Unveiling Geographic Patterns in Critical Congenital Heart Defects: A Spatial Analysis of Selected air Pollutants

by

Ghazal Zargari

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Clinical Epidemiology

School of Public Health
University of Alberta

Abstract

Congenital heart defects (CHDs) are structural abnormalities of the heart present at birth, arising when the heart or nearby blood vessels don't develop properly during fetal growth. These conditions are a major public health concern, particularly for infants and young children, due to their complexity and the severity of their effects. Critical congenital heart defects (cCHDs) are a severe subset that require immediate medical intervention to prevent life-threatening complications. While genetic factors, such as chromosomal abnormalities, play a role in cCHDs, recent research has highlighted the potential impact of environmental factors, including air pollution, on their development. This has led to extensive studies investigating the relationship between air pollution and cCHDs.

In this study, we use geographically weighted multinomial logistic regression (GWMLR) to explore the relationship between exposure to four key air pollutants—particulate matter with a diameter of 2.5 micrometers or less (PM2.5), ozone (O3), nitrogen dioxide (NO2), and air quality smoke (AQSMK)—and various subtypes of cCHDs. By analyzing data from 1,484 infants diagnosed with cCHDs, we examine how air pollutant exposure and cCHD incidence vary across different geographic regions.

Our findings reveal significant patterns. PM2.5 exposure was associated with cCHDs in 0.32% of the locations studied, with these significant associations primarily clustered in Saskatchewan and Manitoba. When we compared the cCHD subtype distribution in these significant locations with the original dataset, we found notable discrepancies, highlighting the importance of accounting for spatial differences in air pollution exposure. These associations were found in small, localized areas, under-

scoring the need for targeted public health interventions.

For ozone (O3) exposure, about 15% of locations showed significant impacts on cCHD subtypes, with most cases occurring in Alberta and a smaller number in Saskatchewan. While there were similarities in the distribution of cCHD subtypes across these regions, significant differences emerged when spatial adjustments were applied, further emphasizing the need to consider geographic variability in environmental health studies.

Interestingly, no significant associations were found for NO2 and AQSMK exposure in any of the locations, suggesting that these pollutants may not have a direct impact on cCHD incidence in the studied population. However, when adjusting for these pollutants, we discovered significant associations between sex and cCHDs across all four pollutants, with AQSMK and NO2 showing a higher number of significant locations. Notably, the odds ratios in these cases were consistently below one, indicating a higher risk of cCHDs among male infants.

The differences between the original dataset and the GWMLR results underscore the importance of using advanced modeling techniques to uncover complex patterns that may be missed by conventional methods. However, interpreting these findings can be challenging, especially given that most odds ratios were below one. This raises questions about the underlying mechanisms, which could include the choice of reference group, genetic predispositions, or undiagnosed chromosomal abnormalities within certain cCHD subtypes.

Additionally, the influence of undiagnosed confounding factors and interactions between air pollutants and other environmental or genetic variables may complicate our understanding of the relationship between air pollution and cCHDs.

This research provides valuable insights into the geographic variability of air pollution exposure and its impact on cCHD incidence. The findings can guide targeted public health interventions and inspire further research into the environmental factors contributing to cCHDs. By gaining a better understanding of how environmental

factors influence CHD development, we can create more effective strategies to reduce the burden of cCHDs in affected communities.

In conclusion, our study highlights the need for comprehensive strategies to address air pollution and its role in CHD incidence. Targeted interventions that consider geographic patterns, along with continued research into the mechanisms behind these associations, are crucial for improving health outcomes for individuals with cCHDs and reducing the global burden of congenital heart defects. Through collaborative and interdisciplinary research, we can advance our understanding of cCHD causes and develop effective prevention and management strategies to protect the health and well-being of vulnerable populations.

Preface

This thesis presents research conducted as part of a collaborative effort with several researchers and institutions. The work draws on data from the Western Canadian Complex Pediatric Therapies Follow-up Program (CPTFP) registry, with key contributions from Dr. Charlene Robertson and Dr. Ari Joffe, who provided essential data and offered invaluable insights into the analysis.

Dr. Payam Amini provided significant expertise in the technical aspects of spatial modeling, particularly in the application of geographically weighted multinomial logistic regression (GWMLR). The air quality data, crucial for examining the relationship between air pollution and critical congenital heart defects (cCHDs), was generously provided by the Canadian Urban Environmental Health Research Consortium. Asim Thapa, now a graduate of the University of Alberta, played a key role in organizing and classifying the air quality data, enabling its effective use in the analysis.

My primary responsibilities in this research included developing the methodological framework, integrating the air quality data with patient information from the CPTFP registry and conducting the data analysis. This work allowed for an exploration of geographic variability in pollutant exposure and its relationship to cCHD incidence. In addition, I was responsible for preparing the manuscript and synthesizing the findings.

Throughout the research and writing process, Dr. Irina Dinu served as my supervisory author, providing critical guidance, contributing to the conceptual development, and offering revisions that helped shape this thesis.

This dissertation is dedicated to all the infants and children around the world affected by congenital heart defects (CHDs), their families, and the healthcare professionals tirelessly working to improve their lives. Your strength, resilience, and unwavering determination inspire us to push the boundaries of knowledge and strive for better outcomes in the fight against CHDs. May this research contribute, in some small way, to the advancement of medical science and the well-being of those impacted by CHDs.

Acknowledgements

Completing this thesis dissertation has been a journey marked by the support and contributions of numerous individuals and organizations, without whom this endeavor would not have been possible. I am deeply indebted to my academic supervisor, Dr. Irina Dinu, for their exceptional guidance, unwavering support, and invaluable mentorship throughout the duration of this research project. Their expertise, encouragement, and constructive feedback have been instrumental in shaping the direction and outcomes of this thesis dissertation.

I extend my sincere appreciation to the dedicated members of the research team, including Dr. Charlene Robertson, Dr. Ari Joffe, and Professor Payam Amini, for their invaluable contributions to various aspects of this study. Their collaboration, expertise, and commitment to excellence have significantly enriched the research process and outcomes. Their invaluable feedback, insights, and guidance throughout the development and execution of this research project have been instrumental in shaping its direction and success.

My gratitude also extends to the Western Canadian Complex Pediatric Therapies Follow-up Program (CPTFP) registry for providing the invaluable patient data utilized in this thesis dissertation. This dataset, derived from patients diagnosed with critical congenital heart disease (cCHD), has been pivotal in shaping the research findings presented herein.

Additionally, heartfelt thanks are extended to the Canadian Urban Environmental Health Research Consortium for their generous provision of data on ambient air quality. This data has been indispensable in exploring the relationship between air

pollution and critical congenital heart defects, enhancing the depth and breadth of the research findings presented herein.

I am profoundly grateful to my family and friends for their unwavering support, encouragement, and understanding throughout this journey. Their love, patience, and belief in me have been a constant source of strength and inspiration. Their unwavering support has been invaluable in sustaining me through the challenges and triumphs of this academic pursuit, and I am forever grateful for their presence in my life.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Congenital heart defects (CHD) are structural abnormalities or malformations in the heart that develop during fetal growth and are present at birth. These anomalies can significantly impact the heart's structure, function, and blood flow, resulting in a spectrum of symptoms ranging from mild to life-threatening. Notably, CHDs rank as the most prevalent type of congenital anomaly, affecting approximately 8-10 out of every 1000 live births.[1] Among these cases, up to one-third are classified as critical congenital heart defects (cCHDs), denoting severe cardiac conditions that necessitate prompt medical or surgical intervention shortly after birth, or may even lead to mortality within the first month of life.[1, 2]

While genetic factors have been extensively scrutinized in the context of CHDs, our comprehension of non-genetic risk factors remains somewhat limited. Although chromosomal abnormalities are recognized contributors to CHDs, the underlying etiology of approximately 70% of all birth defects remains unidentified. It's widely accepted that a combination of genetic and environmental factors plays a pivotal role in the onset of birth defects, with genetic influences accounting for roughly 20% and environmental factors contributing to around 10% of all cases. [3]

Reproductive-related factors and parental characteristics, including advanced parental age, lifestyle habits, maternal drug exposures, and maternal health conditions such as diabetes mellitus, maternal phenylketonuria syndrome, and obesity, have all been

associated with an elevated risk of CHDs in newborns.[4] Moreover, exposure to environmental toxins and pollutants is increasingly recognized as a significant contributor to the development of CHDs. Furthermore, there is growing recognition of the potential role of environmental toxins and pollutants in the development of CHDs. [5]

Exposure to air pollutants, in particular, has garnered attention due to its widespread presence and potential impact on fetal development. Studies have indicated associations between exposure to environmental pollutants and the risk of CHDs, highlighting the importance of understanding and mitigating these risks.

In our study, we focused on four key air pollutants: PM2.5 (particulate matter with a diameter of 2.5 micrometers or smaller), NO2 (nitrogen dioxide), O3 (ozone), and AQSMK (ambient Air Quality Smoke Model). These pollutants are pervasive in urban and industrialized areas, and their detrimental effects on human health are well-documented. However, their potential impact on fetal development and the etiology of congenital heart defects (CHDs) remains an active area of investigation.

PM2.5 (Particulate Matter 2.5): PM2.5 refers to particles suspended in the air that are 2.5 micrometers or smaller in diameter. These fine particles can penetrate deep into the lungs and even enter the bloodstream, posing significant health risks. PM2.5 is generated from various sources, including vehicle emissions, industrial activities, and wildfires. Exposure to PM2.5 can stimulate oxidative stress, leading to inflammation and cellular damage in the respiratory system and potentially affecting cardiovascular health. [6]

Ozone (O3), a reactive gas composed of three oxygen atoms, is a major component of smog. While ozone in the stratosphere protects us from harmful ultraviolet radiation, ground-level ozone can be harmful to human health. It is primarily formed through chemical reactions between nitrogen oxides (NOx) and volatile organic compounds (VOCs) in the presence of sunlight, with sources including vehicle emissions, industrial processes, and chemical solvents. Ozone is a potent oxidant, capable of reacting with and damaging lung tissues and exacerbating respiratory conditions. [7]

NO2 is a gas that forms from the combustion of fossil fuels, particularly from vehicles, power plants, and industrial processes. It is a primary component of vehicle exhaust and emissions from combustion sources. Exposure to NO2 can irritate the airways, aggravate respiratory conditions, and contribute to the formation of other air pollutants such as ozone and particulate matter. [8]

AQSMK is a modelling tool developed by the Environmental Health Services of the BC Centre for Disease Control, based on The Canadian Optimized Statistical Smoke Model (CanOSSEM). It is used to estimate the concentrations of particulate matter in the air resulting from smoke, particularly from sources like wildfires. AQSMK incorporates various factors specific to smoke generation and dispersion, including weather conditions, topography, and fire characteristics. By considering these factors, AQSMK provides valuable insights into the distribution and impact of smoke-related particulate matter on air quality and public health. [9]

1.1 Literature Review

While research has provided insights into the correlation between air pollutants and certain types of CHDs, such as septal defects, evidence regarding more critical CHD types remains less conclusive and subject to debate. Therefore, further investigation into the impact of environmental factors, including air pollutants, on the development of CHDs is imperative for advancing preventive strategies and safeguarding maternal and fetal health.

