Spatial Analysis of Ischemic Heart Disease in Manitoba

by

Justin Dyck

A thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Department of Community Health Sciences

University of Manitoba

Winnipeg

Abstract

Introduction: Chronic diseases rarely follow uniform distributions throughout geographical space and so identifying regions that have frequent occurrences or elevated prevalences is important. The disease of interest for this research project was ischemic heart disease (IHD) which is a highly prevalent disease in Manitoba. This research project was focused on detecting spatial patterns of IHD and assessing the associations of some of the potential causes of these patterns.

Purpose and Objectives: The purpose of this project was to use statistical tools to detect spatial and temporal patterns of IHD in Manitoba. The objectives were to: (1) detect geographic clusters of acute myocardial infarctions (AMI) within Manitoba; (2) assess whether IHD is related to the geographic distribution of some its well-known risk factors; and (3) what relationship IHD has to the temporal dimension throughout geographic space.

Methods: This project used data from the population research data repository housed at the Manitoba Centre for Health Policy to identify persons diagnosed with IHD. The first objective was assessed using the flexible spatial scanner to detect clusters of AMIs. The second objective was assessed using spatial Poisson regression models that modelled the spatial covariance structures with neighborhood conditional autoregressive structures. The third objective was assessed by extending the spatial model to the temporal dimension by modeling the temporal covariance structures with random-walk covariance structures. Space-time interaction effects were assessed to complete the evaluation of the third objective.

Results: One primary and eight secondary disease clusters of AMIs were identified, where the primary cluster occurred in the central Manitoba region. Hypertension prevalence and indigenous

population proportion significantly predicted IHD prevalence while controlling for socio-economic status (SES) and spatial autocorrelation. A 10% increase in hypertension prevalence within a region was found to increase the risk of IHD by 9.3% (95% CI: 5.8%-13.0%), and a 10% increase in the indigenous population proportion within a region was found to increase the risk of IHD by 0.7% (95% CI: 0.1%-1.3%). When controlling for temporal autocorrelation, indigenous population proportion was not a significant predictor of IHD, while hypertension was. The results were within error the same for males and females when stratifying by sex. Modelled IHD prevalence was found to be decreasing over time, but the majority of this occurred in the female sub-group, while the male sub-group was mostly constant. Counter to this finding, IHD prevalence in some regions substantially increased over the study period.

Conclusions: This research identified AMI clusters as well as modelled the spatial and temporal variation in IHD within 96 regions in Manitoba over 23 years. It was found that there were significant associations between IHD and the two covariates of hypertension and indigenous population proportion, and there was indication that similar regions experienced worse outcomes for both IHD prevalence and AMI incidence. The most significant effect was the space-time interaction, suggesting that the temporal patterns in IHD prevalence vary significantly throughout space, with some regions having significantly increasing trends over time counter to the provincial average.

Acknowledgments

First, I would like to thank my thesis advisor Dr. Mahmoud Torabi for his support and knowledge throughout this process. The statistical insight into the world of spatially dependent data has been inspiring and provoking, which has undoubtedly kept my work on the right trajectory. Also, my thesis committee members; Dr. Robert Tate and Dr. Julia Uhanova; who have been very helpful in ensuring that my research is relevant from a community and public health perspective.

This project has been financially backed by the Manitoba Training Program for Health Services Research fellowship, the Natural Sciences and Engineering Research Council of Canada- Discovery Grant (NSERC-DG), and GETS (University of Manitoba) to Dr. Mahmoud Torabi.

I also acknowledge the Manitoba Centre for Health Policy for use of data contained in the Manitoba Population Research Data Repository under project #2018-044 (HIPC#2018/2019-24). The results and conclusions are those of my own and no official endorsement by the Manitoba Centre for Health Policy, or Manitoba Health is intended or should be inferred. Data used in this study are from the Manitoba Population Research Data Repository housed at the Manitoba Centre for Health Policy, University of Manitoba and were derived from data provided by Manitoba Health.

I would also like to thank the team at the George and Fay Yee Centre for Healthcare Innovation (CHI) for providing the opportunity of a student internship. This experience has been extremely rewarding as it has further prepared me for research in healthcare data science. The community at CHI provided valuable insight and resources for aspiring researchers, and I am very grateful for this experience.

Finally, I would like to thank my family; especially my partner for her patience and support, my grandparents for their encouragement and assistance, and my mother and sisters for believing in my capabilities. I would not have been able to complete this without you all.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AMI – Acute Myocardial Infarction (heart attack)

ATC – Anatomic Therapeutic Chemical

BYM – Besag-York-Mollie [spatial model]

CCHS – Canadian Community Health Survey

CI – Credible Interval

CIHI – Canadian Institute for Health Information

CVD – Cardiovascular Disease

DA – [Census] Dissemination area

DB – [Census] Dissemination Block

DIC – Deviance Information Criterion

DPIN – Drug Program Information Network

FSS – Flexible Spatial Scanner

GMRF – Gaussian-Markov-Random-Field

ICAR – Intrinsic Conditionally Auto-Regressive

ICD – International Classification of Diseases

IHD – Ischemic Heart Disease

INLA – Integrated Nested Laplace Approximation

MAUP – Modifiable Areal Unit Problem

MCHP – Manitoba Centre for Health Policy

MCMC – Markov Chain Monte Carlo

PHIN – Personal Health Identification Number

PCCF – Postal Code Conversion File

PRDR – Population Research Data Repository

RHA – Regional Health Authority

RHAD – Regional Health Authority District

RR – Risk Ratio

SEFI – Socio-economic Factor Index

SEFI-2 – [Second version of the] Socio-economic Factor Index

SES – Socio-economic Status

SIR – Standardized Incidence Ratio

SRR – Standardized Risk Ratio

Chapter 1: Introduction

1.1 - Background

Spatial analyses and disease mapping are important aspects of epidemiological studies, as they can provide the answer to the "where?" question that researchers often ask when performing health research. Chronic diseases in general rarely follow uniform distributions throughout geographical space, and so identifying regions that have frequent occurrences or have elevated prevalences is important as it can aid in our understanding of the underlying contributing factors of the illness of interest. Gaining this understanding is possible due to the nature of the risk factors for the chronic disease of interest; these are usually not randomly distributed throughout geographical space either. Therefore, applying a spatial analysis to health data will often reveal important information on the connections between the disease of interest and its suspected risk factors. Further, geographical patterns of the disease could be revealed in the analysis, which would provide information on where our attention is needed to provide equitable care for the citizens in those regions.

The disease of interest for this research project is Ischemic Heart Disease (IHD), which is a highly prevalent chronic disease in much of the world and particularly in Manitoba has been measured to have a prevalence of approximately 7.9% for Manitobans aged 19 and older and 8.5% of all Canadians aged 20+ (Fransoo et al., 2013, pg. 93; Public Health Agency of Canada, 2017). IHD (also known as coronary heart disease), is a condition where reduced blood flow to the heart is present, and usually develops due to blockages or constrictions of the arteries in the heart (Mayo Clinic, 2018). In the extreme case when a coronary artery is blocked completely, the lack of oxygen causes damage to the heart muscles which is known as an acute myocardial infarction (AMI) or heart attack and is often life-threatening.

The leading cause of all deaths and the second leading cause of premature deaths in Manitoba has been measured to be the broader disease classification of circulatory illnesses (Fransoo et al., 2013, pg. 44, 50), of which IHD is a major contributor. Manitobans from many different backgrounds and economic standings are affected by IHD, however, it has been shown that it does not affect people of different demographic or socioeconomic status (SES) stratums equally (Tobe et al, 2015, pg. 1124; Finegold et al, 2013, pg. 939). This is important in the context of this research project, as people tend to live near other persons who are of similar economic standings and demographic classifications. Thus, it should be expected that some of the geographical variation in disease prevalence could be attributed to the geographical distribution of demographic and SES factors. This is compounded by the natural phenomena that nearer regions are in general more highly correlated than further regions (Tobler, 1970), and so many of the statistical assumptions such as independent residuals (random errors) that are made when performing regressions or even simply displaying descriptive statistics, are invalid and could lead to incorrect conclusions. Therefore, accounting for the inherent spatial dependence of chronic diseases is of great importance to preserve accuracy in disease prevalence reporting, which is the central focus of this research project.

1.2 - Purpose and Objectives

The research aim of this project is to use tools within the area of spatial statistics to detect clusters of AMI incidences as well as to identify any geographic and/or temporal patterns in IHD prevalence rates within Manitoba. This is addressed through three main objectives and their corresponding research questions outlined in the succeeding paragraphs.

The first objective is to detect geographic clusters of the acute form of IHD (AMI) within Manitoba, where the corresponding research question is: **Are AMI occurrences clustering within areas that are classified as being lower than the provincial average for SES, have high indigenous population proportions, are urban settings, or are a combination of these factors?**

The second objective is to assess whether the prevalence of IHD is related to the geographic distribution of some well-known risk factors such as hypertension, SES, indigenous status, and pollution/urbanicity. This leads to the second research question: **Are geographic patterns of IHD prevalence related to the spatial distributions of hypertension, indigenous population proportion, SES, and urbanicity?**

The third objective of this project is to discover the relationship of IHD prevalence to the dimension of time. This leads to the third and final research question: **Are there any increasing temporal trends of IHD prevalence in regions with low SES, high hypertension prevalence, urban settings, and large indigenous populations?**

1.3 - Justification

Previous research studies have uncovered spatial disease patterns for cardiovascular diseases (CVD) in other areas of the world. For instance, elevated risks for IHD in Spanish towns that were of lower overall SES standing and higher overall rates of obesity were uncovered (Medrano et al, 2012). As well as high probabilities of elevated hypertension were found in small geographic areas in Germany that also had high rates of social deprivation (Kauhl et al, 2018). However, in the context of Manitoba, very little other than descriptive reports on the state of IHD

within administrative health regions has been done. These reports are important, as they provide valuable information on the state and the dynamics of IHD with respect to time and space. In the "2013 RHA Indicators Atlas" from the Manitoba Centre for Health Policy (MCHP) it has been shown clearly that there are large differences in the geographical distribution of IHD along with a temporal trend (Fransoo et al., 2013, pg. 93). But in order to dig deeper to potentially understand some of the probable causes of these trends as well as the contributing factors underlying the disease distributions, we need to turn to statistical disease models rather than relying on descriptive statistics.

This research project utilizes data from both the Population Research Data Repository (PRDR) housed at the MCHP and the 2016 Canadian Census records. This allows for relatively small levels of geographic aggregation to be obtained as the PRDR records the postal codes of all individual Manitobans with active Manitoba health insurance, and the census data is broken into the 2183 census dissemination areas (DA)'s that make up Manitoba. Linking these datasets at larger regions defined by the 96 administrative health districts is done to preserve the accuracy of the findings, as both the modifiable areal unit problem (MAUP) and the spatially misaligned data problem need to be overcome; these are outlined in the later methods chapter. As health reporting is often done at this district level, the results from this project will have more potential for comparisons to be drawn with previous reports. This enables the usability of these results by healthcare stakeholders when making population level decisions regarding their health regions.

From a policy perspective, this research could be quite beneficial. The granularity of geographical regions can be much finer using spatial models rather than simply stating prevalence rates, as the errors are modelled more accurately and thus less inflated. This could lead to a substantial amount of evidence for the question of "where is this disease most

problematic in Manitoba, where has it been getting worse, and what are some high contributors or indicators of the elevated risk?" The connections between risk factors, the disease in question, and geography presents new evidence that could be used to target either certain geographies, or certain economic or demographic groups with preventative health measures. As equitable healthcare for all Manitobans is a priority from a policy lens, evidence of the type produced in this research project is needed to further this goal.

Chapter 2: Literature Review and Theoretical Background

This chapter is split into two main sections: the first being a review of the literature surrounding IHD. Here the current state of IHD, the social and economic impacts of IHD, and the main contributing factors that increase IHD risk are explored. The second main theme is that of how geography plays a role, and a general theoretical framework is laid for how previous studies have sought to understand IHD's distribution throughout both the geographical and the temporal dimensions.

2.1 - Ischemic Heart Disease: Impacts and Risk Factors

According to the "2013 RHA Indicators Atlas" report from the MCHP, during both the 2002-2006 and the 2007-2011 time periods, circulatory illness has been the leading cause of death in Manitoba and is the second leading cause of premature mortality next to cancers (Fransoo et al., 2013, pg. 44, 50). In particular, IHD; which accounts for roughly half of all circulatory illness deaths (The Conference Board of Canada, 2019); has had a prevalence rate of 7.92% and an incidence rate of 6.73 per 1000 person years for Manitobans aged 19 and over in 2007-2011, and the incidence rate of AMIs was at 4.09 per 1000 persons aged 40 and over in 2007-2011 (Fransoo et al., 2013, pg. 93, 97, 109). Persons with a history of having a heart attack is 2.1% of all Canadians aged 19+, and accounts for approximately a quarter of all persons living with IHD in Canada (Public Health agency of Canada, 2017). For IHD, the general trend has been a decline in both prevalence and incidence from the 2002-2006 to the 2007-2011 time periods (Fransoo et al., 2013, pg. 93, 97), which indicates that IHD is on a downward trend in

Manitoba. This is also true for AMI incidence, which has also seen a down turn over the 2000-2004 to the 2005-2009 time periods (Fransoo et al., 2013, pg. 109).

This decreasing trend in IHD prevalence, IHD incidence, and AMI incidences has been found to be consistent with the trends for Canada at large (Public Health Agency of Canada, 2018, pg. 6-9). However, this reduction may not be consistent across all age-groups, ethnicities, and geographic regions. For instance, there has been a significant increase in the AMI incidence rate in the Interlake-Eastern Health Region in Manitoba from the 2002-2006 to the 2007-2011 time periods (Fransoo et al., 2013, pg. 110). The IHD prevalence and incidence rates have also been shown to be rising significantly over these periods in multiple health districts in Manitoba, such as the Cross Lake and Norway House districts, which indicates that some of the nuances in this disease get masked when taking the population level statistics at face-value (Fransoo et al., 2013, pg. 95, 99). As equitable healthcare is a priority for Manitoba's health care system, understanding how one of the leading causes of premature deaths in Manitoba is affecting different geographic regions more severely than others should be a focus of this priority.

The main impacts of IHD that are looked at more closely in the succeeding paragraphs could be broadly categorized into two main sections: loss of potential years of life, and taxpayer burden in the form of health care system resources. Other societal impacts are of course also present for IHD. However, quality of life and non-financial healthcare system burdens are not explored here as these are more difficult to measure precisely.

The cost of potential years of life lost (PYLL) is high in the context of circulatory diseases. The second highest cause of premature mortality in Manitoba, defined as mortality before reaching the age of 75, is circulatory illnesses taking an estimated 21-22% of all

premature deaths (MB Health, 2017, pg. 20; Fransoo et al., 2013, pg. 50). Although this trend also seems to be declining, it would be important to know if it is declining throughout all ethnicity groups, geographic locations etc. or whether all types or only certain types of circulatory diseases are related to the decrease in premature mortality. This also is confirmed in other first world countries that have reported a decrease in age- and sex-standardized deaths due to IHD (Finegold et al, 2013, pg. 939). This highlights an important risk factor for IHD; SES, as it has been established that there are much higher rates of IHD in those under 60 years of age in countries of lower economic standing (Finegold et al, 2013, pg. 939). This is a theme that is very prominent in epidemiology, as many chronic diseases affect populations from lower economic stratums more severely than others.

The financial impact of circulatory illnesses is substantial. Given that we are in the context of a universal health insurance system in Manitoba which is funded by the taxpayers, this impact is important as it directly influences the costs imposed on the citizens of Manitoba. According to the patient cost estimator provided by the Canadian Institute for Health Information (CIHI), AMIs in Manitoba that required the use of a coronary angiogram cost the health care system $8,467 per case on average, and $7,289 if a coronary angiogram was not needed (CIHI, 2018). When accounting for the typical number of cases for each of these categories there is an average of 69 and 180 cases respectively for the age group 18-59 in Manitoba per year, resulting in a total cost of approximately $1.9 million (CIHI, 2018). But this is just the tip of the iceberg. If other interventions are needed these numbers increase drastically, as for instance, a cardiac valve replacement costs on average $32,432 per case (CIHI, 2018). Overall an estimated total for all circulatory system disease emergencies and interventions costs the health care system in Manitoba an estimated $23 million per year for persons 18-59 years old and $70.8 million when

considering all age groups over 18 years of age (CIHI, 2018). As this is already a substantial investment on the part of taxpayers, any solution that lessens the burden even marginally will benefit Manitoba's Provincial health budget.

It is well understood through extensive health studies that the three general categories that influence the risk of developing most chronic diseases are lifestyle practices, the environment, and genetics or family history. IHD is no exception, there is a wide body of literature supporting the claim that IHD risk is influenced by all three of these factors. This study is concerned primarily with modifiable risk factors such as lifestyle behaviours, SES, and geographic location. Non-modifiable factors such as genetics are explored when necessary; such as accounting for the indigenous populations in Manitoba.

Lifestyle factors such as tobacco use, heavy alcohol consumption, sedentary lifestyles, and unhealthy eating habits all increase the risk of developing CVD (Mayo Clinic, 2018). At a population level these are often hard to study however, as data on every person's lifestyle is not explicitly available. Surveys such as the Canadian Community Health Survey (CCHS) conducted by Statistics Canada could help to bridge this gap, but with only a small number of respondents for Manitoba, this survey's measurement accuracy isn't precise enough for a population level study. It has been shown that these lifestyle factors are linked to SES (Huckle et al, 2010, pg. 1199; Reid et al, 2010, pg. 76; O'Donoghue et al, 2016, pg. 21), and so some of the variation in IHD may be able to be partially explained by SES if one or more of these other risk factors is unmeasured. In fact, life expectancy in general has been shown to be strongly associated with SES (The Health Officers Council of British Columbia, 2013; Public Health Agency of Canada,

2018, pg. 60). Therefore, a surrogate variable for these lifestyle factors in the form of SES indicators is valuable for a population level research project on IHD.

As previously alluded to, the environment also has a role to play for IHD risk. Pollution, specifically particulate matter and nitrogen dioxide have been shown to contribute to CVD risk in general (Pun et al, 2014; Yoo et al, 2018). However, as there are only five air quality monitoring stations across the province this doesn't provide enough information to make accurate inferences about the association between IHD and the air quality. It may be of interest to assume higher pollution rates occur in higher populated areas such as urban settings as pollution is generally more prominent in cities versus rural settings.

The environment can also influence lifestyle behaviours. The link between environmental deprivation such as pollution and SES has been well studied where more environmental deprivation is positively associated with more economic deprivation (Goodman et al., 2011, pg. 772; Batisse et al, 2017, pg. 507). The same can be said of alcohol consumption, sedentary lifestyles, and smoking habits; environments with more economic deprivation are highly associated with these factors (Shimotsu et al, 2013, pg. 455; Eyre et al, 2014, pg. 238; Anderson, 2006, pg. 491). This again points to an underlying SES variable that influences lifestyle behaviours as well as where people live; whether in areas of higher or lower pollution.

Lastly, genetics and family history play a significant role in the risk of IHD (Mayo Clinic, 2018). However, this variable is often not measured as we would need to collect historical information on family history of IHD, etc., so we may substitute a genetic-related variable as a measurement of ethnicity. Recent work has shown that the decreasing rate of CVD is not the same in non-indigenous groups as it is in indigenous groups in Canada, and is in fact decreasing slower for indigenous men and even increasing in indigenous women (Tobe et al, 2015, pg.

1124). In the context of this research project, indigenous is taken to mean anyone self-identifying as either First Nations, Inuit, or Metis. It is hypothesized that much of this difference is due to higher rates of diabetes in indigenous populations (Tobe et al, 2015; Public Health Agency of Canada, 2018, pg. 195), as diabetes is another important risk factor for CVD in general (Mayo Clinic, 2018). Further, indigenous Canadians living off-reserve have been shown to have a higher than average alcohol consumption rate, and higher tobacco smoking and exposure to second-hand smoke rates than the non-indigenous Canadian population (Public Health Agency of Canada, 2018, pg. 290, 312), which puts them more at risk for IHD due to factors that are not entirely related to genetics. There is also a link between SES and indigenous status, as off-reserve indigenous Canadians had higher rates of food insecurity and working poverty status than non-indigenous Canadians (Public Health Agency of Canada, 2018, pg. 381, 404). Since indigenous persons contribute to 18% of Manitoba's total population (Statistics Canada, 2016), this could affect the distribution of IHD significantly and should be accounted for. However, since many of the risk factors for IHD are present at higher levels in the indigenous population, it shouldn't be assumed that the elevation of IHD in indigenous persons is due to genetic factors alone, but rather a combination of intrinsic and extrinsic factors.

Other less extreme health conditions can also contribute significantly to IHD risk, as comorbidities are often found in individuals with IHD and are often diagnosed prior to IHD diagnosis (Berger et al, 2010, pg. 879). This can lead to important variables in an analysis of IHD risk, as diagnoses of other illnesses is tracked by the health system in Manitoba. Hypertension, which is a less severe form of CVD, is an extremely prevalent disease as well as being highly associated with IHD risk. Also known as chronic high blood pressure, hypertension occurs when the pressure exerted by blood on the arterial walls is higher than what is normal in a

healthy individual, which can have detrimental health effects if elevated for too long a period or at too high a rate (Mayo Clinic, 2018). Hypertension is the most prevalent form of CVD in Manitoba at 25.6% prevalence in 2011-2012 (Fransoo et al., 2013, pg. 69), and has been shown to be a risk factor for other forms of CVD including IHD (Spinar, 2012, pg. 434).

Although IHD prevalence, IHD incidence, and AMI incidence rates have been shown to be on the decline in recent years, hypertension prevalence is on the rise (MB Health, 2017, pg. 24, 28; Fransoo et al., 2013, pg. 69, 93, 109). Hypertension is one of the leading risk factors for IHD and AMI (Weber et al, 2016, pg. 468; Spinar, 2012, pg. 434), and is an important factor to measure and account for in an analysis of either of these diseases. However, as hypertension has been shown to be on the rise and IHD and AMI rates to be on the decline, there must be other factors that contribute to the differences in these opposing trends. Further, it has been shown that hypertension is also associated with other risk factors of IHD such as tobacco use and heavy alcohol consumption (Mayo Clinic 2018). Since SES is associated with these lifestyle factors which influence IHD risk as well as hypertension (Subramanyam et al, 2013, pg. 144), it could be expected that SES is major confounder between hypertension and IHD. This confounding could also be present between SES and indigenous, as both of these variables are associated with IHD as well as indigenous persons being more susceptible to lower levels of SES (Public Health Agency of Canada, 2018, pg. 381, 404).

An important variable to track for population level IHD risk is population demographics. Age and sex both play important roles in IHD risk, as it has been shown that the risk is not equal in all age groups or between sexes. Much of the literature supports the claim that women have higher rates of mortality due to IHD than do men and are older at diagnosis (Ashley & Geraci, 2013, pg. 430; Berger et al, 2010, pg. 879; Vaccarino, 1998, pg. 2059). Women also have higher

rates of comorbidities such as hypertension and diabetes than do men, whereas men have higher rates of extrinsic risk factors such as tobacco usage and heavy alcohol consumption (Berger et al, 2010, pg. 879). This confirms that the standard procedure of accounting for these demographic variables; age and sex; in epidemiological studies of chronic diseases is important in the context of IHD as they may confound with other risk factors, comorbidities, or significantly interact with each other.

2.2 – Theoretical Background

Ecological studies can often provide measures of association between diseases and risk factors without the need or prior to needing clinical or laboratory studies. This is especially true when the ecological study is population based, as the statistical power of such a study is potentially large in magnitude. As administrative health records are collected and organized for the population in Manitoba, this allows for population health studies to be conducted for many various diseases or conditions.

Modelling of geographical ecological associations violate the assumption that the residual errors are independently and normally distributed. It is well known that neighborhoods that are closer together often share similar traits, including demographics and economic status. As these and other geographically distributed variables are key covariates when studying chronic diseases, it does not make sense to say that neighboring regions are independently contributing to the outcome of interest. Further, constructing artificial boundaries between areas or neighborhoods may introduce the issue of imposing a false discontinuity into a continuous process. Here, risk factor levels will not naturally stop or change at the geographic boundaries we constructed to

formulate these variables, and so to account for this extra variation it may be of benefit to look into spatial regression models.

An exploratory approach to detecting clusters of diseases using spatial scanners has had some limited usage in identifying clusters of AMIs. These spatial scanners used to detect geographic disease clusters can be a helpful tool for surveillance and exploratory uses, as they provide an easy to implement method for detecting areas that have statistically significant elevated disease rates. This allows the researcher to identify regions that are potentially problematic compared to the overall rates. However, it does not allow for covariates to be modelled, and so the potential causes of the clusters would need to be explored with other methods such as spatial and spatio-temporal models.

Previous literature has demonstrated successful usage of these spatial scanners for identifying clusters of AMIs in the past, and in the Manitoba context these have been used successfully to identify clusters of multiple sclerosis (Torabi et al, 2014). In other contexts, cluster detection has been used to identify nine high risk clusters of AMIs in Texas in a study conducted in 2011, and 211 clusters were similarly identified in Denmark in a 2016 study (Pedigo et al, 2011; Kjaerulf et al, 2016). Spatial analysis of AMIs in the Strasbourg Metropolitan Area also revealed significant clusters of disease as well (Kihal-Talantikite et al, 2017). The identification of these clusters would be useful for health services policy analysts, as it would enable them to focus their time and resources to regions of high need. However, this type of data exploration has not been done in the context of AMI's in Manitoba to date, where other reports have been just descriptive in nature. These descriptive reports, such as MCHP's "2013 RHA Indicators Atlas," state whether health district regions have statistically significant rates of disease, though these do not make use of neighborhood relationships. Cluster detection

uses information from neighboring regions to detect clusters of regions with elevated rates. This allows for more robust reporting, as the ecological fallacy and the MAUP are less likely to be as persistent of an issue.

When using areal level data (data that has been aggregated into geographic regions) for studying disease distributions, we are often dealing with count data. Poisson regression models are the natural choice for this type of data, but as previously alluded to, the independent residual assumption that is needed for this class of models is often violated. The dependency on location and "closeness" in geographic space, as well as any extra spatial variation that is not accounted for by the covariates needs to be addressed. To overcome this issue, random-effects models or mixed models are of value to explore.

In the spatial context, we could define this as a latent spatial process which accounts for extra spatial heterogeneity in the response variable that is not accounted for by the covariates we are modelling (Waller & Gotway, 2004, pg. 384). Corrections of this type may be helpful in the analysis and will be described in more detail in Chapter 3.

Using this (quite broad) class of models for modelling different forms of CVD has been done to much success in the past for other geographic contexts. The Besag-York-Mollie (BYM) model first popularized in the paper by Besag et al in 1991 has been used to correctly identify whether there are high posterior probabilities of having elevated risk ratios of CVD in small geographic areas in a number of studies (Besag et al, 1991). Elevated risk ratios of IHD were found in regions that had higher measured NO_2 pollution levels in Barcelona as well as in Madrid Spain using the BYM model (Domınguez-Berjon et al, 2010, pg. 1089; Barcelo et al, 2009, pg. 5521). The BYM model has also been used to discover the association between IHD risk and regions of high levels of alcohol abuse and obesity in Shenzhen China (Du et al, 2016,

pg. 7). These are just a few studies that illustrate the usefulness of models like the BYM in discovering spatial patterns of disease and their geographical relationship to risk factors.

To gain information on any temporal trend in disease risk, an extension of the spatial models; such as the BYM mentioned above; can be constructed for the dimension of time. Here, the data is aggregated not only by region but also by time such as years or months. Additional random effects terms can be implemented in the model to assess the disease's relationship to time as well as any spatio-temporal interactions that may be present. The reason spatio-temporal models are often used is to see whether the spatial patterns that are present vary over time. Or equivalently, if the temporal patterns that have been identified vary over geographical space.

The usage of spatio-temporal models for detecting IHD patterns has been somewhat limited. However, there have been a few studies linking CVD to ambient pollution levels and SES (Yoo et al, 2018; Hart et al, 2015). A spatio-temporal analysis of women revealed that regions that had higher than average proportions of women who were in the upper age group, were obese, or who had been diagnosed with diabetes had higher sensitivity to ambient particulate matter related risk for CVD (Hart et al, 2015). This was especially true if elevated exposure had occurred within the last 12 months (Hart et al, 2015). This further demonstrates the linkage of lifestyle behaviours, SES, and pollution to heart conditions at the space and time aggregated level.

2.3 – Final Thoughts

Overall, the literature appears to be pointing to a combination of lifestyle, genetic, and environmental factors as being significant epidemiological risk factors for IHD. Confounding

and modification of some of these covariates will need to be examined closely as there has been evidence of SES being related to all three of these factors to varying degrees. An example of confounding in this context could be where both indigenous populations and SES are significantly associated with IHD prevalence, and where SES and indigenous populations are also associated with each other. Modification would be present if out of these two covariates only indigenous is associated with IHD prevalence, where SES affects this association but is not directly related to IHD prevalence itself. Not only this consideration, but measurement of these variables may be difficult if not impossible to do directly and may need measures of ethnicity and SES to account for the variation contributed by these variables. Pollution has been a persistent indicator of CVD in the literature, but unfortunately Manitoba has limited data on particulate matter pollution from multiple geographical locations. Thus, surrogate variables; such as urbanicity in place of pollution; for the unmeasured or sparsely measured risk factors of IHD would need to be constructed.

The latent spatial process that is also not accounted for by the covariates and the extra spatial variation not accounted for by the surrogate variables will need to be predicted as well, where there have been many classes of these models such as the BYM implemented in the context of CVD in the past for this purpose. And finally, the temporal effect of IHD and its risk factors will need to be explored as well as lifestyle patterns, environmental effects, and population demographics are dynamic over time. This change in IHD risk factors will most likely change the way IHD is distributed throughout geographical space, and so this is important to account for.

Chapter 3: Methods

3.1 –Hypotheses

The hypotheses (H_a) for this research project are:

1. AMI incidences are clustering in areas typically associated with:
    a. Lower levels of SES than the provincial average
    b. Higher indigenous population proportion than the provincial average
    c. Are considered geographically isolated
    d. Urban settings
2. After accounting for the heterogenetic spatial variation, and controlling for SES confounding; prevalence of IHD is, at the health district level, associated with:
    a. Increased hypertension risk
    b. Indigenous status
    c. Urbanicity
3. After controlling for SES confounding, the spatio-temporal patterns in IHD prevalence are associated with:
    a. The spatio-temporal distribution of hypertension risk
    b. The spatial distribution of the indigenous population
    c. Urbanicity

3.2 – Data Sources, Study Period, and Cohort

This research project is a population-based ecological study using administrative health records housed at the MCHP, with a linkage to Statistics Canada's 2016 census records. MCHP's

PRDR includes four datasets of interest: the Manitoba health insurance registry, medical/physician claims, hospital abstracts (outpatient records), and the drug program information network (DPIN) datasets. These datasets are then linked at the individual level using the scrambled personal health identification numbers (PHIN)s provided by MB health. This allows for confidentiality while also providing rich information on the individual's interactions with the health care system. It is with this linkage that AMI incidences, IHD prevalence rates, and hypertension rates for regions in Manitoba are defined.

The time-period of interest is from the 1995-1996 fiscal year to the 2017-2018 fiscal year (April 1, 1995 – March 31, 2018), which was chosen due to the availability of the data in the PRDR.

The study cohort is comprised of all Manitobans that have had active Manitoba Health Insurance at some point within the study period. This is further restricted to the age groups of 40 to 85 years of age. This age restriction has been chosen due to the fact that the younger age groups have very few incidences of IHD and so will bias the results towards zero, and the older age groups have been ignored due to many incidences of IHD being present as the cause of old-age related mortality. The focus of this research project is to understand the patterns in IHD risk for the population that is at highest risk of premature mortality due to IHD, and so the 40-85 age group has been chosen accordingly as the average life-expectancy at birth for Canadians is 82 years (Statistics Canada, 2019). Manitobans are counted in this cohort only when they are in this age range for each specific year. This means that for the modelling section described later, a person will age out of the study if they exceed this range at any point from 1995-2018. Conversely, new persons are included in the study if they reach the inclusion age range at any point from 1995-2018.

3.3 – Variable Definitions

3.3.1 – Outcome Measures

As data quality is often an issue with administrative records, it is prudent to obtain a valid definition of disease cases from the administrative data. Algorithms to define a disease case from administrative records have already been explored in this context (Lix et al, 2006, 2008). These algorithms have been externally validated with the CCHS from Statistics Canada (Lix et al, 2006, 2008), and have been in use for other reports in Manitoba such as the "2013 RHA Indicator's Atlas" from MCHP (Fransoo et al, 2013). Therefore, implementation of these algorithms is done for identifying our case groups when evaluating the research hypotheses in order to be consistent with other literature in Manitoba.

To evaluate the first hypothesis, identification of persons who have experienced an AMI incidence is based only on the hospital abstracts database (hospital separations). Here they must have an international classification of diseases (ICD)-9 or ICD-10 code for AMI as the diagnosis. For the second and third research questions and their corresponding hypotheses, IHD prevalence is calculated by identifying individuals that have had at least one instance of IHD in the hospital abstracts or at least two instances in the medical claims dataset, or a combination of at least one instance in the medical claims dataset and two prescriptions in the DPIN dataset within a five year period. After these persons have been identified as having IHD they are counted in the numerator for prevalence until censored out for either death, movement to outside Manitoba, or aging out of the study.

ICD-9 and ICD-10 codes are used to identify instances in the hospital abstracts and medical claims datasets, while anatomic therapeutic chemical (ATC) codes are used for the DPIN dataset. A full list of the relevant codes can be found in the appendix.

Geographic classification of individuals is done with the 6-digit postal code and municipality code as defined by Manitoba Health, Seniors & Active Living. From here, individuals will be assigned and aggregated into the 71 regional health authority districts (RHAD) for regions outside Winnipeg, and the 25 neighborhood clusters for within Winnipeg regions. For the remainder of this project the term RHAD is used to define all 96 regions, where these regions are called areal units. These will be the basis for all subsequent analysis for the evaluation of the three research hypotheses.

AMI incidence is calculated as the incidence rate for the time period of 2011 to 2016 for the 96 areal units. IHD prevalence for the second research question is calculated for the most recent years of data available, which is the 5-year period of 2013 to 2017, for the 96 areal units. The third research question further aggregates the IHD prevalence data into years as well as the 96 areal units. Here, 18 years are explored from 1998 to 2015. Note that each of these 18 yearly aggregates use 5-year time frames as defined earlier, i.e. the 1998 year is the prevalence of IHD for 1996-2000, 1999 is the IHD prevalence for 1997-2001, up to the 2015 year which is the IHD prevalence for 2013-2017. Here a person is counted in the numerator for prevalence starting at the mid-year for the 5-year period for which the case criteria is satisfied, until either death, movement out of Manitoba, aging out of the study, or the study end date. The denominator, or population for the study is the number of Manitobans with active health insurance (aged 40-85) counted uniquely for each year in the study.

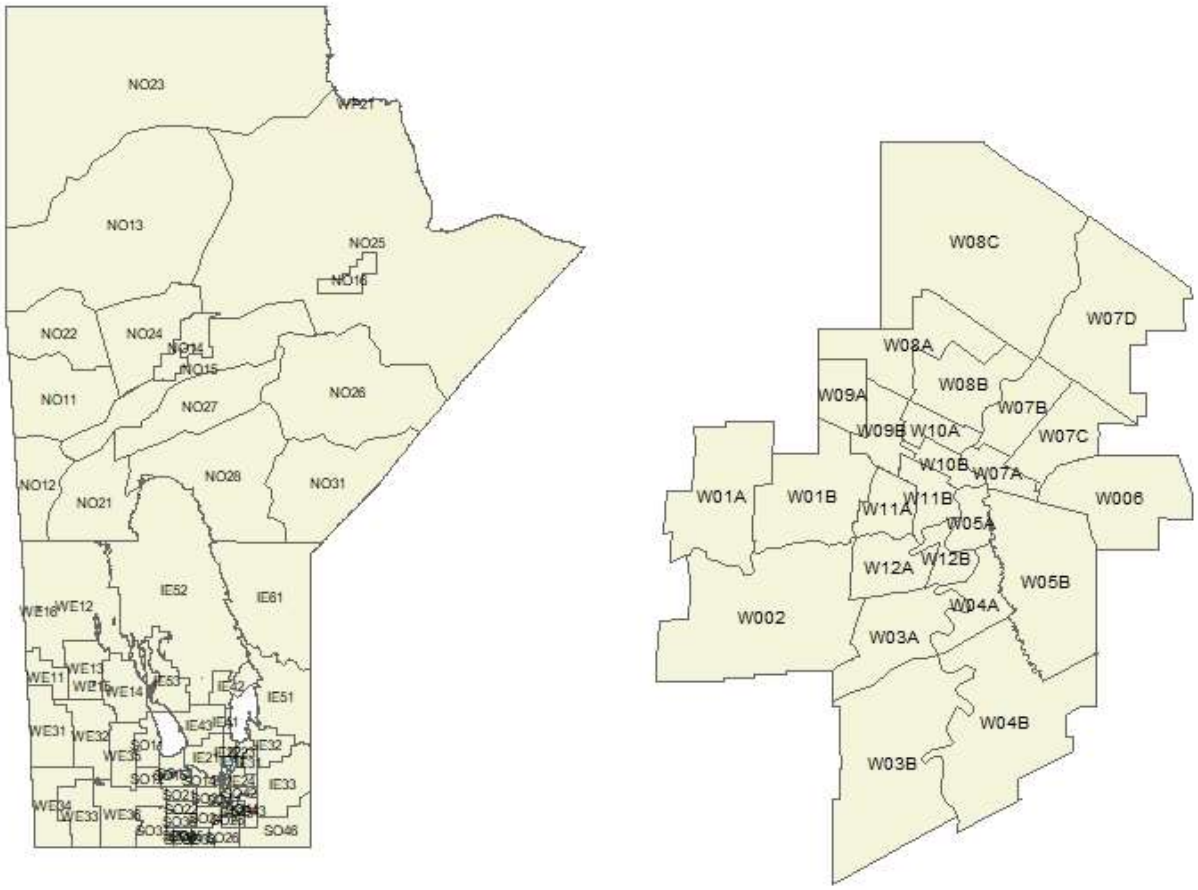

Figure 3.1: RHADs and Winnipeg Neighborhood Clusters. (Region codes given in appendix).

