summary of results for comparison check

summary <- bind_rows(summary(df.model$MEAN.TRUE.NON),
                    summary(df.model$MEAN.TRUE.LUR),
                    summary(df.model$MEAN.TRUE.TEM),
                    summary(df.model$MEDIAN.TRUE.NON),
                    summary(df.model$MEDIAN.TRUE.LUR),
                    summary(df.model$MEDIAN.TRUE.TEM))
model.names <- data.frame(Model = c("MEAN.TRUE.NON",
                                   "MEAN.TRUE.LUR",
                                   "MEAN.TRUE.TEM",
                                   "MEDIAN.TRUE.NON",
                                   "MEDIAN.TRUE.LUR",
                                   "MEDIAN.TRUE.TEM"))

summary <- bind_cols(model.names, summary)
return(summary)
}

summary.a.pm25 <- file.export("pm25")
summary.a.no <- file.export("no")
summary.a.no2 <- file.export("no2")

# Check Transformation -----
check.transform <- function(pollutant) {

  df <- read_csv(paste0("Data/Cleaned/", pollutant, "_model_trimmed.csv"))

  df <- filter(df, Imputed == T)

  # Split models

  mean.non <- filter(df, Measure == "Mean", Truncation == "None")
  mean.lur <- filter(df, Measure == "Mean", Truncation == "LUR")
  mean.tem <- filter(df, Measure == "Mean", Truncation == "Temporal")
  median.non <- filter(df, Measure == "Median", Truncation == "None")
  median.lur <- filter(df, Measure == "Median", Truncation == "LUR")
  median.tem <- filter(df, Measure == "Median", Truncation == "Temporal")

  rm(df)

  # Create summaries

```

```

summary <- (bind_rows(summary(mean.non$Modelled),
                        summary(mean.lur$Modelled),
                        summary(mean.tem$Modelled),
                        summary(median.non$Modelled),
                        summary(median.lur$Modelled),
                        summary(median.tem$Modelled)))
model.names <- data.frame(Model = c("MEAN.TRUE.NON",
                                    "MEAN.TRUE.LUR",
                                    "MEAN.TRUE.TEM",
                                    "MEDIAN.TRUE.NON",
                                    "MEDIAN.TRUE.LUR",
                                    "MEDIAN.TRUE.TEM"))

summary <- bind_cols(model.names, summary)
return(summary)
}

summary.b.pm25 <- check.transform("pm25")
summary.b.no   <- check.transform("no")
summary.b.no2  <- check.transform("no2")

# Check Export -----
check.export <- function(pollutant) {

  df.model <- read_csv(paste0("Data/Cleaned/",
                              pollutant,
                              "_model_trimmed_long.csv"))

  summary <- bind_rows(summary(df.model$MEAN.TRUE.NON),
                      summary(df.model$MEAN.TRUE.LUR),
                      summary(df.model$MEAN.TRUE.TEM),
                      summary(df.model$MEDIAN.TRUE.NON),
                      summary(df.model$MEDIAN.TRUE.LUR),
                      summary(df.model$MEDIAN.TRUE.TEM))
  model.names <- data.frame(Model = c("MEAN.TRUE.NON",
                                      "MEAN.TRUE.LUR",
                                      "MEAN.TRUE.TEM",
                                      "MEDIAN.TRUE.NON",
                                      "MEDIAN.TRUE.LUR",
                                      "MEDIAN.TRUE.TEM"))

  summary <- bind_cols(model.names, summary)
  return(summary)
}

summary.c.pm25 <- check.export("pm25")
summary.c.no   <- check.export("no")
summary.c.no2  <- check.export("no2")

# Compare Summaries -----

summary.a.pm25

all(all.equal(summary.a.pm25, summary.b.pm25) &
    all.equal(summary.b.pm25, summary.c.pm25) &
    all.equal(summary.c.pm25, summary.a.pm25))

summary.a.no

all(all.equal(summary.a.no, summary.b.no) &
    all.equal(summary.b.no, summary.c.no) &
    all.equal(summary.c.no, summary.a.no))

summary.a.no2

all(all.equal(summary.a.no2, summary.b.no2) &
    all.equal(summary.b.no2, summary.c.no2) &
    all.equal(summary.c.no2, summary.a.no2))

```

```
rm(list = ls())
```

## 09\_export\_final\_models.R

```
# Load Libraries and Data -----  
  
library(tidyverse)  
library(lubridate)  
  
pm25 <- read_csv("Data/Cleaned/pm25_model_trimmed_long.csv")  
no <- read_csv("Data/Cleaned/no_model_trimmed_long.csv")  
no2 <- read_csv("Data/Cleaned/no2_model_trimmed_long.csv")  
  
# Drop Unused Temporally Interpolated Models -----  
  
# MEAN.TRUE.LUR  
#  
# Temporal factor: MEAN  
# Imputation: TRUE  
# Truncation: At LUR  
#  
# Temporal adjustment using factors based on a ratio between  
# the monthly mean of air pollutant concentrations across all air monitors  
# for the month of interest over the annual mean of air pollutant  
# concentrations across all air monitors for the LUR year. Missing air  
# monitor data has been imputed prior to developing temporal factors. And  
# LUR models have been truncated to prevent zeros prior to temporal  
# adjustment.  
  
pm25 <- select(pm25, YEAR, MONTH, PC, MEAN.TRUE.LUR) %>%  
  mutate(DATE = ymd(as.Date(paste0(YEAR,"-",MONTH,"-01"))),  
         POSTAL.CODE = PC,  
         PM25 = MEAN.TRUE.LUR) %>%  
  select(DATE, POSTAL.CODE, PM25)  
  
no <- select(no, YEAR, MONTH, PC, MEAN.TRUE.LUR) %>%  
  mutate(DATE = ymd(as.Date(paste0(YEAR,"-",MONTH,"-01"))),  
         POSTAL.CODE = PC,  
         NO = MEAN.TRUE.LUR) %>%  
  select(DATE, POSTAL.CODE, NO)  
  
no2 <- select(no2, YEAR, MONTH, PC, MEAN.TRUE.LUR) %>%  
  mutate(DATE = ymd(as.Date(paste0(YEAR,"-",MONTH,"-01"))),  
         POSTAL.CODE = PC,  
         NO2 = MEAN.TRUE.LUR) %>%  
  select(DATE, POSTAL.CODE, NO2)  
  
# Merge Pollutants -----  
  
model <- full_join(pm25, no, by = c("DATE", "POSTAL.CODE"))  
model <- full_join(model, no2, by = c("DATE", "POSTAL.CODE"))  
  
model <- arrange(model, POSTAL.CODE, DATE)  
  
summary(model)  
  
# Save File -----  
  
write_csv(model, "Results/temporally_adjusted_air_pollution_models.csv")  
  
rm(list = ls())  
  
# Delete Temporal Files -----  
  
file.remove(file.path("Data/Temp/", list.files("Data/Temp/")))
```

## **Appendix D.**

### **Temporally Adjusted LUR Model Estimates: Data File**

The accompanying CSV file contains the temporally adjusted land use regression model estimates for all postal codes in Metro Vancouver from 2003–2009.

#### **Record Layout:**

- DATE: Date formatted as YYYY-MM-DD
- POSTAL.CODE: 6-digit postal code
- PM25: Estimated concentrations of PM<sub>2.5</sub> in µg/m<sup>3</sup>
- NO: Estimated concentrations of NO in ppb
- NO2: Estimated concentrations of NO<sub>2</sub> in ppb

#### **Filename:**

Pagalan\_Lief\_temporally\_adjusted\_lur\_model\_estimates.csv