One notable study by Hall focused specifically on critical forms of CHDs in relation to exposure to PM2.5 and ozone variables. While no overall significant relationship was found during the first trimester of pregnancy, associations were observed when analyzing each week individually from weeks 3 to 8 of pregnancy. [10]

In another study by Buteau et al., first-trimester exposure to PM2.5 and NO2 was investigated in relation to CHDs, revealing a positive association between PM2.5 exposure and tetralogy of Fallot, as well as associations between both PM2.5 and NO2

with coarctation of the aorta. However, confidence intervals encompassed the null, likely due to the limited proportion of patients with critical CHDs in the study.[11]

Stingone et al. found a positive association between areas with the highest levels of PM2.5 exposure and hypoplastic left heart syndrome, while Tanner et al. indicated higher levels of exposure were linked to an increased risk of various critical CHDs including non-isolated truncus arteriosus, total anomalous pulmonary venous return, coarctation of the aorta, and interrupted aortic arch. [12, 13] Despite these findings, separate systematic reviews and meta-analyses conducted by Hall and Chen did not find significant evidence supporting an association between PM2.5 and congenital heart defects. [5, 14]

Regarding ozone exposure, one study found an increased risk of CHDs overall, including tetralogy of Fallot, especially during the second and third trimesters. [15] Another study reported associations between ozone exposure and conotruncal and both pulmonary and aortic valve and artery disorders related to ozone exposure in the second month of pregnancy. [16] However, a negative correlation between exposure to ozone and CHDs was reported in a study by Vinikoor-Imler. [17] Furthermore, a systematic review and meta-analysis led by Vrijheid, covering 10 studies, found no links between ozone exposure and any types of CHDs.[18]

An umbrella review by Michel evaluated 35 studies on maternal exposure to air pollutants during pregnancy and various CHD subtypes, showing a moderate level of confidence in the association between NO2 exposure and coarctation of the aorta, while other associations varied among the studies reviewed. [19] Other independent systematic reviews also supported the relationship between exposure to nitrogen dioxide (NO2) and the development of coarctation of the aorta, as well as associations with pulmonary valve stenosis and tetralogy of Fallot.[12, 15, 18, 20]

Black smoke was weakly associated with cardiac chamber malformations in a case-control study, albeit only when utilized as a continuous variable, with no clear evidence of a dose-response relationship. [21]

In our study, we aimed to explore the potential correlation between prenatal exposure to air pollution and the risk of critical congenital heart defects (cCHDs), while considering spatial variations. To accomplish this, we employed geographically weighted multinomial logistic regression (GWMLR), a sophisticated statistical method enabling the estimation of distinct models for diverse locations within the study area. This approach allows us to capture and analyze the nuanced relationships between a dependent variable with three or more categories, such as different types of CHDs, and a set of independent variables. Specifically, we calculated the odds ratios (ORs) with confidence intervals (CIs) and p-values to assess the association between exposure to different pollutants and the risk of CHDs across different geographical regions within our study area. By employing GWMLR, we were able to account for spatial heterogeneity, thereby offering a more comprehensive understanding of how air pollution may influence the incidence of cCHDs in varying environmental contexts.

Chapter 2

Methodology

2.1 Data Collection and Preparation

We gathered data from a registry of 1,484 patients diagnosed with critical congenital heart disease (cCHD) in the Stollery Children's Hospital, who underwent surgery with cardiopulmonary bypass at 6 weeks or less, shunts up to 6 months, and repair of isomerism defects up to 1 year, between September 1996 and November 2021.

This registry was an integral component of the Western Canadian Complex Pediatric Therapies Follow-up Program (CPTFP), initiated in September 1996. The Stollery Children's Hospital emerged as a central hub for the referral and care of critically ill infants with cCHD from across western Canada. The CPTFP employed a developmental follow-up process in which parents of infants undergoing interventions for cCHD were approached for registration and consent for developmental assessments, with a participation rate of over 99.5

The evaluation process extended to the children's home sites, with all developmental outcomes recorded and centralized at the Glenrose Rehabilitation Hospital in Edmonton. Consequently, the CPTFP has curated a comprehensive database encompassing acute care details and long-term outcomes for all children who underwent surgery or interventions for cCHD at the Stollery since its inception in 1996. [22]

To accurately classify cases of critical congenital heart defects (cCHD), it was necessary to separate those caused by chromosomal and genetic abnormalities from those

with other causes. To achieve this, we utilized the classification system developed by Botto et al., which excluded data from infants with chromosomal and genetic conditions, as well as rare heart defects, vascular anomalies, isolated valve dysplasias and arrhythmias. Botto's classification system provides a comprehensive risk assessment of cCHD based on cardiac phenotype, cardiac complexity, and baby phenotype, and was applied to data from the US National Birth Defect Prevention Study of 4703 infants born between 1997-2002. We employed the level 3 of Botto classification system, which consists of eight categories, to classify cases of congenital heart defects (CHDs). These categories include conotruncal, anomalous pulmonary venous return (APVR), atrioventricular septal defects (AVSD), left ventricular outflow tract obstruction (LVOTO), right ventricular outflow tract obstruction (RVOTO), septal, heterotaxy, and complex. However, for our specific analysis, we have decided to exclude AVSD and septal defects, as they are generally considered non-critical CHDs. [3] Since the Botto classification excluded cases with known chromosomal abnormalities, we added an additional category of reference (group 7) to our database to capture these cases separately and accurately track cases of cCHD caused by chromosomal and other known abnormalities while utilizing the categories of the Botto classification for the rest of the cases. This decision was made to ensure that our analysis focused specifically on the relative impact of air pollutants within the context of cCHDs. By selecting this subgroup as our reference, we ensured that our comparison groups were more homogenous in terms of underlying genetic predispositions, thereby reducing potential confounding factors and enhancing the interpretability of our findings.

To determine the population distribution of our cohort, we utilized postal codes. As the fetal heart begins to form around 3 weeks after fertilization and is fully formed and begins beating by the end of the eighth week of pregnancy [23], we linked the postal codes for the mother's residence during the first and second months of pregnancy separately to exposure data on ambient and industrial air quality. We obtained measurements of ozone (O3: 2002-2015)[24–26], ground-level fine particulate mat-

ter (PM2.5: 2000-2012)[27], nitrogen dioxide (NO2: 1984-2016)[28–30] and AQSMK (2010-2019)[31] from land use regression models developed by the Canadian Environmental Health Research Consortium (CANUE)[9]. AQSMK is a variable generated by the Canadian Optimized Statistical Smoke Exposure Model (CanOSSEM) to estimate national daily fine particulate matter (PM2.5) exposure specifically from wildfires using a machine learning approach.[31]

These pollutants were specifically selected due to their availability of monthly exposure levels, unlike other variables which were only available on an annual basis. The decision to focus on pollutants with monthly data was crucial, particularly considering the critical period of fetal heart development, which occurs between weeks 3 and 8 of pregnancy. During this window, exposure data corresponding to the first two months of pregnancy was deemed essential for capturing potential associations between prenatal pollutant exposure and the development of cCHDs.[24, 32]

All variables were checked for missing data and outliers, and appropriate steps were taken to address these issues before building the model. Missing data were addressed through imputation techniques, while outliers were examined to determine their impact on the analysis and adjusted or removed as necessary to ensure the robustness of our findings. Initially, we considered including the distance from the source of pollution in the model. However, given that the exposure data were relatively close to postal codes, with spatial resolutions ranging from individual postal code locations to approximately 1 square kilometer (less than 5 kilometers for AQSMK), this information was indirectly captured through the geographical identifiers. Therefore, distance from the source of pollution was not included as a separate variable in the analysis. Additionally, our original dataset contained demographic information such as sex, rural-urban status, and gestational week. Nonetheless, we made the decision to include only sex as a confounding factor in the model. This choice was based on several considerations. First, postal codes inherently correspond to specific geographical locations, which indirectly capture the rural-urban status of the study population. This

makes it unnecessary to include a separate variable for rural-urban differences. Additionally, we chose not to include gestational week as a variable because the formation of the fetal heart is generally completed by the eighth week of pregnancy. Given that our focus was on the critical early stages of heart development, gestational week was deemed less relevant for our analysis and was therefore omitted.

2.2 Multinomial Logistic Regression

In order to investigate the relationship between air pollution and various subtypes of cCHDs, Initially we employed multinomial logistic regression, a statistical technique adept at analyzing the connection between multiple independent variables and a dependent variable with three or more unordered categories. Our dependent variable was the classification of cCHD subtypes according to the Botto classification system, encompassing six categories, with an additional category designated for known chromosomal abnormalities serving as the reference group.

To account for potential confounding effects, we included sex as an independent variable in the model. We denoted the dependent variable as Y, representing cCHD subtypes based on the Botto classification system, and the independent variables as X_1: air pollution levels in the first and second months of pregnancy (combined) and X_2: sex (male vs. female).

The multinomial logistic regression model can be written as:

$$\log[p(Y = j|X)] = \beta_{0j} + \beta_{1j}X_1 + \beta_{2j}X_2 \tag{2.1}$$

for $j = 1, 2, ..., 7$, where $p(Y = j|X)$ is the probability of cCHD subtype j given the values of the independent variables X_1, X_2 and $\beta_{0j}, \beta_{1j}, \beta_{2j}$ are the coefficients associated with each independent variable for cCHD subtype j.

This model operates under the assumption that the log-odds of each cCHD subtype are linearly related to the independent variables, and that the error terms follow a

multinomial distribution. Notably, the linearity assumption implies a consistent effect of predictors on the log-odds across all levels of the dependent variable.

To estimate the coefficients, we employed maximum likelihood estimation, a method that seeks the parameter values maximizing the likelihood of observing the data given the model. In essence, maximum likelihood estimation endeavors to find the most plausible parameter values that align with the observed data within the framework of the specified model.

By using multinomial logistic regression, we aimed to explore the complex links between air pollution, sex, and different types of cCHDs. Our goal was to identify potential risk factors and provide insights that could help guide targeted interventions and preventive measures.

2.3 Geographically Weighted Multinomial Logistic Regression

While multinomial logistic regression has been instrumental in uncovering the relationship between air pollution and cCHD subtypes, its assumption of spatial homogeneity might not fully capture the intricate spatial variations present in the data. To overcome this limitation and gain deeper insights into the spatial dynamics of this relationship, we employed geographically weighted multinomial logistic regression (GWMLR).

GWMLR is a sophisticated spatial regression technique that offers a nuanced approach to modeling spatially varying relationships. Unlike traditional regression models, GWMLR acknowledges that the relationship between predictors and outcomes may differ across geographic locations. By estimating distinct regression coefficients for different locations within the study area, GWMLR allows us to capture and analyze spatial heterogeneity effectively.

In GWMLR, we systematically defined hypothetical bins around each data point (X_0), effectively creating a spatially structured dataset. Within these bins, we com-

puted weighted averages of the cCHD subtype values (Y), taking into account the spatial distribution of observations. This approach ensures that the estimation of regression coefficients is informed not only by the overall trend in the data but also by the local characteristics of each spatial unit.

To address spatial autocorrelation and appropriately assign weights to observations, we employed a Gaussian kernel function. This function assigns weights to each observation based on its spatial distance from the kernel's center, which serves as the focal point for the spatial weighting scheme. By giving greater weight to observations closer to the kernel center, GWMLR acknowledges the spatial dependency present in the data, where neighboring observations are more likely to exhibit similar characteristics than those farther apart.

The use of a Gaussian kernel function in GWMLR facilitates the estimation of local coefficients that vary across space. This allows us to uncover spatially varying relationships between air pollution and cCHD subtypes, providing insights into how these relationships manifest across different geographical contexts within the study area. By embracing the complexity of spatial heterogeneity, GWMLR offers a powerful framework for exploring and understanding spatial patterns in health outcomes and environmental exposures.

2.3.1 Gaussian Kernel Function and Distance Metric

The Gaussian kernel function is expressed as:

$$w_i = e^{-\left(\frac{d}{h}\right)^2} \tag{2.2}$$

Where:

- w_i represents the weight assigned to the i-th observation,
- d_i is the distance between the observation and the kernel center, and
- h was the bandwidth controlling the rate of weight decay.

To compute the distance d_i, we utilize the Euclidean distance metric, which measures the dissimilarity or separation between an observation and the kernel center:

$$d_i = \sqrt{\sum_{j=1}^{n}(x_{ij} - c_j)^2} \tag{2.3}$$

Here,

- x_i denotes the j-th feature (variable) for i-th's observation,
- c_j represents the j-th feature of the kernel center.
- n is the total number of features, and
- The kernel center serves as a reference point used in the computation of distances, with each feature of the kernel center influencing the distance calculation.

Additionally, to address biases near the boundary of the study area, an edge correction factor was applied to the Gaussian kernel function, adjusting the weights assigned to observations based on their proximity to the edge of the spatial domain. The goal was to account for biases introduced by observations near the boundary and ensure a more accurate estimation of spatial relationships.

The edge correction factor $q_h(y|W)$ is defined as:

$$q_h(y|W) = h^{-2} \int_W K\left(\frac{u-y}{h}\right) du, \qquad y \in W \tag{2.4}$$

Where,

- $q_h(y|W)$ represents the edge correction factor for an observation at location y within the spatial domain W_i. The factor $q_h(y|W)$ adjusts the weights assigned to observations, taking into account their proximity to the edge of the spatial domain.
- h is the bandwidth or smoothing parameter and

- $\int_W$ is the integral taken over the spatial domain W, representing the region of interest.

- $K\left(\frac{u-y}{h}\right)$ is the Gaussian kernel function, where u represents the spatial coordinates, and y is a scaled distance. The scaling by h, adjusts the distance measure so that it aligns with the bandwidth of the kernel, influencing the width of the kernel and, consequently, the weight assigned to the observation.