3.3.2 – Explanatory Variables

The first research hypothesis is evaluated by comparing regional levels of AMI incidences with regional SES measures, indigenous population proportions, and urban/remote location definitions. The second and third research questions will also need information on SES, indigenous population, and urban locations; and will additionally need information on hypertension. The 2016 Census from Statistics Canada will be used to define variables for SES

measures and indigenous populations, whereas hypertension will be calculated from the PRDR housed at the MCHP.

Hypertension has been one of the leading risk factors of IHD, and is a major comorbidity (Weber et al, 2016, pg. 468; Berger et al, 2010, pg. 879; Spinar, 2012, pg. 434). For this reason, it is of interest to use hypertension; as it is also a highly prevalent chronic condition in Manitoba (MB Health, 2017, pg. 24, 28; Fransoo et al., 2013, pg. 69, 93, 109); as a risk factor for IHD prevalence at the population level. To define hypertension the four databases previously mentioned within the PRDR will be used. Individuals with hypertension are defined as persons showing up at least once in the hospital abstracts, at least once in the medical claims dataset, or at least twice in the DPIN dataset with a hypertension related diagnosis or prescription within a one-year time period. The ICD-9/10 codes, and the ATCs used to identify these diagnoses and prescriptions are given in the appendix. This hypertension prevalence rate is then indirectly age and sex standardized using the overall provincial population demographics into a relative risk for each of the 96 areal units. For the second research question hypertension prevalence is calculated for the 2015 year only, and for the third question it is evaluated by calculating the prevalence rates for each year from 1998 to 2015. Also note that if a person was counted as prevalent in a previous year, they are now counted as prevalent for all subsequent years until excluded from the study.

The next indicator that will be used in the evaluation of all three research questions is that of indigenous status; taken to mean any individual identifying as First Nations, Inuit, or Metis. As Manitoba's population consists of 18% indigenous persons (Statistics Canada, 2016), and it has been shown that the indigenous population suffer from worse health outcomes regarding CVD than the other population groups (Tobe et al, 2015, pg. 1124), understanding how

indigenous status affects IHD risk is important in the Manitoban context. If this population group is at higher risk, this could point to an area that our health care system could address. The creation of an indigenous population variable is done via the 2016 Census, and simply counts the number of indigenous persons in a census DA and divides by the total population to create an indigenous population proportion variable. Only the 2016 census was used to define this covariate as the demographic and ethnic make-up of Manitoba has remained mostly stable throughout time. Census data is also only available in 5-year cycles, so a complete yearly data source isn't available for usage as a covariate, and thus estimation would need to be employed. Also, data collection, compilation, and release methods and protocols change census-to-census, so differences in aggregated findings for census covariates may be due to this factor. To avoid these issues and provide consistent values for census covariates, only the 2016 census was utilized.

For the first research question, a geographical isolation variable needs to be defined. This variable needs to measure geography as a barrier to healthcare access, as it is hypothesized that isolated communities in Manitoba have higher incidences of AMIs due to chronic conditions not being diagnosed and treated before acute events occur when living in relative isolation. For this, MCHP's definition of remote communities is used as it provides a literature backed measure for healthcare access barriers due to geography. In this definition, remote communities are assigned if they have no permanent road access, are a four or more-hour drive from the nearest major hospital, or have rail or fly-in access only. Then, if most communities within a district are designated as remote, the entire district is assigned as remote. A list of these regions is found in the appendix.

As previously studies have identified a significant link between air pollution and CVD risk (Yoo et al, 2018; Pun et al, 2014; Domınguez-Berjon et al, 2010; Barcelo et al, 2009), a measure of pollution as an indicator of IHD risk is used to assess whether pollution is contributing to IHD risk in Manitoba. Since pollution is only measured at few weather stations within Manitoba, a surrogate variable that identifies urban regions will be used to assess the contribution. This is done because of the implicit circumstance that cities generally have higher ambient air pollution levels than rural regions due to higher densities of road vehicles, air and rail transportation, factories, etc. In Manitoba there are only a handful of densely populated cities; the obvious one being Winnipeg, with Brandon being a second. Thus, this indicator variable will code health districts within Winnipeg or Brandon as being urban to address the research questions.

As SES has been shown to be associated with both hypertension and indigenous, as well as being a significant indicator of chronic diseases and premature mortality (Public Health Agency of Canada, 2018, pg. 381, 404; The Health Officers Council of British Columbia, 2013; Public Health Agency of Canada, 2018, pg. 60), it is expected that confounding between these variables is present and significant. Thus, a SES measure is constructed from the 2016 Census, and used to control for this expected confounding. The socio-economic factor index (SEFI) is an index defined by researchers at MCHP that bridges socio-economic factors that are highly influential on potential years of life lost, premature mortality rates, life expectancy, and self rated health (Chateau et al, 2012). The second iteration of this algorithm is known as SEFI-2 and contains information on the four census variables of median household income, proportion of high school graduates, unemployment rate, and proportion of single-parent families (Chateau et al, 2012). This factor index is the standardized factor scores from a factor analysis of these four

variables. This in theory creates a single variable which explains a large amount of the common variation between these 4 variables and allows for information from multiple types of SES variables to be incorporated into a single measure. At the DA level (2183 DA's exist in Manitoba), the amount of common variation between these variables is 52.2%, which means that over half of the variation across these variables is explained by the SEFI-2 factor scores. This index variable measures socio-economic deprivation, where higher SEFI-2 values for a region indicates that this region has more socio-economic deprivation compared with regions with lower SEFI-2 scores. Said another way, higher SEFI-2 scores indicate lower levels of SES.

For both census variables; indigenous population proportion and SEFI-2; a link between the census data which is organized into 2183 census DA's and PRDR data which is organized by postal code and then aggregated into the 96 health districts, needs to be constructed. Since the health districts are defined using postal and municipal codes, and the census uses census blocks and DA's, a one-to-one link unfortunately does not exist. There have been attempts at using different methods to link these data sets in the past, such as the extensive use of the postal code conversion file (PCCF), but for this research project a stochastic method for linking these datasets is used and is defined in detail in section 3.4.4.

3.4 – Statistical Analysis

The outcome for all subsequent statistical models is expressed in terms of either a standardized incidence ratio (SIR) for the AMI cluster detection or a standardized risk ratio (SRR) for the IHD Poisson models. Standardization is done to factor out the effects of age and sex in the analyses, as these are not of primary interest for the research project, but we know that

they do contribute to IHD risk (Ashley & Geraci, 2013, pg. 430; Berger et al, 2010, pg. 879; Vaccarino, 1998, pg. 2059). These are calculated using indirect standardization to the overall Manitoba provincial population. This is implemented by calculating E_i; the expected number of individuals in the i^{th} areal unit with the disease of interest; which is done using (1).

$$E_i = \sum_j \frac{Y_j^{(s)}}{n_j^{(s)}} n_{ij} \qquad (1)$$

where i=1, 2, …, 96 is the index for the areal units, $Y_j^{(s)}$ and $n_j^{(s)}$ are the disease counts and population totals in the standard population for the j^{th} age-sex group, and n_{ij} is the population total for the j^{th} age-sex group in the i^{th} areal unit. Age groups are defined as 40-44, 45-49, …, 80-84 where these are then broken into the two sex categories to make up the j levels (for each year in the spatio-temporal case). E_i is then used as the model offset in the Poisson models rather than using a population count which is often done in the simple case. For AMI incidence the E_i term is calculated with $Y_j^{(s)}$ being the number of new cases of AMI's in the standard population, and for IHD prevalence the $Y_j^{(s)}$ is the number of existing cases of IHD in the standard population. Then the SIR for AMI incidence is given as $SIR_i = \frac{new\ cases\ of\ AMI_i}{E_i}$, and the SRR for IHD prevalence is given as $SRR_i = \frac{Number\ of\ persons\ with\ IHD_i}{E_i}$ for the i^{th} areal unit. Note that AMIs could be repeat cases, i.e. a person could be counted multiple times in the numerator.

For research questions 2 and 3, sex stratification is also employed where 3 separate models are used for each question; an overall model and a model for each sex. Age standardization is still completed for the sex stratified models by modifying the j groups to only break the population into age groups rather than age-sex groups. This stratification is done to gain any insight available into the differences in IHD outcomes across the sexes.

The remainder of this section is broken up into four parts. The first describes the method of cluster detection using the flexible spatial scanner (FSS). This method is applied for the first research question, where AMI clusters are calculated to potentially identify geographic clusters with higher AMI incidences compared to the rest of the province. We can then visually compare the clusters of AMI incidences with the distribution of the SEFI-2 variable, the distribution of the indigenous population, and the geographically isolated and urban communities. The FSS is also employed to detect clusters of IHD prevalence as an exploratory measure to see which regions have significantly elevated prevalences for research questions 2 and 3. The second and third parts of this section describe the models used for both the spatial analysis for research question 2 and the spatio-temporal analysis for research question 3. The final part of this section is dedicated to describing the methods used to overcome the misaligned data problem that occurred between the census and the PRDR data.

3.4.1 – Cluster Detection

Cluster detection using spatial scanners is a method used to identify "hot-spots" or geographic clusters of significantly elevated disease incidences or prevalences compared to the rest of the geographic area in the analysis. Spatial scanners were first proposed by Kulldorff & Nagarwalla in 1995 in its original form of the circular scan statistic (Kulldorff & Nagarwalla, 1995), and have since been modified in 2005 to the FSS which allows for varying shapes of clusters to be detected (Tango & Takahashi, 2005). The underlying idea behind either spatial scanner, is that of measuring which set of neighboring regions contribute to a statistically elevated disease rate as compared to all other regions. This is done by testing the hypothesis in (2).

$$H_0: E(Y_i) = \lambda_i \text{ , } H_a: E(Y_i) > \lambda_i, \quad (2)$$

$$\text{where } Y_i \sim \text{Poisson}(\lambda_i)$$

Where Y_i and λ_i are the disease count and expected disease count respectively in the i^{th} areal unit. With the circular spatial scanner, this hypothesis can be assessed directly, by using either a radius drawn from the centre of each i^{th} region up to a specified distance and using all regions who fall within that radius, or using a nearest neighbor approach and using up to a predefined number of nearest neighbors for each i^{th} region. Each combination of regions is then tested in hypothesis (2) to come up with the most probable disease cluster as well as any non-overlapping secondary disease clusters that are also statistically significant.

The FSS is a bit more robust, as it can take on any other shape in addition to the concentric circles or nearest neighbor method of the circular spatial scanner. This method is preferable in the Manitoba context as both distances between centroids and the number of neighbors for regions is extremely variable and hence could miss otherwise obvious clusters. In the flexible scanner, an irregularly shaped window Z is fit with the neighbors of the i^{th} areal unit, up to a specified maximum number of neighbors. The hypothesis for the flexible scanner changes slightly now to include the window around the i^{th} region as in (3).

$$H_0: E(Y(\mathbf{Z}_{ik})) = \lambda(\mathbf{Z}_{ik}), H_a: E(Y(\mathbf{Z}_{ik})) > \lambda(\mathbf{Z}_{ik}), \quad (3)$$

$$\text{where } Y(\mathbf{Z}_{ik}) \sim \text{Poisson}(\lambda(\mathbf{Z}_{ik}))$$

Where $\mathbf{Z}_{ik}$ is the k^{th} irregularly shaped window around the i^{th} region. The test statistic is constructed using a likelihood ratio and given in (4) where $\mathbf{Z}^0$ is the set of all regions not in $\mathbf{Z}$, $y(\mathbf{Z})$ is the observed number of cases within window $\mathbf{Z}$, and $\mathbf{I}()$ is the indicator function. Here, the model offset is E_i as defined in (1).

$$\sup_{Z \epsilon Z} \left(\frac{y(Z)}{\lambda(Z)}\right)^{y(Z)} \left(\frac{y(Z^0)}{\lambda(Z^0)}\right)^{y(Z^0)} I\left(\left(\frac{y(Z)}{\lambda(Z)}\right) > \left(\frac{y(Z^0)}{\lambda(Z^0)}\right)\right) \quad (4)$$

To find the distribution of (4) under the null hypothesis, Monte Carlo sampling is done until convergence. This method provides the most likely cluster as defined by (4) which is called the primary cluster, as well as any secondary non-overlapping clusters that are also statistically significant. Note that the expected number of cases λ within each window Z is the age and sex standardized expected count of AMI incidences, indirectly standardized from the overall Manitoba population as in (1).

For research question 1, the primary cluster of AMI incidences and the secondary significant clusters for the time frame of 2011-2015 are visually compared to SEFI-2 and indigenous population proportion values in the same regions. Clusters are also assessed to be in remote or urban locations to finish evaluating the hypotheses for this research question. For the second research question, clusters of IHD prevalence for the overall population as well as the sex stratified populations are calculated for the 2013-2017 time period as an exploratory approach to the IHD data. These clusters are then visually compared to the smoothed risk ratios given by the spatial models described in section 3.4.2.

3.4.2 – Poisson and Spatial Poisson Regression Models

Poisson regression models are one of the most common disease models, as it easily relates disease counts for specified regions to regional level covariates. This is specified properly under a few assumptions: the first is that the disease of interest has relatively few counts for each observed areal unit compared with a relatively large population count. Second, when using the

Poisson distribution for modelling we must assume that its mean is equal to its variance. This is fairly restrictive as the variance is often greater than the mean where this problem is called overdispersion. Finally, the assumptions from linear models such as linear relationships between covariates and the response, zero-mean and independent normally distributed residuals, and no perfect multicollinearity between the covariates also need to be met. Some of these assumptions may be loosened if needed by the incorporation of spatially dependent random effects.

For a standard log-linear Poisson regression model, we assume that the disease counts in each region follow a Poisson distribution as $Y_i \sim \text{Poisson}(\lambda_i)$, where $E(Y_i) = \text{var}(Y_i) = \lambda_i$. Then:

$$\log(\lambda_i) = \log(E_i) + \boldsymbol{X_i\beta} \qquad (5)$$

Note here that $E(Y_i \mid \mathbf{x_i}) = \lambda_i = E_i * \exp\{\mathbf{X_i\beta}\}$, where $\mathbf{X_i\beta}$ is the matrix containing the covariate matrix ($\mathbf{X_i}$) and the vector of their model coefficients ($\boldsymbol{\beta}$), and E_i is the age- sex-standardized expected disease count defined by (1) where here it is also called the model off-set. Rearranging this equation, we get that:

$$\log(\lambda_i) - \log(E_i) = \boldsymbol{X_i\beta}$$

$$\log\left(\frac{\lambda_i}{E_i}\right) = \boldsymbol{X_i\beta}$$

$$\frac{\lambda_i}{E_i} = \exp\{\boldsymbol{X_i\beta}\} \quad (6)$$

where $\left(\frac{\lambda_i}{E_i}\right)$ is the SRR for the i^{th} areal unit.

This model specification is the first model tested. However, this specification has some issues. The most important one being the distribution of the residuals appear to be spatially heteroscedastic and correlated (see figures 4.18 and 4.19). This is often the case with spatially

referenced data, as it is expected that close regions carry some of the same variation and be more correlated than regions that are further apart. The spatial auto-correlation that is present in this model will also cause overdispersion which violates the assumption of the Poisson distribution. To overcome this issue, we can turn to adding spatial random-effects to the Poisson regression model.

As previously mentioned, a popular modification to the Poisson regression model specification for dealing with spatially dependent data is the BYM model proposed by Besag et al. This modification includes two additive random effects terms to the Poisson regression model; one spatially correlated heterogeneity term (U_i), and the other a non-correlated heterogeneity term (V_i) as given in (7).

$$Y_i \sim Poisson(\lambda_i)$$

$$\log(\lambda_i) = \log(E_i) + \boldsymbol{X_i\beta} + V_i + U_i \qquad (7)$$

$$SRR_i = \exp\{\boldsymbol{X_i\beta} + V_i + U_i\}$$

where:

$$V_i \sim N(0, \sigma_v^2)$$

$$U_i \mid n_{\delta_i}, U_j; j\epsilon\delta_i \sim N(\overline{U}_i, \sigma_{u,i}^2)$$

Here, $\overline{U}_i = \frac{1}{n_{\delta_i}} \sum_{j\epsilon\delta_i} U_j$ where δ_i is the neighborhood of the ith region, n_{δ_i} is the number of neighbors for the ith areal unit, and $\sigma_{u,i}^2 = var(U_i \mid n_{\delta_i}, U_j; j\epsilon\delta_i)$. The neighborhood of the ith areal unit δ_i, is defined as all areal units who share a border with the ith areal unit, and hence this model is in the class of Gaussian-Markov-random-field (GMRF) models. This model is also

widely known as the intrinsic or improper conditionally autoregressive (ICAR) spatial model, as the joint distribution is a singular normal distribution, and where the conditional distribution of U_i given its neighbors is also normal. Because of the singularity, Bayesian inference is carried out using the integrated nested Laplace approximation (INLA) approach (Rue & Martino, 2007). This method for inference was chosen due to the extreme complexity of these models. A more popular method for Bayesian inference is that of Markov chain Monte Carlo (MCMC) sampling. However, for these models, inference is very time consuming and cumbersome where WinBUGS in particular has issues with handling space-time interactions.

This model specification is made possible by the proof of the *Hammersley-Clifford theorem*, which allows under certain conditions a joint distribution to be approximated by a set of conditional distributions (Waller & Gotway, 2004, pg. 371). This proof was first shown by Besag (Besag, 1974), and allows us to look at the joint distribution of the 96 areal units as a set of 96 conditional distributions; conditional on the neighborhoods of each areal unit. Here, the U_i term is essentially weighted by its neighbors which are defined as areal units who share a border and allows for simultaneous smoothing of the SRR between areal units and their neighbors. This smoothing controls for the spatial autocorrelation between areal units which reduces the effect of the MAUP. The MAUP occurs when attempting to make inferences about a population at a different aggregation level than what is provided. For instance, making inferences at the provincial level for IHD from one of the 96 areal units, or vice versa. Generally speaking, the larger the aggregation level the smaller the variance (as the population n increases, the variance decreases). However, by smoothing the SRR through the use of the ICAR model we account for the spatial covariance between areal units, and thus reduce the overall variance in the SRR estimates. This reduction in variance produces more stable estimates of the SRR for smaller areal

units, or at different aggregation levels, or with differing methods for forming the boundaries between areal units. This smoothing also reduces the affect of the ecological fallacy bias, as individuals within each areal unit are better represented by the aggregated and smoothed SRR. This doesn't mean that it isn't a relevant consideration when making inferences about an individual, only that these inferences will be better as the smoothed rates will have less internal variation than non-smoothed rates do. Thus, we can be more confident that the smoothed SRR's produced by the ICAR model represent the real risk levels within each areal unit regardless of aggregation method or population size.

INLA works for data that is distributed via a Normal or Gaussian distribution, and where the set of parameters are also distributed with a multivariate normal distribution (Blangiardo & Cameletti, 2015, pg 107-112). The components (model parameters) of this latent Gaussian field are also assumed to be conditionally independent, consequently leading to a GMRF. If these conditions are satisfied, the INLA algorithm can be used for Bayesian inference of the model parameters instead of the MCMC sampler. This algorithm produces results in a much shorter time frame than the MCMC approach, which is of interest when modelling with many different specifications and/or covariate combinations. The Laplace approximation is said to have negligible error when the assumptions are satisfied and can be checked via cross-validation. This is done by assessing the conditional predictive ordinate and the probability integral transform (Blangiardo & Cameletti, 2015). Using the R-INLA package in R, spatial and spatio-temporal models are relatively quick to compute and easy to implement, and thus are done so for this research project. Here, priors are selected as uninformatively as possible so as to not bias the results with prior beliefs of the data, and the deviance information criterion (DIC) is used to assess model fit.

3.4.3 – Spatio-Temporal Poisson Regression Models

An extension of the spatial Poisson regression models as described in section 3.4.2 is the class of spatio-temporal Poisson regression models. In this case, the data is aggregated by both space and time, as it is beneficial to explore whether the structure of the data is correlated in both the time and space dimensions. As before, the spatially-correlated heterogeneity term and the spatially-non-correlated heterogeneity terms are added to the Poisson regression model. Now two temporally varying random effects terms are also added. Similar to the spatially varying random effects terms, one of the temporal random effects terms models the temporally-correlated heterogeneity and the other models the non-correlated heterogeneity in the temporal dimension. This specification is given in (8).

$$Y_{it} \sim Poisson(\lambda_{it})$$

$$\log(\lambda_{it}) = \log(E_{it}) + \boldsymbol{X_{it}\beta} + V_i + U_i + \phi_t + \gamma_t \quad (8)$$

Notice in (8) that the index is now consisting of both i and t terms. As before i denotes the i^{th} areal unit; i=1,2,…,96; and t now denotes time in years for t=1998,1999,…,2015. Here ϕ_t is the temporally unstructured random effects term and γ_t is the temporally structured random effects term. These are specified using (9) and (10) respectively.

$$\phi_t \sim N(0, \sigma_\phi^2) \quad (9)$$

$$\gamma_t \mid \gamma_{t-1}, \gamma_{t-2} \sim N(2\gamma_{t-1} + \gamma_{t-2}, \sigma_\gamma^2) \qquad (10)$$

The γ_t term here is also known as a random walk term of second order. What this imposes is a correlation structure where the errors are now assumed to depend on both the regions adjacent to the i^{th} region as well as depend on the previous two years of data. This allows for measuring the

change in time and how one year influences the value of the next. The second order random walk was chosen as it was a better fit for the data than the first order random walk indicating that the temporal covariance structure relies on both the previous year as well as two years previous.

Another type of change that can (and should) be accounted for if present, is that of space-time interactions. This occurs when we assume that the temporal trends vary over space, when the spatial trends vary over time, or both. To account for this, we add another random effects-term to the model which is now measured over both i and t, as given in (11) with the SRR given in (12).

$$\log(\lambda_{it}) = \log(E_{it}) + \boldsymbol{X_{it}\beta} + V_{\boldsymbol{i}} + U_{\boldsymbol{i}} + \phi_t + \gamma_t + \delta_{it} \quad (11)$$

$$SRR_{it} = \exp\{\boldsymbol{X_{it}\beta} + V_{\boldsymbol{i}} + U_{\boldsymbol{i}} + \phi_t + \gamma_t + \delta_{it}\} \quad (12)$$

There are four different types of interactions that can occur as the space-time interaction effect δ_{it}, these are given in table 3.1.

Interaction	Interacting Parameters	Description
I	V_i and ϕ_t	Unstructured interaction.
II	V_i and γ_t	Spatially unstructured, temporally structured interaction.
III	U_i and ϕ_t	Spatially structured, temporally unstructured interaction.
IV	U_i and γ_t	Both spatially and temporally structured interaction.

Table 3.1: Space-time interaction types.

The structure matrices for each interaction type are defined by the Kroneker product of the spatial and temporal structure matrices of the parameters interacting. This implies that there is no structure for interaction type I as the matrix structures for both of these parameters is the

identity matrix, and $\boldsymbol{I} \otimes \boldsymbol{I} = \boldsymbol{I}$. For interaction II the structure comes from the temporal effect, for III it comes from the spatial effect, and in IV it is the product of both the space and time effects. As there is indication that we have both spatially varying temporal effects and temporally varying spatial effects in our IHD data, all four interaction types are assessed for model fit and specification. Inference for these models is, as before, carried out with a Bayesian approach using INLA.

3.4.4 – Misaligned Data Problem

An issue that often appears when using census data in Canada is the misaligned data problem. This occurs because the census is conducted at a census block level at its lowest aggregation level, and healthcare data is most often collected at the postal code level at the lowest level of aggregation. Here, census blocks and postal codes more often than not, do not share borders and so inferring about census rates within postal codes is not readily done as no one-to-one conversion between these data types exist.

Statistics Canada has developed an algorithm in the form of the PCCF, that does attempt a conversion between different census levels, such as the census DA or dissemination block (DB) and the 6-digit postal code which is provided and maintained by the Canada Post Corporation. However, this is a deterministic function and only links about 94% of all postal codes (Statistics Canada, 2015). Hence in places where postal codes and DA's cover a large landmass, such as in Northern Manitoba, the accuracy of the census variables should be called into question.

As mentioned, the aggregation level that this research project is interested in, is the 96 RHADs and neighborhood clusters, which are formed by aggregating postal and municipal codes together. This is chosen due to the organization structures of Manitoba's health system being more-or-less aligned to these areal units. The lowest level of aggregation that is publicly available from Statistics Canada for the census is that of the census DA. In Manitoba there are 2183 census DA's where the majority of these represent 1000 persons or less.

Since the misalignment occurs at a lower level of aggregation, these 2183 regions are also misaligned with the 96 areal units. A map of this misalignment is depicted in figure 3.2. As can be seen, there are a large number of misaligned regions between these two datasets which needs to be overcome if we want to use the census data accurately.

The purpose of this section is to describe a stochastically defined method for linking these two datasets. A few examples of previous work in this area has been done by Mugglin et al and Agarwal et al (Mugglin, Carlin, & Gelfand, 2000; Agarwal, Gelfand, & Silander, 2002). However, these methods are more specific to the data that was used to motivate these methods. The modelling approach outlined in this section was developed specifically for relating Canadian census data to the 96 areal units of interest in Manitoba.

This method will enable more precise measures of the census covariates for the 96 areal units, while also being able to define a measure of error for each of the predicted values. The census variables that are of interest for this project are the four that make up the SEFI-2 factor scores; proportion of lone parent families, unemployment rate, high school education rate, and median household income; and the indigenous population count for each region. For simplicity, the SEFI-2 standardized factor scores are calculated for the 2183 regions prior to conversion to

the 96 administrative regions. This saves some time as we then only need to convert two variables, SEFI-2 and indigenous population count, rather than all five.

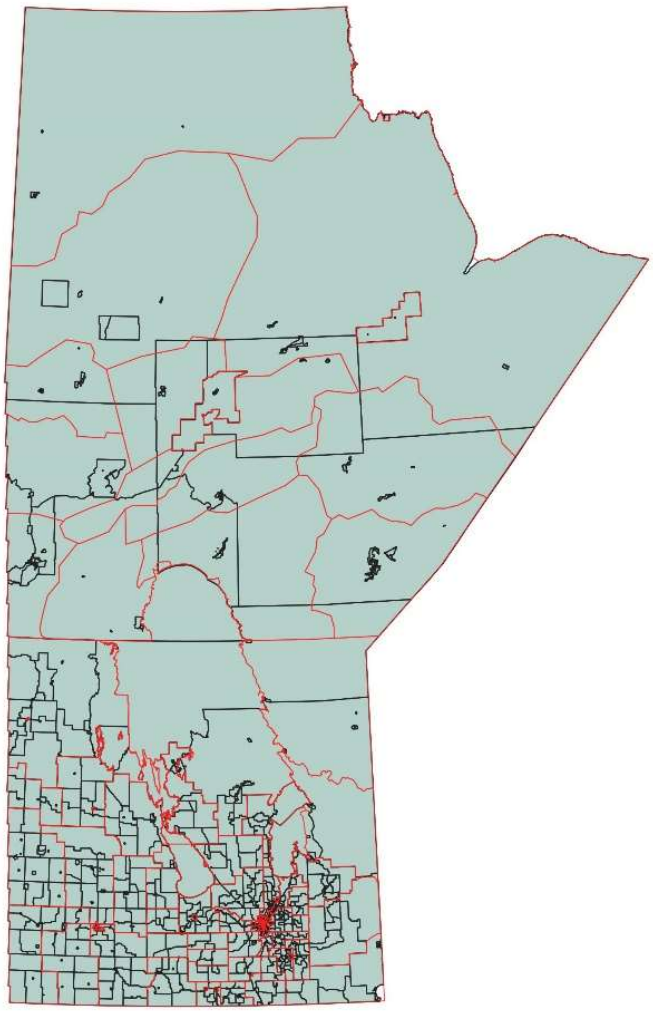

Figure 3. 2: RHADs (red borders), superimposed onto census DAs (black borders).

A naïve approach to this problem would be to simply take a weighted average of the regions, weighted by either the region's area or population size. This however, can be misleading as both the indigenous population composition and the SEFI-2 variable are continuous processes that are "binned" into areal units which effectively discretizes each process. Because of this, a spatial covariance structure is more than likely to occur between neighboring regions. Now since the indigenous population variable is a count variable and the SEFI-2 variable is continuous, the succeeding methods are slightly different. Thus, a step-by-step method for the count variable will be defined first and the modifications for the continuous variable will be provided afterwards.

Method for Count Variable

The first thing that must be done is to define a one-to-one function linking the two aggregation methods. This is done by intersecting the two aggregation methods and defining each region that is created with the intersection as the atomic areal units. When this is done, 3195 atomic units are created. If we let A_i be the i^{th} RHAD (i=1,2,…,96), and B_j as the j^{th} census DA (j=1,2,…,2183), then we can define the intersecting regions from either of these perspectives. Hence there exists an A_{ik} and a B_{jl} for which $A_{ik}=B_{jl}$. Here A_{ik} represents the k^{th} atom within the i^{th} region created from the intersection, and B_{jl} represents the l^{th} atom within the j^{th} DA created from the intersection. Hence, we have coerced the datasets to have a one-to-one conversion.

However, the problem remains; we do not know the value of our count variable for the jl^{th} atom, we rather only know the value for the j^{th} region and so it needs to be estimated. If we let X_j be the indigenous population count for the j^{th} DA, X_{jl} the count for the jl^{th} atom, P_j the population of the j^{th} DA, and P_{jl} the population for the jl^{th} atom, then we can begin estimating as follows:

$$\widehat{X_{jl}} = X_j \frac{|B_{jl}|}{|B_j|} \quad (13)$$

$$\widehat{P_{jl}} = P_j \frac{|B_{jl}|}{|B_j|}$$

Where |C| denotes the area of region |C|. This provides an initial estimate for atomic regions. However, area size as a weight for population counts is often misleading as there are vast expanses of space in parts of Manitoba where there are relatively few or no persons living. This provides the basis for the problem of a deterministic function to link these regions; the covariance between regions will contain some spatial structure that should be accounted for. To overcome this, we will fit the estimated counts from (13) with the previously mentioned BYM model. This will impose a GMRF structure onto the data and smooth the counts by accounting for the neighboring values. Not only this, but it also provides a measure of error for each of the 3195 atoms which can be analyzed to assess if there are atoms with large errors. The full model is given in (14).

$$\widehat{X_{jl}} \sim Poisson(\lambda_{jl})$$

$$\log(\lambda_{jl}) = \log(\widehat{P_{jl}}) + U_{jl} + V_{jl} \quad (14)$$

Here, U_{jl} and V_{jl} are defined as before in the spatial ICAR model of (7). We then collect the fitted values of the model ($\widehat{\lambda_{jl}}$) and convert them using the one-to-one function defined above as $\widehat{\lambda_{jl}} = \widehat{\lambda_{\iota k}}$. The final step is to sum over k to get the count for the i^{th} areal unit as $\widehat{\lambda_\iota} = \sum_k \widehat{\lambda_{\iota k}}$ and then divide by the total population in the i^{th} areal unit to obtain the estimated indigenous population proportion covariate that we need for modelling IHD prevalence.

Continuous Modifications for SEFI-2

Inserting a continuous variable into the methods mentioned above requires a bit of a modification. First, we do not need to weight a continuous variable by area size, as we can initially estimate the value of X_{jl} as X_j for each l^{th} atomic unit. The model in (14) also changes, as now we have a normally distributed continuous variable, rather than a Poisson distributed count variable. If we denote the estimated value for the jl^{th} atom as $\widehat{\mu_{jl}}$ then:

$$\widehat{\mu_{jl}} = U_{jl} + V_{jl} \quad (15)$$

Note that the values for U_{jl} and V_{jl} are not the same in (14) as they are in (15), but are still defined by the ICAR specification and spatially unstructured specification respectively. As before we can convert the fitted values to the ik method of indexing the intersected dataset by simply equating them with the one-to-one function. The main difference here is now instead of simply summing over k, we must take a weighted average of all k intersections, weighted by population size. This is given in (16).

$$\widehat{\mu_i} = \frac{1}{P_i}\sum_k P_{ik}\,\widehat{\mu_{ik}} \quad (16)$$

By using these methods, an estimated value for each variable defined by census data has been obtained, where the estimates have been corrected for their spatial covariance structures. This provides a more accurate measure, while also providing a measure of error for which we can ensure that we retain precise variable values. For this study, all subsequent models using indigenous population proportion and/or the SEFI-2 covariates, use the corrected versions of these variables as defined in this section.

Chapter 4: Statistical Modelling & Results

This chapter begins with some main cohort characteristics and how they relate to the outcomes of interest. Crude and standardized AMI incidences and IHD prevalences are shown by RHAD, and in the case of IHD prevalence also by year and sex. Characteristics of the study covariates are also shown by region for posterity. Following this, the research questions are addressed and evaluated with their corresponding hypotheses directly. In the case of the modelling at the latter half of this chapter, step-by-step model building and assumption checking is done where appropriate.

4.1 – Summary Statistics

Over the time frame of 2011-2016, there were measured to be 13313 AMI events for the age group 40-85, or on average 2219 per year. This translates to a crude incidence rate of 3.79 per 1000 persons per year, and an age- sex-standardized incidence rate of 3.82 per 1000 persons per year. This ranges from 1.95 per 1000 persons in St. Boniface E (Winnipeg Health Region), to 9.87 per 1000 persons in N Seven Regions (Southern Health Region). Figure 4.1 shows the crude incidence rates of AMI for all 96 RHADs.

The overall crude prevalence of IHD for each year and stratified by sex is given in figure 4.2. It can be seen that the general trend is a decrease in prevalence over time (1998-2015), and that females generally have lower prevalence than males. Figure 4.3 shows the 2015 crude prevalence of IHD broken down into the 5 Regional Health Authorities (RHA)'s and stratified by sex. These plots show that there is reason to believe that there are both spatial and temporal trends in the IHD data, which will need to be explored further.

Figure 4.1: 2011-2016 crude AMI incidence rate per 1000 persons per year, for ages 40-85, by RHAD. Vertical line represents provincial rate of 3.79 per 1000 persons per year.

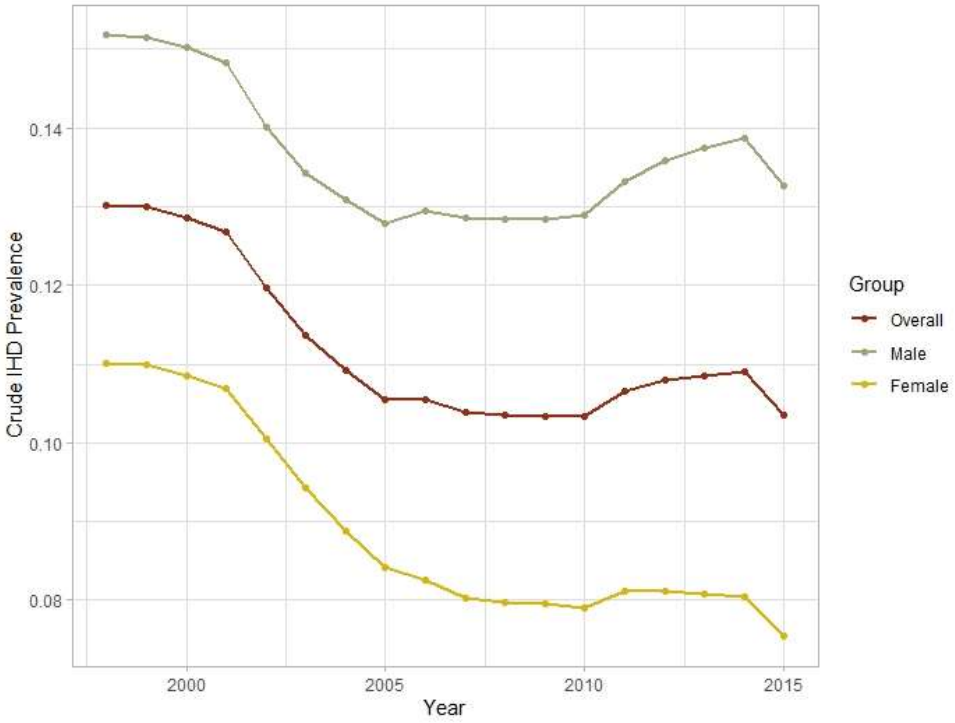

Figure 4. 2: Crude provincial IHD prevalence (proportion) by sex over time (1998-2015).