By incorporating the Gaussian kernel function and edge correction factor, GWMLR enables the exploration of spatially varying relationships between air pollution and cCHD subtypes, providing insights into how these relationships manifest across different geographical contexts within the study area.

2.3.2 Bandwidth Selection

In geographically weighted multinomial logistic regression (GWMLR), choosing the right bandwidth is essential because it defines the spatial range over which observations are grouped to estimate local coefficients. This step is key to accurately capturing how relationships between predictors and outcomes vary across different areas.

To determine the appropriate bandwidth, we used the rule of thumb selector, a commonly applied method where the bandwidth is set to 1/3 of the kernel width. This approach helps balance capturing localized spatial variations with maintaining computational efficiency.

The 1/3 rule is favored for its simplicity and practicality, especially when detailed prior knowledge about the dataset or specific spatial relationships is limited. It provides a straightforward starting point for bandwidth selection, allowing for quick decisions without extensive tuning.

For our study, we used the rule of thumb to set the bandwidth to approximately 1/3 of the kernel width (6σ), which covers 99.7% of the data under a normal distribution

curve. We calculated a fixed bandwidth of about 1208 km at the mean latitude (52.35686 degrees), based on a longitude standard deviation of 7.88357 degrees.

By selecting the bandwidth based on this rule, we ensured that the GWMLR model accurately reflects the spatial variability in the relationship between air pollution and cCHD subtypes. This careful bandwidth selection enhances the robustness of our spatial analysis, helping us uncover localized patterns and trends that might be missed by traditional regression models.

2.3.3 Model Fitting

During the model fitting stage, addressing collinearity among independent variables was crucial for ensuring the robustness of our analysis. To manage collinearity, we used a stratified approach by running separate models for each pollutant, stratified by sex. This allowed us to evaluate the unique impact of each pollutant on different cCHD subtypes and to explore potential interactions between pollutants and sex.

Noting that pollutant levels were similar in the first and second months of pregnancy, and seeing consistent results across individual models, we combined the data for a more powerful analysis. This integration enhanced the statistical strength of our study and reinforced the reliability of our findings on the relationship between air pollution and cCHD subtypes.

The geographically weighted multinomial logistic regression model was then formulated, expressing the relationship between air pollution and cCHD subtypes based on the Botto classification system. This model considers six categories of cCHD subtypes, with an additional category for known chromosomal abnormalities serving as the reference point.

Let Y be the dependent variable representing cCHD subtypes, and $X_1, X_2, ..., X_k$ be the independent variables, including air pollution data and sex, the geographically

weighted multinomial logistic regression model can be expressed as:

$$\log[\frac{P(Y=i)}{P(Y=K)}] = \beta_{i0}(s) + \beta_{i1}(s)X_1 + \beta_{i2}(s)X_2 + ... + \beta_{ik}(s)X_k \tag{2.5}$$

where:

- i represents the category of cCHD subtypes,
- K is the reference category for known chromosomal abnormalities,
- s denotes the location of the observation, and
- $\beta_{i0}(s), \beta_{i1}(s), \beta_{i2}(s), ..., \beta_{ik}(s)$ are the location-specific coefficients for the intercept and independent variables, respectively.

In this formulation, the logit of the probability of being in category i versus the reference category K is modeled as a linear combination of the independent variables, with coefficients varying spatially. The kernel function assigns weights to each observation based on its distance from the center of the kernel, determining the influence of spatial location on these coefficients.

We computed local standard errors for each coefficient, allowing us to calculate p-values. These p-values help determine the statistical significance of the coefficients at different locations. Additionally, we calculated odds ratios for each coefficient to interpret the impact of the independent variables on the likelihood of different cCHD subtypes. An odds ratio greater than 1 indicates an increased probability of a specific cCHD subtype relative to the reference category as the predictor variable increases, while an odds ratio less than 1 indicates a decreased probability.

A positive coefficient for an independent variable X_k implies that as X_k increases, the probability of observing a particular cCHD subtype i, relative to the reference category K, also increases at that location, assuming other variables remain constant. Conversely, a negative coefficient suggests a higher likelihood of being in the reference category K as X_k increases.

The GWMLR approach allows for location-specific coefficients by computing weighted regressions at each geographic point, resulting in detailed insights into spatial variations. By spatially disaggregating the data, each longitude and latitude coordinate is associated with multiple cCHD subtype categories, allowing us to capture the distribution of cases accurately.

Through this method, we uncover nuanced spatial patterns in the relationship between air pollution and cCHD subtypes, offering valuable insights for targeted public health interventions and policy decisions.

2.4 Model Assessment

To evaluate our conventional multinomial logistic regression model, we carefully examined the residual distribution and performed a likelihood ratio test. We employed McFadden's R-squared as a global measure of goodness of fit, which is calculated by comparing the log-likelihood of the fitted model to that of a null model containing only an intercept. Despite these efforts, the conventional model exhibited consistently low McFadden's R-squared values, indicating a limited ability to explain the variance across all four pollutants. This limitation underscored the need for a more sophisticated approach capable of capturing the spatial differences in the relationship between predictors and cCHD subtypes.

To address this, we employed the Geographically Weighted Multinomial Logistic Regression (GWMLR) model, which allows for spatially varying coefficients. While the conventional model's fit could be assessed using McFadden's R-squared, the spatial nature of GWMLR required us to focus on local pseudo-R-squared values. These local values were calculated for specific geographic regions, offering insights into how well the model fit the data in those areas.

The visualization of local pseudo-R-squared values on maps provided a clear picture of where the GWMLR model performed well, particularly in regions with distinct environmental or demographic characteristics. This spatial assessment revealed clus-

ters of high model fit, indicating areas where the relationship between air pollution and cCHD subtypes was most accurately captured by the model.

By shifting our focus from a global to a local assessment of model fit, we uncovered nuanced spatial patterns in the data, offering valuable insights that can inform targeted interventions and policy decisions aimed at reducing the impact of air pollution on congenital heart defects.

2.5 Visualization

To comprehensively explore the spatial patterns and relationships between variables, we utilized GIS software to craft informative 2-dimensional plots depicting odds ratios and p-values. These visualizations allowed us to gain insights into the geographical distribution of associations between air pollution and the incidence of cCHD subtypes.

Our approach involved overlaying these maps with the locations of the study population. Given that exposure data were closely linked to postal codes, we leveraged this spatial information rather than referencing monitoring stations. This enabled us to delve into the spatial distribution of pollution exposure and its correlation with cCHD incidence, offering a detailed perspective on potential environmental health risks across geographic regions.

Visualizing the corresponding p-values allowed us to assess the statistical significance of observed associations, highlighting regions where the relationships between air pollution exposure and cCHD incidence were particularly noteworthy. The plotted odds ratios provided a visual representation of the strength and direction of association between air pollution and each cCHD subtype.

By integrating spatial analysis with visual representation, our visualizations not only facilitated the identification of spatial patterns and trends but also provided a means to communicate findings effectively to stakeholders and policymakers. Through intuitive visualizations, we aimed to empower decision-makers with actionable insights to address environmental health challenges and safeguard public well-being.

Chapter 3

Results

3.1 Descriptive Results

Table 3.1 summarizes the characteristics of 1484 patients diagnosed with cCHD who underwent surgical procedures involving cardiopulmonary bypass within 6 weeks, shunts up to 6 months, and repair of isomerism defects up to 1 year between September 1996 and November 2021. Among these patients, 183 cases (12.33%) had known chromosomal abnormalities, while the remaining 1301 patients were classified into groups 1-6 of the Botto classification. These patients either were identified as not having chromosomal etiology or were not yet diagnosed to have a chromosomal basis with current available methods.

The dataset consisted of 60% male and 40% female patients, with the majority (85%) having a gestational age of 37 weeks or more. 68% of the patients were from urban areas, while 33% were from rural areas. Alberta was the most common province of origin, accounting for 57% of the total patient population, followed by Manitoba (18%) and Saskatchewan (18%). British Columbia contributed to a smaller percentage of cases (3.5%), while patients from Yukon, Northwest Territories, Nunavut, Ontario, New Brunswick, and Nova Scotia collectively accounted for 3.3% of the total. The mean and standard deviation of several air pollutants, including NO2, O3, PM2.5 and AQSMK, for each Botto group and for cases with known chromosomal abnormalities is displayed in Table 3.1. The table shows that the mean values for each pollutant are

relatively consistent across the different Botto groups and for cases with chromosomal abnormalities. However, there are slight variations in the standard deviation of each pollutant between the groups.

Table 3.1: Characteristics of Patients with Critical Congenital Heart Disease (cCHD) and Air Pollutant Levels in Western Canada, 1996-2021

Botto group	1	2	3	4	5	6	Chromosomal
Male %	60.37	65.85	62.64	52.5	56.25	59.81	52.46
Female %	39.62	34.15	37.36	47.5	43.75	40.19	47.54
Alberta	63.56	48.78	58.05	56.88	38.75	53.27	61.2
Manitoba	18.09	24.39	13.51	20.62	25	20.56	15.3
Saskatchewan	13.83	18.7	18.39	16.88	26.25	19.63	18.58
British Columbia	1.33	2.44	7.76	1.25	3.75	3.74	2.19
Other	3.19	5.69	2.3	4.38	6.25	2.8	2.73
PM2.5 ($\mu g/m^3$)	6.83 ± 1.70	6.38 ± 1.81	7.06 ± 1.84	6.69 ± 2.13	6.08 ± 1.63	6.67 ± 1.50	6.67 ± 1.68
O3 (ppb)	19.73±8.18	19.99 ± 7.90	20.87 ± 8.06	22.09 ± 7.35	20.78 ± 7.46	20.25 ± 7.82	22.66 ± 6.81
NO2 (ppb)	13.81±7.36	13.30 ± 8.20	13.76 ± 7.07	12.59 ± 6.63	13.18 ± 7.50	13.18 ± 7.50	12.53 ± 6.14
AQSMK ($\mu g/m^3$)	7.69 ± 1.81	8.02 ± 2.64	7.86 ± 2.41	7.64 ± 2.04	7.94 ± 1.72	7.53 ± 2.30	7.85 ± 2.65

3.2 Analytic results

3.2.1 Ground-level fine particulate matter (PM2.5)

The conventional multinomial logistic regression model analysis on ground-level particulate matter (PM2.5) initially highlighted a significant correlation with the incidence of non-chromosomal based subtypes of cCHD (LR chi2(12) = 23.17, P-value = 0.026). However, the resulting pseudo-R squared value of 0.0033 underscored the model's limited interpretability, as PM2.5's p-values across some categories failed to reach significance.

The employment of the Geographically Weighted Multinomial Logistic Regression (GWMLR) model led to the discovery of intriguing insights, particularly in specific geographic locations. The analysis included data on PM2.5 exposure levels and spatial disaggregation, with each longitude and latitude coordinate associated with six different points, corresponding to specific congenital heart disease (cCHD) subtype categories. A total of 5665 locations were available for analysis, with 16 locations (0.32%) exhibiting noteworthy p-values following adjustments for sex. These findings revealed odds ratios below 1 for all significant locations.

Remarkably, nine of these significant instances were associated with category 1 cCHD subtype, primarily clustered in Saskatchewan, while the remaining six cases pertained to category 4, primarily observed in Manitoba (Figure 3.1). It's noteworthy that despite Alberta (6.23 ± 3.17) and British Columbia (5.92 ± 2.23) reporting higher PM2.5 emission levels, Saskatchewan and Manitoba recorded even lower emissions.

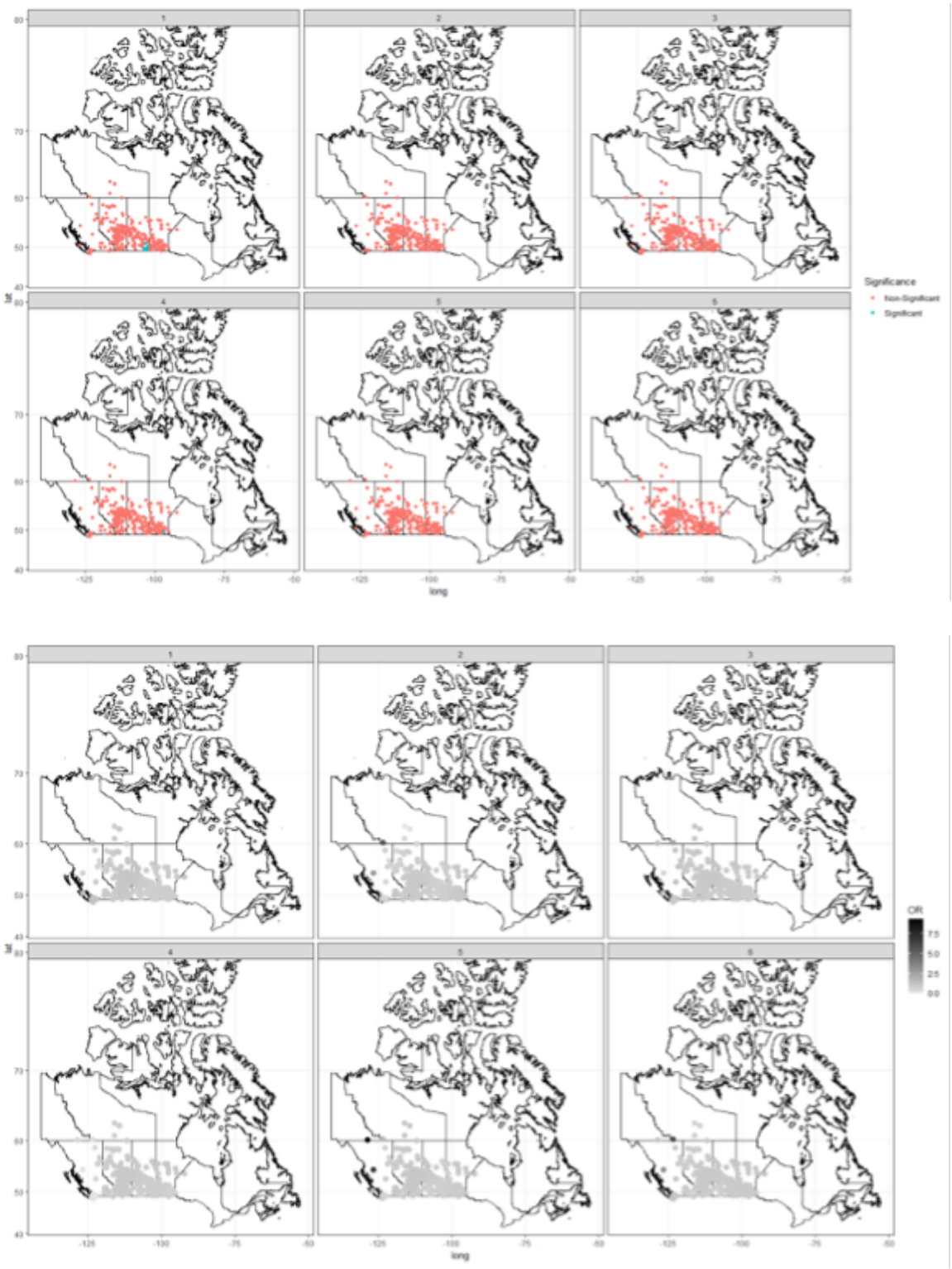

Figure 3.1: Exploring the Impact of PM2.5 Exposure on the Incidence of cCHD Subtypes: Accounting for Spatial Variability and Adjusting for Sex. Up: significance (Blue: significant, Red: non-significant), Down: Odds ratios.

Comparing the cCHD subtype distribution in significant locations post-spatial adjustment with the original dataset revealed intriguing disparities. Despite varied distributions in the original dataset (Table 3.2), all significant Saskatchewan locations post-adjustment were linked to category 1 cCHD subtypes. Similarly, in Manitoba, all significant locations post-adjustment were associated with category 4 cCHD subtypes. Notably, the regions identified in Saskatchewan and Manitoba exhibited relatively compact spatial extents, with Saskatchewan covering approximately 166 kilometers (longitude) by 184 kilometers (latitude) and Manitoba spanning about 22.2 kilometers (longitude) by 23.8 kilometers (latitude). These confined spatial scopes underscored the localized nature of the observed associations, emphasizing the necessity of scrutinizing environmental risk factors at fine spatial scales.

Furthermore, our investigation delved into potential etiologies or demographic patterns that might elucidate these observations. In Saskatchewan, areas with significant associations between PM2.5 exposure and cCHDs showcased diverse environmental stressors such as oil and gas exploration, coal mining, and agricultural production. Similarly, in Manitoba, industrial activities including small industry, manufacturing, and specific factories likely contributed to localized variations in air pollution levels, consequently impacting cCHD prevalence.

Table 3.2: Distribution of cCHD Subtypes in Significant Locations for PM2.5 After Spatial Adjustment based on original dataset.

	1 Conotruncal	2 APVR	3 LVOTO	4 RVOTO	5 Heterotaxy	6 Complex	7 Chromosomal
Sasketchewan	4	1	1	2	1	1	1
Manitoba	1	-	-	2	1	2	-

Further adjustment for PM2.5 led to 45 locations (0.8%) displaying significant p-values for the sex variable, with an overwhelming majority (95%) of these instances yielding ORs below 1 (Figure 3.2). These findings consistently point towards a lower prevalence of cCHD among females, suggesting a potential interplay between PM2.5

exposure and sex in shaping cCHD occurrences.

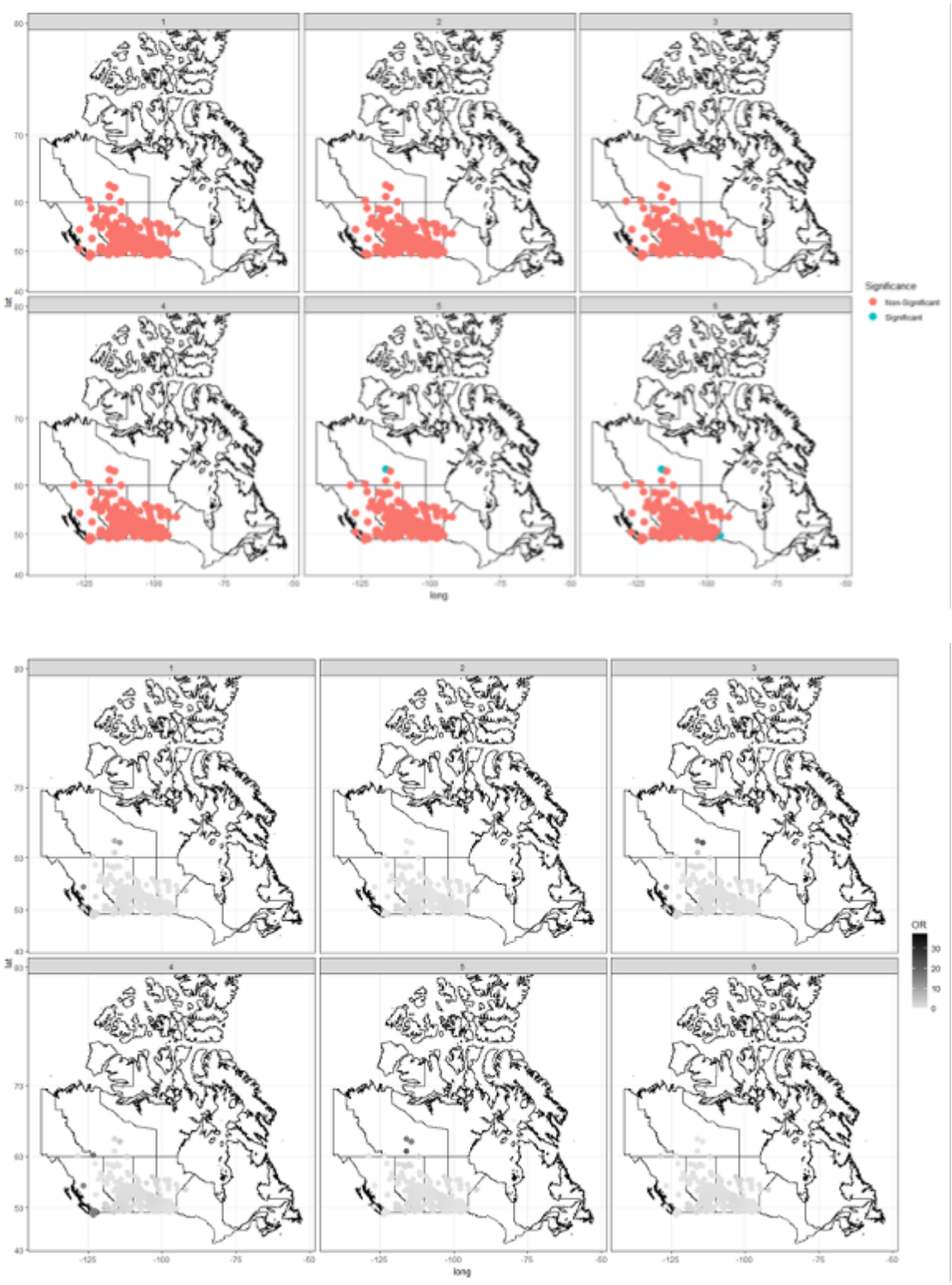

Figure 3.2: Exploring the Impact of Sex on the Incidence of cCHD Subtypes: Accounting for Spatial Variability and Adjusting for PM 2.5. Up: significance (Blue: significant, Red: non-significant), Down: Odds ratios.

3.2.2 Ozone (O3)

The analysis on ozone (O3) using conventional MLR underscores its significance, supported by a LR chi-squared statistic of 44.90 and a p-value of < 0.001, affirming the overall model's statistical robustness. However, the modest Pseudo R-squared value of 0.0082 implies that the model captures only a fraction of the dependent variable's variability.

Transitioning to the GWMLR analysis, after spatially adjusting each point for six categories, data for 4290 locations were available for examination. Remarkably, approximately 14.2% of these locations (609 locations) exhibited noteworthy p-values, highlighting the spatial variability in the relationship between O3 and the incidence of cCHD subtypes.

Intriguingly, among these locations, 362 cases were linked to category 6, with a remarkable concentration of 98% observed in Alberta. Additionally, 247 cases were associated with category 1, primarily distributed across Alberta (81%) and Saskatchewan (18%). Notably, all locations with significant p-values for ozone demonstrated odds ratios below 1, except for one location, indicating a consistent trend. Visualization maps have highlighted most significant locations in an area spanning between east Alberta and west Saskatchewan (Figure 3.3).