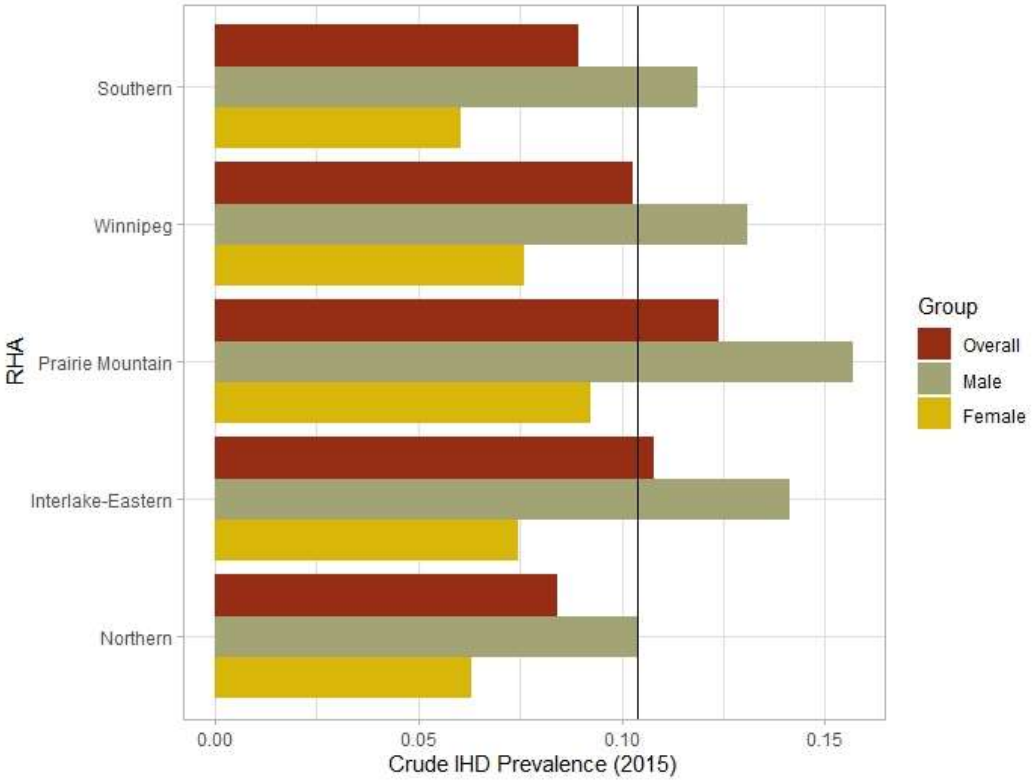

Figure 4. 3: 2015 crude IHD prevalence (proportion) by RHA

The remaining figures in this section (4.4 – 4.6) show regional rates of the study covariates. Figure 4.4 shows the SEFI-2 values, 4.5 shows the indigenous population proportion levels, and 4.6 shows the standardized hypertension prevalence risk ratio (RR). Regional variation is present, which is expected. Note that figure 4.6 showing the standardized hypertension prevalence RR looks to be more uniform than the other covariate figures. This is most likely due to the fact that hypertension is a very prevalent condition in Manitoba regardless of place. Therefore, it may be a covariate that does not carry as much of the spatial variation in IHD prevalence as some of the other covariates. Table 4.1 gives some general statistics regarding these covariates as well.

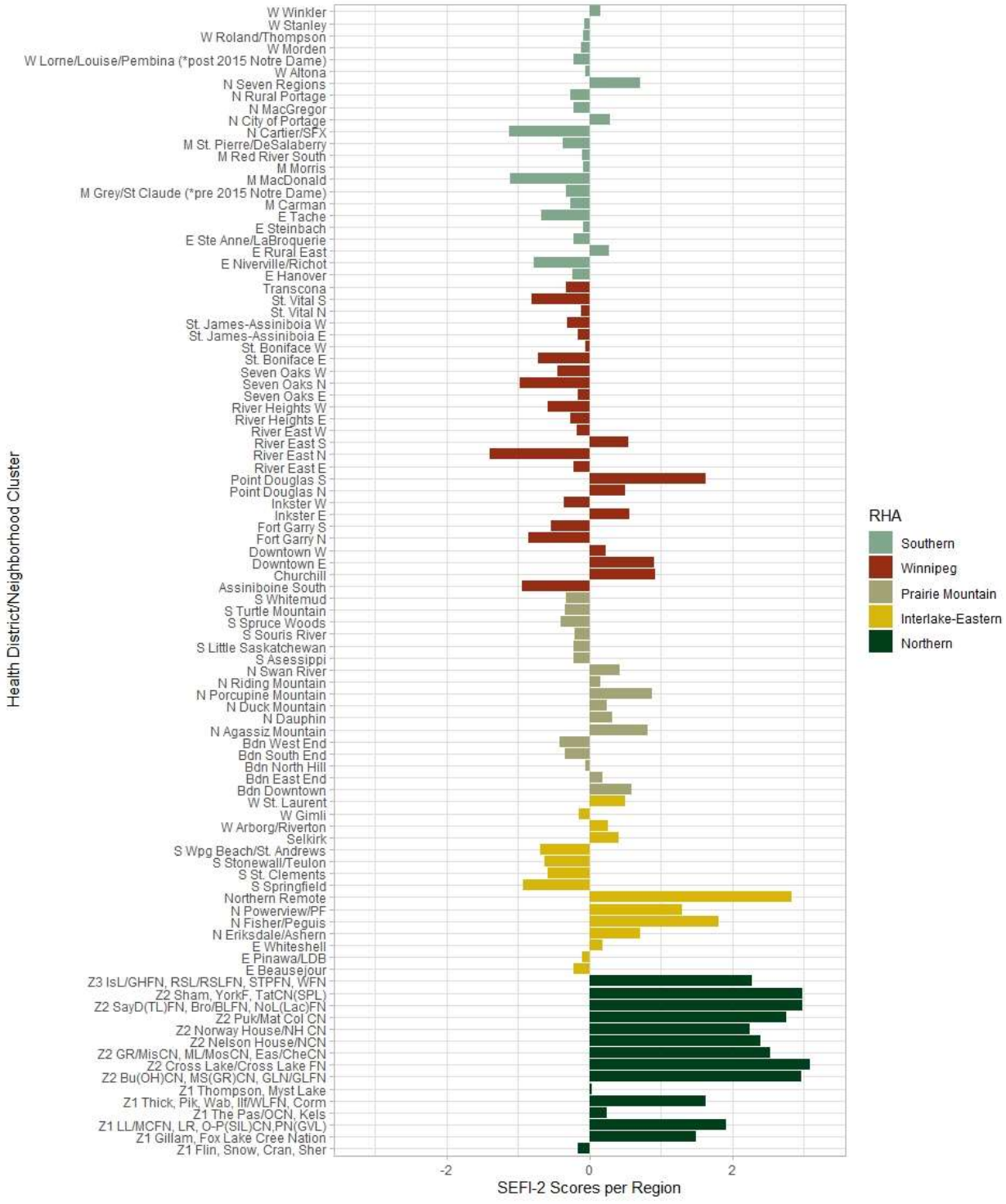

Figure 4. 4: SEFI-2 scores for each RHAD using the 2016 census. Negative values indicate higher SES (less deprivation), whereas positive values indicate lower SES (more deprivation) within regions.

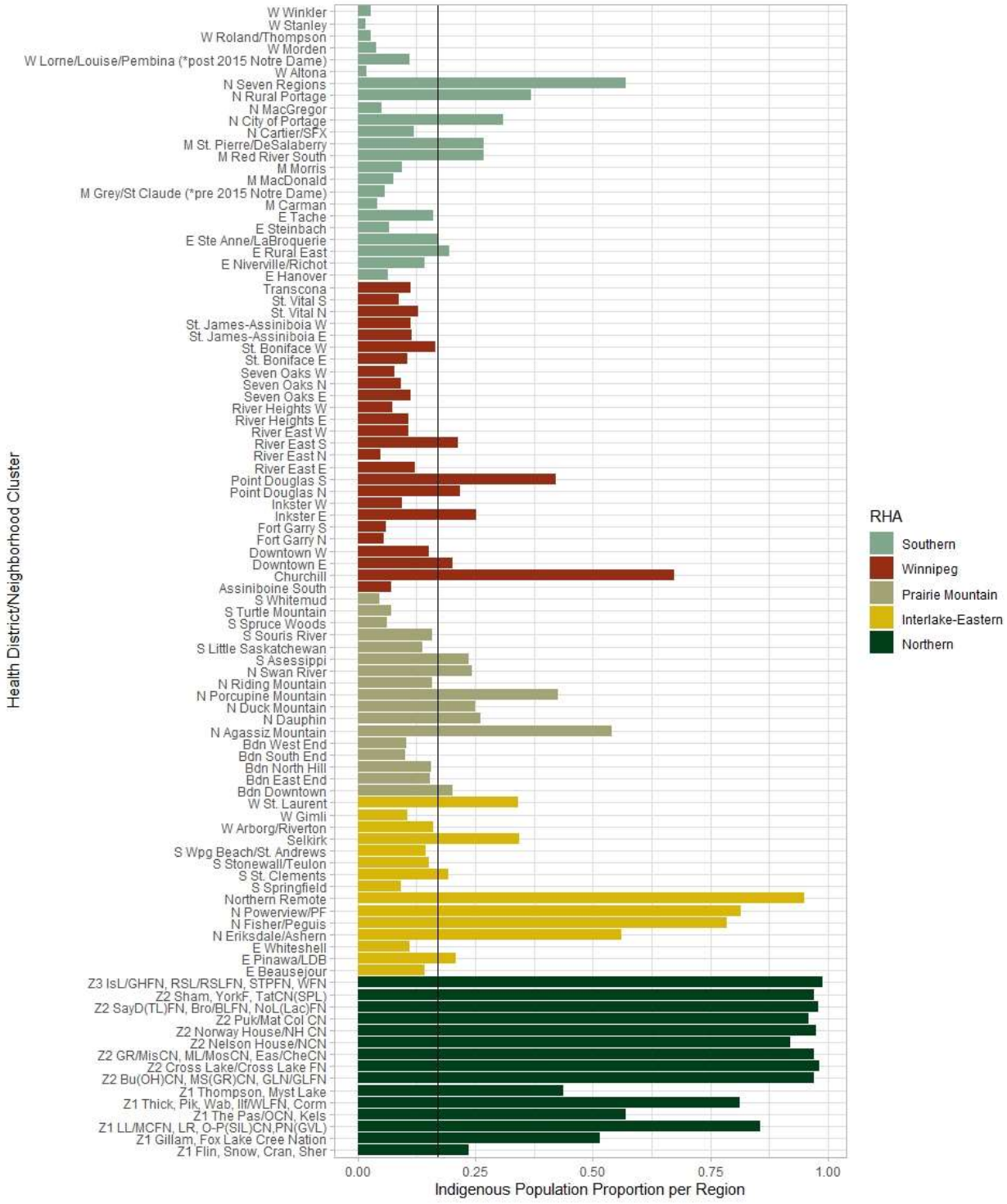

Figure 4. 5: Indigenous population proportion for each RHAD using the 2016 census. Vertical line represents the provincial proportion of 0.18.

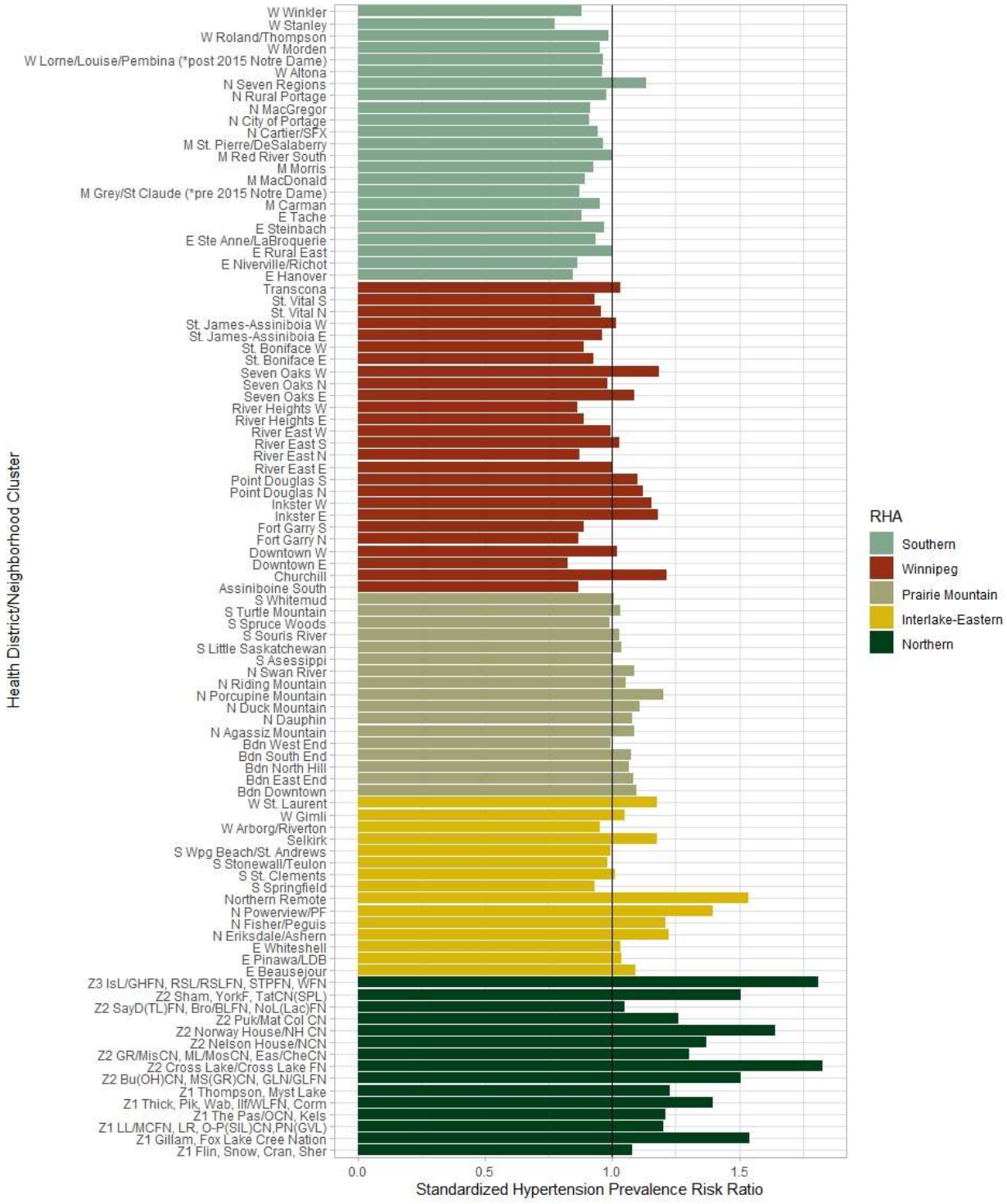

Figure 4.6: Standardized hypertension RR for 2015. (RR of 1 means there is no elevated or reduced risk associated with a particular region compared to the provincial average).

Variable	Mean(sd)	Provincial Rate	Median	Range
Hypertension Prevalence RR (2015)	1.074(0.20)	1.000	1.023	(0.776, 1.827)
Indigenous Population Proportion	0.290(0.30)	0.180	0.158	(0.017, 0.989)
SEFI-2	0.269(1.05)	NA	-0.098	(-1.391, 3.079)
Urban Regions	NA	30 regions	NA	NA

Table 4. 1: Summary statistics of study covariates. Hypertension is calculated for 2015, and the indigenous and SEFI-2 covariates are calculated from the 2016 census. Note that the mean and sd is the mean and sd across the 96 areal units. The Provincial rate is the overall rate when we do not consider breaking the province into regions.

4.2 – AMI Cluster Detection

For the AMI cluster detection, we identify groups of regions that have an elevated AMI age- sex-SIRs. Figure 4.7 shows the AMI SIRs by RHAD, with corresponding map in figure 4.8. It can be seen that there are spatial patterns present, and that exploring these further with the FSS should be done. The FSS will detect clusters of regions that are "hotspots" for AMI incidences. Checking whether lower levels of SES (more socio-economic deprivation), indigenous populations, or are occurring in remote or urban regions is done by visually comparing the maps of the geographical distributions of these variables.

The FSS identified a most likely cluster as well as 8 secondary non-overlapping significant clusters. These are shown in figure 4.9. To evaluate the first hypothesis, comparisons of these maps are visually compared to maps of remote regions, regions that have higher than the average indigenous populations, and regions that have higher than the average socio-economic deprivation scores. It appears that the primary cluster does contain some regions that have higher indigenous populations and more socio-economic deprivation. However, this cluster does not contain any regions that are geographically isolated (remote regions), or urban regions (Portage

la Prairie is the largest city within this cluster). Some of the secondary clusters do seem to be associated with regions of more socio-economic deprivation and higher indigenous populations. In Winnipeg for instance, the regions with higher levels of socio-economic deprivation and higher indigenous populations are mostly contained within a secondary cluster, but this cluster also contains other regions beyond these.

Figure 4.7: AMI age- sex-SIRs by RHAD for 2011-2016.

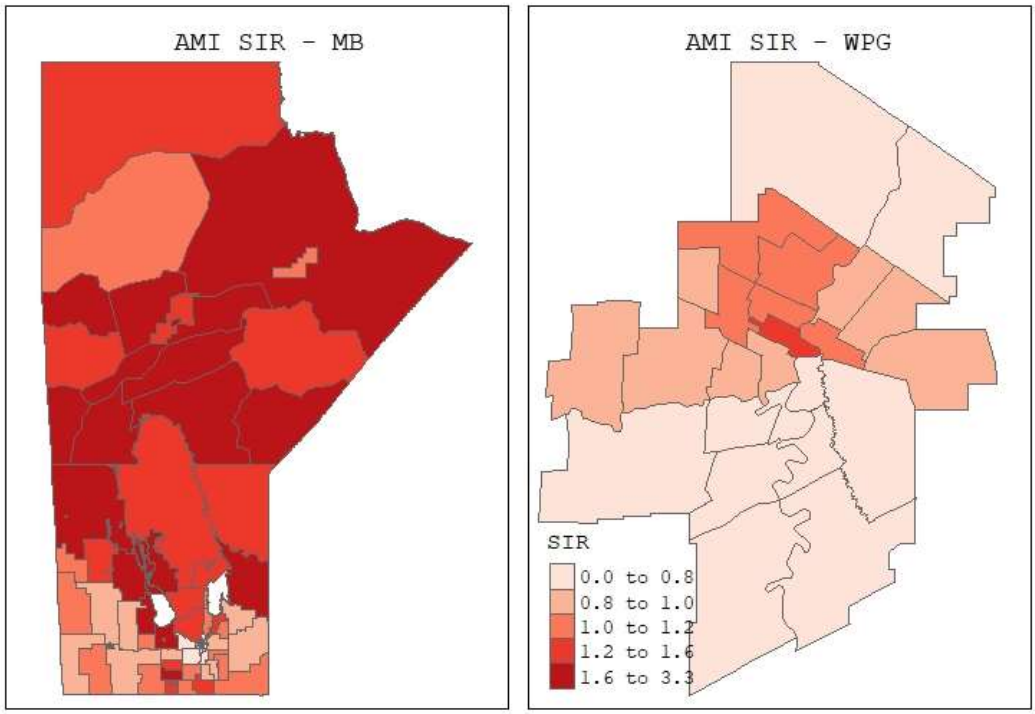

Figure 4.8: Map of AMI age- sex-SIRs for 2011-2016.

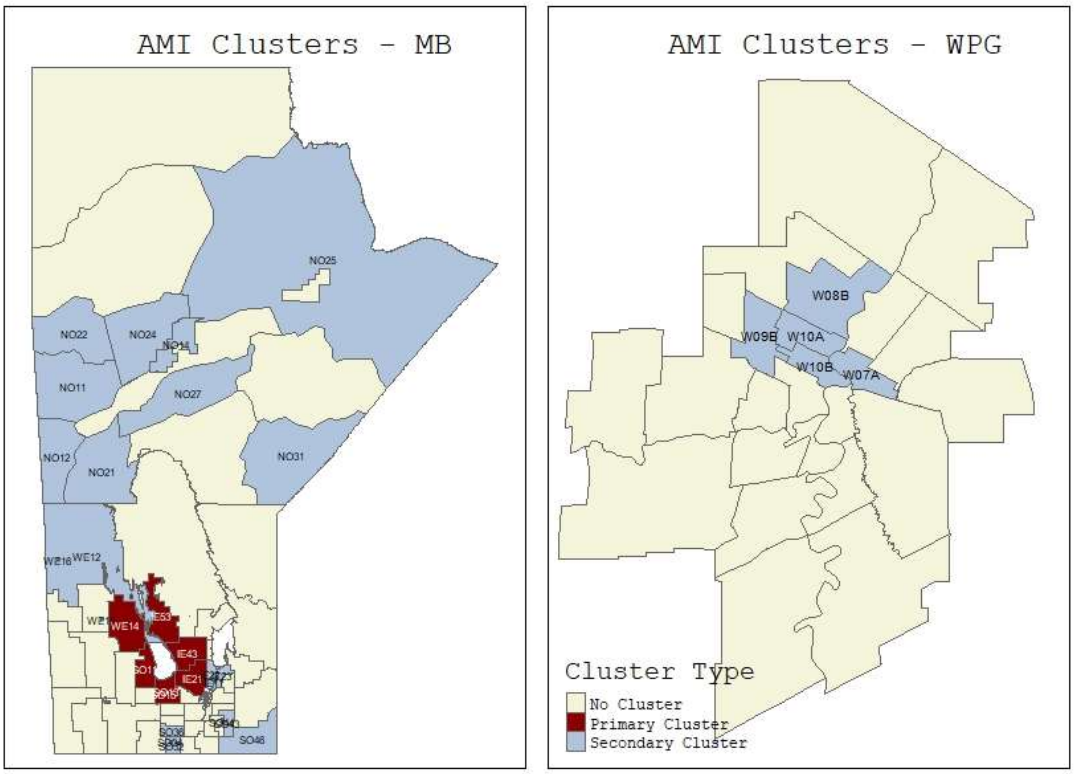

Figure 4. 9: Clusters of AMI incidences (2011-2016).

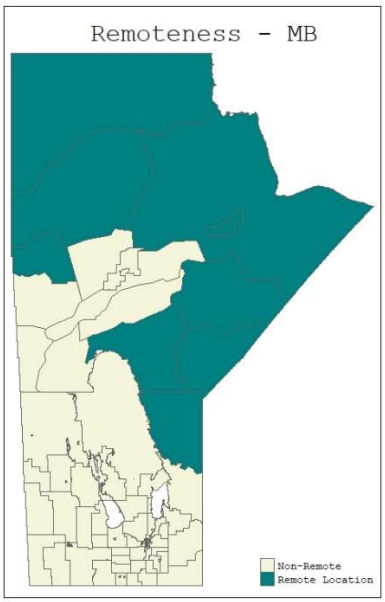

Figure 4. 10: Map of regions classified as remote.

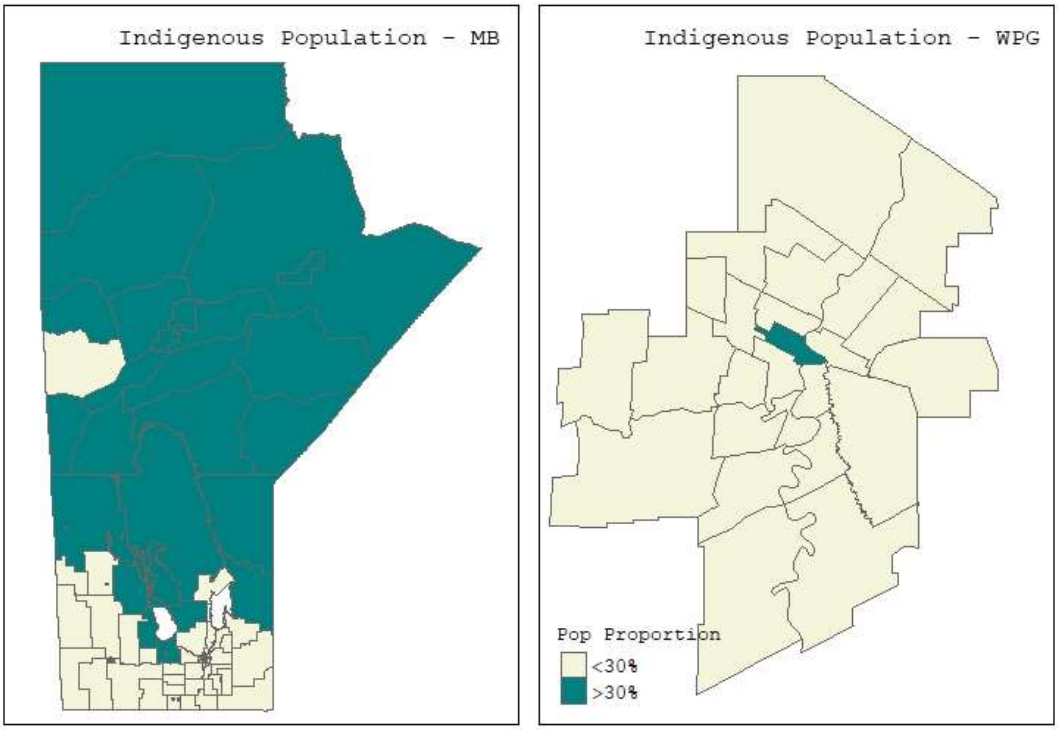

Figure 4. 11: Map of regions that have higher than the average proportion of indigenous persons. Average of 30% is calculated as the mean of the proportions among the 96 RHADs.

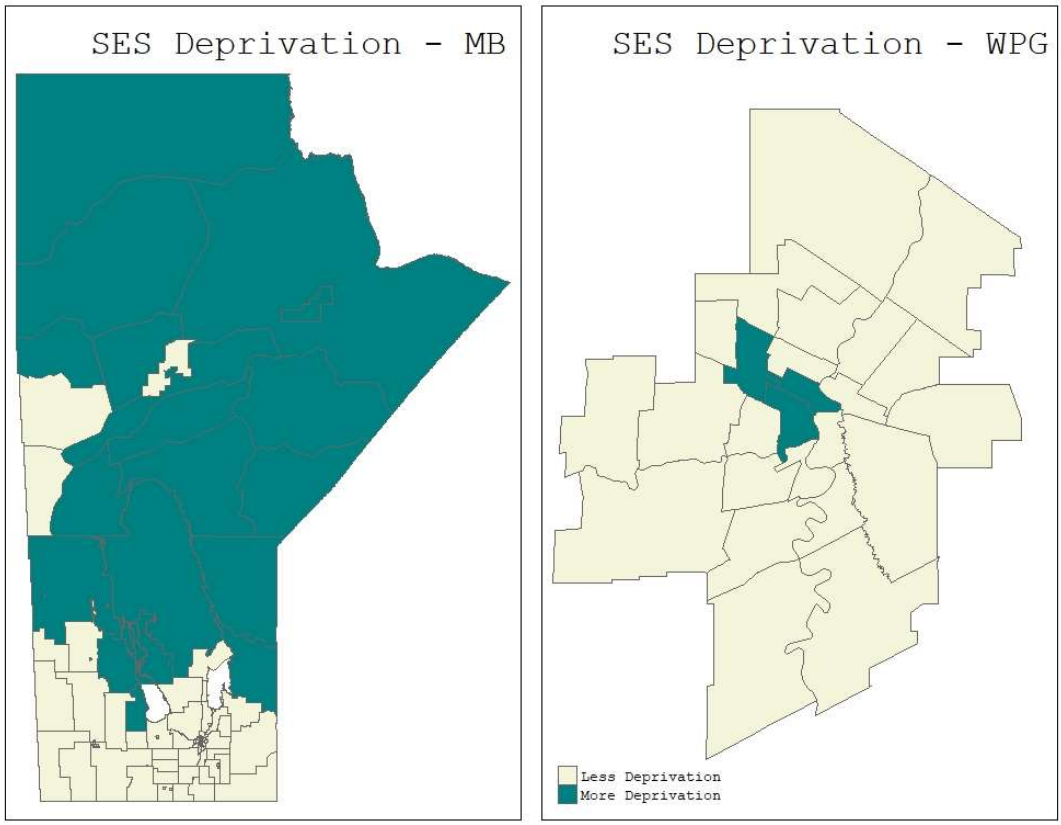

Figure 4. 12: Regions that have higher than the average socio-economic deprivation scores.

4.3 – Poisson and Spatial Poisson Regression Modelling

This section starts with some initial data exploration. First, we conduct a cluster analysis for IHD prevalence for the overall data and the sex-stratified data. Next, we move to checking the Poisson model assumptions and then finally performing some initial modelling with the Poisson model for the overall dataset using INLA for inference. A case will then be made to move to the spatial Poisson model, where the overall data as well as the sex-stratified data will be modelled using the ICAR/BYM specification with INLA.

Data Exploration with FSS Cluster Detection

Cluster detection is done for the overall data and the sex-stratified data and is displayed in figures 4.13-4.15. It can be seen that the clusters are all following the same geographical patterns. It seems to be that the primary cluster is in the central-western region of Manitoba, with secondary clusters in Winnipeg, north-eastern Manitoba, and a cluster in Selkirk as a lone-region cluster in the overall data as well as the sex-stratified data. Females seem to have larger clustering of regions whereas males seem to be clustering in smaller groups of regions.

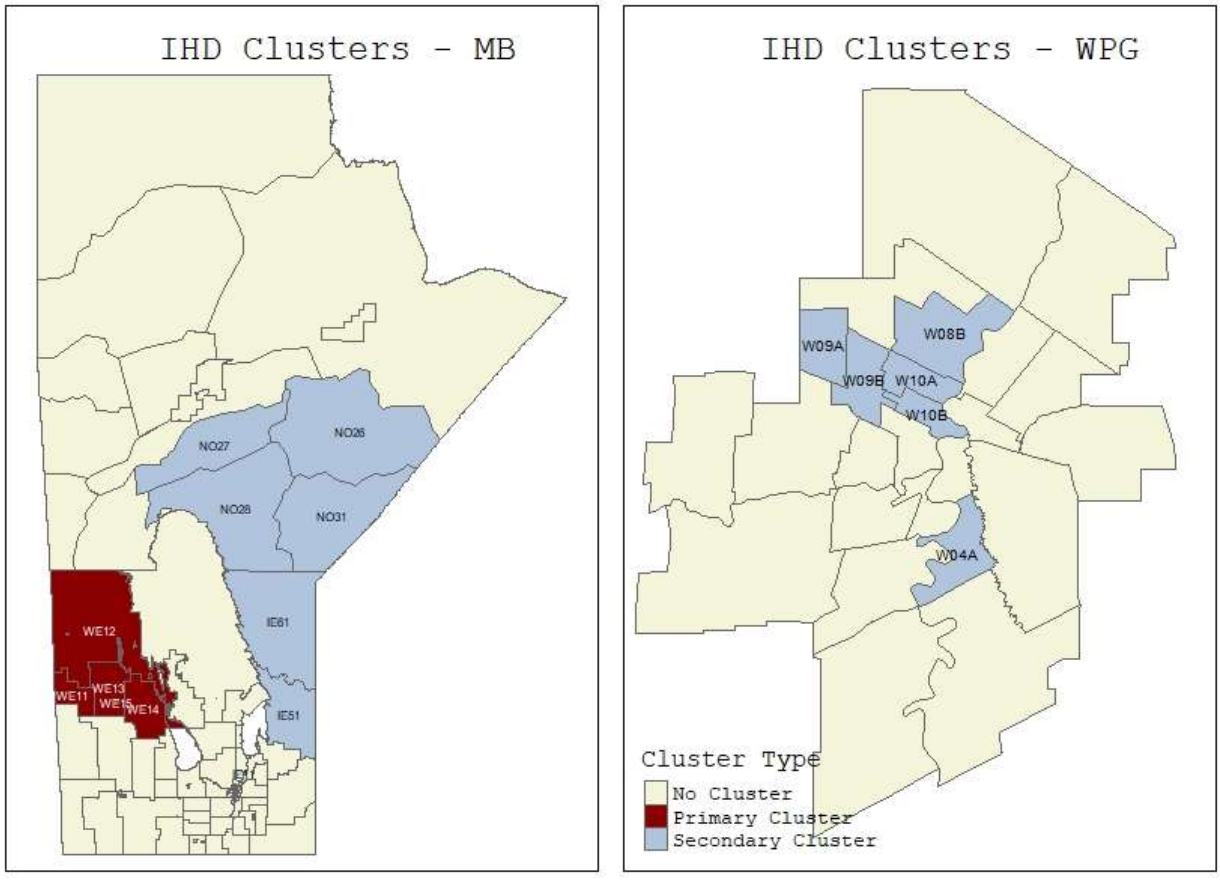

Figure 4. 13: Map of IHD clusters, overall data (2015)

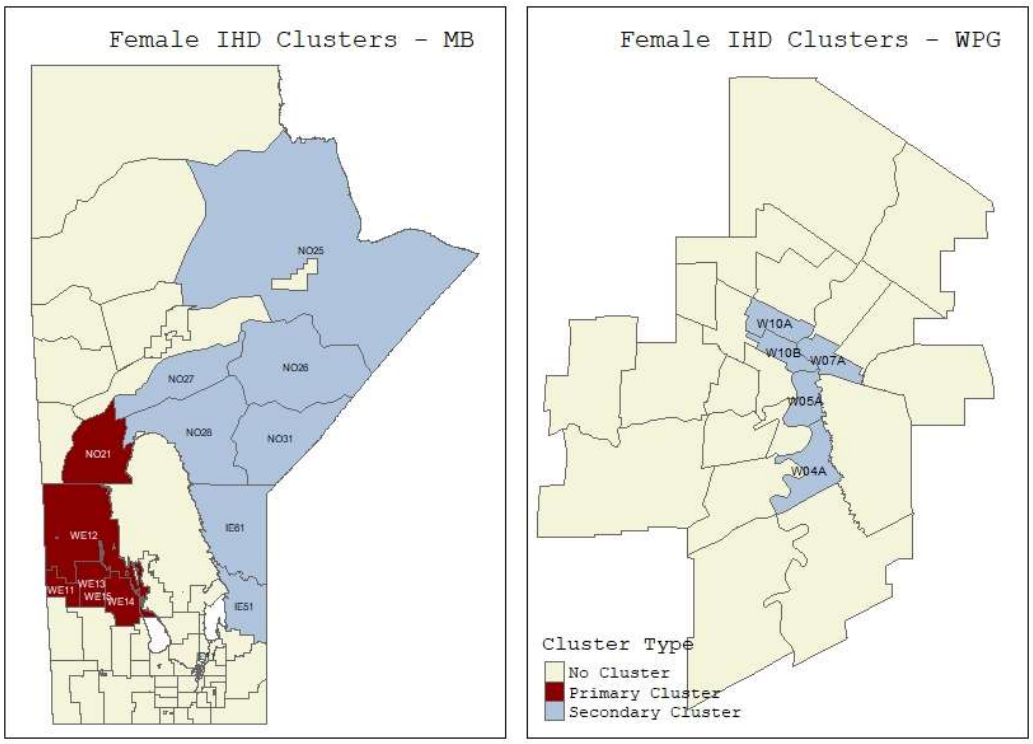

Figure 4. 14: Female IHD clusters (2015)

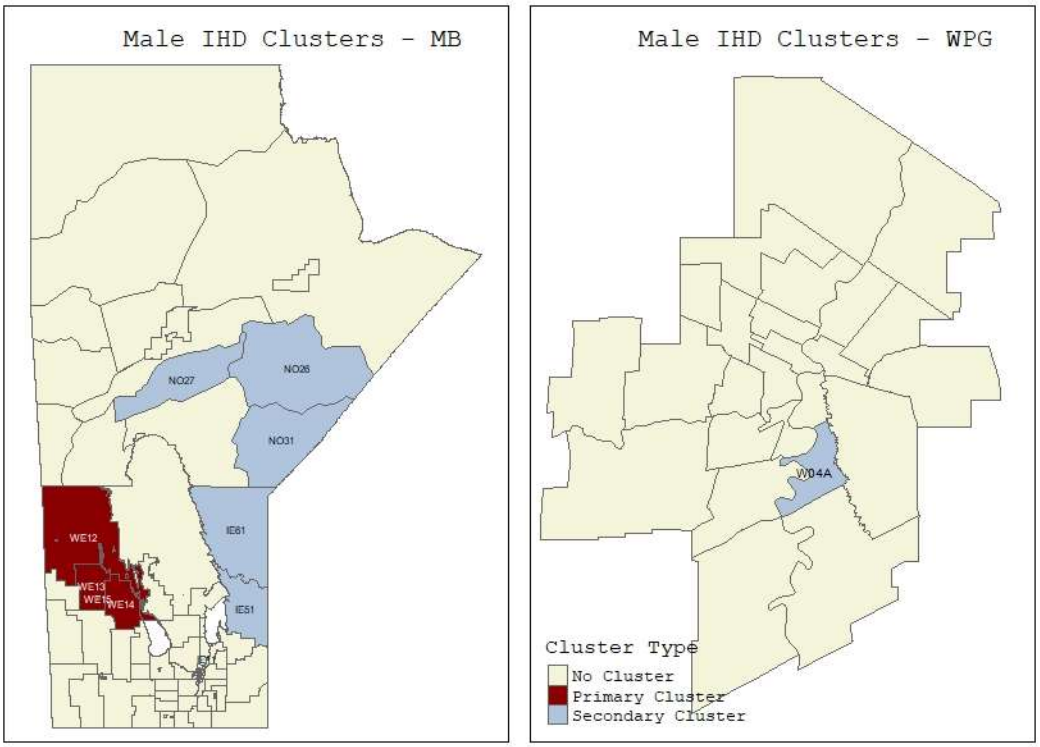

Figure 4. 15: Male IHD clusters (2015)

Poisson Regression Model Building

In order to model the data with a Poisson regression model, there are a few assumptions that need to be met. The first is that the response ($\log(\lambda_i)$) needs to be normally distributed. Figure 4.16 is a histogram of log(IHD RR) to check this normality. This shows that the response is approximately normally distributed which is what we need.