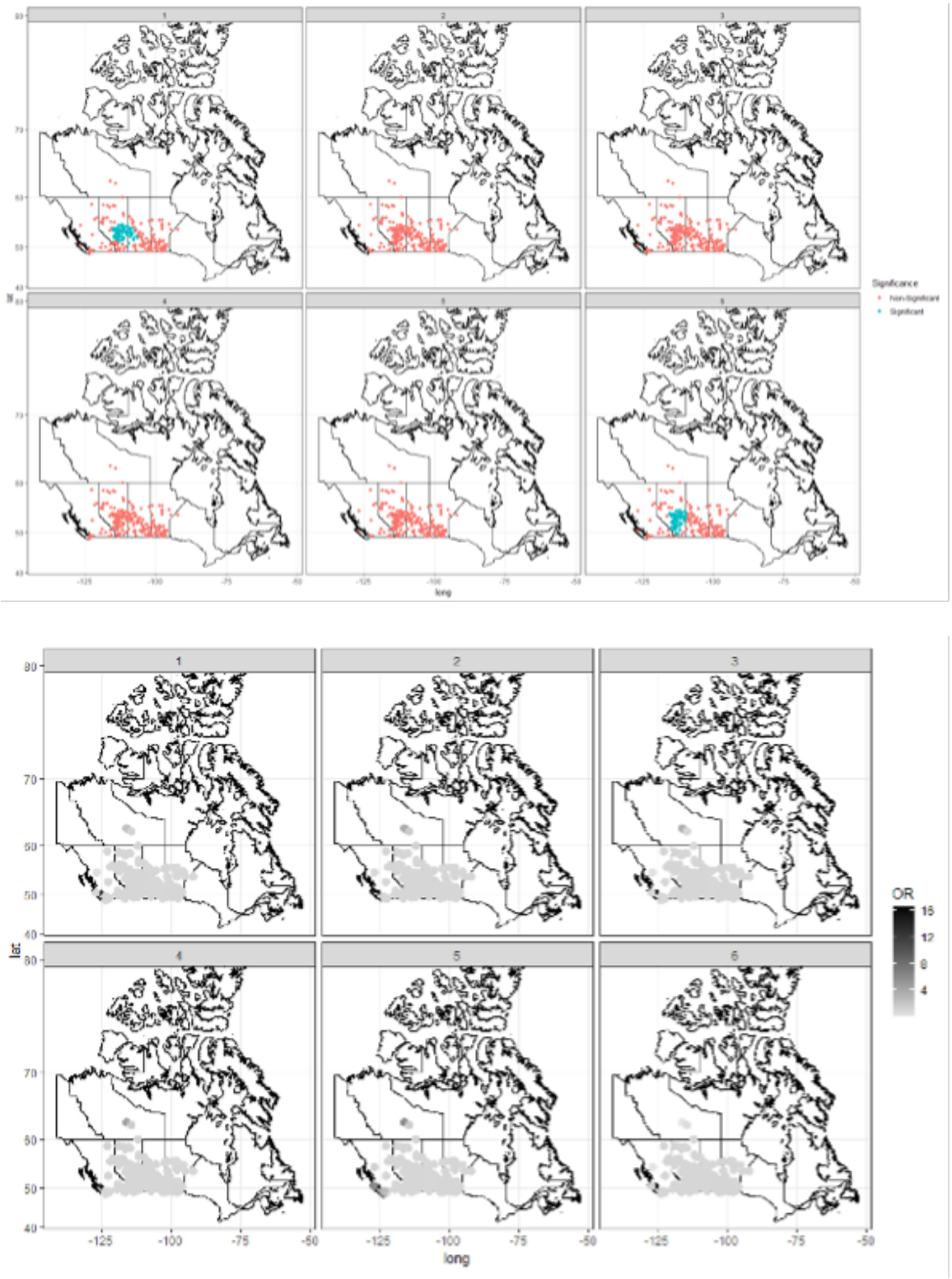

Figure 3.3: Exploring the Impact of Ozone Exposure on the Incidence of cCHD Subtypes: Accounting for Spatial Variability and Adjusting for Sex. Up: significance (Blue: significant, Red: non-significant), Down: Odds ratios.

In Alberta, the original dataset revealed a diverse array of cCHD subtypes. However, after spatial adjustment, the majority of cases (65%) were associated with category 6, while the remainder (35%) were linked to category 1. In Saskatchewan, a similar trend was observed in the distribution of cases across categories. However, discrepancies arose regarding the significant associations with locations, with the majority (88%) of cases now associated with Category 1 after adjustments, and the remaining 12% linked to Category 6 (Table 3.3).

Interestingly, the spatial distribution of significant ozone-related cCHD cases appeared more distinct compared to PM2.5, with several locations aligning with areas characterized by unique cultural and geographical features. These observations subtly hint at the presence of underlying demographic nuances shaping the observed associations, further complicating the interpretation due to the interplay of distinct social and environmental dynamics within these regions.

Table 3.3: Distribution of cCHD Subtypes in Significant Locations for Ozone After Spatial Adjustment based on original dataset.

	1 Conotruncal	2 APVR	3 LVOTO	4 RVOTO	5 Heterotaxy	6 Complex	7 Chromosomal
Alberta	151	43	133	58	25	92	68
Sasketchewan	15	6	12	7	3	10	9

After incorporating adjustments for ozone levels, a further subset of 23 locations (0.53%) exhibited a significant p-value for the sex variable. Remarkably, around 70% of these locations demonstrated ORs below 1 for sex, indicating a notable association between ozone exposure, gender, and the occurrence of cCHD. (Figure 3.4)

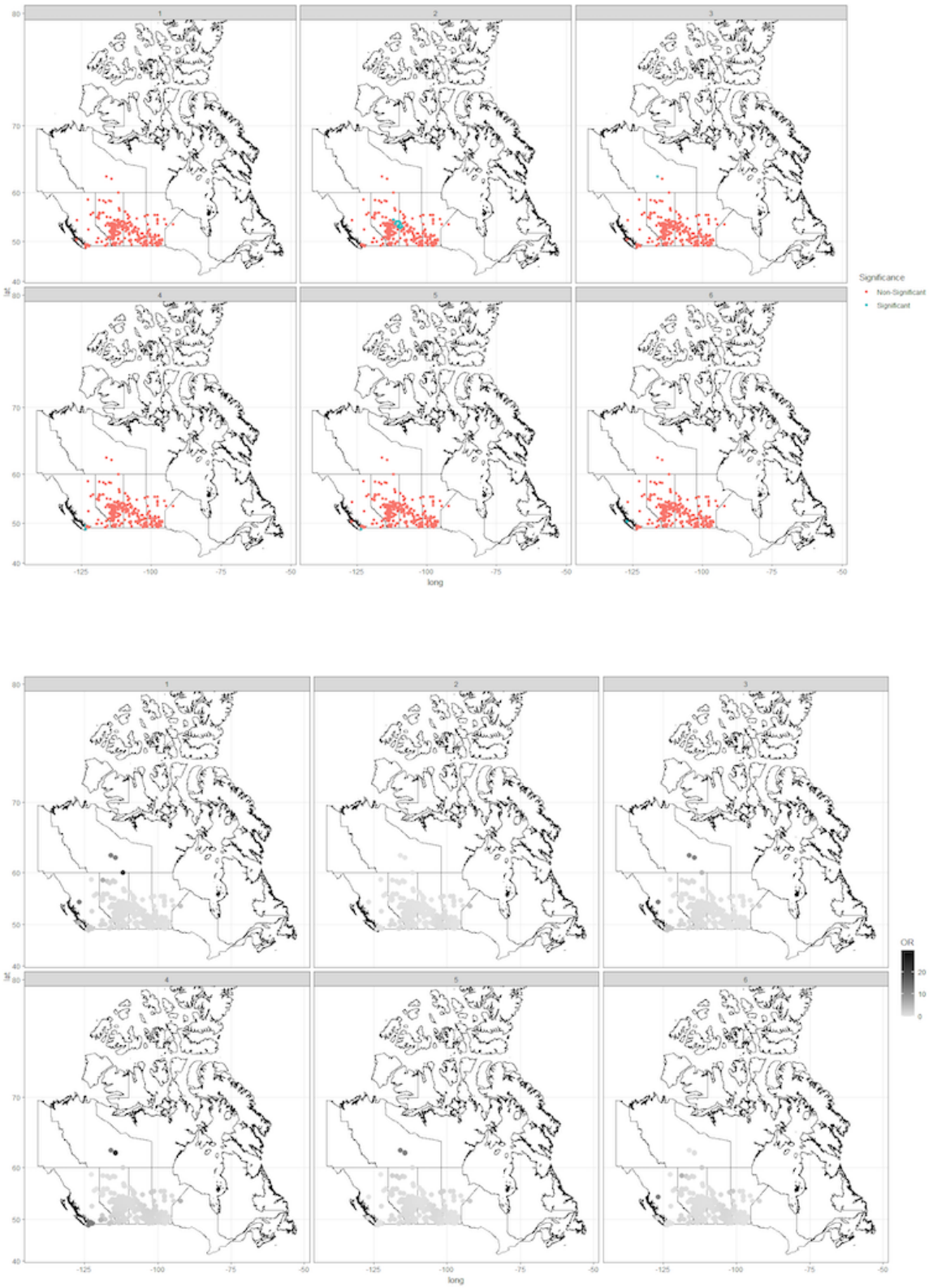

Figure 3.4: Exploring the Impact of Sex on the Incidence of cCHD Subtypes: Accounting for Spatial Variability and Adjusting for Ozone. Up: significance (Blue: significant, Red: non-significant), Down: Odds ratios.

3.2.3 Nitrogen dioxide (NO2)

For nitrogen dioxide (NO2), traditional MLR revealed a statistically significant association, highlighted by LR chi2(12) = 57.43 and p-value < 0.001. Although these results indicate a relationship between NO2 and cCHD subtypes, the model's explanatory power was modest, with a Pseudo R2 of 0.0076.

Following spatial adjustment of each point for six categories, we had data for 4856 locations. Surprisingly, the application of GWMLR yielded no significant associations with NO2 in any locations. This discrepancy underscores the complexity of NO2's spatial impact on cCHD subtypes and suggests that while a detectable overall association exists, it may lack substantial strength or be influenced by nuanced spatial dynamics not captured by traditional modeling approaches. (Figure **??**)

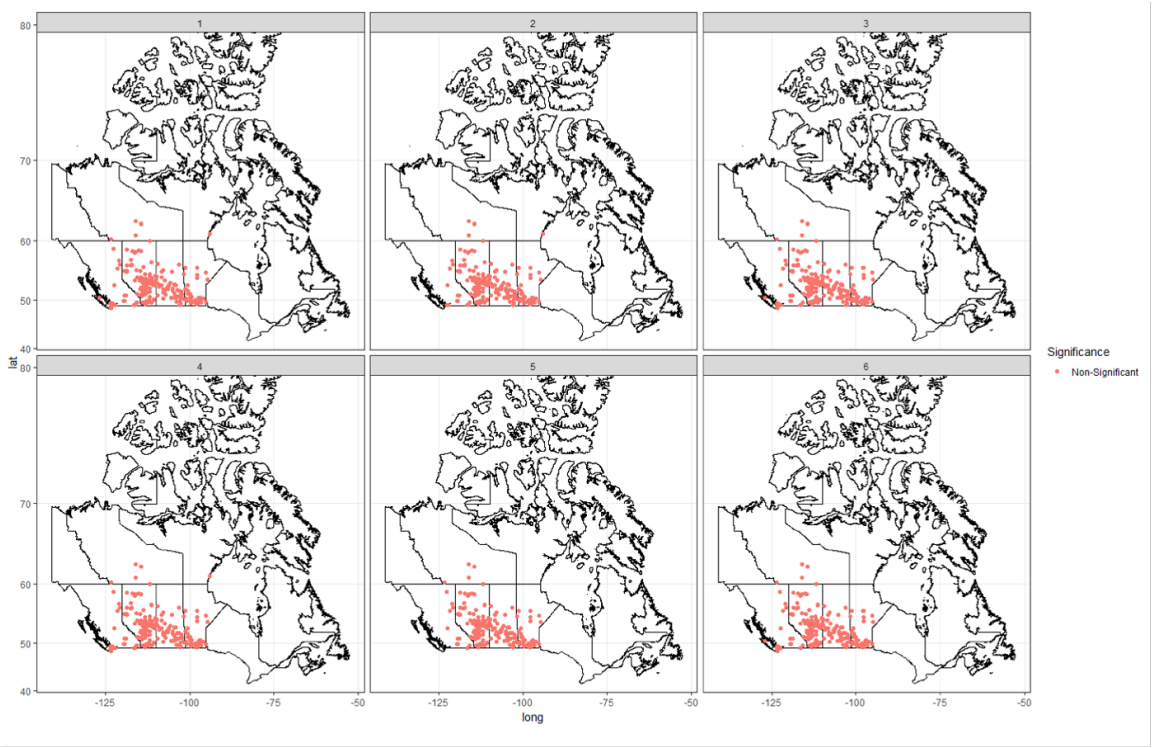

Figure 3.5: Exploring the Impact of NO2 Exposure on the Incidence of cCHD Subtypes: Accounting for Spatial Variability and Adjusting for Sex (Red: non-significant).