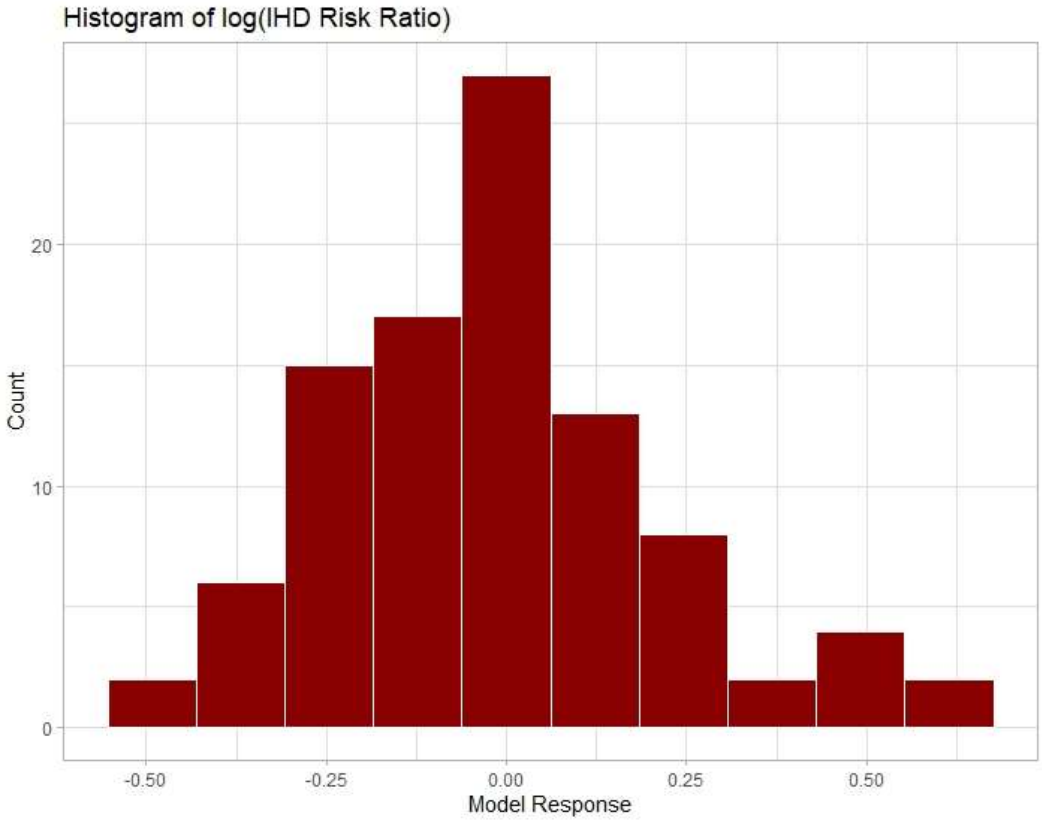

Figure 4. 16: Histogram of log(IHD RR).

The next assumption we need to check, is that of linearity in the predictors. For this part, linearity of the hypertension and the indigenous population proportion predictors is checked in figure 4.17. It can be seen that the covariates are linear to the response after transforming the indigenous population proportion variable with the logarithm transformation.

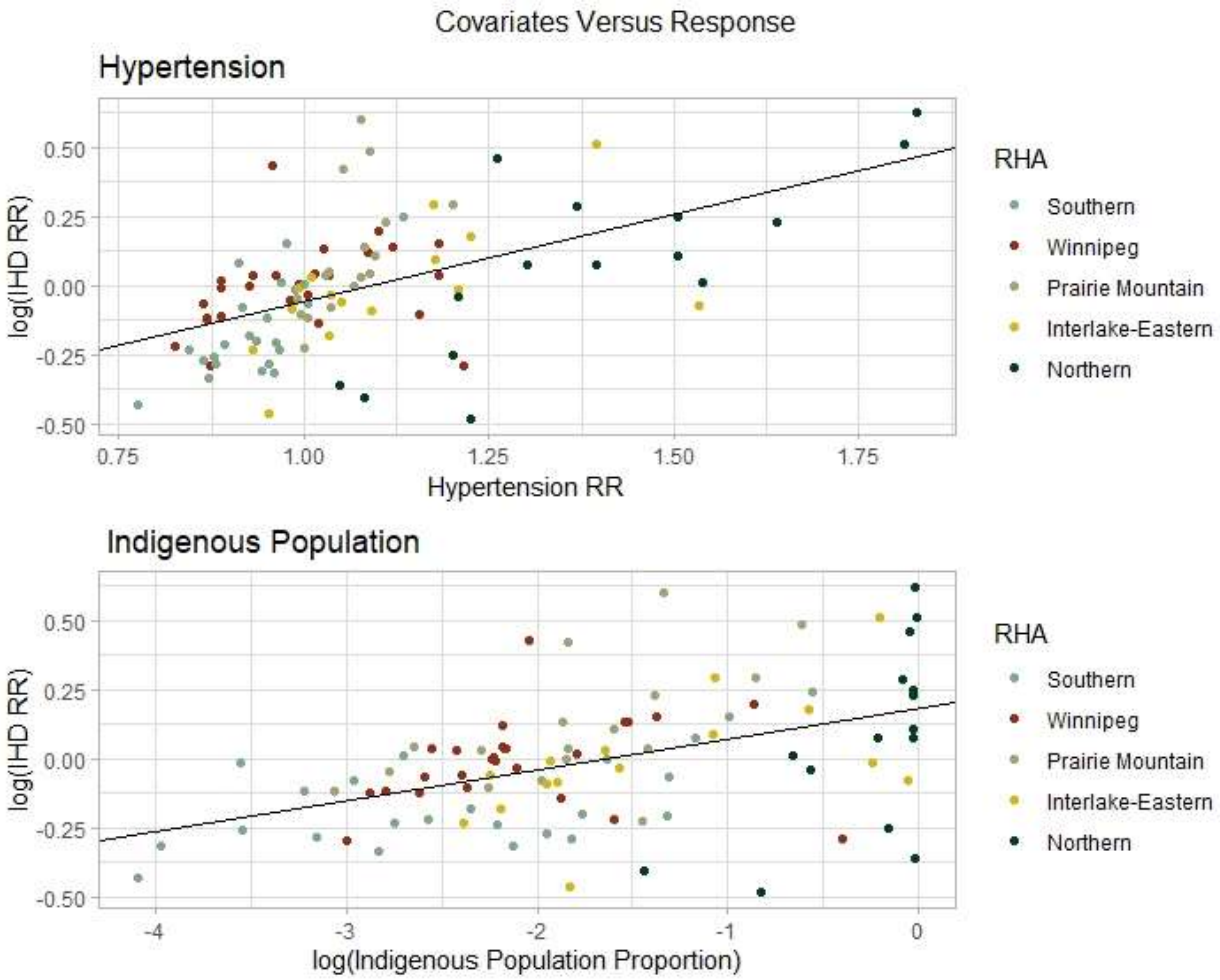

Figure 4. 17: Check for linearity of the hypertension and indigenous population covariates.

Another assumption that needs to be satisfied is the independence of the observations. If this is not satisfied it will cause overdispersion and lead to heteroscedasticity of the residuals.

The first model to be run is the Poisson regression model with hypertension RR, log of indigenous population proportion, and the urban categorical variable as covariates. The model summary is given in table 4.2. It is shown here that all three covariates of interest are significant predictors of IHD prevalence. The next assumption to check is whether the residuals are independently distributed. The residuals are displayed on the map in figure 4.18 which shows a

large amount of spatial correlation between regions. This will cause overdispersion and violates the assumption of independent and identically distributed residuals for the Poisson regression model.

Poisson Generalized Linear Model – IHD SRR					
Covariate	**Mean**	**sd**	**RR**	**95% CI (LL)**	**95% CI (UL)**
(Intercept)	-0.360	0.052	0.698	0.630	0.773
Hypertension RR	0.477	0.040	1.049[1]	1.041[1]	1.057[1]
log(Indigenous)	0.084	0.008	1.008[2]	1.007[2]	1.009[2]
Urban	0.085	0.009	1.088	1.070	1.107
DIC	2279.83				

Table 4. 2: Poisson regression model output. Mean is the mean of the predicted posterior density for each covariate and sd is its standard deviation. RR is the transformed mean in terms of a risk ratio, and the credible interval (CI) for the RR is given in the following 2 columns. All following model output contains these measures.

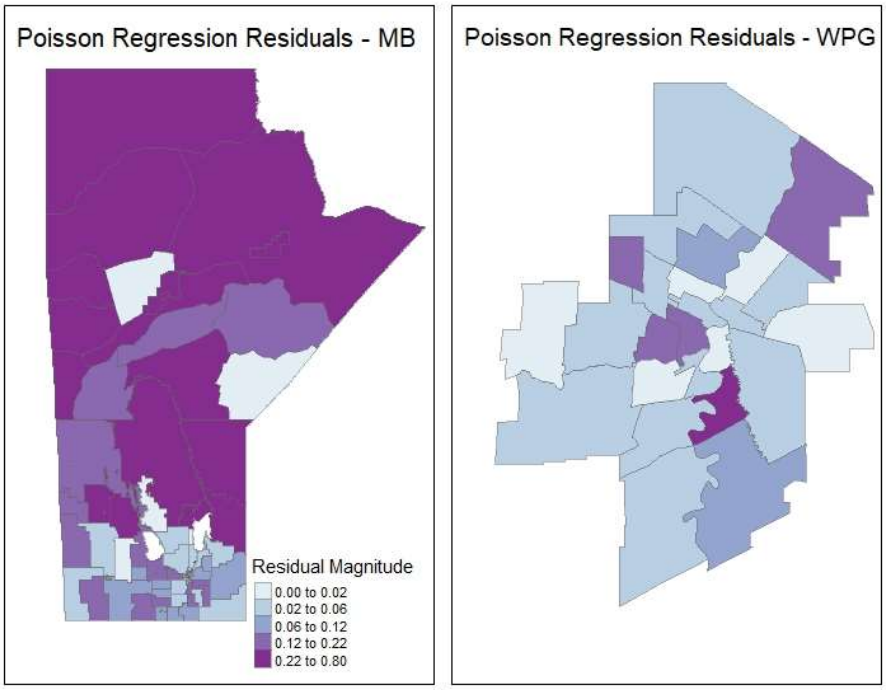

Figure 4. 18: Residual plot for Poisson regression model, (each color bin represents a quantile of the output).

[1] Based off a 0.1 increase in the hypertension RR value holding all other variables constant. All following models report hypertension estimates in this way.

[2] Based off a relative 10% increase in indigenous population proportion. E.g. increasing the proportion from 0.5 to 0.55, or 0.2 to 0.22. All following models report indigenous population proportion estimates in this way.

Spatial Poisson Regression Model Building

As the residuals are very clearly correlated and spatially heteroscedastic, the Poisson regression model is mis-specified. This will cause overdispersion which means that the data are not being modelled adequately. To account for this, we now turn to the spatial Poisson regression model with the BYM specification. Table 4.3 shows the model output for the initial spatial Poisson regression model.

Spatial Poisson Model – IHD SRR					
Covariate	**mean**	**sd**	**RR**	**95% CI (LL)**	**95% CI (UL)**
(Intercept)	-0.922	0.193	0.398	0.271	0.581
Hypertension RR	0.960	0.155	1.101	1.068	1.135
log(Indigenous)	0.086	0.029	1.008	1.003	1.014
Urban	0.095	0.056	1.100	0.985	1.228
DIC	929.945				

Table 4. 3: Initial spatial Poisson regression model output.

Here it appears that when accounting for the spatial covariance in the data that urban regions no longer significantly contribute to the response. Also, of importance to note is that the DIC is now under half of what it was in the non-spatial Poisson regression model. A map of the spatial regression model's residuals is given in figure 4.19, where it now appears as though the residuals are more independently distributed and much lower in magnitude than the non-spatial model's residuals were. This indicates a much better model fit.

Table 4.3 shows that if we increase the hypertension prevalence in a region by 10% more than the provincial average, we should expect the IHD prevalence in that region to be 10.1% (95% CI: 6.8%-13.5%) higher than the provincial average. We should also expect that a 10% relative increase in the proportion of indigenous population from one region to another will result in a 0.8% (95% CI: 0.3%-1.4%) increase in IHD prevalence.

It is now prudent to also account for one of the suspected confounders of these models; that of SES. Here we use the SEFI-2 covariate as a measure of SES in order to account for any SES confounding between indigenous persons or hypertension to the response. First, a check for linearity to the response is performed in figure 4.20, which shows a linear relationship between SEFI-2 and the response, as needed.

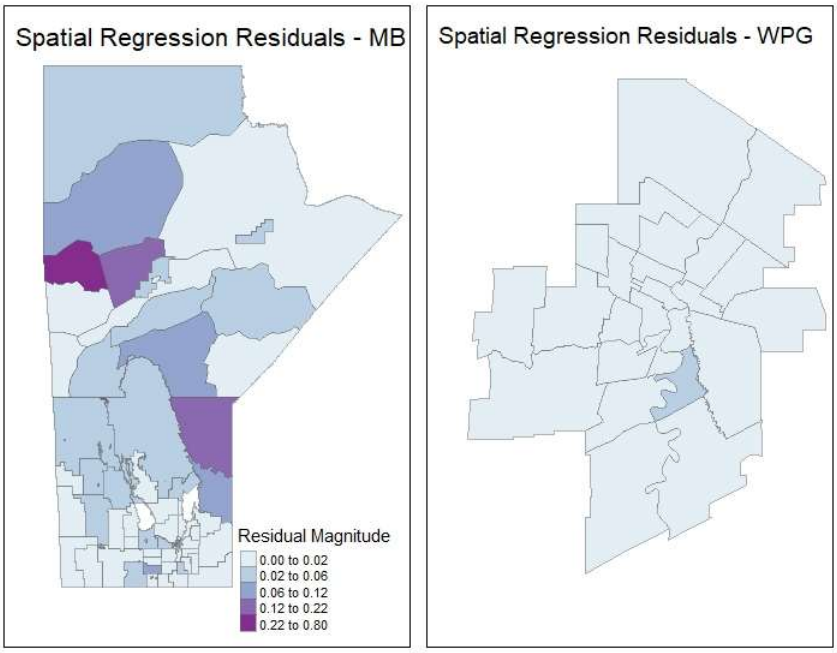

Figure 4. 19: Residual plot for initial spatial Poisson regression model.

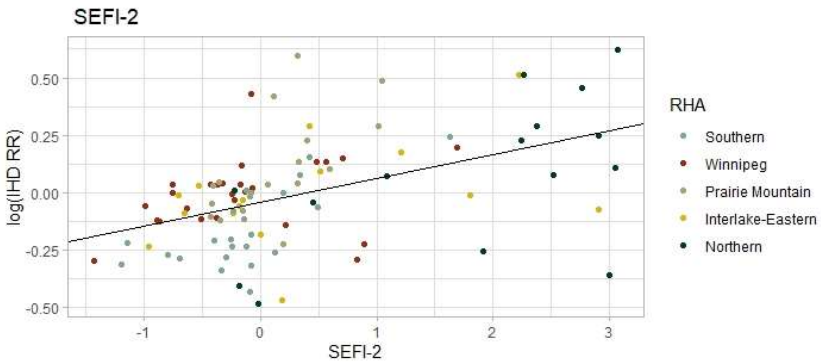

Figure 4. 20: Check for linearity of SEFI-2 covariate.

Table 4.4 shows the model output for the spatial Poisson regression model with the added SEFI-2 covariate. Here it can be seen that the model is more-or-less the same as the previous one in terms of fit, assessed by the DIC. Hypertension and indigenous population proportion are still significant in the presence of SEFI-2 while urban is again not significant. The estimates for hypertension RR and indigenous population proportion attenuated toward zero by 7.1% and 16% respectively. Therefore, even though the SEFI-2 variable isn't significant in the model, it affects the results which means there is some confounding or effect modification occurring between SEFI-2 and indigenous population proportion, and to a lesser extent with hypertension. For this reason, SEFI-2 should be kept in the model, as we can then control for these effects.

Spatial Poisson Model – IHD SRR					
Covariate	**mean**	**sd**	**RR**	**95% CI (LL)**	**95% CI (UL)**
(Intercept)	-0.881	0.198	0.415	0.280	0.610
Hypertension RR	0.892	0.168	1.093	1.058	1.130
log(Indigenous)	0.072	0.032	1.007	1.001	1.013
Urban	0.089	0.056	1.094	0.979	1.221
SEFI-2	0.035	0.033	1.035	0.970	1.105
DIC	929.969				

Table 4. 4: Spatial Poisson regression model output, controlling for SES.

Table 4.5 shows the variance of the random-effects components of the model, as well as the proportion for each out of the total of the model's variation. Here, the majority of the model's variation is coming from the spatially structured random effects term, indicating that the spatial structure is a significant component of the model.

Variance Component	Variance Estimate	Proportion of Model Variance
Spatially Structured	6.68E-2	0.986
Spatially Unstructured	9.39-04	0.014

Table 4. 5: Estimated variance components for spatial regression model controlling for SES.

Figure 4.21 maps the observed IHD standardized risk ratios (SRR)'s, and figure 4.22 maps the fitted and smoothed IHD SRR's. This smoothing has produced some changes in the appearance of these maps, most notably in central Manitoba where the observed SRR's obtained from the data may be lower than what they are in reality. Figure 4.23 shows regions that have significantly elevated IHD prevalence (95% credible interval). These are calculated by checking if the fitted probability distribution's 2.5% quantiles are greater than RR=1 for each RHAD. Comparing the map of IHD clusters in figure 4.13 with the elevated regions in figure 4.23, it can be seen that many of the same regions that were reported as part of a cluster have also been estimated to have significantly elevated prevalences of IHD.

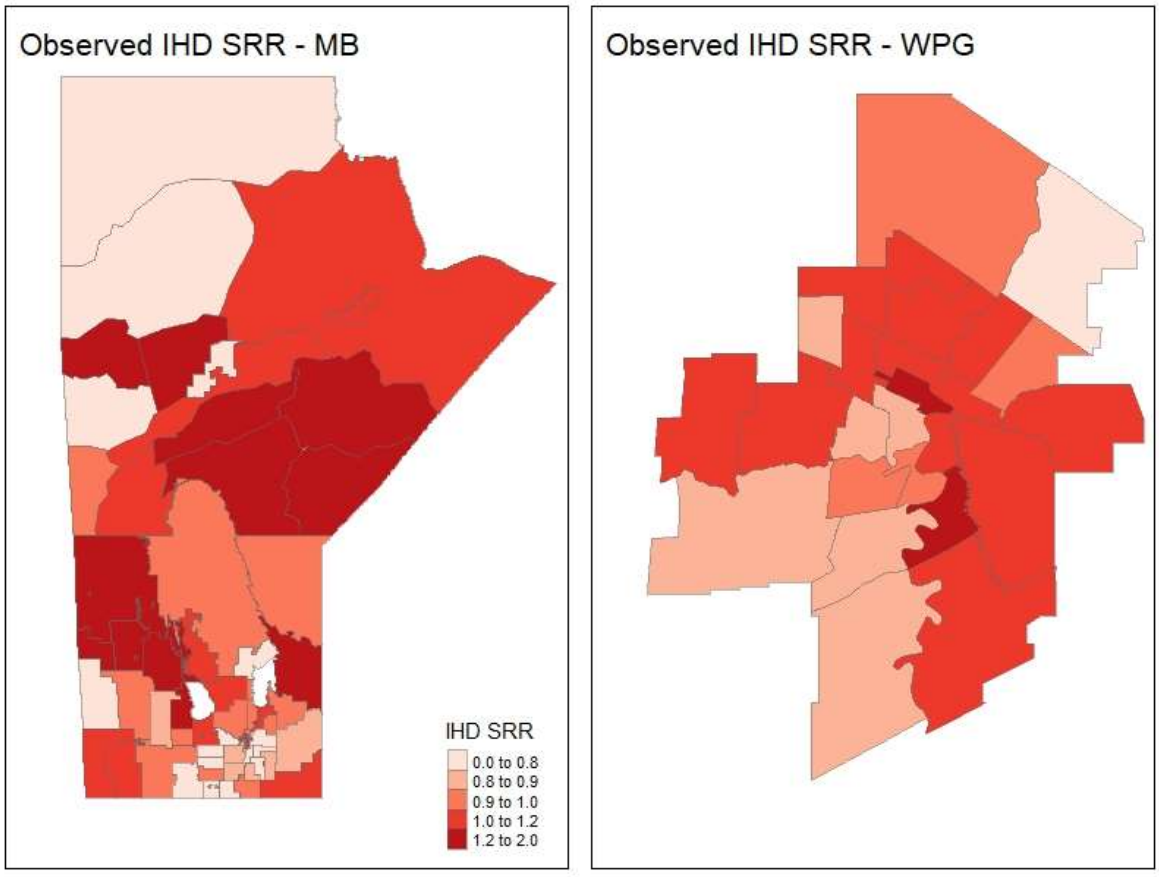

Figure 4. 21: Observed IHD SRR's (2015).

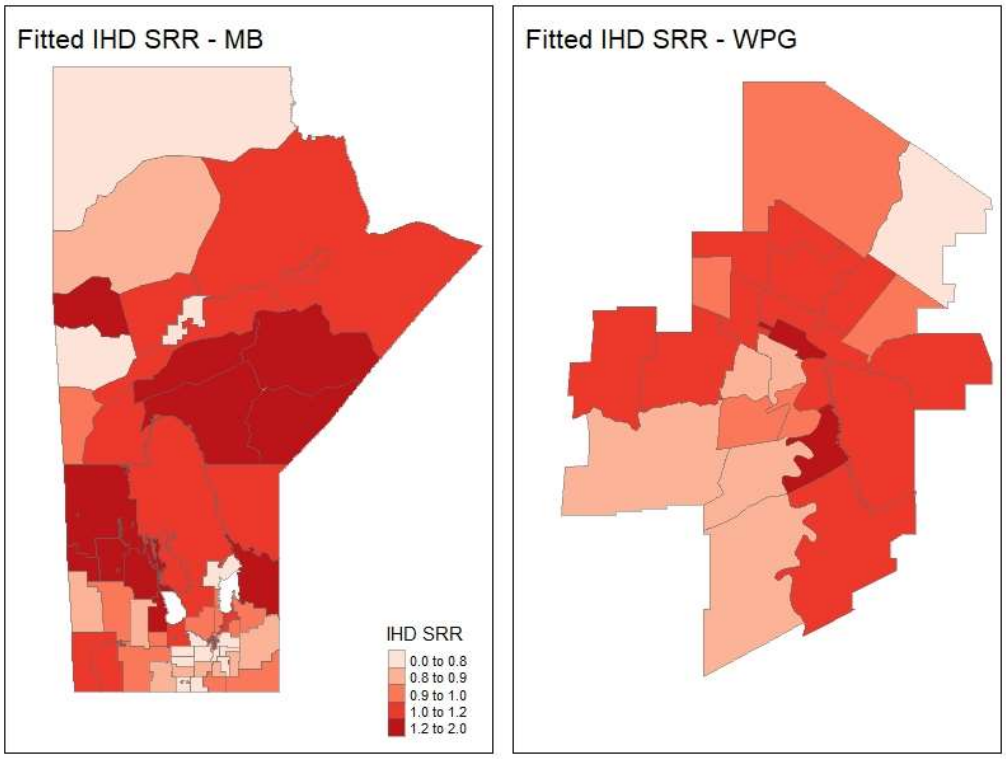

Figure 4. 22: Fitted/smoothed IHD SRR's (2015). Fitted with the model from table 4.4.

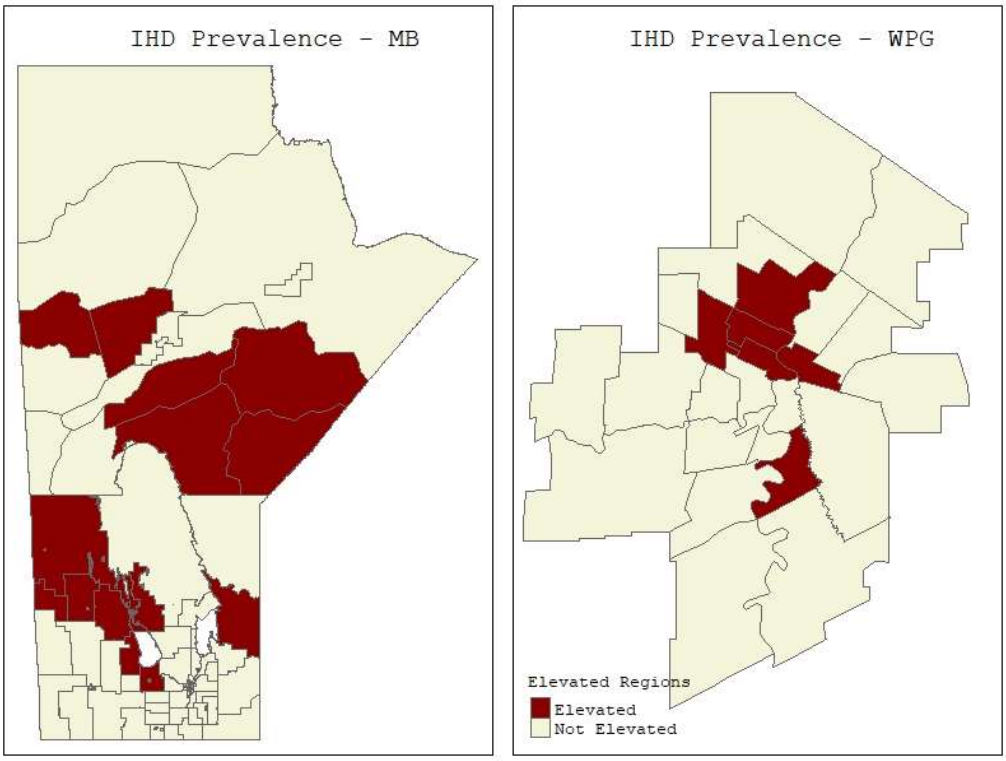

Figure 4. 23: Regions with significantly elevated levels of IHD prevalence at the 95% credible interval (2015). Estimated by the model in table 4.4.

Sex-stratified Modelling

The next avenue to explore, is to apply the model from table 4.8 to the sex-stratified data to see if there are any differences in how the sexes are affected by IHD. Model outputs for females and males are given in tables 4.6 and 4.7 respectively.

Spatial Poisson Model – IHD SRR Females					
Covariate	**mean**	**sd**	**RR**	**95% CI (LL)**	**95% CI (UL)**
(Intercept)	-0.786	0.247	0.456	0.279	0.738
Hypertension RR	0.833	0.205	1.087	1.044	1.132
log(indigenous)	0.106	0.042	1.010	1.002	1.018
Urban	0.113	0.069	1.120	0.977	1.284
SEFI-2	0.069	0.046	1.072	0.980	1.172
DIC	821.622				

Table 4. 6: Spatial regression model output for female data.

Spatial Poisson Model – IHD SRR Males					
Covariate	**Mean**	**sd**	**RR**	**95% CI (LL)**	**95% CI (UL)**
(Intercept)	-0.793	0.182	0.453	0.316	0.647
Hypertension RR	0.807	0.159	1.084	1.051	1.119
log(indigenous)	0.062	0.031	1.006	1.000	1.012
Urban	0.081	0.053	1.084	0.976	1.203
SEFI-2	0.024	0.031	1.024	0.964	1.088
DIC	878.320				

Table 4. 7: Spatial regression model output for male data.

It appears that for females, indigenous population proportion is more associated whereas hypertension is less associated with IHD prevalence than for the overall population. There also appears to be less association with both of these variables for males than there is in the overall data. The regions with significantly elevated IHD prevalence are given in figures 4.24 and 4.25.

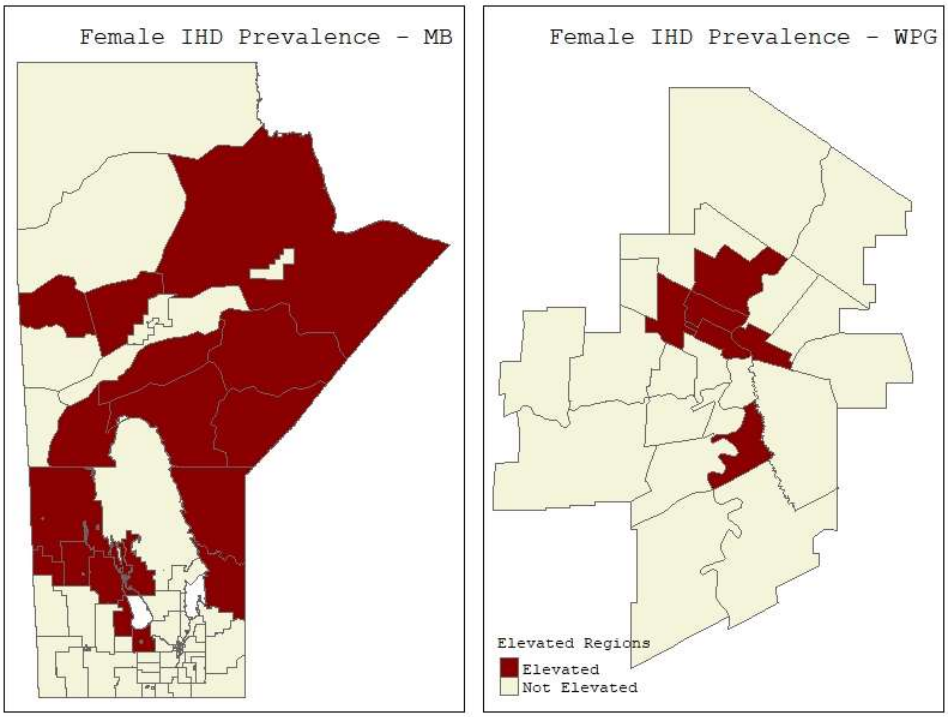

Figure 4. 24: Significantly elevated levels of IHD prevalence for females at the 95% credible interval (2015). Estimated by the model in table 4.6.

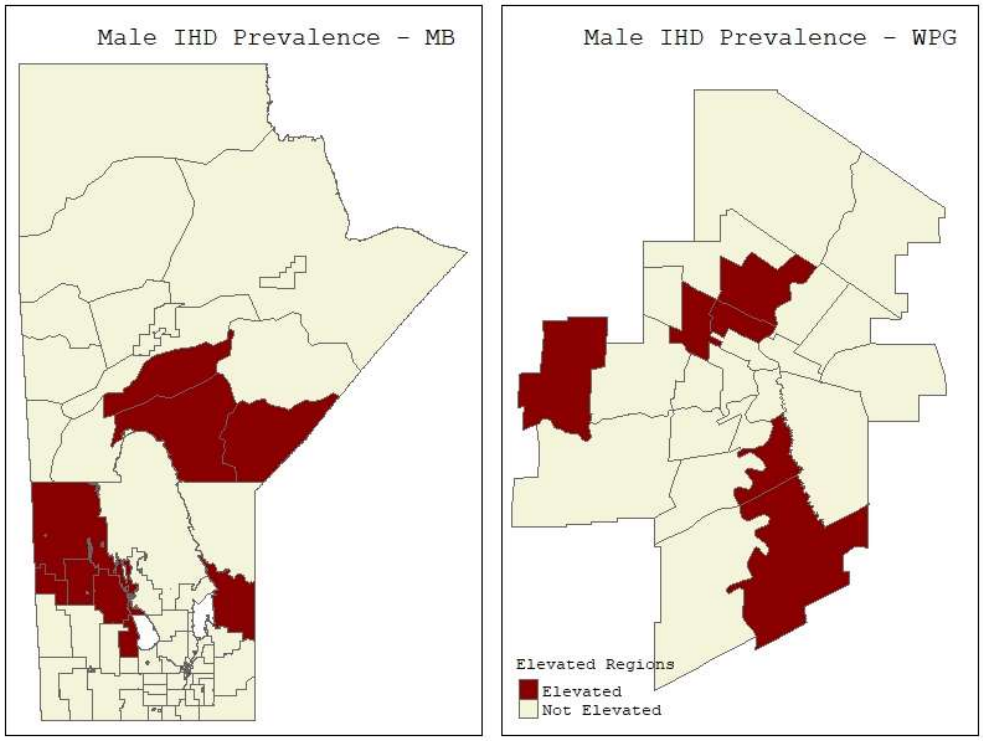

Figure 4. 25: Significantly elevated levels of IHD prevalence for males at the 95% credible interval (2015). Estimated by the model in table 4.7.

Although there are some differences between the cluster detection plots and the elevated IHD prevalence plots, the patterns are closely following each other. Females again have more regions that have elevated rates, whereas males are more tightly clustered. In Winnipeg, the regions that have elevated IHD prevalence in males seem to be more scattered than the cluster detection method was able to detect.

Thus, after accounting for the heterogenetic spatial variation, and controlling for SES confounding; IHD prevalence is, at the RHAD level, associated with increased hypertension risk and indigenous population proportion. This is true for the overall data as well as when the data is sex-stratified. The association between urban regions and IHD prevalence is not conclusive, as this covariate was not significant in any of the models after accounting for the spatial covariance. There are some clear disease patterns for both the overall data and the sex-stratified data, as many of the same regions have been identified as having an elevated IHD prevalence and being detected by the FSS as being part of a disease cluster.

4.4 – Spatio-Temporal Regression Modelling

The next aspect for this research project that is of interest is to assess how IHD is changing over time as well as space. To motivate this, the temporal pattern of the disease is explored in more detail. Figure 4.2 shows the overall crude prevalence of IHD over the time period of interest (1998-2015) aggregated by year and sex-stratified. This shows that the overall temporal pattern is very similar between the sexes and the overall data, with the obvious difference being that males have higher prevalences than females. Figure 4.26 shows the overall

observed SRR for IHD over this time period as well as the observed IHD SRR broken down by RHA.

It can be seen from figure 4.26 that there are differences in the temporal patterns of IHD between the RHA's. This leads to the hypothesis that IHD contains a space-time pattern that spatially varies over time, or temporally varies over space. For this section there are five different models to be considered. The first is the spatio-temporal Poisson regression model without any space-time interaction, and the subsequent four are the models with the 4 different types of space-time interactions (defined in section 3.4.3). The model fits are assessed through the use of the DIC, where the model with the best fit is kept. The results of this comparison are given in table 4.11 for the overall data as well as the sex-stratified data.

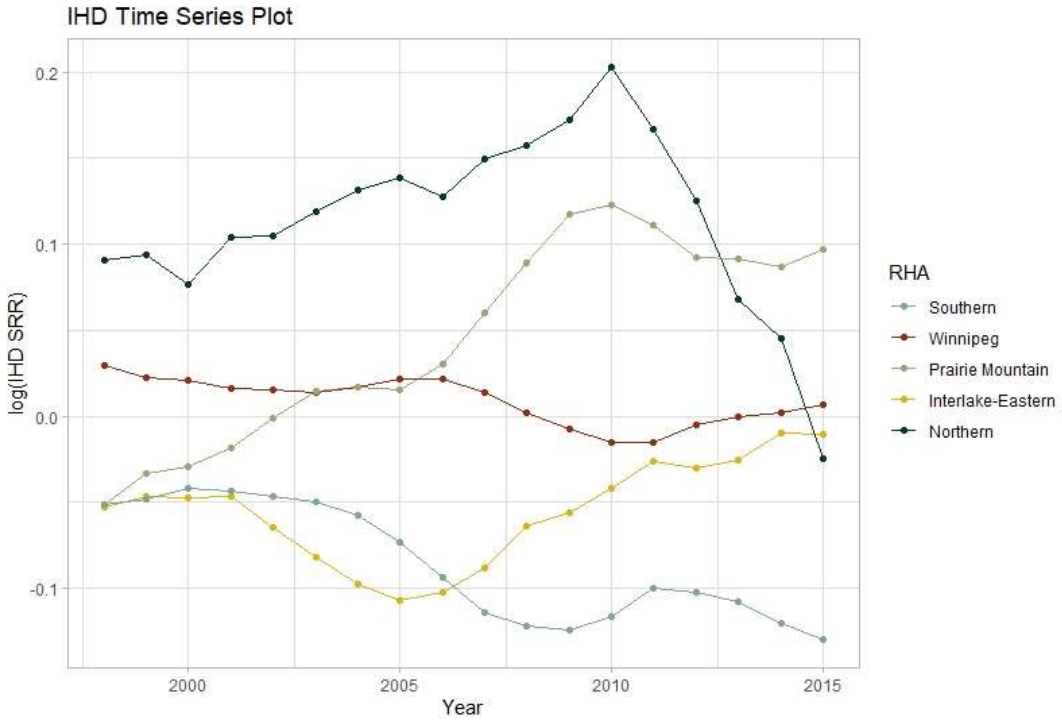

Figure 4. 26: Time series plot of log transformed IHD SRR's.

Interaction Type	DIC		
	Overall Data	Female Data	Male Data
No Interaction	17884.66	15242.00	15047.19
Type I	16213.18	14457.61	14767.03
Type II	**14907.22**	**13208.11**	**13861.09**
Type III	16017.77	14293.01	14606.79
Type IV	14912.09	13213.77	13856.97

Table 4. 8: Spatio-temporal regression model fitting, assessed with the DIC.

As it is seen in table 4.8, all three models with the type II interaction; temporally structured and spatially unstructured correlation structures; are the best fitting for all three datasets. The models with type IV interactions; temporally and spatially structured correlation structures; are almost as good but it does not seem that the spatial structure is necessary in the interaction term. This is most likely due to how the temporal patterns vary throughout space, where the spatial patterns are not varying throughout time. The model outputs for the overall and sex-stratified models with type II interactions are given in tables 4.9-4.11.

Spatio-Temporal Poisson Model – IHD SRR					
Covariate	**mean**	**sd**	**RR**	**95% CI (LL)**	**95% CI (UL)**
(Intercept)	-0.103	0.152	0.902	0.670	1.214
Hypertension RR	0.289	0.053	1.029	1.019	1.040
log(Indigenous)	0.056	0.076	1.005	0.991	1.020
Urban	-0.035	0.245	0.966	0.597	1.562
SEFI-2	0.010	0.122	1.010	0.795	1.283
DIC	14907.220				

Table 4. 9: Spatio-temporal Poisson regression model output for overall data.

Spatio-Temporal Poisson Model – IHD SRR Females					
Covariate	**mean**	**sd**	**RR**	**95% CI (LL)**	**95% CI (UL)**
(Intercept)	-0.110	0.153	0.896	0.663	1.209
Hypertension RR	0.333	0.066	1.034	1.020	1.047
log(indigenous)	0.066	0.076	1.006	0.992	1.021
Urban	-0.048	0.248	0.953	0.586	1.549
SEFI-2	0.032	0.123	1.032	0.810	1.315
DIC	13208.110				

Table 4. 10: Spatio-temporal Poisson regression model output for female data.

Spatio-Temporal Poisson Model – IHD SRR Males					
Covariate	**mean**	**sd**	**RR**	**95% CI (LL)**	**95% CI (UL)**
(Intercept)	-0.103	0.150	0.903	0.673	1.210
Hypertension RR	0.258	0.053	1.026	1.015	1.037
log(indigenous)	0.051	0.075	1.005	0.991	1.019
Urban	-0.024	0.238	0.976	0.612	1.558
SEFI-2	-0.004	0.118	0.996	0.790	1.255
DIC	13861.090				

Table 4. 11: Spatio-temporal Poisson regression model output for male data.

An important finding to note is that in these models, indigenous population proportion is not a significant contributor to the response. Another finding is that as before in the spatial models, the magnitude of the association between hypertension and IHD looks to be marginally larger for females than males, although this increase is within error where the measures of association between the covariates and response are all within error the same for the overall data and the sex-stratified data.

Table 4.12 shows the amount of the model's variance coming from each of the random effects terms. It's clear to see that after accounting for the temporally structured heterogeneity in the data, that the space-time interaction term carries the majority of the variation. Here, this is the temporally structured and spatially unstructured random effects term. What this means is that the temporal trend varies throughout space and the covariance between time points for each areal unit are significant factors.

Variance Component	**Variance Estimate**	**Proportion of Model Variance**
Spatially Structured	1.4E-3	0.271
Spatially Unstructured	1.4E-3	0.270
Temporally Structured	2.1E-5	0.006
Temporally Unstructured	1.7E-5	0.005
Space-Time Interaction (type II)	2.5E-3	0.447

Table 4. 12: Estimated variance components for spatio-temporal regression model for overall data.

Figures 4.27-4.32 maps the IHD SRR's for the overall data and for the sex-stratified data. Visually comparing these maps, it appears that females have worse IHD prevalences than males especially in northern Manitoba. This could be somewhat misleading however, as the standard population in the SRR formula would be the provincial female population for the female data and the provincial male population for the male data. Therefore, comparisons across strata should be made with caution. It does appear that the SRR for all groups is decreasing with time in most regions, although there are some outliers that appear to be gradually increasing with respect to time e.g. St. Vital North.

The plot of the temporally structured random effect over time in figure 4.33 shows the adjustments made to the SRR by the temporally structured random effects for each of the models. The model for the male stratified data shows the least deviance from zero, which indicates that over time there was little adjustment needed and thus little covariance between the years of data existed. The female stratified data shows greater change, and the overall data model represents the average of these two models. Overall, the values for the structured temporal random effects were all on average decreasing over time. This indicates that the adjustment the model made here was to smooth the rates to lower values in more recent years, and indicates that we would have been overestimating the SRR's in the later years compared with the earlier years if we did not account for the structured temporal random effect. Also, if we hold all covariates and spatial effects constant, this shows that over time the provincial IHD prevalence is on the decline over these years of data, with the largest declines in the female stratified data.

The plot in figure 4.33 could be replicated for each of the 96 areal units by taking plotting the overall temporal trend term from figure 4.33 and adding the space-time interaction term for each areal unit. As this is the most significant term in terms of how much model variance it

carries, this would be the most meaningful plot. However, this is obviously a large and potentially messy plot to display, and so instead figure 4.34 shows the six RHADs that had the largest increases in the model fitted IHD RR over time. These regions had a 30% or greater increase in IHD RR from 1998 to 2015. This plot accounts for any temporal change in the hypertension covariate, as well as any changes in the temporally unstructured, temporally structured, and temporally structured space-time interaction random effects terms. What this shows is that even though the trend over time is a decrease in IHD prevalence, especially for females, that there are some regions that have had substantial increases in prevalence.

Thus, after accounting for the spatial and temporal heterogeneity as well as the temporally structured space-time interaction effect, it appears that the IHD prevalence is decreasing over time. This is true for females, whereas the decrease for males has been more marginal. After accounting for both the spatial and temporal effects as well as SES through the SEFI-2 variable, it appears that indigenous population proportion is no longer a significant predictor of IHD prevalence and urbanicity is still not significant. However, the spatio-temporal distribution of hypertension prevalence is still a significant contributor to IHD prevalence for all three datasets.

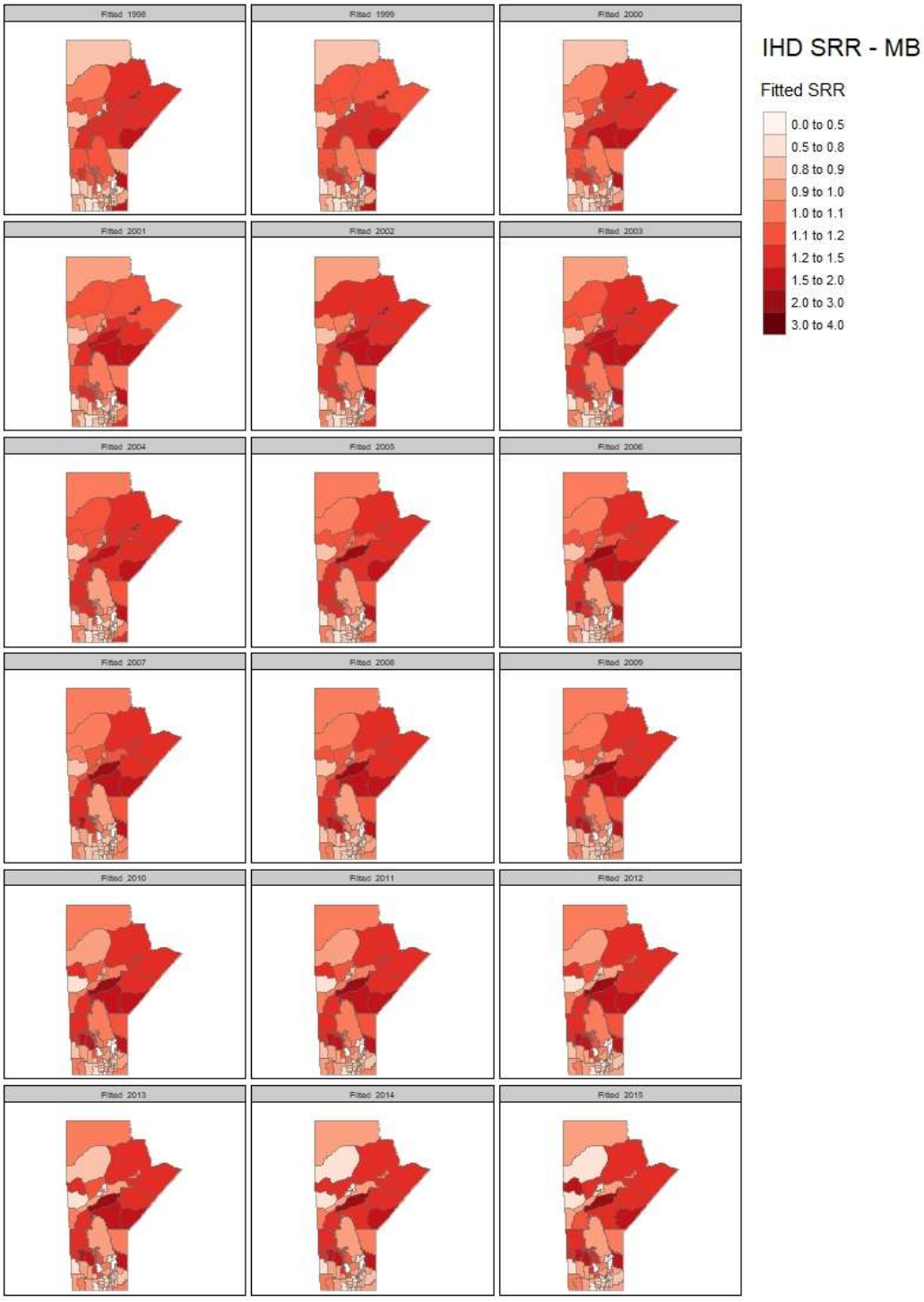

Figure 4. 27: Fitted/smoothed IHD SRR's by year for the overall data. Fitted with the model in table 4.9.

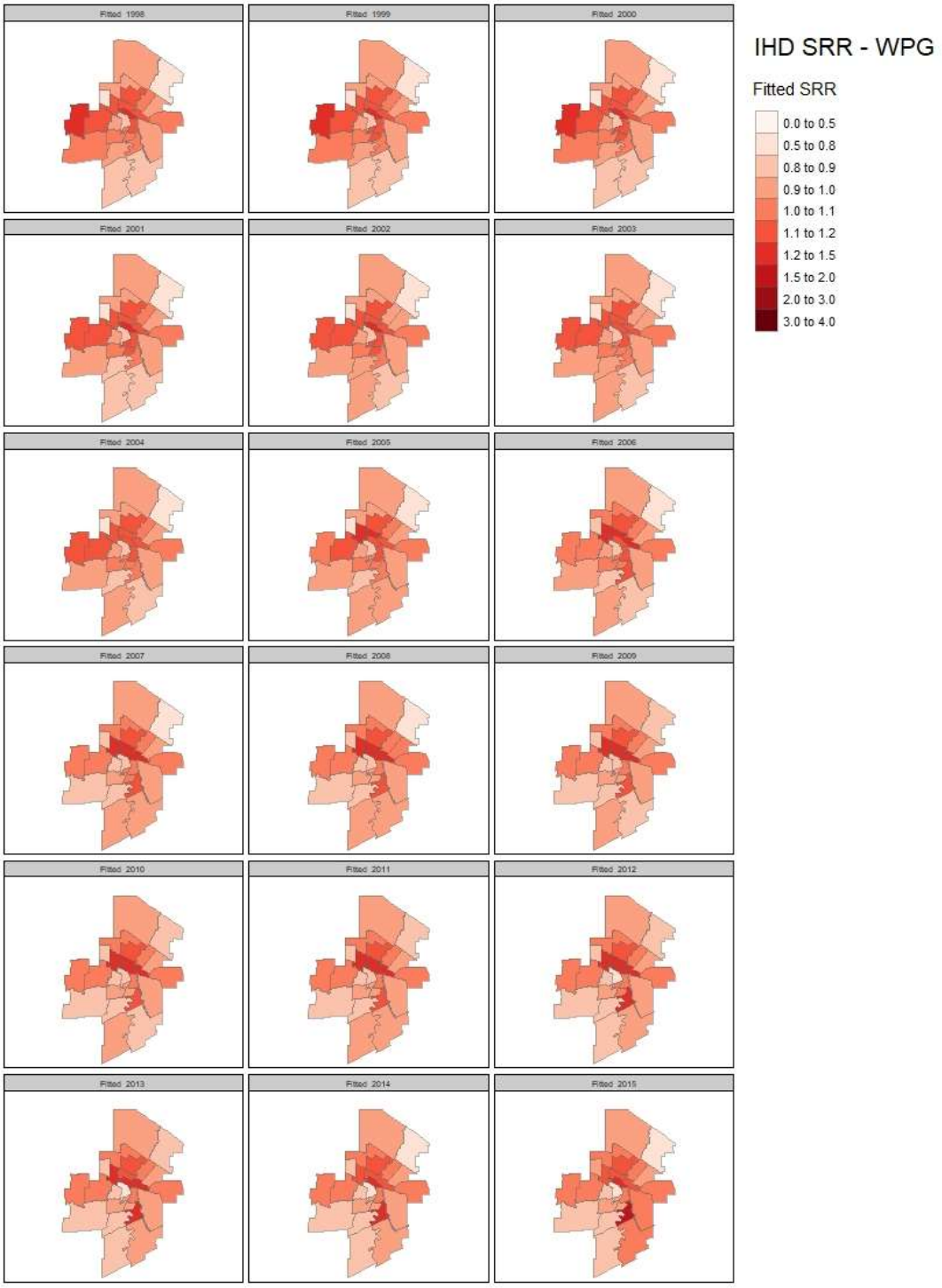

Figure 4. 28: Fitted/smoothed IHD SRR's by year for the overall data. Fitted with the model in table 4.9.

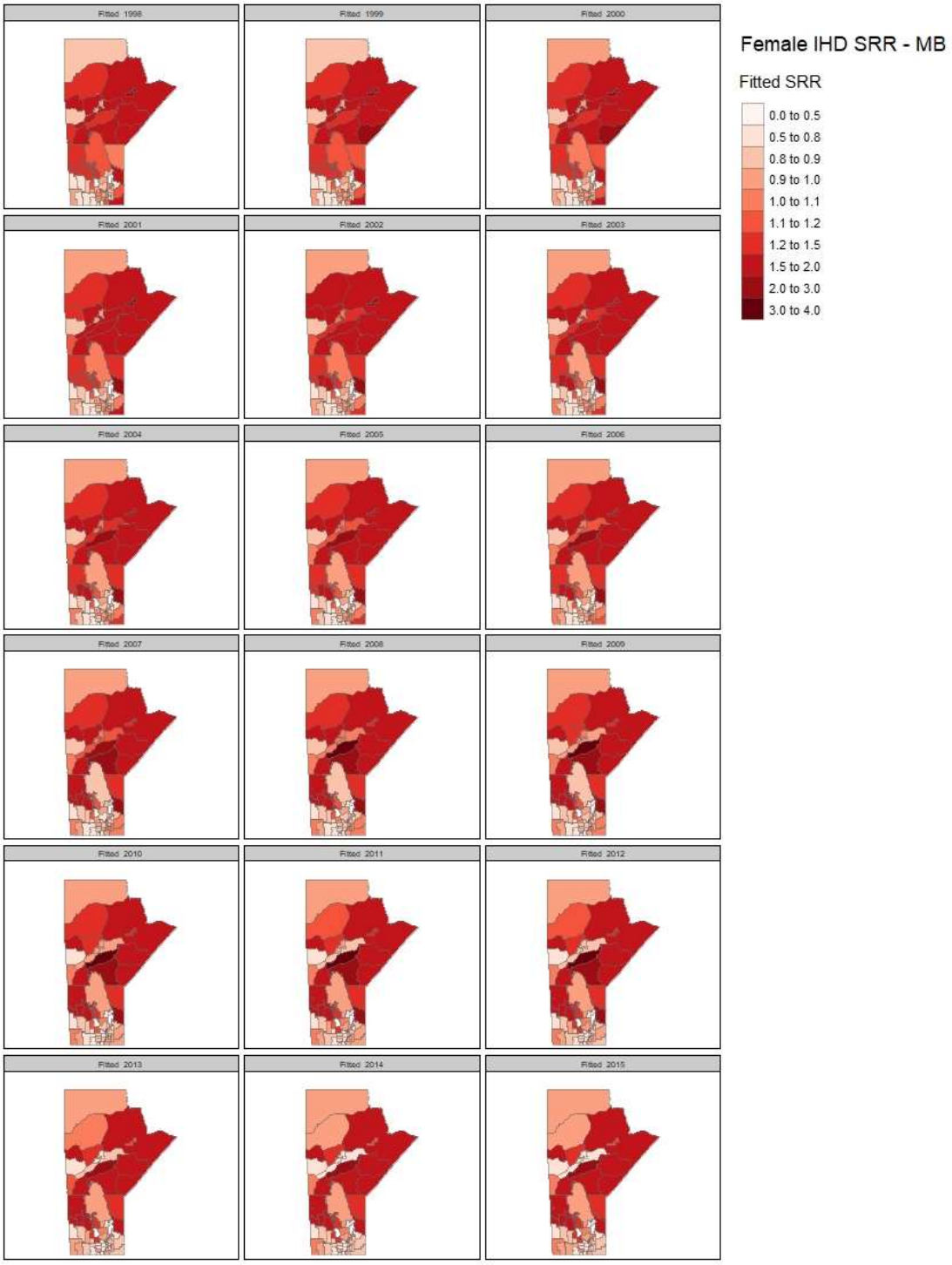

Figure 4. 29: Fitted/smoothed IHD SRR's by year for the female data. Fitted with the model in table 4.10.

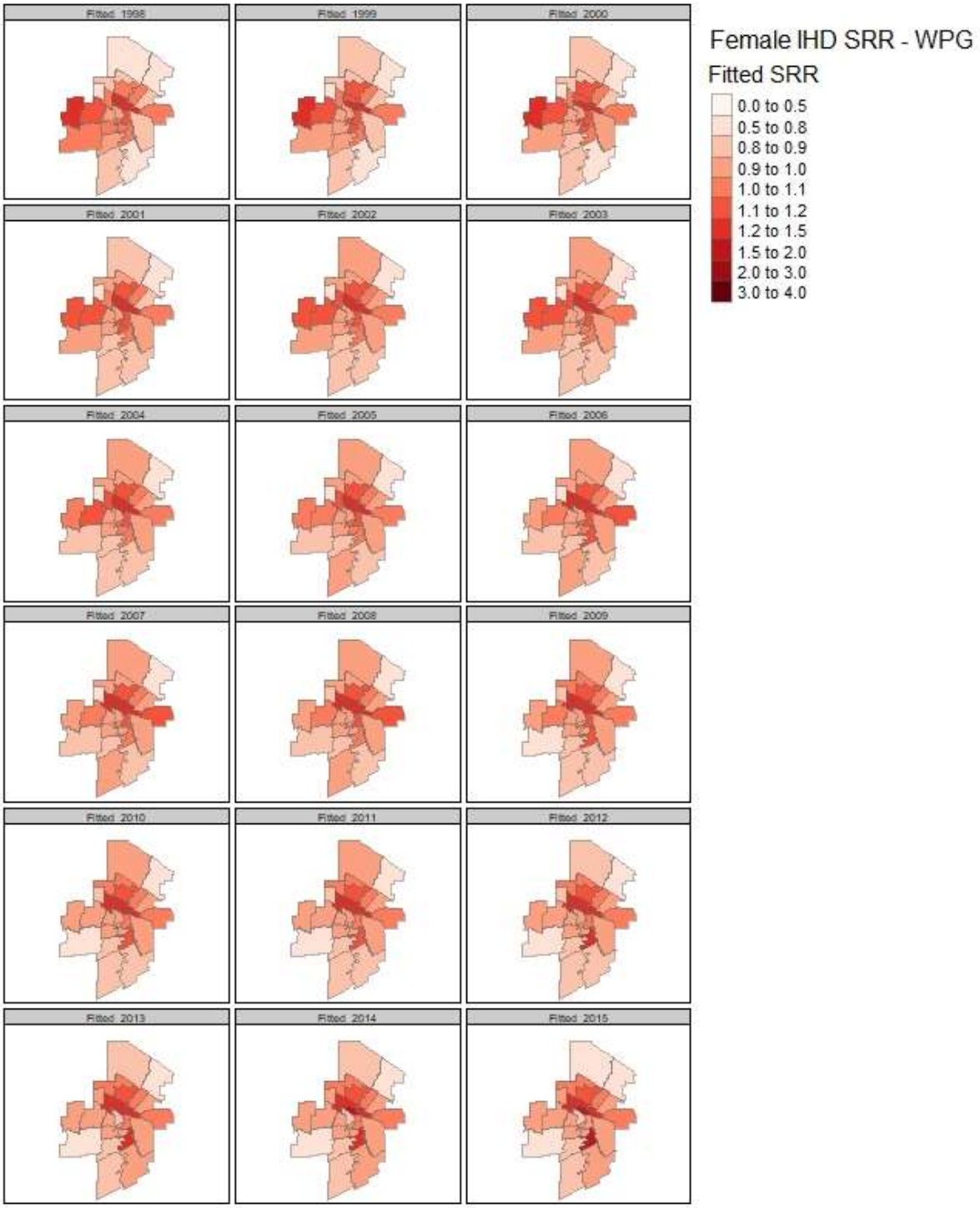

Figure 4. 30: Fitted/smoothed IHD SRR's by year for the female data. Fitted with the model in table 4.10.

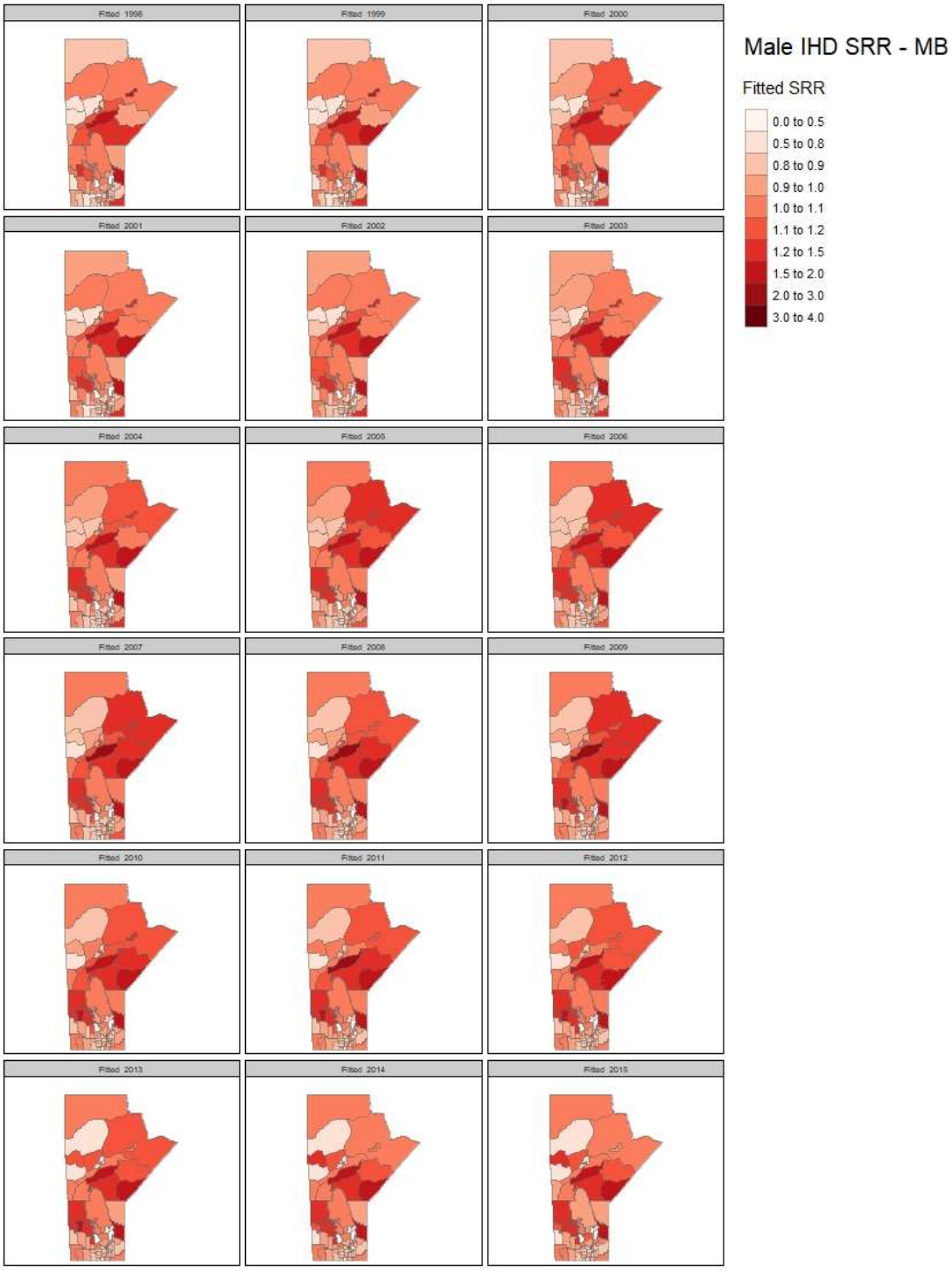

Figure 4. 31: Fitted/smoothed IHD SRR's by year for the male data. Fitted with the model in table 4.11.

Figure 4. 32: Fitted/smoothed IHD SRR's by year for the male data. Fitted with the model in table 4.11.

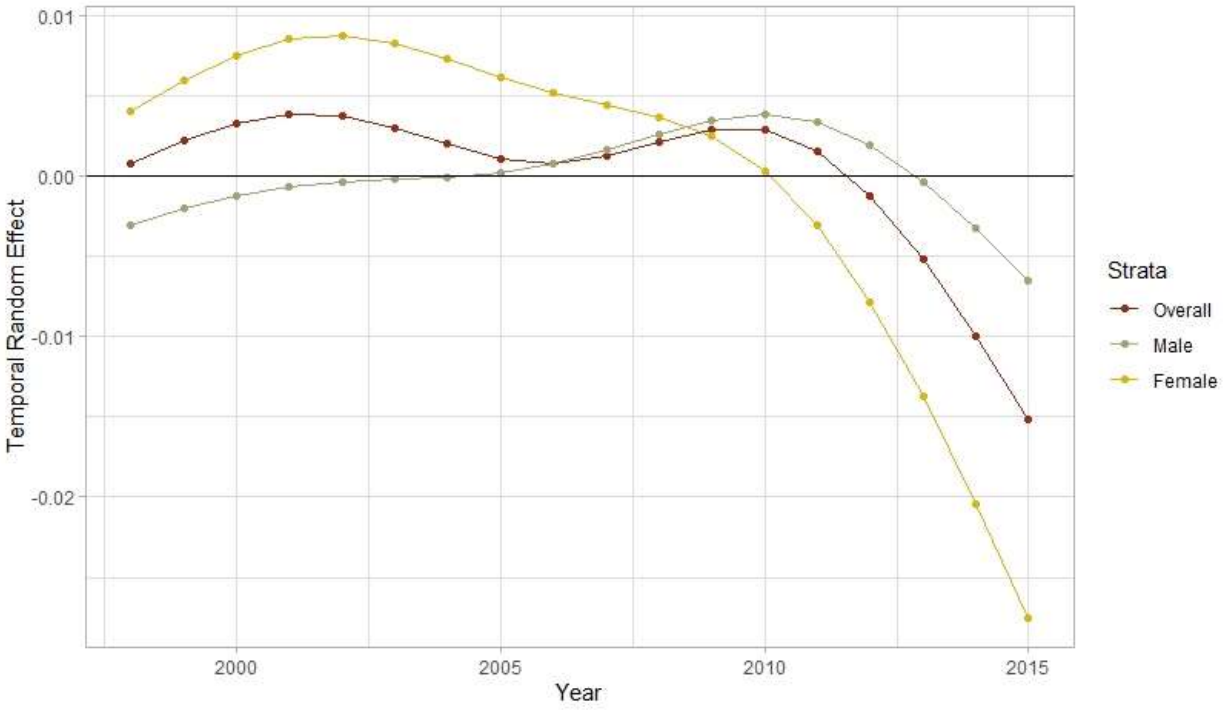

Figure 4. 33: Temporally structured random effect over time. Obtained from the models in tables 4.9-4.11.

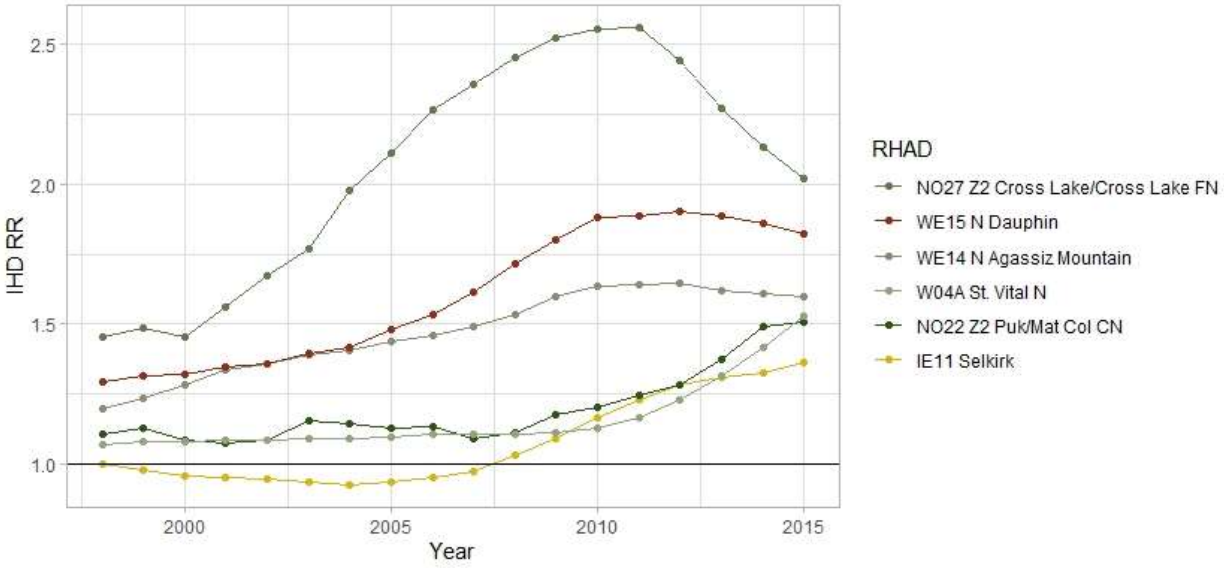

Figure 4. 24: Model fitted IHD RR for six RHADs that had the largest increases over the time period (1998-2015). This is fitted with the model from table 4.9. Note that the legend order is the order of the RR's for each RHAD for 2015 (highest to lowest). Each of these RHADs had a 30% or larger increase in IHD RR from 1998 to 2015.

Chapter 5 – Discussion & Conclusions

This research project has been focused on AMI incidence and IHD prevalence for the 40-85 age group in Manitoba. The main objective was to assess whether there were any spatial and/or spatio-temporal patterns in this disease in Manitoba, and if so, identify some probable causes for these patterns. Part of this objective was to identify potential clusters of one of the more acute outcomes of IHD, that of AMIs. To evaluate these objectives, the first method employed was to use the FSS to identify clusters or hotspots of AMI disease incidences. Next, the FSS was used to explore the IHD data for 2015 and check whether some of the same patterns could be identified for the chronic condition as were discovered for the acute condition. The next step was to model the IHD data with the Poisson and spatial Poisson regression models to uncover any patterns and to identify potential drivers of these patterns. Finally, the IHD data from 1998 to 2015 was modelled using a spatio-temporal Poisson regression model to uncover any temporal and space-time interaction patterns, and again, identify any potential causes of these patterns.

The first part of this chapter is dedicated to discussing the results obtained from the cluster detection and regression modelling outlined in previous chapters. Here, the results are interpreted and evaluated against the current literature pertaining to AMI's and IHD. The next section discusses the strengths and limitations of the data and the analysis. This is followed by an implications and final conclusions section.

5.1 – Discussion

5.1.1 – AMI Cluster Detection

The first main objective of this research project was to identify clusters of one of the acute forms of IHD, that of AMI's within Manitoba. The application of the FSS cluster detection method produced one primary cluster and eight secondary clusters of AMI incidences for the 2011-2016 time-frame. To evaluate the first research question's hypotheses, a visual comparison between each cluster and the geographical distributions of the SEFI-2 variable, the indigenous population, and remote and urban regions is done.

For the primary cluster three out of the seven regions in the cluster also are regions with lower levels of SES than the average in Manitoba. Four out of seven regions have large indigenous populations (>30% indigenous population proportion), and none of the regions are either remote or considered urban. This indicates that judging based on this cluster alone, that the evidence is rather inconclusive with regards to the research hypotheses for this question.

For the secondary clusters north of the primary cluster, the regions within these clusters are almost all associated with more socio-economic deprivation than the average. Only three regions are exceptions to this, these being NO11, NO12, and NO14 which do not have higher levels of socio-economic deprivation. The regions within these clusters also have higher than 30% indigenous populations, with only NO11 not being included in this category, where the proportion is still above the provincial proportion of 18% at 24%.

Secondary clusters east and south of the Primary cluster and not within Winnipeg are less conclusive. These clusters are not associated with higher levels of socio-economic deprivation,

higher indigenous populations, urban regions, or remote regions. Thus, for these we cannot say that there is any evidence to support any of the hypotheses.

The secondary cluster within Winnipeg, does seem to be associated with both more socio-economic deprivation and higher indigenous populations with three out of five regions in this cluster being associated with higher levels of socio-economic deprivation, and one out of five being associated with higher indigenous populations; the only region with higher than the average indigenous population within Winnipeg. Obviously, this cluster is not in a remote location but rather is within Winnipeg, so can be said to be associated with urbanicity but not remoteness.

Overall, six out of nine clusters look to be associated with higher than average socio-economic deprivation and higher than average indigenous populations. Thus, it could be said that hypotheses a and b for research question 1 are in fact true. There are three clusters in the Southern and Interlake-Eastern RHA's that do not seem to be associated with these variables however, where these should be taken into consideration. The urban and remote regions both look to be less associated with the AMI clusters, and thus hypotheses c and d are most likely false.

Unfortunately, the FSS does not provide an explicit measure of association between predictor variables and the clusters, so these conclusions should be examined with caution. In terms of the first main objective, clusters have been clearly identified and are shown in figure 4.9.

5.1.2 – Spatial Patterns of IHD Prevalence

The second main objective of this research was to identify any spatial patterns of IHD prevalence within Manitoba. To do this, exploration of the 2015 IHD data was done using the FSS, where one primary and four secondary clusters were identified for the overall data, and one primary and three secondary clusters were identified for each of the stratified datasets. It is then possible to visually compare these cluster maps with the clusters identified for AMI's to see whether these are occurring in similar locations. It can be seen that the primary clusters for the IHD datasets are in similar locations to the primary cluster of AMI's, with the primary IHD clusters being just a bit further north than the primary AMI cluster. The secondary clusters do seem to be in similar locations, but nothing is very clear, as there are few common borders between these.

To evaluate the hypotheses for the second research question, the spatial Poisson regression model that controls for the SEFI-2 covariate is used. Table 4.4 shows this model for the overall data and tables 4.6 and 4.7 show the models for the sex-stratified data. For all three models, the hypertension and indigenous measures are significant predictors in the model for IHD even after controlling for SEFI-2, but urbanicity and SEFI-2 are not. Between the datasets there is little variation between these measures of association, where any variation is within error. This shows that out of the covariates selected for modelling IHD prevalence, there is little difference in how they predict IHD across sexes with both male and female groups having similar estimates for each significant variable in the models. Here a 10% increase in hypertension risk within a region increases the risk for IHD by 9.3% (95% CI: 5.8%-13.0%) for the overall population, 8.7% (95% CI: 4.4%-13.2%) for the female population, and 8.4% (95% CI: 5.1%-11.9%) in the male population when holding the other covariates in the model constant.

A 10% relative increase in the indigenous population proportion covariate within a region increases the risk for IHD by 0.7% (95% CI: 0.1%-1.3%) in the overall population, 1.0% (95% CI: 0.2%-1.8%) in the female population, and 0.6% (95% CI: 0.0%-1.2%) in the male population while holding all other covariates constant.

The indigenous population proportion covariate had moderate confounding with the SEFI-2 variable, dropping towards zero by 16%. But since the indigenous covariate is still significant, it implies that there is still some variation in IHD prevalence that is being explained by the indigenous population covariate and not the SEFI-2 variable in these models. This confirms other findings, as indigenous populations in Canada have been shown to have worse health outcomes overall, where CVD is no exception (Tobe et al, 2015; Public Health Agency of Canada, 2018, pg. 195). As SES has also been shown in the literature to significantly contribute to worse health outcomes (The Health Officers Council of British Columbia, 2013; Public Health Agency of Canada, 2018, pg. 60), the fact that the indigenous population covariate is significant in the presence of the confounding SES covariate is noteworthy. Especially in Manitoba where 18% of the population are indigenous peoples, this finding is important as it indicates a large population subgroup at higher risk for a chronic and life-threatening disease.

Analyzing the maps of regions with significantly elevated IHD SRR's, it is shown that many of the same patterns identified with the FSS method are present when using the spatial Poisson regression model. This is a good indication of which regions are experiencing far worse outcomes than others. In particular, the regions that are part of the primary clusters of both AMI's and IHD prevalences in all three datasets are also associated with significantly higher prevalence of IHD as obtained from the spatial Poisson regression models. The clusters of IHD

and AMIs within Winnipeg also match up closely to the significantly elevated IHD rate maps. This again appears to be a pattern of IHD and AMI's occurring in similar locations.

These patterns are expected, as chronic IHD can over time lead to more acute events such as AMI's. The interesting regions are the regions for which AMI clusters occurred, but where there are no clusters or significant elevation in IHD SRRs. As most of these regions are located either in the north, the Interlake, and the southwest corner of Manitoba, these could potentially be regions where barriers to healthcare services are occurring. If IHD prevalence is in reality higher in these regions than what has been observed in the data; where a large unobserved portion of cases are potentially going undiagnosed, i.e. we are undercounting the number of persons with IHD in these areas; this could lead to higher rates of AMI's in these regions as the condition has worse outcomes if it has gone untreated (Heart and Stroke Foundation, 2018). If barriers to healthcare services; such as social, cultural, or physical barriers; in these areas are in fact reducing the diagnosis rate of IHD in these regions, this should be identified in order to reduce the risk of acute events such as AMI's.