After adjustment for NO2, a significant p-value for the sex variable was observed in a subset of 272 locations (5.6%). Impressively, nearly 99% of these locations had odds ratios below 1, suggesting a considerable association wherein males were more affected than females by NO2 exposure. This highlights a distinct gender-specific impact of NO2 on the occurrence of cCHD, emphasizing the importance of considering gender disparities in environmental health research. (Figure 3.6)

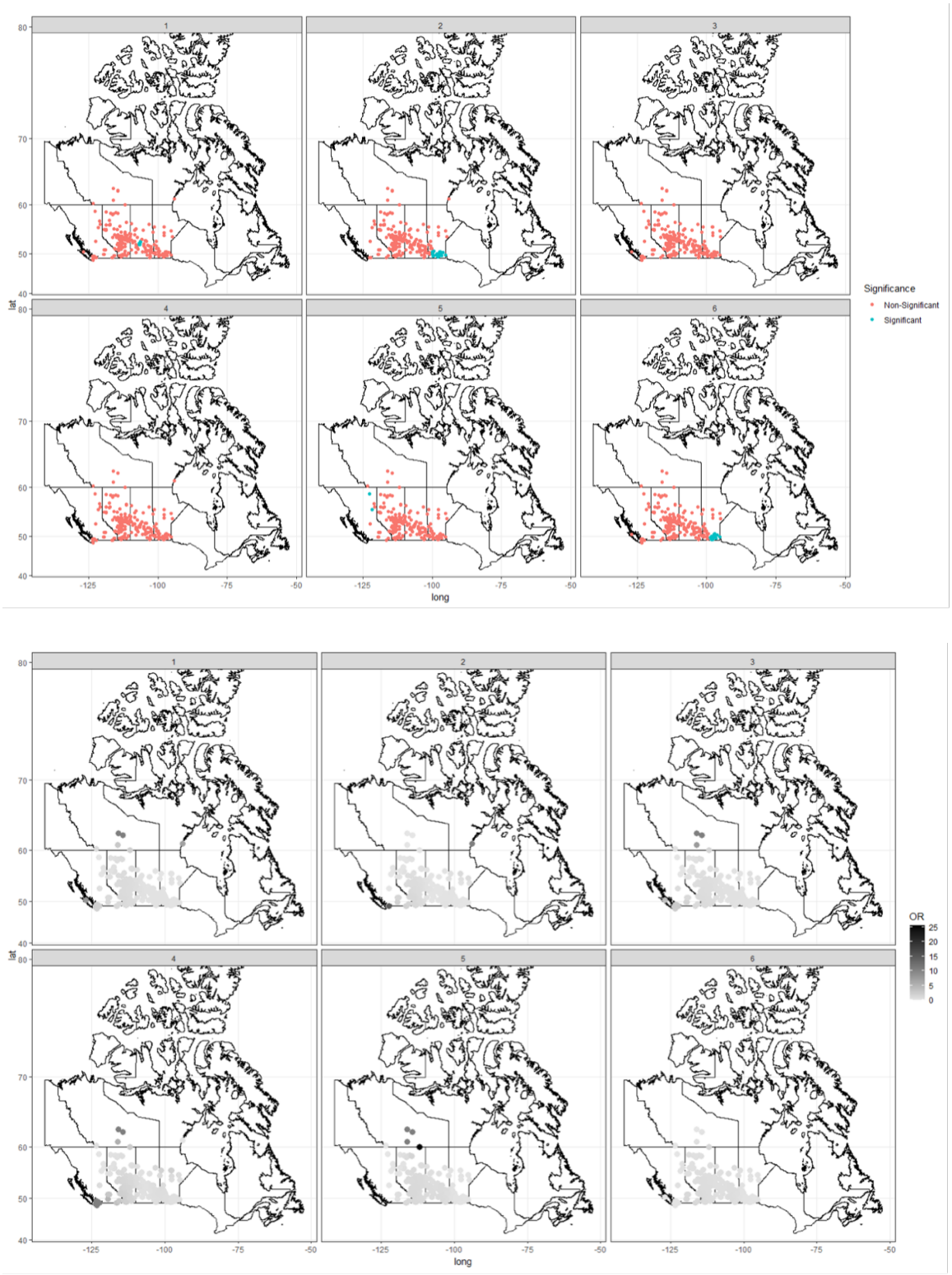

Figure 3.6: Exploring the Impact of Sex on the Incidence of cCHD Subtypes: Accounting for Spatial Variability and Adjusting for NO2. Up: significance (Blue: significant, Red: non-significant), Down: Odds ratios.

3.2.4 Air quality smoke model (AQSMK)

As previously mentioned, AQSMK is a variable generated by the Canadian Optimized Statistical Smoke Exposure Model (CanOSSEM) to estimate national daily fine particulate matter (PM2.5) exposure specifically from wildfires using a machine learning approach. [31] In relation to AQSMK, the analysis using the MLR model revealed no noteworthy connections with specific subtypes of cCHD, as indicated by LR chi2(12)=19.58 and a P-value of 0.0754.

Data were available for 4284 locations after spatial adjustment of each point for six categories. However, similar to the results obtained from the traditional MLR model, the GWMLR model did not identify any significant associations across various locations. These findings suggest that AQSMK exposure may not have a discernible impact on the incidence of specific cCHD subtypes, at least in the context of the variables considered in this analysis. (Figure 3.7)

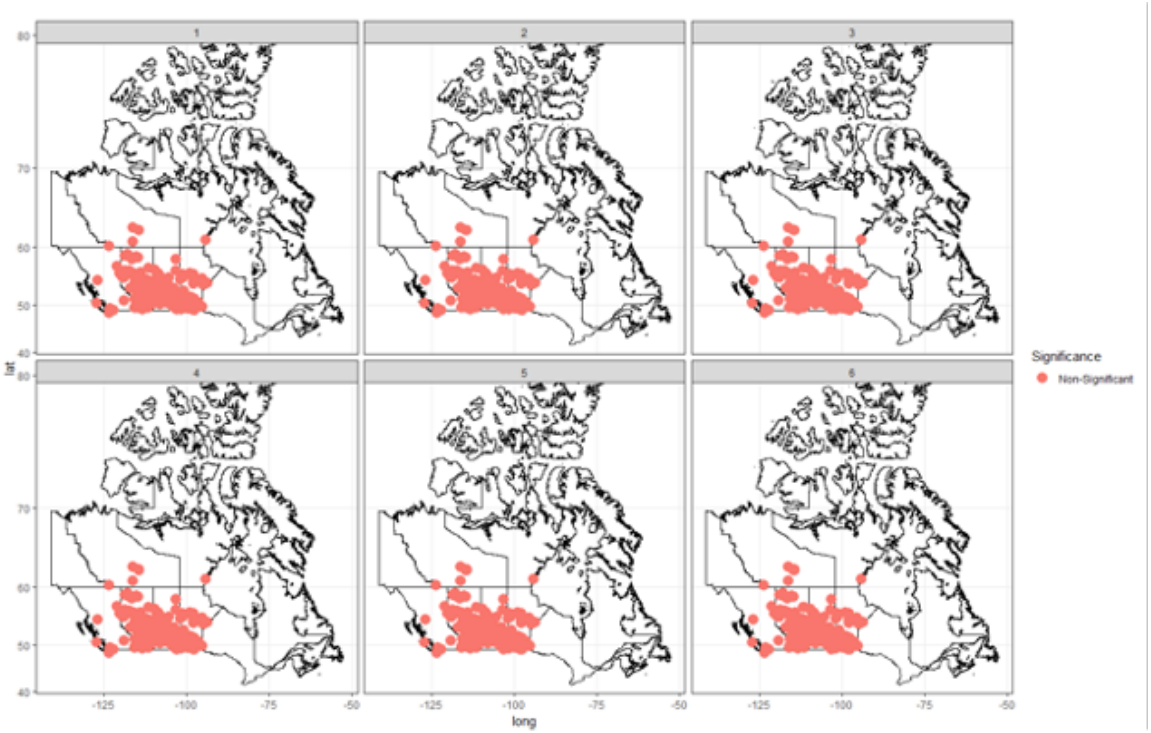

Figure 3.7: Exploring the Impact of AQSMK Exposure on the Incidence of cCHD Subtypes: Accounting for Spatial Variability and Adjusting for Sex (Red: non-significant).

However, when examining the sex variable after adjustment for AQSMK, there were 299 locations (7% of the dataset) where significant p-values were observed. Intriguingly, in all of these locations, the odds ratios were consistently below 1. This suggests a consistent association between AQSMK exposure and gender in relation to the occurrence of cCHD, with males potentially being more affected than females. (Figure 3.8)

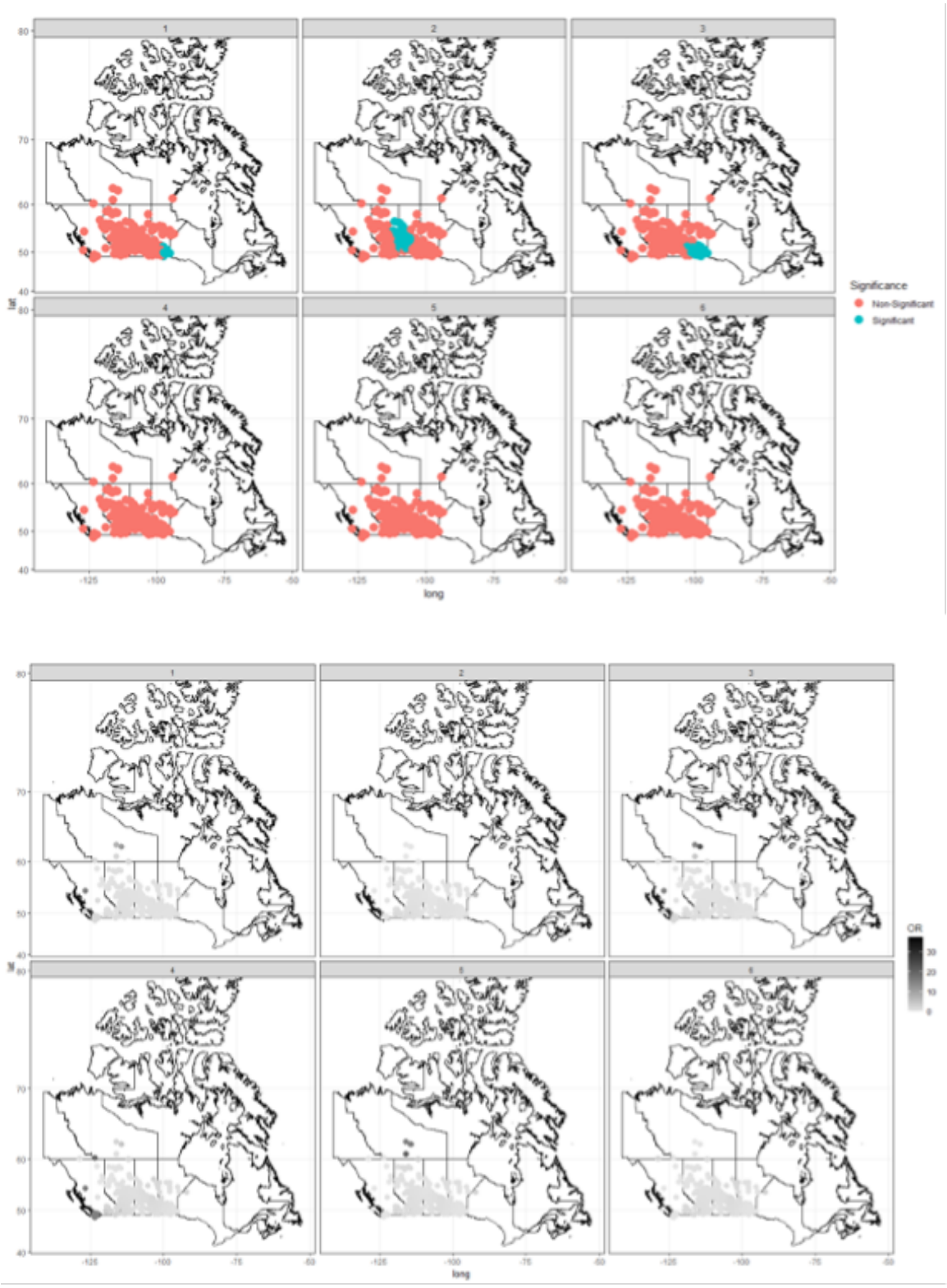

Figure 3.8: Exploring the Impact of Sex on the Incidence of cCHD Subtypes: Accounting for Spatial Variability and Adjusting for AQSMK. Up: significance (Blue: significant, Red: non-significant), Down: Odds ratios.

Chapter 4

Discussion

In our study, we employed geographically weighted multinomial logistic regression (GWMLR) to comprehensively investigate the association between exposure to four different air pollutants (PM2.5, ozone, NO2, and AQSMK) and various subtypes of critical congenital heart defects (cCHDs). Our aim was to provide nuanced insights into the underlying mechanisms and spatial variations of this association, thereby contributing to the current understanding of cCHD etiology and the role of environmental factors.

4.1 Ground-level fine particulate matter (PM2.5)

In our investigation of ground-level fine particulate matter (PM2.5) exposure and its association with critical congenital heart defects (cCHDs), we observed significant associations in 0.32% of the locations. However, it's crucial to note that the subtypes exhibiting significant associations differed from those observed in the original dataset for some locations, highlighting the complexity and spatial heterogeneity inherent in the relationship between air pollutant exposure and cCHD incidence.