5.1.3 – Spatio-Temporal Patterns of IHD Prevalence

The third and final objective was to identify any temporal and/or space-time interaction patterns in the IHD data for the 1998-2015 years of data. This was done by extending the spatial Poisson regression model to now incorporate both temporal random effects terms and a space-time interaction random effects term. The model that fit the data best was the one that had a temporally structured and spatially unstructured space-time interaction random effects term. This means that the spatial correlation is not significantly varying throughout time, but the temporal

correlation does vary throughout space. This is most likely due to the regions in Manitoba being very diverse in terms of population demographics, mobility, and population density; where these factors could be influencing the trajectories of IHD prevalence throughout time.

The largest difference in covariate estimates between the spatial regression models and the spatio-temporal models, is that of the now insignificant indigenous population variable. When taking account of the temporal changes in IHD, this covariate now isn't significantly predicting the response. This is likely due to the large changes in IHD prevalence in some regions, whereas the indigenous population proportion covariate is fixed with respect to time. This was chosen because over time Manitoba's population is relatively stationary, rather than using previous census results to construct this variable. Considering that even in the spatial models this variable was only marginally significant, after accounting for temporal changes the indigenous population proportion in a region does not seem to indicate elevated IHD risk.

The only significant covariate in the spatio-temporal models was the measure of hypertension. This covariate varied over both time and space, and so accounted for the variation in the covariate over both dimensions. Although there weren't any significant differences between sexes for this covariate, females did have a slightly higher RR increase due to an increase in hypertension prevalence. Here a 10% increase in hypertension prevalence would predict a 2.9% (95% CI: 1.9%-4.0%) increase in IHD prevalence for the overall population, a 3.4% (95% CI: 2.0%-4.7%) increase in IHD prevalence for the female population, and a 2.6% (95% CI: 1.5%-3.7%) increase in IHD prevalence for the male population. This finding is supported by the literature, where hypertension is a leading risk factor and comorbidity for IHD (Weber et al, 2016, pg. 468; Spinar, 2012, pg. 434), so it was expected to be a predictor of IHD over both the spatial and temporal dimensions. Also note that these predicted model coefficients

are significantly smaller in the spatio-temporal model than in the spatial model for 2015. Thus, when accounting for the temporal changes in IHD and hypertension prevalence, hypertension does not predict as large of an increase in IHD prevalence. Thus, the coefficients in the Poisson regression model and the spatial Poisson regression model could be said to be too "optimistic" in predicting IHD, and that some of the prediction error is now being accounted for in the spatial and temporal random effects.

An interesting characteristic of the fitted spatio-temporal model was also that the spatial covariance structure seemed to shrink in the presence of the temporal covariance structure. This is most likely due to larger changes and more persistent patterns over time than over geographical space. Using smaller aggregation sizes for regions may introduce more spatial auto-correlation, and so if in the future dis-aggregating into smaller areal units is considered helpful, this could come into play more significantly. Regardless, the temporal patterns do vary significantly over space, as indicated by the type II interaction term used in the best fitting model, and so the spatial pattern is such that shouldn't be ignored.

This is especially true when considering figures 4.33 and 4.34. It is clearly seen that the overall provincial trend in IHD is a decreasing one, especially for the female sub-population. However, there are some regions that have substantial increases over the same time period, which indicates that this disease is affecting certain areas of Manitoba more than others. Having now identified these regions, this information could be used to mitigate these patterns in order to more reflect the temporal patterns in other parts of Manitoba.

Evaluation of the third research question is done using the spatio-temporal model coefficients, as well as analyzing the regions identified in figure 4.34. In terms of the spatio-temporal model coefficients, there is no evidence to support hypothesis b or c, as both the

indigenous and urban covariates were insignificant in all three spatio-temporal models. Hypothesis a is supported by the evidence generated by these models, as the spatio-temporal trend in hypertension prevalence does significantly predict the spatio-temporal trend in IHD prevalence. The six regions identified in figure 4.34 are then analyzed closer to see if there are any similarities between them. Table 5.1 shows some descriptive statistics regarding these six regions.

RHAD	Indigenous Pop (%)	SEFI-2	Urban
NO27	98.2	3.08	F
WE15	26.3	0.32	F
WE14	54.1	0.81	F
W04A	13.0	-0.12	T
NO22	96.0	2.75	F
IE11	34.4	0.40	F

Table 5. 1: Covariate values for the six RHAD's that had the largest increases in IHD prevalence over the study period (1998-2015). Indigenous population (%) and SEFI-2 were obtained from the 2016 census. Urban denotes regions within Winnipeg or Brandon.

Out of these six regions, five had higher than the provincial proportion of indigenous peoples, and five had more socio-economic deprivation than the average. There were only two regions that had large measures of socio-economic deprivation, substantially greater than the average, where the other four were very close to the provincial average. The single urban region identified in this group looks to be an odd region in this group, as it has lower indigenous populations and socio-economic deprivation levels than the other regions. Hence, it does appear that there could be some patterns here, as these regions do look to have large indigenous populations aside from the region in Winnipeg. SEFI-2 is still inconclusive and with only one urban region in this group of six, these covariates still appear to not be affecting these increases in IHD prevalence over time.

5.2 – Strengths & Limitations

Some of the main limitations stem from the usage of aggregated data to model IHD prevalence. Here, the underlying continuous process of disease occurrences is discretized into 96 aggregated areal units. By using aggregated data this project is an ecological study and careful consideration of region and population sizes is needed in order to not commit ecological fallacy. This occurs when one or more of the variables of interest gets 'averaged out' within each aggregated region. Here, the region may misrepresent individuals as a central tendency and extremes on either side of the distribution may go unnoticed. However, usage of the autoregressive models described in preceding chapters that smooth the disease rates alleviates the effects of this issue while having the ability to retain small regions. This limitation may still be problematic when considering Northern Manitoba where the individual effect of covariates on the response could be extremely sensitive since the regions are quite large due to low populations. The neighborhood structure in areas such as northern Manitoba will be different than in Winnipeg for example, so the accuracy of the models may be less precise when compared to other regions in Manitoba.

The ecological aspect of this study also influences how we can talk about the covariates, and their relationship to the response in these models. As hypertension, indigenous status, or SEFI-2 values were not linked at the individual level to the response variables we can not say that causation at the individual level occurs. All we can say is that if the variables are significant in the models, is that these covariates are associated to the response at the areal level. As we do not measure whether a person is first diagnosed with hypertension, and then IHD, we also can not say that at the individual level that hypertension predicts future onset of IHD. What we can say however, is that regions with higher hypertension prevalence are associated with higher

levels of IHD prevalence. Here, the data points are areal units not individuals, so inferences about the direct association at the individual level is not advised.

The number of areal units used in the models was also a bit limiting. As IHD is still a relatively rare disease when looking at our cohort of the 40-85 age group, the areal units needed to be large enough to contain enough counts for each year of data. There are methods for dealing with low counts and zero-inflation so this could be mitigated to some extent, but the assumption is still that the log-response in the models is distributed normally. As the 96 areal units used in this study were selected both for this reason, as well as because health service delivery is generally administrated considering these regions; a case could be made for using smaller regions. The disease counts in most areal units were well above the minimal threshold, so disaggregating further could have produced higher spatial resolution. This would also have allowed for more covariates to be used in the modelling, as 96 observations is fairly limiting in terms of the number of covariates that could be used without running the risk of over-fitting the model(s). The final decision to not disaggregate further for this research project was made to preserve the ability to compare the results in this project to other studies and reports that have been extensively generated using these RHADs with this method of aggregation.

The selection of urbanicity as a surrogate variable for pollution is undoubtedly over simplistic. Not having an adequate covariate for pollution is a main weakness for this study as pollution has been shown to be associated with CVD in previous studies (Pun et al, 2014; Yoo et al, 2018). As regions in Winnipeg or Brandon would naturally have variation in pollution levels between them as well as between regions outside of these cities, much of this information is lost when we only use a binary variable categorizing regions as urban or not. This is most likely the reason why this study did not corroborate the literature and produce significant results regarding

an association between urbanicity/pollution and IHD prevalence levels in either the spatial or spatio-temporal case.

Another aspect of the modelling that will need to be considered, especially when looking at expected versus observed numbers of diagnoses, is that of the quality of the administrative data itself. It has been shown that non-fee-for-service physicians, specifically shadow billed physicians, are much less likely to report their patient's records as accurately as their fee-for-service counterparts (Alshammari & Hux., 2009, pg. 474). Alshammari & Hux were able to show that for diabetes, the reporting done by non-fee-for-service physicians was biased towards extreme cases (Alshammari & Hux, 2009, pg. 474). Although it is more important to capture these cases, the incompleteness of the data could result in biasing the estimated disease rates for these regions towards the average when in fact they could be elevated. As the trend has been to increase the numbers of physicians on non-fee-for-service pay schedules in Northern Manitoba, these regions could end up having biased estimates. This problem will need to be overcome however, as this project is mainly interested in areas where access to health care services is more of an issue, i.e. in northern and remote Manitoba.

Finally, another main limitation stems from not having data on undiagnosed cases of IHD for individuals who have had mortality due to IHD. This data; found in the vital statistics dataset; was not available for this research project, and so only non-mortal cases of IHD were considered. This could be concerning when assessing IHD prevalence or AMI incidences in Northern Manitoba, as physical barriers to healthcare services may mean that more persons are living with undiagnosed IHD in these locations. If IHD goes undiagnosed the risk for death increases, where these cases would not be captured in this study.

The biggest strength of this research is the ability of the models to capture the spatial covariance between areal units in the spatial models, as well as the spatial, temporal, and space-time interaction effects in the spatio-temporal models. Modelling the spatial and temporal covariance structures addresses the issue of non-heteroscedastic residuals which occurs when ignoring the spatial correlation of disease prevalence. These models produce far more accurate measures of association between the covariates and the response as well, as it was shown that when the above assumptions are not met that these measures were much too optimistic in prediction magnitude.

The models also produce more precise estimates of the IHD SRRs for each region. This smoothing shrinks the inherent variability of the observed estimates within each region by borrowing information from the neighboring regions. Modelling these covariance structures also produces a higher degree of accuracy for the identification of areal units that had significantly elevated IHD prevalence. This is important, especially when analyzing the results from regions with lower populations, as the variability in the predicted IHD SRRs could be too large to make conclusions without shrinking this variance using the autoregressive models.

Another main strength is the usage of a model-based approach for the misaligned data problem between the administrative health and the census data. As the PCCF is only accurate for 94% of the Canadian postal codes, it means that up to 6% of geographical space could be missed entirely (Statistics Canada, 2015). The model-based approach outlined in Chapter 3 has 100% coverage, as well as being stochastically defined so that measures of error are readily produced for analysis. Thus, the same variance shrinking done in the main models is done here, where these estimates could be said to be more precisely measuring the real values within each of the areal units. This reduces the risk of measurement error, as well as providing more accurate

measures of association between the census defined covariates and the response for the main models.

5.3 – Future Directions

This research project provides an initial starting point for spatial analyses of IHD in Manitoba. Future studies could start by including vital statistics data in the analyses for more completeness of the case event data. Also, other covariates such as comorbidities could be used in the analysis, as for example diabetes has been shown to be related to IHD (Mayo Clinic, 2018). A covariate that defines pollution more completely, rather than just an urbanicity variable would also be useful, as it would be interesting to see if different types of pollution affect the development of IHD to differing extents. As the current climate change issue is becoming more important, discovering how our changing climate or the pollution that is associated with climate change affects the development of chronic diseases like IHD could be very important to explore in the future. Regarding the mapping of disease rates and the detection of spatial patterns of IHD, future studies explore other aggregation structures that increase the resolution of the maps, in order to better understand how the patterns are structured.

5.4 – Implications & Conclusions

5.4.1 - Implications

The results produced by this research project could be useful for health system policy makers and planners when making decisions about health care delivery. The identification of

AMI clusters provides valuable evidence for where there are significantly more AMIs occurring than in the rest of the province. Some of these clusters may come with little surprise, but others may indicate regions that would have otherwise been overlooked. This along with the identification of IHD clusters and areal units with significantly elevated IHD prevalence, can provide a better picture of where IHD is affecting Manitobans with greater magnitude than the rest of the province. Also, the identification of regions whose temporal trends were counter to the provincial decreasing trend of IHD prevalence, could aid in evaluating where certain policies that were designed to decrease heart disease are working and where more work is needed. As reducing health disparities among Manitobans is one of the six priorities of Manitoba Health, Seniors & Active Living (Manitoba Health Senior & Active Living, 2016), identifying areas where one of the most deadly diseases is more prevalent than in others should be important. This evidence directly reflects this priority and could be used to back health care policies that aim to reduce disparities among the Manitoban population.

As it was shown that after accounting for SES and the spatial and temporal covariance structures that hypertension prevalence is significantly associated with IHD prevalence, policies to reduce hypertension could indirectly also affect IHD. As hypertension prevalence has been shown to be increasing for Manitobans in recent years (MB Health, 2017, pg. 24, 28; Fransoo et al., 2013, pg. 69, 93, 109), funding programs to mitigate this trend could result in better outcomes downstream when it comes to IHD or other CVD conditions.

Another of Manitoba Health, Seniors & Active Living's priorities is to "Lead advances in health service delivery with First Nations, Métis, and Inuit Manitobans, through policy and programs with a focus on prevention, primary health care, public health, and education." (Manitoba Health Senior & Active Living, 2016). As the spatial models for the 2015 data

showed that regions with larger indigenous populations had more risk of having elevated IHD prevalences as well as five out of the six regions with the largest increases over time being regions with high indigenous populations, this research points towards a health outcome where disparities between indigenous populations and the rest of Manitobans exist. Thus, to meet this priority, a reduction in IHD among the indigenous population in Manitoba is necessary. This research provides information on which regions policy could be focused on and prioritized to achieve the greatest impact.

The research products that have been and will be produced by this research project will be useful from a policy lens as well as from an ecological health research perspective. The method for dealing with spatially misaligned data in section 3.4.4 is novel in the Manitoban context and is generalizable to other diseases and other Canadian geographies or provinces. Currently under development is a R-package that will allow users to convert a census covariate of interest using the method outlined in section 3.4.4 to the RHAD level of aggregation from the census DA level of aggregation. Since early simulation results have shown that this method is substantially better than the one currently employed by many researchers conducting ecological studies in Manitoba, this has the potential to improve the accuracy of the census variables that are used in these research projects.

Health surveillance of chronic diseases will be increasingly important as new technology and analysis methods become available. Surveillance projects are helpful to policy analysts and decision makers in the healthcare system, as they provide up-to-date evidence regarding health outcomes within the Province. This allows for better resource allocation, better preparedness for the future of our health system, and an increased probability for avoiding an epidemic before it occurs. The results from this research project could be a start to an ongoing surveillance project

that captures new clusters of IHD prevalence or AMI incidence as they occur, or track the temporal trends of IHD prevalence as they develop throughout Manitoba.

An example of a user-friendly application that could be used as a surveillance tool has been developed which maps the fitted IHD prevalence using the spatio-temporal models from sections 3.4.3 and 4.4. This allows for a more complete set of model visualizations, as the maps in this application can be zoomed in to any specific location, temporal trends for any region can be analyzed (not just the six selected in sections 4.4 and 5.1.3), any year can be selected to map, and exceedance probabilities (the probability that a region has a SRR greater than 1) can be set to any value the user desires. This will allow for policy decision makers to discover the results of this project for themselves, allowing for focused insight into aspects that are required for the policy in question which will increase the likelihood that the results are considered in the policy making process. Not only could this tool be valuable in the IHD and AMI context, but it could also be generalized to other diseases if necessary, and would be easily updated as new and more recent data becomes available.

5.4.2 - Conclusions

This research project was able to identify clusters of AMI incidences, as well as model IHD prevalence throughout space as well as through the time period of 1998-2015. The associations between SES, indigenous populations, remoteness, or urbanicity to the AMI clusters were largely inconclusive. Minor associations between SES and indigenous populations to AMI clusters were noted as there were some visual patterns in the results.

Accounting for the spatial and temporal covariance structures allowed for better model specifications and hence more accurate results than using the naïve Poisson regression model. It was found that when accounting for the spatial covariance structure, the 2015 data showed significant associations between IHD prevalence and the two covariates of hypertension prevalence and indigenous population proportion even when accounting for SES confounding. The results were within error the same for males and females when stratifying the data by sex.

There were some spatial disease patterns that were occurring for both IHD prevalence and AMI incidences as there were a few shared borders between AMI clusters, IHD clusters, and regions with elevated IHD prevalence. This indicates that the same regions are experiencing worse health outcomes for both the chronic and the acute case of IHD, which is expected as the chronic form of IHD can often lead to more adverse events. Central-west Manitoba in particular had worse outcomes for both AMI and IHD than the rest of the province.

When accounting for the temporal and space-time effects along with the spatial effect, it was found that the indigenous population proportion was no longer significant in the models. Here again, the differences in model predictions were within error the same for both males and females. Thus, overall it appears that out of the chosen study covariates, there was little difference in how these affected IHD prevalence in either males or females. Females had more regions within their AMI clusters and also more regions that had elevated IHD prevalence than males. However, overall males still had larger prevalences than females, where the differences between them could be explained by males possibly having more concentrated prevalences in fewer areas than females.

The silver lining in this research is that the previous work done which showed that the overall trend in IHD prevalence has been decreasing over time is confirmed (Fransoo et al.,

2013, pg. 93). However, there are some nuances to this finding. It appears that the decrease was mostly in the female data, where overall males have been relatively constant throughout the study period. Also, as the space-time interaction effect was the most significant, it shows that the temporal patterns in IHD prevalence varies significantly throughout space. The regions that had the largest increase in IHD prevalence over the time period were shown in figure 4.34, where further investigation into these proved to show little similarities regarding SEFI-2 or urbanicity, but five of these six regions did show larger than average indigenous population proportions. Thus, further investigation into the main drivers behind these increases, and specifically why indigenous populations have much larger increases of IHD prevalence over time within these regions should be done in order to reverse this trend and decrease the risk of it occurring in other regions.

References

Agarwal D., Gelfand A., Silander J., (2002), Investigating tropical deforestation using two-stage spatially misaligned regression models, *Journal of Agricultural, Biological, and Environmental Statistics 7,* (3), 420-439.

Alshammari A. M., Hux, J. E., (2009), The Impact of Non-Fee-For-Service Reimbursment on Chronic Disease Surveillance Using Administrative Data, *Canadian Public Health Association 100*, (6), 472-474

Anderson P., (2006), Global use of alcohol, drugs and tobacco, *Drug and Alcohol Review 25,* 489-502.

Ashley K. E., Geraci S. A., (2013), Ischemic heart disease in women, *Southern Medical Journal 106,* (7) 427-433.

Barcelo M. A., Saez M., Saurina C., (2009), Spatial variability in mortality inequalities, socio-economic deprivation, and air pollution in small areas of the Barcelona metropolitan region, Spain, *Science of the Total Environment 407,* 5501-5523.

Baptiste A., (2017), gridExtra: Miscellaneous functions for "Grid" graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra

Batisse E., Goudreau S., Baumgartner J., Smargiassi A., (2017), Socio-economic inequalities in exposure to industrial air pollution emissions in Quebec public schools, *Canadian Journal of Public Health 108,* (5/6) 503-509.

Berger S. J., Elliot L., Gallup D., Roe M., Granger C. B., Armstrong P. W., Simes R. J., White H. D., Van de Werf F., Topol E. J., Hochman J. S., Newby L. K., Harrington R. A., Califf R. M., Becker R. C., Douglas P. S., (2010), Sex differences in mortality following acute coronary syndromes, *JAMA 302,* (8) 874-882.

Besag J., (1974), Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society. Series B (Methodological) 36,* (2) 192-236.

Besag J., York J., Mollie A., (1991), Bayesian image restoration with two applications in spatial statistics, *Annals of the Institute of Statistical Mathematics 43,* 1-20.

Bivand R., Keitt T., Rowlingson B., (2019). rgdal: Bindings for the 'geospatial' data abstraction library. R package version 1.4-4. https://CRAN.R-project.org/package=rgdal

Bivand R., Lewin-Koh N., (2019). maptools: Tools for handling spatial objects. R package version 0.9-5. https://CRAN.R-project.org/package=maptools

Bivand R., Rundel C., (2019), rgeos: Interface to geometry engine - Open Source ('GEOS'). R package version 0.4-3. https://CRAN.R-project.org/package=rgeos

Bivand R. S., Wong D. W. S., (2018) Comparing implementations of global and local indicators of spatial association, *TEST 27,* (3) 716-748. https://doi.org/10.1007/s11749-018-0599-x

Blangiardo M., Cameletti M., (2015), *Spatial and Spatio-temporal Bayesian Models with R-INLA,* West Sussex, UK: John Wiley & Sons Ltd.

Caroll R., Lawson A. B., Faes C., Kirby R. S., Aregay M., Watjou K., (2016), Spatio-temporal Bayesian model selection for disease mapping, *Environmetrics 27,* 466-478.

Canadian Institute for Health Information, (2018). Patient cost estimator, Ottawa: Canadian Institute for Health Information; Retrieved from: https://www.cihi.ca/en/patient-cost-estimator

Chateau D., Metge C., Prior H., Soodeen R., (2012), Learning from the census: The socio-economic factor index (SEFI) and health outcomes in Manitoba, *Canadian Journal of Public Health 103,* (2) 23-27.

Domınguez-Berjon M. F., Gandarillas A., Segura del Pozo J., Zorrilla B., Soto M. J., Lopez L., Duque I., Marta M. I., Abad I., (2010), Census tract socioeconomic and physical environment and cardiovascular mortality in the region of Madrid (Spain), *J Epidemiol Community Health 64,* 1086-1093.

Du Q., Zhang M., Li Y., Luan H., Liang S., Ren F., (2016), Spatial patterns of ischemic heart disease in Shenzhen, China: A Bayesian multi-disease modelling approach to inform health planning policies, *International Journal of Environmental Research and Public Health 13,* 436, doi:10.3390

Du Z., Hao Y., (2018). FlexScan: Flexible scan statistics. R package version 0.1.0. https://CRAN.R-project.org/package=FlexScan

Eyre E. L. J., Duncan M. J., Birch S. L., Cox V., Blackett M., (2014), Physical activity patterns of ethnic children from low socio-economic environments within the UK, *Journal of Sports Sciences 33,* (3) 232-242.

Gamerman D., Lopes H. F., (2006), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference,* Boca Raton, FL: Chapman & Hall/CRC.

Goodman A., Wilkinson P., Stafford M., Tonne C., (2011), Characterising socio-economic inequalities in exposure to air pollution: A comparison of socio-economic markers and scales of measurement, *Health & Place 17,* 767-774.

Finegold J. A., Asaria P., Francis D. P., (2013), Mortality from ischaemic heart disease by country, region, and age: Statistics from World Health Organization and United Nations, *International Journal of Cardiology 168,* 934-945

Fransoo R, Martens P, *The Need to Know Team*, Prior H, Burchill C, Koseva I, Bailly A, Allegro E., (2013), *The 2013 RHA Indicators Atlas*. Winnipeg, MB: Manitoba Centre for Health Policy

Hart J. E., Puett R. C., Rexrode K. M., Albert C. M., Laden F., (2015), Effect modification of long-term air pollution exposures and the risk of incident cardiovascular disease in US women, *Journal of the American Heart Association*, doi: 10.1161/JAHA.115.002301

Heart and Stoke Foundation of Canada, (2018), *Coronary Artery Disease,* retrieved from: https://www.heartandstroke.ca/heart/conditions/coronary-artery-disease.

Huckle T., You R. Q., Casswell S., (2010), Socio-economic status predicts drinking patterns but not alcohol-related consequences independently, *Addiction 105,* 1192-1202.

Kauhl B., Maier W., Schweikart J., Keste A., Moskwyn M., (2018), Exploring the small-scale spatial distribution of hypertension and its association to area deprivation based on health insurance claims in Northeastern Germany, *BMC Public Health 18.*

Kihal-Talantikite W., Weber C., Pedrono G., Segala C., Arveiler D., Sabel C. E., Deguen S., Bard D., (2017), Developing a data-driven spatial approach to assessment of neighborhood influences on the spatial distribution myocardial infarction, *International Journal of Health Geographics 16,* (22) doi: 10.1186/s12942-017-0094-8

Kirby R. S., Liu J., Lawson A. B., Choi J., Cai B., Hossain M., (2011), Spatio-temporal patterning of small area low birth weight incidence and its correlates: A latent spatial structure approach, *Spatial and Spatio-temporal Epidemiology 2,* 265-271.

Kjaerulff T. M., Ersboll A. K., Gislason G., Schipperjin J., (2016), Geographical clustering of incident acute myocardial infarction in Denmark: a spatial analysis approach, *Spatial and Spatio-Temporal Epidemiology 19,* 46-59.

Kulldorff M., Nagarwalla N., (1995), Spatial disease clusters: detection and inference, *Statistics in Medicine 14,* (8) 799-810.

Lawson A. B., (2013), *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*, Boca Raton, FL: Chapman & Hall/CRC

Lindgren F., Rue H., (2015), Bayesian spatial modelling with R-INLA. *Journal of Statistical Software 63,* (19) 1-25. URL http://www.jstatsoft.org/v63/i19/.

Lix L., Yogendran M., Burchill C., Metge C., McKeen N., Moore D., Bond R., (2006), *Defining and validating chronic diseases: An administrative data approach,* Winnipeg, MB: Manitoba Centre for Health Policy

Lix L., Yogendran M., Mann J., (2008), *Defining and validating chronic diseases: An administrative data approach. An update with ICD-10-CA,* Winnipeg, MB: Manitoba Centre for Health Policy.

Lunn D.J., Thomas A., Best N., Spiegelhalter, D., (2000) WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing 10*, 325–337.

Manitoba Health, Seniors and Active Living, (2017). *Annual Statistics 2015-2016,* Retrieved from: https://www.gov.mb.ca/health/annstats/index.html

Manitoba Health, Seniors and Active Living, (2016). *Vision and Mission,,* Retrieved from: https://www.gov.mb.ca/health/documents/priorities_goals.pdf

Mayo Clinic, (2018), *Myocardial Ischemia,* retrieved from: https://www.mayoclinic.org/diseases-conditions/myocardial-ischemia/symptoms-causes/syc-20375417

Mayo Clinic, (2018), *High blood pressure (hypertension),* retrieved from: https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/symptoms-causes/syc-20373410

Medrano M. J., Boix R., Palmera A., Ramis R., Galan I., Lopez-Abente G., (2012), Towns with extremely low mortality due to ischemic heart disease in Spain, *BMC Public Health 12*

Mugglin A., Carlin B., Gelfand A., (2000), Fully model-based approaches to spatially misaligned data, *Journal of the American Statistical Association 95,* (451), 877-887.

O'Donoghue G., Perchoux C., Mensah K., Lakerveld J., Van der Pleog H., … Nazare J., (2016), A systematic review of correlates of sedentary behaviours in adults aged 18-65 years: a socio-ecological approach, *BMC Public Health 16,* (163).

Pedigo A., Aldrich T., Odoi A., (2011), Neighborhood disparities in stroke and myocardial infarction mortality: a GIS and spatial scan statistics approach, *BMC Public Health 11*

Plummer M., Best N., Cowles K., Vines K., (2006). CODA: Convergence diagnosis and output analysis for MCMC*, R News 6*, 7-11

Public Health Agency of Canada, (2018), *Key health inequalities in Canada: A national portrait,* Retrieved from: https://www.canada.ca/en/public-health/services/publications/science-research-data/understanding-report-key-health-inequalities-canada.html

Public Health Agency of Canada, (2018), *Report from the Canadian chronic disease surveillance system: Heart disease in Canada 2018,* Retrieved from: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/report-heart-disease-Canada-2018.html

Public Health Agency of Canada, (2017), *Heart Disease in Canada: Highlights from the Canadian Chronic Disease Surveillance Program,* Retrieved from: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/heart-disease-canada-fact-sheet.html

Pun V. C., Yu I. T., Ho K., Qiu H., Sun Z., Tian L., (2014), Differential effects of source-specific particulate matter on emergency hospitalizations for ischemic heart disease in Hong Kong, *Environmental Health Perspectives 122,* (4) 391-396.

QGIS Development Team, (2019), QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

R Core Team (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, URL https://www.R-project.org/.

Ram K., Wickham H., (2018), wesanderson: A Wes Anderson palette generator. R package version 0.3.6. https://CRAN.R-project.org/package=wesanderson

Reid J. L., Hammond D., Driezen P., (2010), Socio-economic status and smoking in Canada, 1999-2006: Has there been any progress in disparities in tobacco use? *Canadian Journal of Public Health 101,* (1) 73-78

Rue H., Martino S., (2007), Approximate Bayesian inference for hierarchical Gaussian Markov random field models, *Journal of Statistical Planning and Inference 137,* 3177-3192.

Shimotsu S. T., Jones-Webb R. J., MacLehose R.F., Nelson T. F., Forster J. L., Lytle L. A., (2013), Neighborhood socioeconomic characteristics, the retail environment, and alcohol consumption: A multilevel analysis, *Drug and Alcohol Dependence 132,* 449-456.

Spinar, J., (2012), Hypertension and ischemic heart disease, *Cor et Vasa 54,* 433-438.

Statistics Canada, (2019), *Life expectancy and other elements of the life table, Canada, all provinces except Prince Edward Island*, retrieved from: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310011401

Statistics Canada, (2016), *Aboriginal Peoples: Fact Sheet for Manitoba,* retrieved from: https://www150.statcan.gc.ca/n1/pub/89-656-x/89-656-x2016008-eng.htm

Statistics Canada, (2015), *Postal Code Conversion File (PCCF), Reference Guide, 2013.* Statistics Canada Catalogue no. 92-154-G. https://www150.statcan.gc.ca/n1/pub/92-154-g/2013001/overview-apercu-eng.htm

Sturtz S., Ligges U., Gelman A., (2005), R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software 12,* (3) 1-16.

Subramanyam M. A., James S. A., Diez-Roux A. V., Hickson D. A., Sarpong D., Sims M., Taylor H. A., Wyatt S. B., (2013), Socioeconomic status, John Henryism and blood pressure among African-Americans in the Jackson Heart Study, *Social Science & Medicine 93,* 139-146.

Takahashi K., Yokoyama T., Tango T., (2007), FleXScan user guide for version 2.0, Wako, Japan: National Institute of Public Health.

Tango T., Takahashi K., (2005), A flexibly shaped spatial scan statistic for detecting clusters, *International Journal of Health Geographics 4,* doi:10.1186/1476-072X4-11

Tennekes M., (2018), tmap: Thematic maps in R, *Journal of Statistical Software 84,* (6) 1-39. doi:10.18637/jss.v084.i06.

The Conference Board of Canada (2019), *Mortality Due to Circulatory Diseases*, retrieved from: https://www.conferenceboard.ca/hcp/Details/Health/mortality-circulatory-diseases.aspx?AspxAutoDetectCookieSupport=1#ftn3-ref

The Health Officers Council of British Columbia, (2013), *Health inequalities in BC,* retrieved from: https://www.healthofficerscouncil.net/2013/04/13/health-inequities-in-british-columbia

Tobe S. W., Maar M., Roy M. A., Warburton D. E. R., Preventing cardiovascular and renal disease in Canada's aboriginal populations, *Canadian Journal of Cardiology 31,* 1124-1129.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography 46*, 234–240.

Torabi, M., Green, C., Yu, N., Marrie, R. A., (2014). Application of Three Focused Cluster Detection Methods to Study Geographic Variation in the Incidence of Multiple Sclerosis in Manitoba, Canada. *Neuroepidemiology*, *43,* 38-48.

Vaccarino V., Horwitz R. I., Meehan T. P., Petrillo M. K., Radford M. J., Krumholz H. M., (1998), Sex differences in mortality after myocardial infarction, *Archives of Internal Medicine 158,* (18) 2054-2062.

Waller L. A., Gotway C. A., (2004), *Applied spatial statistics for public health data*, Hoboken, NJ: John C. Wiley and Sons Inc.

Weber T., Lang I., Zweiker R., Horn S., Wenzel R. R., Watschinger B., … Metzler B., (2016), Hypertension and coronary artery disease: epidemiology, physiology, effects of treatment, and recommendations, *The Central European Journal of Medicine 128,* 467-479.

Wickham H., (2017), tidyverse: Easily install and load the 'Tidyverse'. R package version 1.2.1. https://CRAN.R-project.org/package=tidyverse

Yoo E., Brown P., Eum Y., (2018), Ambient air quality and spatio-temporal patterns of cardiovascular disease emergency department visits, *International Journal of Health Geographics 17,* doi: https://doi.org/10.1186/s12942-018-0138-8

Appendix I: Disease Definitions

An AMI incidence was defined as persons being identified in the hospital abstracts database with an ICD-9 or ICD-10 code for AMI as the diagnosis. These codes are:

- ICD-9-CM-410
- ICD-10-CA-I21

IHD prevalence was defined using individuals that had at least one instance of IHD in the hospital abstracts or at least two instances in the medical claims dataset, or a combination of at least one instance in the medical claims dataset and two prescriptions in the DPIN dataset within a five year period. After these persons were identified they were counted in the numerator for prevalence until censored out for either death, movement to outside Manitoba, or aging out of the study. The ICD-9 and ICD-10 codes for the hospital abstract and medical claims datasets are:

- ICD-9-CM-(410-414.9)
- ICD-10-CA-(I20-I25)

The ATC codes for identification in the DPIN dataset are:

- ATC-C01
- ATC-C07
- ATC-C08
- ATC-C09

Hypertension prevalence was defined using individuals that had at least one instance of hypertension in the hospital abstracts, at least one in the medical claims dataset, or at least two in the DPIN dataset with a hypertension related diagnosis or prescription within a one-year time

period. After these persons were identified they were counted in the numerator for prevalence until censored out for either death, movement to outside Manitoba, or aging out of the study. The ICD-9 and ICD-10 codes for the hospital abstract and medical claims datasets are:

- ICD-9-CM-(401-405)
- ICD-10-CA-(I10-I15)

The ATC codes for identification in the DPIN dataset are:

- ATC-C02
- ATC-C03
- ATC-C07
- ATC-C08
- ATC-C09

Appendix II: Region Codes

Region codes for all 96 areal units are given in the below Table. Note that the first two characters of the code denote which RHA the areal unit belongs to. Here IE – Interlake-Eastern Health Authority, NO – Northern Regional Health Authority, SO – Southern Regional Health Authority, WE – Prairie Mountain Health Authority, and W0 and W1 – Winnipeg Regional Health Authority. Whether these regions were considered urban or remote in the analysis is also given in this table.