Upon adjustment, all significant locations in Saskatchewan were linked to category 1 cCHD subtypes, while in Manitoba, all significant locations were associated with category 4 cCHD subtypes. Category 1 primarily encompasses conotruncal abnormalities such as tetralogy of Fallot (TOF), transposition of the great arteries (TGA),

and truncus arteriosus. These findings are consistent with previous studies by Buteau et al. and Tanner et al., which reported associations between PM2.5 exposure and the incidence of tetralogy of Fallot and non-isolated truncus arteriosus, respectively. [11, 13]

Our classification includes right ventricular outflow tract obstruction (RVOTO) diseases such as Pulmonary and Tricuspid atresia, Ebstein anomaly, and pulmonary valve stenosis in category 4. However, there are no existing associations reported in the literature between the incidence of these types of cCHDs and exposure to PM2.5. This underscores the need for further research to elucidate the specific mechanisms underlying the observed associations and to explore potential links between PM2.5 exposure and a broader range of cCHD subtypes.

Remarkably, the regions identified in Saskatchewan and Manitoba exhibited relatively confined spatial extents. These compact spatial scopes underscore the localized nature of the observed associations, emphasizing the importance of scrutinizing environmental risk factors at fine spatial scales.

Furthermore, our investigation delved into potential etiologies or demographic patterns that might elucidate these observations. In Saskatchewan, areas with significant associations between PM2.5 exposure and cCHDs showcased diverse environmental stressors such as oil and gas exploration, coal mining, and agricultural production. Similarly, in Manitoba, industrial activities including small industry, manufacturing, and specific factories likely contributed to localized variations in air pollution levels, consequently impacting cCHD prevalence. Such localized patterns highlight the significance of considering specific environmental contexts when assessing the impact of air pollution on cCHD incidence.

4.2 Ozone (O3)

The GWMLR analysis revealed significant associations between ozone exposure and cCHDs in 14.2% of the locations. Interestingly, the geographical range of signifi-

cant locations was broader than that observed for PM2.5, primarily spanning the area between eastern Alberta and southern Saskatchewan. Notably, despite Alberta not having the highest ozone levels among all provinces, it exhibited a notably high number of significant locations. Similar to PM2.5, the cCHD subtypes exhibiting significant associations varied across different locations, highlighting nuanced patterns in the relationship between air pollutant exposure and cCHD incidence.

Similar to PM2.5, our analysis revealed discrepancies between the cCHD subtypes present in the original dataset and those exhibiting significant associations in the GWMLR results. This highlights nuanced patterns in the relationship between air pollutant exposure and cCHD incidence, wherein certain subtypes may be more strongly influenced by environmental factors in specific locations. Upon closer examination of demographic characteristics within these locations, diverse population profiles were observed. While we refrain from specifying particular demographic groups out of ethical considerations, the presence of similar population compositions underscores the complexity of factors influencing cCHD incidence. Further exploration into the socio-economic and environmental context of these populations may offer valuable insights into the observed associations.

Our analysis revealed significant associations primarily with categories 1 (conotruncal) and 6 (complex) cCHDs. Interestingly, while most studies have noted an overall increase in cHD incidence with ozone exposure, some have specifically mentioned associations with conotruncal anomalies and tetralogy of Fallot, aligning with our findings for category 1. Notably, our study focused exclusively on the first two months of pregnancy, differing from research highlighting increased risks during the second and third trimesters. Category 6 typically encompasses a combination of conotruncal or left ventricular outflow tract obstruction (LVOTO) anomalies with other types of cCHDs.

In another study, associations with pulmonary and aortic valvular and arterial anomalies were reported, which would correspond to our categories 3 and 4, respec-

tively. However, we did not find any significant associations with these categories in our results. Additionally, some studies have reported null or inverse relationships between ozone exposure and cCHD incidence. [16–18]

4.3 Nitrogen dioxide (NO2)

In our investigation into the relationship between nitrogen dioxide (NO2) exposure and congenital heart defects (cCHDs), traditional multinomial logistic regression (MLR) initially indicated a statistically significant association between NO2 levels and specific subtypes of cCHDs. However, despite this initial finding, the explanatory power of the MLR model was modest, reflected by a Pseudo R-squared value of 0.0076. This suggests that although a significant association was identified, the model could only explain a small proportion of the variability in cCHDs.

Interestingly, upon employing the Geographically Weighted Multinomial Logistic Regression (GWMLR) modeling approach, the previously observed significant association between NO2 exposure and cCHDs did not persist. This discrepancy may arise from various factors. For instance, GWMLR has the capability to account for complex interactions and nonlinear relationships between NO2 exposure and cCHDs, aspects that the traditional MLR model might have overlooked. Moreover, GWMLR can address spatial or temporal dependencies within the data, factors that could influence the association between NO2 and cCHDs. The disparity in results between the two models emphasizes the importance of exploring alternative modeling techniques and thoroughly evaluating the reliability of findings in epidemiological studies.

While previous studies have demonstrated a moderate level of confidence in the association between NO2 exposure and coarctation of the aorta, other studies have reported varying levels of association with tetralogy of Fallot and pulmonary valve stenosis, albeit with lower confidence. [12, 15, 18–20] However, our GWMLR model did not reveal any significant associations with any of these categories. This suggests the need for further investigation and consideration of additional factors in under-

standing the relationship between NO2 exposure and cCHDs.

4.4 Air quality smoke model (AQSMK)

Despite the comprehensive modeling provided by AQSMK, a modeling tool used to estimate the concentrations of particulate matter in the air resulting from smoke, particularly from sources like wildfires, neither traditional multinomial logistic regression nor geographically weighted multinomial logistic regression models have demonstrated any significant associations with any of the cCHD subtypes. This suggests that, despite its ability to estimate smoke-related particulate matter concentrations, AQSMK does not show a clear link between prenatal exposure to wildfire smoke and the risk of specific cCHD subtypes in our study.

While PM2.5 measurements showed some level of significance in association with cCHD subtypes, the lack of significant associations observed with AQSMK warrants further investigation. The differences in results between PM2.5 and AQSMK may stem from several factors.

Firstly, PM2.5 measurements represent direct observations of particulate matter concentrations, providing a more tangible measure of exposure compared to AQSMK estimates, which rely on modeling techniques. The inherent variability and uncertainties associated with modeling approaches could attenuate associations with cCHD subtypes in AQSMK estimates, leading to non-significant results.

Additionally, PM2.5 measurements may capture localized variations in pollution levels more accurately than AQSMK estimates, particularly in areas with high wildfire activity. This spatial variability in exposure levels could contribute to the observed differences in results between PM2.5 and AQSMK.

Moreover, PM2.5 measurements may include particulate matter from various sources beyond wildfire smoke, whereas AQSMK focuses specifically on smoke-related particulate matter. Other pollutants present in PM2.5 measurements, may contribute to the observed associations with cCHD subtypes, whereas AQSMK estimates solely

capture particulate matter from smoke.

Overall, the discrepancies in results between PM2.5 and AQSMK highlight the importance of considering the limitations and potential biases of different exposure assessment methods when interpreting epidemiological findings. Integrating multiple exposure metrics and refining modeling approaches could provide a more comprehensive understanding of the relationship between wildfire smoke exposure and cCHD risk.

4.5 Interpretation of odds ratios

Interpreting the odds ratios (OR) uncovered in our analysis revealed an unexpected trend where nearly all significant locations reported ORs below 1. Conventionally interpreted as indicative of a protective effect, this finding contradicts prevailing conclusions in similar studies.

Several explanations may account for this observation. Our decision to use known chromosomal critical congenital heart defects (cCHDs) as the reference group aimed to mitigate the influence of underlying genetic predispositions, reducing potential confounding factors and facilitating result interpretation. However, it's essential to acknowledge that our findings suggest the baseline risk associated with known chromosomal cCHDs might be higher than that of other groups not yet identified to have a chromosomal etiology. This difference in baseline risk could introduce a downward bias to the odds ratios, resulting in values below 1, as observed in our analysis.

While this approach could potentially induce selection bias, it was necessary because comparing the subtypes with a reference group of unaffected children might not have been appropriate or informative given the complexities of critical congenital heart defects and the genetic factors at play. Therefore, selecting a reference group with known chromosomal cCHDs allowed us to better isolate and understand the specific effects of different subtypes of critical congenital heart defects.

Another significant factor to consider is the possibility that some cases within other

subtypes may have a chromosomal basis that went undiagnosed. Studies by Helm and Buckley have highlighted that septal and atrioventricular septal defect (AVSD) groups are the most common cCHD defects for which their genetic basis can be diagnosed by available tests. [33, 34] However, as explained before these two categories were excluded from our study because of their noncritical nature of malformations.

This limitation could potentially bias the strength and direction of the odds ratios due to detection bias. Our reference group comprised diagnosed chromosomal etiology cCHDs, yet there may be cases within other subtypes with undiagnosed chromosomal abnormalities that tests couldn't detect. This discrepancy in detection could lead to an underestimation of the odds ratios, resulting in values below 1. Thus, the possibility of the presence of undiagnosed chromosomal abnormalities within other subtypes highlights the complexity of our analysis and underscores the need for improved diagnostic methods to accurately identify such cases in future studies.

Additionally, confounding factors may also contribute to the observed trend. Despite our efforts to control for confounders, there may still be unmeasured variables influencing the association between exposure and the incidence of different subtypes of cCHDs. These unaccounted factors could distort the interpretation of the odds ratios, leading to values below 1. Therefore, it is essential to acknowledge the potential impact of confounding on our findings and consider additional analyses or adjustments in future research to better understand the relationship between maternal exposure to air pollutants and the incidence of critical congenital heart defects.

Furthermore, it's crucial to acknowledge that air pollution may interact with other environmental or genetic factors in complex ways, potentially modifying its impact on the incidence of cCHDs. These interactions could either exacerbate or mitigate the individual effects of each factor on cCHD risk. While our analysis concentrated on investigating the associations between individual pollutants and cCHD subtypes separately to circumvent issues of multicollinearity, it's essential to recognize that in real-world scenarios, air pollutants frequently co-occur and may interact synergisti-

cally or antagonistically. Thus, future research should consider the intricate interplay between air pollutants and other factors to gain a comprehensive understanding of their combined influence on cCHD risk.

Moreover, it's worth noting that many of the studies reviewed have reported null or even inverse associations regarding the pollutants and the incidence of cCHD subtypes, indicating the complexity and variability of the relationship between air pollution and cCHD risk across different populations and geographic regions.

The significance observed in small clusters or localized areas, despite odds ratios less than 1 and discrepancies in categories of statistically significant locations with the original dataset, suggests that air pollution may still have a significant influence on cCHDs, albeit not necessarily as a protective factor. This finding provides valuable insights into the spatial variability of the relationship between air pollution and health outcomes, highlighting areas where interventions may be most needed. Ultimately, the magnitude and spatial distribution of the effect of air pollution on cCHD incidence are more critical considerations than the direction of the association.

4.6 Gender Disparities after adjustment for effects of pollutants

After adjusting the model for the pollutant variables, significant associations were observed between sex and cCHDs across all four pollutants. Interestingly, the number of locations with statistical significance varied, with AQSMK and NO2 showing a higher number of significant locations. Notably, in all cases, the odds ratios were below one, indicating that males were more affected than females. This aligns with previous studies that have noted a higher risk of severe CHDs at birth among males, while females tend to be associated with milder CHD subtypes.[35] These findings offer valuable insights for further investigations into underlying mechanisms driving gender disparities in CHD development.

4.7 Limitations

While our study offers valuable insights into the association between air pollution and cCHDs, several limitations must be acknowledged to interpret the results accurately.

Significant limitations due to study design were discussed in detail in Section 4.5. As previously mentioned, undetected confounding factors and unaccounted interactions with other pollutants or environmental and genetic variables could further influence the observed effects. Additionally, while using known chromosomal cCHD cases as the reference group was essential for meaningful comparison, it may have introduced selection bias, potentially resulting in lower odds ratios. Detection bias is another concern, as undiagnosed chromosomal abnormalities in other cCHD subtypes could have led to an underestimation of associations.

Moreover, in regions with smaller sample sizes, the reliability and generalizability of the results may be compromised, increasing the likelihood that observed patterns are influenced more by sampling variability than by genuine spatial effects.

The GWMLR model also brings its own set of challenges. One significant issue is the assumption of spatial stationarity, which means the relationships between variables are assumed to be consistent across different locations. However, in reality, spatial relationships can change dynamically, and this assumption might not always hold true, potentially affecting the accuracy of the model's predictions.

Interpreting statistical significance in GWMLR models can also be complex. Unlike global models that provide a single set of results for the entire dataset, GWMLR fits separate models for each location or subgroup. This localized approach means there isn't a single overall p-value representing the entire study area. Instead, results are specific to each location, which can make it difficult to draw broad conclusions.