Code	Region	Urban/Remote
IE11	Selkirk	
IE21	S Stonewall/Teulon	
IE22	S Wpg Beach/St. Andrews	
IE23	S St. Clements	
IE24	S Springfield	
IE31	E Beausejour	
IE32	E Pinawa/LDB	
IE33	E Whiteshell	
IE41	W Gimli	
IE42	W Arborg/Riverton	
IE43	W St. Laurent	
IE51	N Powerview/PF	
IE52	N Fisher/Peguis	
IE53	N Eriksdale/Ashern	
IE61	Northern Remote	Remote
NO11	Z1 Flin, Snow, Cran, Sher	
NO12	Z1 The Pas/OCN, Kels	
NO13	Z1 LL/MCFN, LR, O-P(SIL)CN,PN(GVL)	Remote
NO14	Z1 Thompson, Myst Lake	

NO15	Z1 Thick, Pik, Wab, Ilf/WLFN, Corm	
NO16	Z1 Gillam, Fox Lake Cree Nation	Remote
NO21	Z2 GR/MisCN, ML/MosCN, Eas/CheCN	
NO22	Z2 Puk/Mat Col CN	Remote
NO23	Z2 SayD(TL)FN, Bro/BLFN, NoL(Lac)FN	Remote
NO24	Z2 Nelson House/NCN	
NO25	Z2 Sham, YorkF, TatCN(SPL)	Remote
NO26	Z2 Bu(OH)CN, MS(GR)CN, GLN/GLFN	Remote
NO27	Z2 Cross Lake/Cross Lake FN	
NO28	Z2 Norway House/NH CN	Remote
NO31	Z3 IsL/GHFN, RSL/RSLFN, STPFN, WFN	Remote
SO11	N Seven Regions	
SO12	N MacGregor	
SO13	N Rural Portage	
SO14	N Cartier/SFX	
SO15	N City of Portage	
SO21	M Grey/St Claude (*pre 2015 Notre Dame)	
SO22	M Carman	
SO23	M MacDonald	
SO24	M Morris	
SO25	M St. Pierre/DeSalaberry	
SO26	M Red River South	
SO31	W Lorne/Louise/Pembina (*post 2015 Notre Dame)	
SO32	W Stanley	
SO33	W Altona	
SO34	W Morden	
SO35	W Winkler	
SO36	W Roland/Thompson	
SO41	E Niverville/Richot	

SO42	E Tache	
SO43	E Ste Anne/LaBroquerie	
SO44	E Steinbach	
SO45	E Hanover	
SO46	E Rural East	
W002	Assiniboine South	Urban
W006	Transcona	Urban
W01A	St. James-Assiniboia W	Urban
W01B	St. James-Assiniboia E	Urban
W03A	Fort Garry N	Urban
W03B	Fort Garry S	Urban
W04A	St. Vital N	Urban
W04B	St. Vital S	Urban
W05A	St. Boniface W	Urban
W05B	St. Boniface E	Urban
W07A	River East S	Urban
W07B	River East W	Urban
W07C	River East E	Urban
W07D	River East N	Urban
W08A	Seven Oaks W	Urban
W08B	Seven Oaks E	Urban
W08C	Seven Oaks N	Urban
W09A	Inkster W	Urban
W09B	Inkster E	Urban
W10A	Point Douglas N	Urban
W10B	Point Douglas S	Urban
W11A	Downtown W	Urban
W11B	Downtown E	Urban
W12A	River Heights W	Urban

W12B	River Heights E	Urban
WP21	Churchill (Part of the WRHA)	
WE11	N Duck Mountain	
WE12	N Porcupine Mountain	
WE13	N Riding Mountain	
WE14	N Agassiz Mountain	
WE15	N Dauphin	
WE16	N Swan River	Urban
WE21	Bdn West End	Urban
WE22	Bdn North Hill	Urban
WE23	Bdn Downtown	Urban
WE24	Bdn South End	Urban
WE25	Bdn East End	
WE31	S Asessippi	
WE32	S Little Saskatchewan	
WE33	S Turtle Mountain	
WE34	S Souris River	
WE35	S Whitemud	
WE36	S Spruce Woods	

Appendix III – R-Code

```
---
title: "Appendix"
author: "Justin Dyck"
date: "October 2, 2019"
output: word_document
---

```{r}
#Packages used
library(spdep)
library(rgeos)
library(rgdal)
library(maptools)
library(coda)
library(R2WinBUGS)
library(tidyverse)
library(tmap)
library(FlexScan)
library(gridExtra)
library(INLA)
library(wesanderson)
```

```{r}
###Read in Data#####
mb = readOGR(dsn="C:/School/Grad School/Thesis/Data/Shape/comb", layer="combined")
IHD = read.csv("IHD.csv",head=T)
IHD = IHD[match(mb@data[["Region"]], substr(IHD$rhad,1,4)),]
IHDM = read.csv("IHD_m.csv",head=T)
IHDM = IHDM[match(mb@data[["Region"]], substr(IHDM$rhad,1,4)),]
IHDF = read.csv("IHD_f.csv",head=T)
IHDF = IHDF[match(mb@data[["Region"]], substr(IHDF$rhad,1,4)),]
AMI = read.csv("ami.csv",head=T)
AMI = AMI %>%
  mutate(count_2011 = ifelse(is.na(count_2011),3,count_2011),
        count_2012 = ifelse(is.na(count_2012),3,count_2012),
        count_2013 = ifelse(is.na(count_2013),3,count_2013),
        count_2014 = ifelse(is.na(count_2014),3,count_2014),
        count_2015 = ifelse(is.na(count_2015),3,count_2015),
        count_2016 = ifelse(is.na(count_2016),3,count_2016),
        count = count_2011+count_2012+count_2013+count_2014+count_2015+count_2016,
        exp = E_2011+E_2012+E_2013+E_2014+E_2015+E_2016,
        pop = pop_2011+pop_2012+pop_2013+pop_2014+pop_2015+pop_2016)
AMI = AMI[match(mb@data[["Region"]], substr(AMI$rhad,1,4)),]

#Census data read in and clean#
```
```

```
inter = readOGR(dsn="C:/School/Grad School/Thesis/Data/Shape/intersected", layer="intersected1")

CENS = read.csv("abi.csv",head=T) %>%
  mutate(pop = ifelse(is.na(Total),20,Total),
         ind = ifelse(is.na(AB),20,AB))
sef = read.csv("sefi2.csv",head=T) %>%
  mutate(DA = Region)
CENS = merge(CENS,sef,by="DA") %>%
  select(-c(Total,AB,Region))

CENS = CENS %>%
  mutate(ind = ifelse(ind==0,5,ind))
CENS = CENS %>%
  mutate(indpr = ind/pop)

cens = readOGR(dsn="C:/School/Grad School/Thesis/Data/Shape/census", layer="lda_000b16a_e")

CENS = CENS[match(cens@data[["DAUID"]], as.factor(CENS$DA)),]

####Intersect the census data with the RHAD data#####

shape_data = inter@data
shape_data = shape_data %>%
  mutate(DA = as.numeric(as.character(DAUID)),
         ID = paste(DAUID,Region))

pops = merge(shape_data,CENS,by='DA',all.x=T)

pops = pops %>%
  mutate(ID = paste(DAUID,Region))

pops = pops[match(shape_data$ID, pops$ID),]

pops = pops %>%
  mutate(Pjl = pop*(area_2/area),
         Xjl = ind *(area_2/area))

pops = pops %>%
  mutate(Pjl = Pjl+1)

####Get adjacency matrices#####

int_nb = poly2nb(inter, queen = F)
int_adj = unlist(int_nb)
int_adj = as.numeric(int_adj)
int_num = matrix(nrow=3195,ncol=1)
for(i in 1:3195){
  int_num[i] = length(int_nb[[i]])
```

```
}

int_num = as.numeric(int_num)
int_snn = sum(int_num)

ihd_nb = poly2nb(mb, queen = F)
ihd_adj = unlist(ihd_nb)
ihd_adj = as.numeric(ihd_adj)
ihd_num = matrix(nrow=96, ncol=1)
for (i in 1:96){
  ihd_num[i] = length(ihd_nb[[i]])
}
ihd_num = as.numeric(ihd_num)
ihd_snn = sum(ihd_num)
```

```{r}
#Misaligned Section

##Intercept model 1 (To get smoothed rates for 3195 regions) - For each census variable required, just change the data list parameters. For count variables, also change the model specification to include the Poisson distribution

mis_model<-function(){

for (i in 1:N) {
   X[i]~dnorm(mu[i],sig[i])

   mu[i] <- V[i] +U[i]

   sig[i] ~ dgamma(0.5,0.005)
   V[i]~dnorm(0,tau.V)

}
 U[1:N] ~ car.normal(adj[], weights[], num[], tauomega.U)
for(k in 1:snn) {
weights[k] <- 1
}

tau.T ~ dgamma(0.5,0.005)
p ~ dbeta(1,1)
sigma.Z <- sqrt(p/tau.T)
omega.U <- sigma.Z/sqrt(1.164)
sigma.V <- sqrt((1-p)/tau.T)
tau.V <- 1/(sigma.V*sigma.V)
tauomega.U <- 1/(omega.U*omega.U)
sd.U <- sd(U[1:N])
vratio <- sd.U*sd.U/(sd.U*sd.U+sigma.V*sigma.V)
```

```
}
data = list(
  N = length(int_num),
  X = pops$sefi2,
  num = int_num,
  adj = int_adj,
  snn = int_snn
)

inits<-list(
  list(U=c(rep(0,3195)), V=c(rep(0,3195)), tau.T = 0.02, p=0.5, sig=c(rep(1,3195))),
  list(U=c(rep(0,3195)), V=c(rep(0,3195)), tau.T = 0.01, p=0.5, sig=c(rep(1,3195))))

parameters <- c("mu","tau.T","sd.U","vratio","omega.U","sigma.V")

mis_model_out<-bugs(data=data, parameters.to.save=parameters, model=mis_model, n.chains=2,
n.iter=10000,n.thin=2,n.burnin=1000,  DIC=F,  inits=inits, debug=TRUE, bugs.directory =
"C:/WinBUGS14")

#####################################################################

#Summarize the smoothed counts into rates to compare in next model.

INDs = mis_model_out$median$L
INDs = as.vector(INDs)

SEFIs = mis_model_out$median$mu
SEFIs = as.vector(INDs)


SM = cbind(region = as.character(pops$Region), p = pops$Pjl, INDs, SEFIs)
SM = data.frame(SM)

write.csv(SM,"smoothed_medians.csv")

```

```{r}

#Prep data for final model

SM = read.csv("smoothed_medians.csv",head=T)

SM = SM %>%
  mutate(p = pops$Pjl)

SM = SM %>%
  group_by(region) %>%
```

```
  summarize(p = sum(p),
            SEFIs = sum(SEFIs),
            INDs = sum(INDs))

SM = SM %>%
  mutate(SEFIs = SEFIs/p,
         INDs = INDs/p)

SM = SM[match(mb@data[["Region"]], SM$region),]

#Urban and remote binary covariates

urb = ifelse(IHD[,74]=="UR",1,0)
rem = ifelse(IHD[,74]=="RE",1,0)
```

```{r}
#Summary Stats and graphs
AMI1 = AMI %>%
  mutate(RHA = ifelse(substr(rhad,1,2)=="SO","Southern",
                      ifelse(substr(rhad,1,2)=="WE","Prairie Mountain",
                             ifelse(substr(rhad,1,2)=="IE","Interlake-Eastern",
                                    ifelse(substr(rhad,1,2)=="NO","Northern","Winnipeg")))),
         Region = substr(AMI$rhad,6,52)) %>%
  arrange(match(RHA, c("Northern","Interlake-Eastern","Prairie
Mountain","Winnipeg","Southern")),Region)

ggplot(data=AMI1)+
  geom_col(aes(y=IR,x=reorder(Region,match(RHA, c("Northern","Interlake-Eastern","Prairie
Mountain","Winnipeg","Southern"))),fill=RHA))+
  coord_flip() +
  scale_fill_manual(breaks = c("Southern","Winnipeg","Prairie Mountain","Interlake-
Eastern","Northern"),
                    values=wes_palette(n=5, name="Cavalcanti1"))+
  labs(y = "AMI Incidence Rate /1000 persons", x = "Health District/Neighborhood Cluster")+
  geom_hline(yintercept=3.79,show.legend=T)+
  theme_light()

ggplot(data=AMI1)+
  geom_col(aes(y=count/exp,x=reorder(Region,match(RHA, c("Northern","Interlake-Eastern","Prairie
Mountain","Winnipeg","Southern"))),fill=RHA))+
  coord_flip() +
  scale_fill_manual(breaks = c("Southern","Winnipeg","Prairie Mountain","Interlake-
Eastern","Northern"),
                    values=wes_palette(n=5, name="Cavalcanti1"))+
  labs(y = "AMI SIR", x = "Health District/Neighborhood Cluster")+
  geom_hline(yintercept=1,show.legend=T)+
  theme_light()
```

```
cr_ihd = matrix(nrow=18,ncol=3)

for(i in 1:18){
cr_ihd[i,1] = sum(ihda[,i])/sum(popa[,i])
cr_ihd[i,2] = sum(ihdf[,i])/sum(popf[,i])
cr_ihd[i,3] = sum(ihdm[,i])/sum(popm[,i])
}

crude_ihd = data.frame(cbind(Year = seq(from=1998, to=2015, by=1), Overall = cr_ihd[,1], Female =
cr_ihd[,2], Male = cr_ihd[,3]))

ihd_pl = data.frame(cbind(rhad = as.character(IHD[,1]), Overall_count = IHD[,70], Overall_pop =
IHD[,71], Female_count = IHDF[,70], Female_pop = IHDF[,71], Male_count = IHDM[,70], Male_pop =
IHDM[,71])) %>%
  mutate(RHA = ifelse(substr(rhad,1,2)=="SO","Southern",
                ifelse(substr(rhad,1,2)=="WE","Prairie Mountain",
                     ifelse(substr(rhad,1,2)=="IE","Interlake-Eastern",
                          ifelse(substr(rhad,1,2)=="NO","Northern","Winnipeg")))),
       Region = substr(AMI$rhad,6,52)) %>%
  arrange(match(RHA, c("Northern","Interlake-Eastern","Prairie
Mountain","Winnipeg","Southern")),Region)

ihd_pl = ihd_pl %>%
  mutate(Overall_count = as.numeric(as.character(Overall_count)),
       Overall_pop = as.numeric(as.character(Overall_pop)),
       Female_count = as.numeric(as.character(Female_count)),
       Female_pop = as.numeric(as.character(Female_pop)),
       Male_count = as.numeric(as.character(Male_count)),
       Male_pop = as.numeric(as.character(Male_pop)))

ihd_pl = ihd_pl %>%
  group_by(RHA) %>%
  summarize(Overall_count = sum(Overall_count),
          Overall_pop = sum(Overall_pop),
          Female_count = sum(Female_count),
          Female_pop = sum(Female_pop),
          Male_count = sum(Male_count),
          Male_pop = sum(Male_pop))

ihd_pl = ihd_pl %>%
  mutate(Overall = Overall_count/Overall_pop,
       Female = Female_count/Female_pop,
       Male = Male_count/Male_pop) %>%
  gather(Group, Prevalence, c(Overall, Female, Male))

ggplot(data=ihd_pl)+
```

```
 geom_col(aes(y=Prevalence,x=reorder(RHA,match(RHA, c("Northern","Interlake-Eastern","Prairie
Mountain","Winnipeg","Southern"))),fill=Group),position = 'dodge')+
 coord_flip() +
 scale_x_discrete(breaks = c("Southern","Winnipeg","Prairie Mountain","Interlake-
Eastern","Northern"))+
 labs(y = "Crude IHD Prevalence (2015)", x = "RHA")+
 scale_fill_manual(breaks = c("Overall", "Male", "Female"),
                   values=wes_palette(n=3, name="Cavalcanti1", type="continuous"))+
 geom_hline(yintercept = 0.104)+
 theme_light()

crude_ihd = crude_ihd %>%
 gather(Group, Prevalence, c(Overall, Female, Male))

ggplot(data=crude_ihd)+
 geom_line(aes(x=Year,y=Prevalence,color=Group),size=1)+
 geom_point(aes(x=Year,y=Prevalence,color=Group))+
 labs(y="Crude IHD Prevalence")+
 scale_color_manual(breaks = c("Overall", "Male", "Female"),
                    values=wes_palette(n=3, name="Cavalcanti1", type="continuous"))+
 theme_light()

cov_pl = SM %>%
 mutate(hyp = IHD[,73],
       region = as.character(region),
       RHA = ifelse(substr(region,1,2)=="SO","Southern",
                 ifelse(substr(region,1,2)=="WE","Prairie Mountain",
                      ifelse(substr(region,1,2)=="IE","Interlake-Eastern",
                           ifelse(substr(region,1,2)=="NO","Northern","Winnipeg")))),
       Region = substr(AMI$rhad,6,52))%>%
 arrange(match(RHA, c("Northern","Interlake-Eastern","Prairie
Mountain","Winnipeg","Southern")),Region)

ggplot(data=cov_pl)+
 geom_col(aes(y=sefis,x=reorder(Region,match(RHA, c("Northern","Interlake-Eastern","Prairie
Mountain","Winnipeg","Southern"))),fill=RHA))+
 scale_y_continuous(limits = c(-3.25, 3.25))+
 coord_flip() +
  scale_fill_manual(breaks = c("Southern","Winnipeg","Prairie Mountain","Interlake-
Eastern","Northern"),
                   values=wes_palette(n=5, name="Cavalcanti1"))+
 labs(y = "SEFI-2 Scores per Region", x = "Health District/Neighborhood Cluster")+
 theme_light()

ggplot(data=cov_pl)+
 geom_col(aes(y=INDs,x=reorder(Region,match(RHA, c("Northern","Interlake-Eastern","Prairie
Mountain","Winnipeg","Southern"))),fill=RHA))+
 coord_flip() +
```

```
 scale_fill_manual(breaks = c("Southern","Winnipeg","Prairie Mountain","Interlake-
Eastern","Northern"),
                   values=wes_palette(n=5, name="Cavalcanti1"))+
 labs(y = "Indigenous Population Proportion per Region", x = "Health District/Neighborhood Cluster")+
 theme_light()+
 geom_hline(yintercept = 0.17)

ggplot(data=cov_pl)+
 geom_col(aes(y=hyp,x=reorder(Region,match(RHA, c("Northern","Interlake-Eastern","Prairie
Mountain","Winnipeg","Southern"))),fill=RHA))+
 coord_flip() +
 scale_fill_manual(breaks = c("Southern","Winnipeg","Prairie Mountain","Interlake-
Eastern","Northern"),
                   values=wes_palette(n=5, name="Cavalcanti1"))+
 labs(y = "Standardized Hypertension Prevalence Risk Ratio", x = "Health District/Neighborhood
Cluster")+
 theme_light()+
 geom_hline(yintercept = 1)

#Mapping observed values using tmap pkg

mb_AMItmap = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
             stringsAsFactors=F) %>%
 mutate(IRR = AMI$count/AMI$exp,
      RHA = ifelse(substr(AMI$rhad,1,2)=="SO","Southern",
                ifelse(substr(AMI$rhad,1,2)=="WE","Prairie Mountain",
                    ifelse(substr(AMI$rhad,1,2)=="IE","Interlake-Eastern",
                        ifelse(substr(AMI$rhad,1,2)=="NO","Northern","Winnipeg")))))

wpg_AMItmap = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
             stringsAsFactors=F) %>%
 mutate(IRR = AMI$count/AMI$exp) %>%
 filter(substr(Region,1,2)=="W0" | substr(Region,1,2)=="W1")

cut = c(0,0.8,1,1.2,1.6,3.3)

mb_ami = tm_shape(mb_AMItmap) +
 tm_fill("IRR",
       title = "Cluster Type",
       palette="Reds",
       title.fontfamily="mono",
       breaks = cut) +
 tm_layout(title = "AMI SIR - MB",
         inner.margins = c(0.02,0.02,0.08,0.02),
         title.fontfamily="mono",
         title.position=c("center","top"),
         legend.show=F,
         title.size=1.2) +
```

```
  tm_borders(lwd = 1, lty = "solid")

wpg_ami = tm_shape(wpg_AMItmap) +
  tm_fill("IRR",
          title = "SIR",
          palette="Reds",
          title.fontfamily="mono",
          breaks =  cut) +
  tm_layout(title = "AMI SIR - WPG",
            inner.margins = c(0.02,0.02,0.08,0.07),
            title.fontfamily="mono",
            title.position=c("center","top"),
            legend.title.fontfamily="mono",
            legend.text.fontfamily="mono",
            legend.position=c("left","bottom"),
            legend.title.size = 1.2,
            legend.text.size = 0.9,
            title.size=1.2) +
  tm_borders(lwd = 1, lty = "solid")

tmap_arrange(mb_ami,wpg_ami,ncol=2)
```

```{r}
#Cluster Detection#
fs = flexscan(map=mb, case=AMI$count,pop=AMI$exp)

clusters = list(
  c(34,46,52,53,55,84,91),
  c(49,67,70,73,77,81,85,92,96),
  c(82),
  c(36),
  c(7,8,9,10,17),
  c(44,71,76),
  c(57,58,59),
  c(95),
  c(56,61,62)
)

c1 = substr(AMI$rhad[clusters[[1]]],1,4)
c2 = substr(AMI$rhad[clusters[[2]]],1,4)
c3 = substr(AMI$rhad[clusters[[3]]],1,4)
c4 = substr(AMI$rhad[clusters[[4]]],1,4)
c5 = substr(AMI$rhad[clusters[[5]]],1,4)
c6 = substr(AMI$rhad[clusters[[6]]],1,4)
c7 = substr(AMI$rhad[clusters[[7]]],1,4)
c8 = substr(AMI$rhad[clusters[[8]]],1,4)
c9 = substr(AMI$rhad[clusters[[9]]],1,4)
```

```
Cluster = data.frame(Region = substr(IHD$rhad,1,4), cl1 = rep(0,96),cl2 = rep(0,96),cl3 = rep(0,96),cl4 =
rep(0,96),cl5 = rep(0,96),cl6 = rep(0,96),cl7 = rep(0,96),cl8 = rep(0,96), cl9 = rep(0,96))

for (i in 1:96){
  Cluster$cl1[i] = ifelse(Cluster$Region[i] %in% c1,1,0)
  Cluster$cl2[i] = ifelse(Cluster$Region[i] %in% c2,1,0)
  Cluster$cl3[i] = ifelse(Cluster$Region[i] %in% c3,1,0)
  Cluster$cl4[i] = ifelse(Cluster$Region[i] %in% c4,1,0)
  Cluster$cl5[i] = ifelse(Cluster$Region[i] %in% c5,1,0)
  Cluster$cl6[i] = ifelse(Cluster$Region[i] %in% c6,1,0)
  Cluster$cl7[i] = ifelse(Cluster$Region[i] %in% c7,1,0)
  Cluster$cl8[i] = ifelse(Cluster$Region[i] %in% c8,1,0)
  Cluster$cl9[i] = ifelse(Cluster$Region[i] %in% c9,1,0)
}
Cluster = Cluster %>%
  mutate(cnum = ifelse(cl1 == 1, 1,
                  ifelse(cl2 == 1, 2,
                       ifelse(cl3 ==1, 3,
                            ifelse(cl4 == 1, 4,
                                 ifelse(cl5 == 1, 5,
                                      ifelse(cl6 == 1, 6,
                                           ifelse(cl7 == 1, 7,
                                                ifelse(cl8 == 1, 8,
                                                     ifelse(cl9 == 1, 9, 0)))))))),
         prime = ifelse(cnum==1,1,
                   ifelse(cnum>1,2,0)))

#Mapping clusters with tmap

mb_clustertmap = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
              stringsAsFactors=F) %>%
  mutate(prime = as.factor(ifelse(Cluster$prime==1,"Primary Cluster",
                           ifelse(Cluster$prime==2,"Secondary Cluster","No Cluster"))),
         depr = as.factor(ifelse(SM$sefis>0.55,"More Deprivation","Less Deprivation")),
         remote = as.factor(ifelse(rem==1,"Remote Location","Non-Remote")),
         indi = as.factor(ifelse(SM$INDs>0.30,">30%","<30%")),
         reg = ifelse(substr(Region,1,2)=="W1" | substr(Region,1,2)=="W0" | prime=="No
Cluster","",Region))

wpg_clustertmap = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
              stringsAsFactors=F) %>%
  mutate(prime = as.factor(ifelse(Cluster$prime==1,"Primary Cluster",
                           ifelse(Cluster$prime==2,"Secondary Cluster","No Cluster"))),
         depr = as.factor(ifelse(SM$sefis>0.55,"More Deprivation","Less Deprivation")),
         remote = as.factor(ifelse(rem==1,"Remote Location","Non-Remote")),
         indi = as.factor(ifelse(SM$INDs>0.30,">30%","<30%")),
         reg = ifelse(prime=="No Cluster","",Region)) %>%
```

```
 filter(substr(Region,1,2)=="W0" | substr(Region,1,2)=="W1")

clustcols <- c("#F5F5DC", "#8B0000", "#B0C4DE")

mb_clust = tm_shape(mb_clustertmap) +
 tm_fill("indi",
        title = "Cluster Type",
        palette=cols,
        title.fontfamily="mono") +
 tm_layout(title = "Indigenous Population - MB",
          inner.margins = c(0.02,0.02,0.08,0.02),
          title.fontfamily="mono",
          title.position=c("center","top"),
          legend.show=F,
          title.size=1.5) +
 tm_borders(lwd = 1, lty = "solid")

wpg_clust = tm_shape(wpg_clustertmap) +
 tm_fill("indi",
        title = "Population Proportion",
        palette=cols,
        title.fontfamily="mono") +
 tm_layout(title = "Indigneous Population - WPG",
          inner.margins = c(0.02,0.02,0.08,0.07),
          title.fontfamily="mono",
          title.position=c("center","top"),
          legend.title.fontfamily="mono",
          legend.text.fontfamily="mono",
          legend.position=c("left","bottom"),
          legend.title.size = 1.5,
          legend.text.size = 0.75,
          title.size=1.5) +
 tm_borders(lwd = 1, lty = "solid")

tmap_arrange(mb_clust,wpg_clust,ncol=2)

#Mapping variables for comparison

cols = c("#F5F5DC","#008080")

mb_depr = tm_shape(mb_clustertmap) +
 tm_fill("depr",
        title = "",
        palette=cols,
        title.fontfamily="mono") +
 tm_layout(title = "SES Deprivation - MB",
          inner.margins = c(0.02,0.02,0.08,0.02),
          title.fontfamily="mono",
```

```
        title.position=c("center","top"),
        legend.show=F,
        title.size=2) +
 tm_borders(lwd = 1, lty = "solid")

wpg_depr = tm_shape(wpg_clustertmap) +
 tm_fill("depr",
       title = "",
       palette=cols,
       title.fontfamily="mono") +
 tm_layout(title = "SES Deprivation - WPG",
        inner.margins = c(0.02,0.02,0.08,0.07),
        title.fontfamily="mono",
        title.position=c("center","top"),
        legend.title.fontfamily="mono",
        legend.text.fontfamily="mono",
        legend.position=c("left","bottom"),
        legend.title.size = 1.5,
        legend.text.size = 0.9,
        title.size=2) +
 tm_borders(lwd = 1, lty = "solid")

mb_rem = tm_shape(mb_clustertmap) +
 tm_fill("remote",
       title = "",
       palette=cols,
       title.fontfamily="mono") +
 tm_layout(title = "Remoteness - MB",
        inner.margins = c(0.02,0.02,0.08,0.02),
        title.fontfamily="mono",
        title.position=c("center","top"),
        legend.title.fontfamily="mono",
        legend.text.fontfamily="mono",
        legend.position=c("right","bottom"),
        legend.title.size = 1.5,
        legend.text.size = 0.9,
        title.size=2) +
 tm_borders(lwd = 1, lty = "solid")

mb_ind = tm_shape(mb_clustertmap) +
 tm_fill("indi",
       title = "",
       palette=cols,
       title.fontfamily="mono") +
 tm_layout(title = "Indigenous Population - MB",
        inner.margins = c(0.02,0.02,0.08,0.02),
        title.fontfamily="mono",
        title.position=c("center","top"),
```

```
          legend.show=F,
          title.size=1) +
  tm_borders(lwd = 1, lty = "solid")

wpg_ind = tm_shape(wpg_clustertmap) +
  tm_fill("indi",
          title = "Pop Proportion",
          palette=cols,
          title.fontfamily="mono") +
  tm_layout(title = "Indigenous Population - WPG",
            inner.margins = c(0.02,0.02,0.08,0.11),
            title.fontfamily="mono",
            title.position=c("center","top"),
            legend.title.fontfamily="mono",
            legend.text.fontfamily="mono",
            legend.position=c("left","bottom"),
            legend.title.size = 1,
            legend.text.size = 0.9,
            title.size=1) +
  tm_borders(lwd = 1, lty = "solid")

tmap_arrange(mb_clust, wpg_clust, ncol = 2)
tmap_arrange(mb_depr, wpg_depr, ncol = 2)
tmap_arrange(mb_ind, wpg_ind, ncol = 2)
```

```{r}
#Cluster Detection for IHD as a data exploration tool#
fsihd = flexscan(map=mb, case=IHD[,70],pop=IHD[,72])

clusters_ihd = list(
  c(35, 36, 42, 49, 84),
  c(23),
  c(69, 75, 79, 80, 81, 82),
  c(6, 7, 8, 9, 17),
  c(58)
)

ci1 = substr(IHD$rhad[clusters_ihd[[1]]],1,4)
ci2 = substr(IHD$rhad[clusters_ihd[[2]]],1,4)
ci3 = substr(IHD$rhad[clusters_ihd[[3]]],1,4)
ci4 = substr(IHD$rhad[clusters_ihd[[4]]],1,4)
ci5 = substr(IHD$rhad[clusters_ihd[[5]]],1,4)

Cluster_ihd = data.frame(Region = substr(IHD$rhad,1,4), cl1 = rep(0,96),cl2 = rep(0,96),cl3 =
rep(0,96),cl4 = rep(0,96),cl5 = rep(0,96))

for (i in 1:96){
```

```
  Cluster_ihd$cl1[i] = ifelse(Cluster_ihd$Region[i] %in% ci1,1,0)
  Cluster_ihd$cl2[i] = ifelse(Cluster_ihd$Region[i] %in% ci2,1,0)
  Cluster_ihd$cl3[i] = ifelse(Cluster_ihd$Region[i] %in% ci3,1,0)
  Cluster_ihd$cl4[i] = ifelse(Cluster_ihd$Region[i] %in% ci4,1,0)
  Cluster_ihd$cl5[i] = ifelse(Cluster_ihd$Region[i] %in% ci5,1,0)

}
Cluster_ihd = Cluster_ihd %>%
  mutate(cnum = ifelse(cl1 == 1, 1,
                  ifelse(cl2 == 1, 2,
                       ifelse(cl3 ==1, 3,
                            ifelse(cl4 == 1, 4,
                                 ifelse(cl5 == 1, 5,0))))),
         prime = ifelse(cnum==1,1,
                   ifelse(cnum>1,2,0)))
####################################

fsihd_f = flexscan(map=mb, case=IHDF[,70],pop=IHDF[,72])


clusters_ihd_f = list(
  c(35, 36, 42, 49, 70, 84),
  c(8, 9, 10, 19, 23),
  c(69, 75, 79, 80, 81, 82, 95),
  c(58)
)

ci1f = substr(IHD$rhad[clusters_ihd_f[[1]]],1,4)
ci2f = substr(IHD$rhad[clusters_ihd_f[[2]]],1,4)
ci3f = substr(IHD$rhad[clusters_ihd_f[[3]]],1,4)
ci4f = substr(IHD$rhad[clusters_ihd_f[[4]]],1,4)

Cluster_ihd_f = data.frame(Region = substr(IHD$rhad,1,4), cl1 = rep(0,96),cl2 = rep(0,96),cl3 =
rep(0,96),cl4 = rep(0,96))

for (i in 1:96){
  Cluster_ihd_f$cl1[i] = ifelse(Cluster_ihd_f$Region[i] %in% ci1f,1,0)
  Cluster_ihd_f$cl2[i] = ifelse(Cluster_ihd_f$Region[i] %in% ci2f,1,0)
  Cluster_ihd_f$cl3[i] = ifelse(Cluster_ihd_f$Region[i] %in% ci3f,1,0)
  Cluster_ihd_f$cl4[i] = ifelse(Cluster_ihd_f$Region[i] %in% ci4f,1,0)
}

Cluster_ihd_f = Cluster_ihd_f %>%
  mutate(cnum = ifelse(cl1 == 1, 1,
                  ifelse(cl2 == 1, 2,
                       ifelse(cl3 ==1, 3,
                            ifelse(cl4 == 1, 4,0)))),
         prime = ifelse(cnum==1,1,
```

```
                ifelse(cnum>1,2,0)))
###################################

fsihd_m = flexscan(map=mb, case=IHDM[,70],pop=IHDM[,72])


clusters_ihd_m = list(
  c(35, 36, 49, 84),
  c(23),
  c(69, 75, 79, 81, 82),
  c(58)
)

ci1m = substr(IHD$rhad[clusters_ihd_m[[1]]],1,4)
ci2m = substr(IHD$rhad[clusters_ihd_m[[2]]],1,4)
ci3m = substr(IHD$rhad[clusters_ihd_m[[3]]],1,4)
ci4m = substr(IHD$rhad[clusters_ihd_m[[4]]],1,4)

Cluster_ihd_m = data.frame(Region = substr(IHD$rhad,1,4), cl1 = rep(0,96),cl2 = rep(0,96),cl3 =
rep(0,96),cl4 = rep(0,96))

for (i in 1:96){
  Cluster_ihd_m$cl1[i] = ifelse(Cluster_ihd_m$Region[i] %in% ci1m,1,0)
  Cluster_ihd_m$cl2[i] = ifelse(Cluster_ihd_m$Region[i] %in% ci2m,1,0)
  Cluster_ihd_m$cl3[i] = ifelse(Cluster_ihd_m$Region[i] %in% ci3m,1,0)
  Cluster_ihd_m$cl4[i] = ifelse(Cluster_ihd_m$Region[i] %in% ci4m,1,0)
}

Cluster_ihd_m = Cluster_ihd_m %>%
  mutate(cnum = ifelse(cl1 == 1, 1,
                ifelse(cl2 == 1, 2,
                     ifelse(cl3 ==1, 3,
                          ifelse(cl4 == 1, 4, 0)))),
       prime = ifelse(cnum==1,1,
                ifelse(cnum>1,2,0)))

#map these clusters

mb_cluster_ihd_tmap = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
              stringsAsFactors=F) %>%
  mutate(all = as.factor(ifelse(Cluster_ihd$prime==1,"Primary Cluster",
                          ifelse(Cluster_ihd$prime==2,"Secondary Cluster","No Cluster"))),
       fem = as.factor(ifelse(Cluster_ihd_f$prime==1,"Primary Cluster",
                          ifelse(Cluster_ihd_f$prime==2,"Secondary Cluster","No Cluster"))),
       mal = as.factor(ifelse(Cluster_ihd_m$prime==1,"Primary Cluster",
                          ifelse(Cluster_ihd_m$prime==2,"Secondary Cluster","No Cluster"))),
       rega = ifelse(substr(Region,1,2)=="W1" | substr(Region,1,2)=="W0" | all=="No Cluster","",Region),
       regf = ifelse(substr(Region,1,2)=="W1" | substr(Region,1,2)=="W0" | fem=="No Cluster","",Region),
```

```
      regm = ifelse(substr(Region,1,2)=="W1" | substr(Region,1,2)=="W0" | mal=="No
Cluster","",Region))

wpg_cluster_ihd_tmap = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
            stringsAsFactors=F) %>%
 mutate(all = as.factor(ifelse(Cluster_ihd$prime==1,"Primary Cluster",
                        ifelse(Cluster_ihd$prime==2,"Secondary Cluster","No Cluster"))),
      fem = as.factor(ifelse(Cluster_ihd_f$prime==1,"Primary Cluster",
                        ifelse(Cluster_ihd_f$prime==2,"Secondary Cluster","No Cluster"))),
      mal = as.factor(ifelse(Cluster_ihd_m$prime==1,"Primary Cluster",
                        ifelse(Cluster_ihd_m$prime==2,"Secondary Cluster","No Cluster"))),
      rega = ifelse(all=="No Cluster","",Region),
      regf = ifelse(fem=="No Cluster","",Region),
      regm = ifelse(mal=="No Cluster","",Region)) %>%
 filter(substr(Region,1,2)=="W0" | substr(Region,1,2)=="W1")

clustcols <- c("#F5F5DC", "#8B0000", "#B0C4DE")

mb_ihd_clust = tm_shape(mb_cluster_ihd_tmap) +
 tm_fill("all",
       title = "Cluster Type",
       palette=clustcols,
       title.fontfamily="mono") +
 tm_layout(title = "IHD Clusters - MB",
         inner.margins = c(0.02,0.02,0.08,0.02),
         title.fontfamily="mono",
         title.position=c("center","top"),
         legend.show=F,
         title.size=1.5) +
 tm_borders(lwd = 1, lty = "solid")+
 tm_text("rega",size=0.5)

wpg_ihd_clust = tm_shape(wpg_cluster_ihd_tmap) +
 tm_fill("all",
       title = "Cluster Type",
       palette=clustcols,
       title.fontfamily="mono") +
 tm_layout(title = "IHD Clusters - WPG",
         inner.margins = c(0.02,0.02,0.08,0.07),
         title.fontfamily="mono",
         title.position=c("center","top"),
         legend.title.fontfamily="mono",
         legend.text.fontfamily="mono",
         legend.position=c("left","bottom"),
         legend.title.size = 1.25,
         legend.text.size = 0.8,
         title.size=1.5) +
 tm_borders(lwd = 1, lty = "solid")+
```

```
 tm_text("rega",size=0.65)

tmap_arrange(mb_ihd_clust, wpg_ihd_clust, ncol = 2)

mb_ihd_clust_f = tm_shape(mb_cluster_ihd_tmap) +
 tm_fill("fem",
        title = "Cluster Type",
        palette=clustcols,
        title.fontfamily="mono") +
 tm_layout(title = "Female IHD Clusters - MB",
          inner.margins = c(0.02,0.02,0.08,0.02),
          title.fontfamily="mono",
          title.position=c("center","top"),
          legend.show=F,
          title.size=1.25) +
 tm_borders(lwd = 1, lty = "solid")+
 tm_text("regf",size=0.5)

wpg_ihd_clust_f = tm_shape(wpg_cluster_ihd_tmap) +
 tm_fill("fem",
        title = "Cluster Type",
        palette=clustcols,
        title.fontfamily="mono") +
 tm_layout(title = "Female IHD Clusters - WPG",
          inner.margins = c(0.02,0.02,0.08,0.07),
          title.fontfamily="mono",
          title.position=c("center","top"),
          legend.title.fontfamily="mono",
          legend.text.fontfamily="mono",
          legend.position=c("left","bottom"),
          legend.title.size = 1.2,
          legend.text.size = 0.8,
          title.size=1.25) +
 tm_borders(lwd = 1, lty = "solid")+
 tm_text("regf",size=0.65)

tmap_arrange(mb_ihd_clust_f, wpg_ihd_clust_f, ncol = 2)

mb_ihd_clust_m = tm_shape(mb_cluster_ihd_tmap) +
 tm_fill("mal",
        title = "Cluster Type",
        palette=clustcols,
        title.fontfamily="mono") +
 tm_layout(title = "Male IHD Clusters - MB",
          inner.margins = c(0.02,0.02,0.08,0.02),
          title.fontfamily="mono",
          title.position=c("center","top"),
          legend.show=F,
```

```
        title.size=1.25) +
 tm_borders(lwd = 1, lty = "solid")+
 tm_text("regm",size=0.5)

wpg_ihd_clust_m = tm_shape(wpg_cluster_ihd_tmap) +
 tm_fill("mal",
       title = "Cluster Type",
       palette=clustcols,
       title.fontfamily="mono") +
 tm_layout(title = "Male IHD Clusters - WPG",
         inner.margins = c(0.02,0.02,0.08,0.07),
         title.fontfamily="mono",
         title.position=c("center","top"),
         legend.title.fontfamily="mono",
         legend.text.fontfamily="mono",
         legend.position=c("left","bottom"),
         legend.title.size = 1.2,
         legend.text.size = 0.8,
         title.size=1.25) +
 tm_borders(lwd = 1, lty = "solid")+
 tm_text("regm",size=0.65)

tmap_arrange(mb_ihd_clust_m, wpg_ihd_clust_m, ncol = 2)
```

```{r}
#Example map

mb_tmap = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
             stringsAsFactors=F) %>%
 mutate(reg = ifelse(substr(Region,1,2)=="W1" | substr(Region,1,2)=="W0" |
substr(Region,1,3)=="WE2","",Region),
       dummy = rep(1,96))

wpg_tmap = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
             stringsAsFactors=F) %>%
 filter(substr(Region,1,2)=="W0" | substr(Region,1,2)=="W1")%>%
 mutate(dummy = rep(1,25))

mbtmap = tm_shape(mb_tmap) +
 tm_fill("dummy",
       palette="#F5F5DC") +
 tm_layout(title = "",
         inner.margins = c(0.02,0.02,0.02,0.02),
         frame=F,
         legend.show=F) +
 tm_borders(lwd = 1, lty = "solid")+
 tm_text("reg",size=0.5)
```

```
wpgtmap = tm_shape(wpg_tmap) +
 tm_fill("dummy",
       title = "",
       palette="#F5F5DC",
       title.fontfamily="mono") +
 tm_layout(title = "",
         inner.margins = c(0.02,0.02,0.2,0.02),
         legend.show=F,
         frame=F) +
 tm_borders(lwd = 1, lty = "solid")+
 tm_text("Region",size=0.65)
tmap_arrange(mbtmap,wpgtmap,ncol=2)
```

```{r}
#Assumption check - Normality of response and linear in predictors

pdat = data.frame(IHD = IHD[,70], E = IHD[,72], hyp = IHD[,73], sefis = SM$sefis, ind = SM$IND, inds =
SM$INDs,urb, sefi = SM$sefi) %>%
 mutate(rhad = substr(AMI[,1],1,4),
      RHA = ifelse(substr(rhad,1,2)=="SO","Southern",
            ifelse(substr(rhad,1,2)=="WE","Prairie Mountain",
                ifelse(substr(rhad,1,2)=="IE","Interlake-Eastern",
                    ifelse(substr(rhad,1,2)=="NO","Northern","Winnipeg")))))

hist(log(pdat$IHD,pdat$E),breaks=15)

ggplot(data=pdat)+
 geom_histogram(mapping = aes(log(IHD/E)),bins=10,color="white",fill="#8B0000")+
 labs(x = "Model Response", y = "Count")+
 ggtitle("Histogram of log(IHD Risk Ratio)")+
 theme_light()

hp = ggplot(data = pdat) +
 geom_point(mapping = aes(x=hyp,y=log(IHD/E),color=RHA))+
 labs(x = "Hypertension RR", y = "log(IHD RR)")+
 geom_abline(intercept=-0.6847,slope=0.6281)+
 scale_color_manual(breaks = c("Southern","Winnipeg","Prairie Mountain","Interlake-
Eastern","Northern"),
            values=wes_palette(n=5, name="Cavalcanti1", type="continuous"))+
 ggtitle("Hypertension")+
 theme_light()
hp
sp = ggplot(data = pdat) +
 geom_point(mapping = aes(x=sefi, y=log(IHD/E), color=RHA))+
 labs(x = "SEFI-2", y = "log(IHD RR)")+
 geom_abline(intercept=-0.04099,slope=0.10461)+
```

```
 scale_color_manual(breaks = c("Southern","Winnipeg","Prairie Mountain","Interlake-
Eastern","Northern"),
                values=wes_palette(n=5, name="Cavalcanti1", type="continuous"))+
 ggtitle(" SEFI-2")+
 theme_light()
sp
ip = ggplot(data = pdat) +
 geom_point(mapping = aes(x=log(inds), y=log(IHD/E),color=RHA))+
 labs(x = "log(Indigenous Population Proportion)", y = "log(IHD RR)")+
 geom_abline(intercept=0.18039,slope=0.11097)+
 ggtitle(" Indigenous Population")+
 scale_color_manual(breaks = c("Southern","Winnipeg","Prairie Mountain","Interlake-
Eastern","Northern"),
                values=wes_palette(n=5, name="Cavalcanti1", type="continuous"))+
 theme_light()

grid.arrange(hp,ip, nrow = 2,top = "Covariates Versus Response")

```

```{r}
#Poisson and Spatial Poisson - INLA

temp = poly2nb(mb)
nb2INLA("MBINLA", temp)
MB.adj = paste(getwd(), "/MBINLA",sep="")
H = inla.read.graph(filename="MBINLA")

#Poisson Model

IHD_2015 = IHD %>%
 select(rhad, count_2015,E_2015,ht_srr_2015)%>%
 mutate(ind = log(SM$INDs),
       urban = urb,
       region = seq(1:96),
       reg = seq(1:96),
       sef = SM$sefis) %>%
 rename(E=E_2015, IHD=count_2015, HT=ht_srr_2015)

form.pois = IHD ~ HT + ind + urban


mod.pois = inla(form.pois, family="poisson", data=IHD_2015, E=E,
                control.compute=list(dic=T,cpo=T))
mod.pois$summary.fixed
mod.pois$dic$dic

write.csv(mod.pois$summary.fixed,"poisson.csv")
```

```
res = abs((IHD_2015$IHD/IHD_2015$E) - mod.pois$summary.fitted.values$mean)

#Initial Spatial Model

form.spat1 = IHD ~ f(region, model="bym", graph=MB.adj) + HT + ind + urban

mod.spat1 = inla(form.spat1, family="poisson", data=IHD_2015, E=E,
                 control.compute=list(dic=T,cpo=T))

mod.spat1$summary.fixed
mod.spat1$dic$dic

res_spat1 = abs((IHD_2015$IHD/IHD_2015$E) - mod.spat1$summary.fitted.values$mean)

write.csv(mod.spat1$summary.fixed,"spat1.csv")

#control for SES

form.spat2 = IHD ~ f(region, model="besag", graph=MB.adj) + f(reg, model="iid", graph=MB.adj) + HT +
ind + urban + sef

mod.spat2 = inla(form.spat2, family="poisson", data=IHD_2015, E=E,
                 control.compute=list(dic=T,cpo=T))

mod.spat2$summary.fixed
mod.spat2$dic$dic

write.csv(mod.spat2$summary.fixed,"spat2.csv")

res_spat2 = abs((IHD_2015$IHD/IHD_2015$E) - mod.spat2$summary.fitted.values$mean)

#Females

IHD_2015_F = IHDF %>%
  select(rhad, count_2015,E_2015,ht_srr_2015)%>%
  mutate(ind = log(SM$INDs),
         urban = urb,
         region = seq(1:96),
         sef = SM$sefis) %>%
  rename(E=E_2015, IHD=count_2015, HT=ht_srr_2015)

form.spat2f = IHD ~ f(region, model="bym", graph=MB.adj) + HT + ind + urban + sef

mod.spat2f = inla(form.spat2f, family="poisson", data=IHD_2015_F, E=E,
                  control.compute=list(dic=T,cpo=T))
```

```
mod.spat2f$summary.fixed
mod.spat2f$dic$dic

write.csv(mod.spat2f$summary.fixed,"spat2f.csv")

#males

IHD_2015_M = IHDM %>%
 select(rhad, count_2015,E_2015,ht_srr_2015)%>%
 mutate(ind = log(SM$INDs),
       urban = urb,
       region = seq(1:96),
       sef = SM$sefis) %>%
 rename(E=E_2015, IHD=count_2015, HT=ht_srr_2015)

form.spat2m = IHD ~ f(region, model="bym", graph=MB.adj) + HT + ind + urban + sef

mod.spat2m = inla(form.spat2m, family="poisson", data=IHD_2015_M, E=E,
              control.compute=list(dic=T,cpo=T))

mod.spat2m$summary.fixed
mod.spat2m$dic$dic

write.csv(mod.spat2m$summary.fixed,"spat2m.csv")

#Mapping Residuals fitted values and elevated regions.

mb_tmap = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
             stringsAsFactors=F) %>%
 mutate(res = res,
       res_spat1 = res_spat1,
       res_spat2 = res_spat2,
       obs = IHD_2015$IHD/IHD_2015$E,
       fitt = mod.spat2$summary.fitted.values$mean,
       sig = ifelse(mod.spat2$summary.fitted.values[[3]]>1,"Elevated","Not Elevated"),
       sigf = ifelse(mod.spat2f$summary.fitted.values[[3]]>1,"Elevated","Not Elevated"),
       sigm = ifelse(mod.spat2m$summary.fitted.values[[3]]>1,"Elevated","Not Elevated"))

wpg_tmap = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
             stringsAsFactors=F) %>%
 mutate(res = res,
       res_spat1 = res_spat1,
       res_spat2 = res_spat2,
       obs = IHD_2015$IHD/IHD_2015$E,
       fitt = mod.spat2$summary.fitted.values$mean,
       sig = ifelse(mod.spat2$summary.fitted.values[[3]]>1,"Elevated","Not Elevated"),
       sigf = ifelse(mod.spat2f$summary.fitted.values[[3]]>1,"Elevated","Not Elevated"),
       sigm = ifelse(mod.spat2m$summary.fitted.values[[3]]>1,"Elevated","Not Elevated")) %>%
```

```
 filter(substr(Region,1,2)=="W0" | substr(Region,1,2)=="W1")

cuts = c(0, 0.8, 0.9, 1, 1.2, 2)

fitmb = tm_shape(mb_tmap) +
 tm_polygons("fitt",
          breaks = cuts,
          palette="Reds",
          border.alpha=0.5,
          title = "IHD SRR") +
 tm_layout(title = "Fitted IHD SRR - MB",
        inner.margins = c(0.02,0.1,0.02,0.02))

fitwpg = tm_shape(wpg_tmap) +
 tm_polygons("fitt",
          breaks = cuts,
          palette="Reds",
          border.alpha=0.5,
          title = "") +
 tm_layout(title = "Fitted IHD SRR - WPG",
        inner.margins = c(0.02,0.1,0.02,0.02),
        legend.show=F)

tmap_arrange(fitmb,fitwpg,ncol=2)

sigcols <- c("#8B0000","#F5F5DC")

mb_sig = tm_shape(mb_tmap) +
 tm_fill("sigm",
       title = "Cluster Type",
       palette=sigcols,
       title.fontfamily="mono") +
 tm_layout(title = "Male IHD Prevalence - MB",
        inner.margins = c(0.02,0.02,0.08,0.02),
        title.fontfamily="mono",
        title.position=c("center","top"),
        legend.show=F,
        title.size=1.25) +
 tm_borders(lwd = 1, lty = "solid")

wpg_sig = tm_shape(wpg_tmap) +
 tm_fill("sigm",
       title = "Elevated Regions",
       palette=sigcols,
       title.fontfamily="mono") +
 tm_layout(title = "Male IHD Prevalence - WPG",
        inner.margins = c(0.02,0.02,0.08,0.07),
        title.fontfamily="mono",
```

```
        title.position=c("center","top"),
        legend.title.fontfamily="mono",
        legend.text.fontfamily="mono",
        legend.position=c("left","bottom"),
        legend.title.size = 1,
        legend.text.size = 0.8,
        title.size=1.25) +
 tm_borders(lwd = 1, lty = "solid")

tmap_arrange(mb_sig,wpg_sig,ncol=2)
```

```{r}
#spatio-temporal

#data prep

w = seq(from=3, to=71, by=4)
x = seq(from=2, to=70, by=4)
y = seq(from=5, to=73, by=4)
z = seq(from=4, to=72, by=4)

ihda = as.matrix(IHD[,x])
hypa = as.matrix(IHD[,y])
expa = as.matrix(IHD[,z])
popa = as.matrix(IHD[,w])

ihdf = as.matrix(IHDF[,x])
hypf = as.matrix(IHDF[,y])
expf = as.matrix(IHDF[,z])
popf = as.matrix(IHDF[,w])

ihdm = as.matrix(IHDM[,x])
hypm = as.matrix(IHDM[,y])
expm = as.matrix(IHDM[,z])
popm = as.matrix(IHDM[,w])

#Spatio-temporal graphs

ihd_pl = data.frame(ihda) %>%
 mutate(rhad = substr(IHD[,1],1,4),
      RHA = ifelse(substr(rhad,1,2)=="SO","Southern",
            ifelse(substr(rhad,1,2)=="WE","Prairie Mountain",
                ifelse(substr(rhad,1,2)=="IE","Interlake-Eastern",
                    ifelse(substr(rhad,1,2)=="NO","Northern","Winnipeg")))))
exp_pl = data.frame(expa) %>%
 mutate(rhad = substr(IHD[,1],1,4),
      RHA = ifelse(substr(rhad,1,2)=="SO","Southern",
```

```
                ifelse(substr(rhad,1,2)=="WE","Prairie Mountain",
                    ifelse(substr(rhad,1,2)=="IE","Interlake-Eastern",
                        ifelse(substr(rhad,1,2)=="NO","Northern","Winnipeg")))))

WP_ihd = ihd_pl %>%
 filter(RHA=="Winnipeg")%>%
 select(-RHA,-rhad)

WP_exp = exp_pl %>%
 filter(RHA=="Winnipeg")%>%
 select(-RHA,-rhad)

WE_ihd = ihd_pl %>%
 filter(RHA=="Prairie Mountain")%>%
 select(-RHA,-rhad)

WE_exp = exp_pl %>%
 filter(RHA=="Prairie Mountain")%>%
 select(-RHA,-rhad)

IE_ihd = ihd_pl %>%
 filter(RHA=="Interlake-Eastern")%>%
 select(-RHA,-rhad)

IE_exp = exp_pl %>%
 filter(RHA=="Interlake-Eastern")%>%
 select(-RHA,-rhad)

SO_ihd = ihd_pl %>%
 filter(RHA=="Southern")%>%
 select(-RHA,-rhad)

SO_exp = exp_pl %>%
 filter(RHA=="Southern")%>%
 select(-RHA,-rhad)

NO_ihd = ihd_pl %>%
 filter(RHA=="Northern")%>%
 select(-RHA,-rhad)

NO_exp = exp_pl %>%
 filter(RHA=="Northern")%>%
 select(-RHA,-rhad)

std_ihd = matrix(nrow=18,ncol=5)

for(i in 1:18){
std_ihd[i,1] = log(sum(WP_ihd[,i])/sum(WP_exp[,i]))
```

```
std_ihd[i,2] = log(sum(WE_ihd[,i])/sum(WE_exp[,i]))
std_ihd[i,3] = log(sum(IE_ihd[,i])/sum(IE_exp[,i]))
std_ihd[i,4] = log(sum(SO_ihd[,i])/sum(SO_exp[,i]))
std_ihd[i,5] = log(sum(NO_ihd[,i])/sum(NO_exp[,i]))
}

stand_ihd = data.frame(cbind(Year = seq(from=1998, to=2015, by=1), WP = std_ihd[,1], WE =
std_ihd[,2], IE = std_ihd[,3]), SO = std_ihd[,4], NO = std_ihd[,5])

stand_ihd = stand_ihd %>%
 gather(RHA, SRR, c(WP, WE, IE, SO, NO)) %>%
 mutate(RHA = ifelse(RHA=="WP","Winnipeg",
                ifelse(RHA=="WE","Prairie Mountain",
                    ifelse(RHA=="IE","Interlake-Eastern",
                        ifelse(RHA=="SO","Southern","Northern")))))

ggplot(data=stand_ihd)+
 geom_line(aes(x=Year,y=SRR,color=RHA))+
 geom_point(aes(y=SRR,x=Year,color=RHA))+
 labs(y="log(IHD SRR)", x="Year")+
 ggtitle("IHD Time Series Plot")+
 scale_color_manual(breaks = c("Southern","Winnipeg","Prairie Mountain","Interlake-
Eastern","Northern"),
                values=wes_palette(n=5, name="Cavalcanti1"))+
 theme_light()

#Data prep for INLA

ihdINLA = as.vector(ihda)
eINLA = as.vector(expa)
hypINLA = as.vector(hypa)


INLA = data.frame(cbind(rhad = rep(as.character(IHD[,1]),18), year = rep(1998:2015,each=96), IHD =
ihdINLA, E = eINLA, HT = hypINLA, region = rep(1:96,18))) %>%
 mutate(rhad = substr(rhad,1,4),
       year = as.numeric(as.character(year)),
       year1 = as.numeric(as.character(year)),
       yr = rep(1:18,each=96),
       yr1 = rep(1:18,each=96),
       IHD = as.numeric(as.character(IHD)),
       E = as.numeric(as.character(E)),
       HT = as.numeric(as.character(HT)),
       region = as.numeric(as.character(region)),
       reg = as.numeric(as.character(region)),
       inter = seq(from=1,to=1728),
       sefis = rep(SM$sefis,18),
       inds = rep(log(SM$INDs),18),
```

```
       urban = rep(urb,18))%>%
  mutate(rhad = as.factor(rhad))

temp = poly2nb(mb)
nb2INLA("MBINLA", temp)
MB.adj = paste(getwd(), "/MBINLA",sep="")
H = inla.read.graph(filename="MBINLA")
image(inla.graph2matrix(H))

#Model Without Interaction

form.IHD = IHD ~ f(region, model="bym", graph=MB.adj) +
                 f(yr, model="rw2") +
                 f(yr1, model="iid") +
                 HT +
                 inds +
                 urban +
                 sefis

mod.IHD = inla(form.IHD, family="poisson", data=INLA, E=E,
               control.compute=list(dic=T,cpo=T))


#Interaction type I (temporally and spatially independent)

form.IHD.intI = IHD ~ f(region, model="bym", graph=MB.adj) +
                  f(yr, model="rw2") +
                  f(yr1, model="iid") +
                  f(inter, model="iid") +
                  HT +
                  inds +
                  urban +
                  sefis

mod.IHD.intI = inla(form.IHD.intI, family="poisson", data=INLA, E=E,
               control.compute=list(dic=T,cpo=T))

#Interaction type II (temporally dependent)

area.int = INLA$region
year.int = INLA$yr

form.IHD.intII = IHD ~ f(region, model="besag", graph=MB.adj) + f(reg, model="iid", graph=MB.adj) +
                  f(yr, model="rw2") +
                  f(yr1, model="iid") +
                  f(area.int, model="iid", group=year.int,
                    control.group=list(model="rw2")) +
                  HT +
```

```
                inds +
                urban +
                sefis

mod.IHD.intII = inla(form.IHD.intII, family="poisson", data=INLA, E=E,
              control.compute=list(dic=T,cpo=T))

#Interaction type III (Spatially dependent)

form.IHD.intIII = IHD ~ f(region, model="bym", graph=MB.adj) +
                  f(yr, model="rw2") +
                  f(yr1, model="iid") +
                  f(year.int, model="iid", group=area.int,
                    control.group=list(model="besag", graph=MB.adj)) +
                  HT +
                  inds +
                  urban +
                  sefis

mod.IHD.intIII = inla(form.IHD.intIII, family="poisson", data=INLA, E=E,
              control.compute=list(dic=T,cpo=T))

#Interaction type IV (Spatially and temporally dependent)

form.IHD.intIV = IHD ~ f(region, model="bym", graph=MB.adj) +
                  f(yr, model="rw2") +
                  f(yr1, model="iid") +
                  f(year.int, model="rw2",
                    group=area.int,
                    control.group=list(model="besag", graph=MB.adj)) +
                  HT +
                  inds +
                  urban +
                  sefis

mod.IHD.intIV = inla(form.IHD.intIV, family="poisson", data=INLA, E=E,
              control.compute=list(dic=T,cpo=T))

#Female stratified data

ihdINLAf = as.vector(ihdf)
eINLAf = as.vector(expf)
hypINLAf = as.vector(hypf)

INLAf = data.frame(cbind(rhad = rep(as.character(IHD[,1]),18), year = rep(1998:2015,each=96), IHD =
ihdINLAf, E = eINLAf, HT = hypINLAf, region = rep(1:96,18))) %>%
  mutate(rhad = substr(rhad,1,4),
         year = as.numeric(as.character(year)),
```

```
      year1 = as.numeric(as.character(year)),
      yr = rep(1:18,each=96),
      yr1 = rep(1:18,each=96),
      IHD = as.numeric(as.character(IHD)),
      E = as.numeric(as.character(E)),
      HT = as.numeric(as.character(HT)),
      region = as.numeric(as.character(region)),
      inter = seq(from=1,to=1728),
      sefis = rep(SM$sefis,18),
      inds = rep(log(SM$INDs),18),
      urban = rep(urb,18))%>%
  mutate(rhad = as.factor(rhad))

#Without Interaction

mod.IHDf = inla(form.IHD, family="poisson", data=INLAf, E=E,
                control.compute=list(dic=T,cpo=T))

#Interaction type I

mod.IHD.intIf = inla(form.IHD.intI, family="poisson", data=INLAf, E=E,
                control.compute=list(dic=T,cpo=T))

#Interaction type II (temporally dependent)

mod.IHD.intIIf = inla(form.IHD.intII, family="poisson", data=INLAf, E=E,
                control.compute=list(dic=T,cpo=T))

#Interaction type III

mod.IHD.intIIIf = inla(form.IHD.intIII, family="poisson", data=INLAf, E=E,
                control.compute=list(dic=T,cpo=T))

#Interaction type IV

mod.IHD.intIVf = inla(form.IHD.intIV, family="poisson", data=INLAf, E=E,
                control.compute=list(dic=T,cpo=T))

#male stratified data

ihdINLAm = as.vector(ihdm)
eINLAm = as.vector(expm)
hypINLAm = as.vector(hypm)

INLAm = data.frame(cbind(rhad = rep(as.character(IHD[,1]),18), year = rep(1998:2015,each=96), IHD =
ihdINLAm, E = eINLAm, HT = hypINLAm, region = rep(1:96,18))) %>%
  mutate(rhad = substr(rhad,1,4),
      year = as.numeric(as.character(year)),
```

```
      year1 = as.numeric(as.character(year)),
      yr = rep(1:18,each=96),
      yr1 = rep(1:18,each=96),
      IHD = as.numeric(as.character(IHD)),
      E = as.numeric(as.character(E)),
      HT = as.numeric(as.character(HT)),
      region = as.numeric(as.character(region)),
      inter = seq(from=1,to=1728),
      sefis = rep(SM$sefis,18),
      inds = rep(log(SM$INDs),18),
      urban = rep(urb,18))%>%
 mutate(rhad = as.factor(rhad))

#Without Interaction

mod.IHDm = inla(form.IHD, family="poisson", data=INLAm, E=E,
                control.compute=list(dic=T,cpo=T))

#Interaction type I

mod.IHD.intIm = inla(form.IHD.intI, family="poisson", data=INLAm, E=E,
                control.compute=list(dic=T,cpo=T))

#Interaction type II (temporally dependent)

mod.IHD.intIIm = inla(form.IHD.intII, family="poisson", data=INLAm, E=E,
                control.compute=list(dic=T,cpo=T))

#Interaction type III

mod.IHD.intIIIm = inla(form.IHD.intIII, family="poisson", data=INLAm, E=E,
                control.compute=list(dic=T,cpo=T))

#Interaction type IV

mod.IHD.intIVm = inla(form.IHD.intIV, family="poisson", data=INLAm, E=E,
                control.compute=list(dic=T,cpo=T))

#Check DIC's to obtain best model (Type II interaction)
mod.IHD$dic$dic;mod.IHD.intI$dic$dic;mod.IHD.intII$dic$dic;mod.IHD.intIII$dic$dic;mod.IHD.intIV$dic$
dic
mod.IHDf$dic$dic;mod.IHD.intIf$dic$dic;mod.IHD.intIIf$dic$dic;mod.IHD.intIIIf$dic$dic;mod.IHD.intIVf$
dic$dic
mod.IHDm$dic$dic;mod.IHD.intIm$dic$dic;mod.IHD.intIIm$dic$dic;mod.IHD.intIIIm$dic$dic;mod.IHD.in
tIVm$dic$dic
```

```{r}
```

```
#output

write.csv(mod.IHD.intII$summary.fixed, "spatempo.csv")
write.csv(mod.IHD.intIIf$summary.fixed, "spatempf.csv")
write.csv(mod.IHD.intIIm$summary.fixed, "spatempm.csv")

fittso = mod.IHD.intII$summary.fitted.values$mean
fittsf = mod.IHD.intIIf$summary.fitted.values$mean
fittsm = mod.IHD.intIIm$summary.fitted.values$mean

fittso = matrix(fittso,nrow=96,ncol=18)
fittsf = matrix(fittsf,nrow=96,ncol=18)
fittsm = matrix(fittsm,nrow=96,ncol=18)

fittso = data.frame(fittso)
fittsf = data.frame(fittsf)
fittsm = data.frame(fittsm)

colnames(fittso) = seq(1998,2015)
colnames(fittsf) = seq(1998,2015)
colnames(fittsm) = seq(1998,2015)

fittso = data.frame(cbind(Region = as.character(substr(IHD[,1],1,4)),fittso))
fittsf = data.frame(cbind(Region = as.character(substr(IHD[,1],1,4)),fittsf))
fittsm = data.frame(cbind(Region = as.character(substr(IHD[,1],1,4)),fittsm))

signifo = mod.IHD.intII$summary.fitted.values[[3]]
signiff = mod.IHD.intIIf$summary.fitted.values[[3]]
signifm = mod.IHD.intIIm$summary.fitted.values[[3]]

signifo = matrix(signifo,nrow=96,ncol=18)
signiff = matrix(signiff,nrow=96,ncol=18)
signifm = matrix(signifm,nrow=96,ncol=18)

for(i in 1:96){
  for(j in 1:18){
    signifo[i,j] = ifelse(signifo[i,j]>1,1,0)
    signiff[i,j] = ifelse(signiff[i,j]>1,1,0)
    signifm[i,j] = ifelse(signifm[i,j]>1,1,0)
  }
}
signifo = data.frame(signifo)
signiff = data.frame(signiff)
signifm = data.frame(signifm)

colnames(signifo) = seq(1998:2015)
colnames(signiff) = seq(1998:2015)
colnames(signifm) = seq(1998:2015)
```

```
signifo = data.frame(cbind(Region = as.character(substr(IHD[,1],1,4)),signifo))
signiff = data.frame(cbind(Region = as.character(substr(IHD[,1],1,4)),signiff))
signifm = data.frame(cbind(Region = as.character(substr(IHD[,1],1,4)),signifm))

#tmap for visualizing

mbfito = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
             stringsAsFactors=F) %>%
  mutate(Fitted_1998 = fittso[,2],
         Fitted_1999 = fittso[,3],
         Fitted_2000 = fittso[,4],
         Fitted_2001 = fittso[,5],
         Fitted_2002 = fittso[,6],
         Fitted_2003 = fittso[,7],
         Fitted_2004 = fittso[,8],
         Fitted_2005 = fittso[,9],
         Fitted_2006 = fittso[,10],
         Fitted_2007 = fittso[,11],
         Fitted_2008 = fittso[,12],
         Fitted_2009 = fittso[,13],
         Fitted_2010 = fittso[,14],
         Fitted_2011 = fittso[,15],
         Fitted_2012 = fittso[,16],
         Fitted_2013 = fittso[,17],
         Fitted_2014 = fittso[,18],
         Fitted_2015 = fittso[,19]) %>%
  gather(year, fitted, c(Fitted_1998, Fitted_1999, Fitted_2000, Fitted_2001, Fitted_2002, Fitted_2003,
Fitted_2004,
                    Fitted_2005, Fitted_2006, Fitted_2007, Fitted_2008, Fitted_2009, Fitted_2010,
Fitted_2011,
                    Fitted_2012, Fitted_2013, Fitted_2014, Fitted_2015))

wpgfito = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
             stringsAsFactors=F) %>%
  mutate(Fitted_1998 = fittso[,2],
         Fitted_1999 = fittso[,3],
         Fitted_2000 = fittso[,4],
         Fitted_2001 = fittso[,5],
         Fitted_2002 = fittso[,6],
         Fitted_2003 = fittso[,7],
         Fitted_2004 = fittso[,8],
         Fitted_2005 = fittso[,9],
         Fitted_2006 = fittso[,10],
         Fitted_2007 = fittso[,11],
         Fitted_2008 = fittso[,12],
         Fitted_2009 = fittso[,13],
         Fitted_2010 = fittso[,14],
```

```
      Fitted_2011 = fittso[,15],
      Fitted_2012 = fittso[,16],
      Fitted_2013 = fittso[,17],
      Fitted_2014 = fittso[,18],
      Fitted_2015 = fittso[,19]) %>%
 gather(year, fitted, c(Fitted_1998, Fitted_1999, Fitted_2000, Fitted_2001, Fitted_2002, Fitted_2003,
Fitted_2004,
                  Fitted_2005, Fitted_2006, Fitted_2007, Fitted_2008, Fitted_2009, Fitted_2010,
Fitted_2011,
                  Fitted_2012, Fitted_2013, Fitted_2014, Fitted_2015)) %>%
 filter(substr(Region,1,2)=="W0" | substr(Region,1,2)=="W1")

mbfitf = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
            stringsAsFactors=F) %>%
 mutate(Fitted_1998 = fittsf[,2],
      Fitted_1999 = fittsf[,3],
      Fitted_2000 = fittsf[,4],
      Fitted_2001 = fittsf[,5],
      Fitted_2002 = fittsf[,6],
      Fitted_2003 = fittsf[,7],
      Fitted_2004 = fittsf[,8],
      Fitted_2005 = fittsf[,9],
      Fitted_2006 = fittsf[,10],
      Fitted_2007 = fittsf[,11],
      Fitted_2008 = fittsf[,12],
      Fitted_2009 = fittsf[,13],
      Fitted_2010 = fittsf[,14],
      Fitted_2011 = fittsf[,15],
      Fitted_2012 = fittsf[,16],
      Fitted_2013 = fittsf[,17],
      Fitted_2014 = fittsf[,18],
      Fitted_2015 = fittsf[,19]) %>%
 gather(year, fitted, c(Fitted_1998, Fitted_1999, Fitted_2000, Fitted_2001, Fitted_2002, Fitted_2003,
Fitted_2004,
                  Fitted_2005, Fitted_2006, Fitted_2007, Fitted_2008, Fitted_2009, Fitted_2010,
Fitted_2011,
                  Fitted_2012, Fitted_2013, Fitted_2014, Fitted_2015))

wpgfitf = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
            stringsAsFactors=F) %>%
 mutate(Fitted_1998 = fittsf[,2],
      Fitted_1999 = fittsf[,3],
      Fitted_2000 = fittsf[,4],
      Fitted_2001 = fittsf[,5],
      Fitted_2002 = fittsf[,6],
      Fitted_2003 = fittsf[,7],
      Fitted_2004 = fittsf[,8],
      Fitted_2005 = fittsf[,9],
```

```
      Fitted_2006 = fittsf[,10],
      Fitted_2007 = fittsf[,11],
      Fitted_2008 = fittsf[,12],
      Fitted_2009 = fittsf[,13],
      Fitted_2010 = fittsf[,14],
      Fitted_2011 = fittsf[,15],
      Fitted_2012 = fittsf[,16],
      Fitted_2013 = fittsf[,17],
      Fitted_2014 = fittsf[,18],
      Fitted_2015 = fittsf[,19]) %>%
 gather(year, fitted, c(Fitted_1998, Fitted_1999, Fitted_2000, Fitted_2001, Fitted_2002, Fitted_2003,
Fitted_2004,
                  Fitted_2005, Fitted_2006, Fitted_2007, Fitted_2008, Fitted_2009, Fitted_2010,
Fitted_2011,
                  Fitted_2012, Fitted_2013, Fitted_2014, Fitted_2015)) %>%
 filter(substr(Region,1,2)=="W0" | substr(Region,1,2)=="W1")

mbfitm = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
            stringsAsFactors=F) %>%
 mutate(Fitted_1998 = fittsm[,2],
      Fitted_1999 = fittsm[,3],
      Fitted_2000 = fittsm[,4],
      Fitted_2001 = fittsm[,5],
      Fitted_2002 = fittsm[,6],
      Fitted_2003 = fittsm[,7],
      Fitted_2004 = fittsm[,8],
      Fitted_2005 = fittsm[,9],
      Fitted_2006 = fittsm[,10],
      Fitted_2007 = fittsm[,11],
      Fitted_2008 = fittsm[,12],
      Fitted_2009 = fittsm[,13],
      Fitted_2010 = fittsm[,14],
      Fitted_2011 = fittsm[,15],
      Fitted_2012 = fittsm[,16],
      Fitted_2013 = fittsm[,17],
      Fitted_2014 = fittsm[,18],
      Fitted_2015 = fittsm[,19]) %>%
 gather(year, fitted, c(Fitted_1998, Fitted_1999, Fitted_2000, Fitted_2001, Fitted_2002, Fitted_2003,
Fitted_2004,
                  Fitted_2005, Fitted_2006, Fitted_2007, Fitted_2008, Fitted_2009, Fitted_2010,
Fitted_2011,
                  Fitted_2012, Fitted_2013, Fitted_2014, Fitted_2015))

wpgfitm = st_read("C:/School/Grad School/Thesis/Data/Shape/comb/combined.shp",
            stringsAsFactors=F) %>%
 mutate(Fitted_1998 = fittsm[,2],
      Fitted_1999 = fittsm[,3],
      Fitted_2000 = fittsm[,4],
```

```
      Fitted_2001 = fittsm[,5],
      Fitted_2002 = fittsm[,6],
      Fitted_2003 = fittsm[,7],
      Fitted_2004 = fittsm[,8],
      Fitted_2005 = fittsm[,9],
      Fitted_2006 = fittsm[,10],
      Fitted_2007 = fittsm[,11],
      Fitted_2008 = fittsm[,12],
      Fitted_2009 = fittsm[,13],
      Fitted_2010 = fittsm[,14],
      Fitted_2011 = fittsm[,15],
      Fitted_2012 = fittsm[,16],
      Fitted_2013 = fittsm[,17],
      Fitted_2014 = fittsm[,18],
      Fitted_2015 = fittsm[,19]) %>%
  gather(year, fitted, c(Fitted_1998, Fitted_1999, Fitted_2000, Fitted_2001, Fitted_2002, Fitted_2003,
Fitted_2004,
                  Fitted_2005, Fitted_2006, Fitted_2007, Fitted_2008, Fitted_2009, Fitted_2010,
Fitted_2011,
                  Fitted_2012, Fitted_2013, Fitted_2014, Fitted_2015)) %>%
  filter(substr(Region,1,2)=="W0" | substr(Region,1,2)=="W1")

cuts = c(0, 0.5, 0.8, 0.9, 1, 1.1, 1.2, 1.5, 2, 3, 4)

tm_shape(mbfitm) +
  tm_polygons("fitted",
          breaks = cuts,
          palette="Reds",
          border.alpha=0.5,
          title = "Fitted SRR") +
  tm_layout(title = "Male IHD SRR - MB",
        inner.margins = c(0.02,0.02,0.1,0.02))  +
  tm_facets(by = "year", free.coords=T, ncol=3)

tm_shape(wpgfitm) +
  tm_polygons("fitted",
          breaks = cuts,
          palette="Reds",
          border.alpha=0.5,
          title = "Fitted SRR") +
  tm_layout(title = "Male IHD SRR - WPG",
        inner.margins = c(0.02,0.02,0.1,0.02))  +
  tm_facets(by = "year", free.coords=T, ncol=3)

#Graph of Temporal random effects term (Overall change in IHD prevalence over time, holding all else
constant)

phi = data.frame(mod.IHD.intII$summary.random$yr$mean)
```

```
phif = mod.IHD.intIIf$summary.random$yr$mean
phim = mod.IHD.intIIm$summary.random$yr$mean

phidat = phi%>%
  mutate(Year = seq(1998,2015,1),
         Overall = phi[,1],
         Female = phif,
         Male = phim)%>%
  select(-mod.IHD.intII.summary.random.yr.mean) %>%
  gather(Strata, PHI, c(Overall, Female, Male))

ggplot(data=phidat)+
  geom_line(aes(x=Year,y=exp(PHI),color=Strata))+
  geom_point(aes(y=exp(PHI),x=Year,color=Strata))+
  labs(y="Temporal Random Effect", x="Year")+
  scale_color_manual(breaks = c("Overall", "Male", "Female"),
                     values=wes_palette(n=3, name="Cavalcanti1", type="continuous"))+
  theme_light() +
  geom_hline(yintercept=1)

#Ascertain regions with largest predicted increase and map their trends
#Found 8 regions that had a 25% or greater increase in predicted IHD SRR

fitted = data.frame(mod.IHD.intII$summary.fitted.values$mean) %>%
  mutate(rhad = rep(IHD[,1],18),
         Year = rep(1998:2015,each=96)) %>%
  rename(fit=mod.IHD.intII.summary.fitted.values.mean)

fitted1 = fitted %>%
  spread(key=Year,value=fit)

fitted1 = fitted1 %>%
  mutate(increasing = ifelse(fitted1[,2]+0.3*fitted1[,2]<fitted1[,19],1,0))

fitted1 = fitted1 %>%
  filter(increasing==1) %>%
  select(-increasing)

colnames(fitted1) = c("rhad","a","b","c","d","e","f","g","h","i","j","k","l","m","n","o","p","q","r")

fitted1 = fitted1 %>%
  gather("a","b","c","d","e","f","g","h","i","j","k","l","m","n","o","p","q","r",key="Year",value="RR")

fitted1 = fitted1 %>%
  mutate(year=rep(1998:2015,each=6))

ggplot(data=fitted1)+
```

```
  geom_line(aes(x=year,y=RR,color=rhad))+
  geom_point(aes(x=year,y=RR,color=rhad))+
  labs(y="IHD RR", x="Year",color="RHAD")+
  scale_color_manual(breaks = c("NO27 Z2 Cross Lake/Cross Lake FN", "WE15 N Dauphin", "WE14 N
Agassiz Mountain", "W04A St. Vital N", "NO22 Z2 Puk/Mat Col CN", "IE11 Selkirk"
                     values=wes_palette(n=6, name="Cavalcanti1", type="continuous"))+
  theme_light() +
  geom_hline(yintercept=1)
```