However it should be noted that, unlike traditional methods that fit multiple models, GWMLR customizes a unique model for each location rather than fitting multiple models in the conventional sense. Thus, it avoids the issue of multiple testing.

Another challenge is the sensitivity of GWMLR results to the choice of spatial scale or unit of analysis. Different resolutions or spatial boundaries can lead to varying results, highlighting the need to carefully select the spatial scale for analysis and consider how it affects the interpretation of findings.

In addition to the limitations related to study design and the challenges inherent in the GWMLR model, it is important to consider the impact of severe congenital heart defects (cCHDs) that may lead to spontaneous pregnancy loss, selective abortions, or stillbirth. These outcomes add complexity to understanding cCHD risk and highlight the need for a comprehensive approach when examining the interplay between pollutants, genetics, and health outcomes.

Therefore, caution must be exercised when interpreting the findings to ensure that potential biases are adequately addressed and accounted for in the analysis. While our study provides valuable insights into the relationship between air pollution and cCHDs, acknowledging and addressing these limitations is essential to ensure the robustness and validity of our findings. Further research efforts should aim to mitigate these limitations and refine the methodology to improve the accuracy and reliability of future studies in this field.

Chapter 5

Conclusion

Our study provides significant insights into the complex relationship between prenatal exposure to air pollutants and the risk of critical congenital heart defects (cCHDs), offering a clearer understanding of how these environmental factors interact with health outcomes. By focusing on spatial variability, we aim to contribute to more precise and effective research efforts and public health interventions.

Through our detailed analysis, we identified important associations between specific air pollutants—such as PM2.5, ozone—and the incidence of cCHDs in particular regions. However, while these findings are important, it is crucial to approach them with caution. Variations in how cCHD subtypes are distributed across locations and odds ratios slightly below 1 suggest that these associations are influenced by several factors, including the choice of reference groups, genetic predispositions, and environmental variables. These complexities make it essential to interpret the data with nuance, recognizing that different cCHD subtypes may behave differently across various regions.

Moreover, discrepancies emerged when comparing the subtypes of cCHDs that were found to be significant in certain areas with the subtype prevalence in the original dataset, highlighting the importance of understanding disease subtypes when assessing spatial associations. For example, clusters of significance for PM2.5 exposure may be connected to the unique environmental or demographic characteristics of these re-

gions, while areas with significant ozone exposure showed demographic patterns that warrant further investigation. These insights point to a need for more detailed studies that consider not only the pollutants themselves but also the context in which exposure occurs.

The odds ratios observed in our study, some falling just below 1, also draw attention to the ongoing debates within the literature about the effects of air pollutants on congenital heart defects. While some of our findings are consistent with previous studies, others differ, indicating that the relationship between air pollutants and cCHDs is complex and multifaceted. The differences in results across studies likely stem from variations in study design, population characteristics, pollutant exposure levels, and the methodologies used to assess risk. This highlights the challenge of drawing definitive conclusions in this field and underscores the need for more research to clarify these relationships.

Despite these challenges, our study provides critical information that can inform public health efforts. The identification of specific regions with significant associations between air pollutants and cCHDs offers a foundation for targeted interventions. By focusing on these areas, public health authorities can implement strategies aimed at reducing pollutant exposure and improving health outcomes for at-risk populations. These strategies could include air quality monitoring, stricter regulations on emissions, and community-based health initiatives designed to raise awareness and mitigate risks.

Our findings underscore the urgent need for more spatially focused research that accounts for the complex interactions between environmental exposures and health outcomes. Future studies should continue to explore how different pollutants interact with each other and with various demographic and genetic factors to influence the risk of cCHDs. Additionally, addressing gaps in our knowledge about how pollutant exposure impacts vulnerable populations, such as pregnant women and children, will be essential for developing effective public health policies.

In conclusion, our study advances the understanding of the intricate relationship between prenatal air pollutant exposure and the incidence of cCHDs. By emphasizing the importance of spatial variability, we provide valuable insights for both research and public health interventions. These findings call for continued collaborative efforts and further investigation to ensure that effective strategies are developed to protect public health, reduce the incidence of congenital heart defects, and improve the well-being of communities affected by environmental pollution. Through sustained research and targeted action, we can work towards minimizing the impact of air pollution on vulnerable populations and fostering healthier environments for future generations.

Bibliography

[1] G. M. Blue *et al.*, "Insights into the genetic architecture underlying complex, critical congenital heart disease.," *American heart journal*, vol. 254, pp. 166–171, 2022.

[2] M. K. Bakker *et al.*, "Prenatal diagnosis and prevalence of critical congenital heart defects: An international retrospective cohort study," *BMJ open*, vol. 9, no. 7, e028139, 2019.

[3] L. D. Botto, A. E. Lin, T. Riehle-Colarusso, S. Malik, and A. Correa, "Seeking causes: Classifying and evaluating congenital heart defects in etiologic studies," *Birth Defects Research Part A: Clinical and Molecular Teratology*, vol. 79, no. 10, pp. 714–727, 2007.

[4] R. Kučienė and V. Dulskienė, "Selected environmental risk factors and congenital heart defects," *Medicina*, vol. 44, no. 11, p. 827, 2008.

[5] E. K.-C. Chen, D. Zmirou-Navier, C. Padilla, and S. Deguen, "Effects of air pollution on the risk of congenital anomalies: A systematic review and meta-analysis," *International journal of environmental research and public health*, vol. 11, no. 8, pp. 7642–7668, 2014.

[6] P. Thangavel, D. Park, and Y.-C. Lee, "Recent insights into particulate matter (pm2. 5)-mediated toxicity in humans: An overview," *International journal of environmental research and public health*, vol. 19, no. 12, p. 7511, 2022.

[7] J. Zhang, Y. Wei, and Z. Fang, "Ozone pollution: A major health hazard worldwide," *Frontiers in immunology*, vol. 10, p. 2518, 2019.

[8] D. M. Stieb *et al.*, "Systematic review and meta-analysis of cohort studies of long term outdoor nitrogen dioxide exposure and mortality," *PloS one*, vol. 16, no. 2, e0246451, 2021.

[9] *Canadian urban environmental health (canue) research consortium. annual datasets on greenness, neighbourhood, air quality, and weather [digital data]. 2021; available from:* https://www.canuedata.ca/metadata.php. Accessed: 2010-09-30.

[10] K. C. Hall, J. C. Robinson, W. H. Cooke III, A. S. Parnell, L. Zhang, and L. Northington, "Relationship between environmental air quality and congenital heart defects," *Nursing research*, vol. 71, no. 4, pp. 266–274, 2022.

[11] S. Buteau, P. Veira, M. Bilodeau-Bertrand, and N. Auger, "Association between first trimester exposure to ambient pm 2.5 and no 2 and congenital heart defects: A population-based cohort study of 1,342,198 live births in canada," *Environmental Health Perspectives*, vol. 131, no. 6, p. 067 009, 2023.

[12] J. A. Stingone *et al.*, "Maternal exposure to criteria air pollutants and congenital heart defects in offspring: Results from the national birth defects prevention study," *Environmental health perspectives*, vol. 122, no. 8, pp. 863–872, 2014.

[13] J. P. Tanner *et al.*, "Associations between exposure to ambient benzene and pm2. 5 during pregnancy and the risk of selected birth defects in offspring," *Environmental research*, vol. 142, pp. 345–353, 2015.

[14] K. C. Hall and J. C. Robinson, "Association between maternal exposure to pollutant particulate matter 2.5 and congenital heart defects: A systematic review," *JBI Evidence Synthesis*, vol. 17, no. 8, pp. 1695–1716, 2019.

[15] C.-Y. Hu *et al.*, "Maternal air pollution exposure and congenital heart defects in offspring: A systematic review and meta-analysis," *Chemosphere*, vol. 253, p. 126 668, 2020.

[16] B. Ritz, F. Yu, S. Fruin, G. Chapa, G. M. Shaw, and J. A. Harris, "Ambient air pollution and risk of birth defects in southern california," *American journal of epidemiology*, vol. 155, no. 1, pp. 17–25, 2002.

[17] L. C. Vinikoor-Imler, T. G. Stewart, T. J. Luben, J. A. Davis, and P. H. Langlois, "An exploratory analysis of the relationship between ambient ozone and particulate matter concentrations during early pregnancy and selected birth defects in texas," *Environmental pollution*, vol. 202, pp. 1–6, 2015.

[18] M. Vrijheid *et al.*, "Ambient air pollution and risk of congenital anomalies: A systematic review and meta-analysis," *Environmental health perspectives*, vol. 119, no. 5, pp. 598–606, 2011.

[19] S. Michel, A. Atmakuri, and O. S. von Ehrenstein, "Prenatal exposure to ambient air pollutants and congenital heart defects: An umbrella review," *Environment International*, p. 108 076, 2023.

[20] R. Nicoll, "Environmental contaminants and congenital heart defects: A re-evaluation of the evidence," *International journal of environmental research and public health*, vol. 15, no. 10, p. 2096, 2018.

[21] P. Dadvand, J. Rankin, S. Rushton, and T. Pless-Mulloli, "Ambient air pollution and congenital heart disease: A register-based study," *Environmental research*, vol. 111, no. 3, pp. 435–441, 2011.

[22] C. M. Robertson *et al.*, "The registry and follow-up of complex pediatric therapies program of western canada: A mechanism for service, audit, and research after life-saving therapies for young children," *Cardiology Research and Practice*, vol. 2011, 2011.

[23] C. M. J. Tan and A. J. Lewandowski, "The transitional heart: From early embryonic and fetal development to neonatal life," *Fetal diagnosis and therapy*, vol. 47, no. 5, pp. 373–386, 2020.

[24] *Air quality research division, t., canada. , environment and climate change canada. 2017.*

[25] A. Robichaud, R. Ménard, Y. Zaïtseva, and D. Anselmo, "Multi-pollutant surface objective analyses and mapping of air quality health index over north america," *Air Quality, Atmosphere & Health*, vol. 9, pp. 743–759, 2016.

[26] A Robichaud and R Ménard, "Multi-year objective analyses of warm season ground-level ozone and pm 2.5 over north america using real-time observations and canadian operational air quality models," *Atmospheric chemistry and physics*, vol. 14, no. 4, pp. 1769–1800, 2014.

[27] A. Van Donkelaar, R. V. Martin, C. Li, and R. T. Burnett, "Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors," *Environmental science & technology*, vol. 53, no. 5, pp. 2595–2611, 2019.

[28] *Suite, c.p.c., [computer file] markham: Dmti spatial inc. 2015.*

[29] P. Hystad *et al.*, "Creating national air pollution models for population exposure assessment in canada," *Environmental health perspectives*, vol. 119, no. 8, pp. 1123–1129, 2011.

[30] J. R. Brook, E. M. Setton, E. Seed, M. Shooshtari, and D. Doiron, "The canadian urban environmental health research consortium–a protocol for building a national environmental exposure data platform for integrated analyses of urban form and health," *BMC public health*, vol. 18, pp. 1–15, 2018.

[31] N. Paul, J. Yao, K. E. McLean, D. M. Stieb, and S. B. Henderson, "The canadian optimized statistical smoke exposure model (canossem): A machine learning approach to estimate national daily fine particulate matter (pm2. 5) exposure," *Science of The Total Environment*, vol. 850, p. 157 956, 2022.

[32] *Environment and climate change canada (eccc). national air pollution surveillance (naps) program 2021 (1974-2020); available from:* https://www.canada.ca/en/environment-climate-change/services/air-pollution/monitoring-networks-data/national-air-pollution-program.html, Accessed: 2010-09-30.

[33] B. M. Helm, B. J. Landis, and S. M. Ware, "Genetic evaluation of inpatient neonatal and infantile congenital heart defects: New findings and review of the literature," *Genes*, vol. 12, no. 8, p. 1244, 2021.

[34] J. R. Buckley, M. N. Kavarana, S. M. Chowdhury, and M. A. Scheurer, "Current practice and utility of chromosome microarray analysis in infants undergoing cardiac surgery," *Congenital heart disease*, vol. 10, no. 3, E131–E138, 2015.

[35] F. Pugnaloni, A. Felici, A.-F. Corno, B. Marino, P. Versacci, and C. Putotto, "Gender differences in congenital heart defects: A narrative review," *Translational Pediatrics*, vol. 12, no. 9, p. 1753, 2023.