
Geographical Variation and Factors Associated with Gastric Cancer in
Manitoba

by

Oluwagbenga A. Fakanye

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: Gastric cancer (GC) is one of the deadliest diseases as most of the cases are diagnosed at late stages, thereby making treatment almost impossible. GC incidence in North America has been reported to be decreasing over the years. Nevertheless, the primary concern is whether the decrease is valid for all communities?

Purpose and Objectives: The purpose of this study is to identify high-risk GC hotspots in Manitoba and investigate factors associated with GC in Manitoba. The objectives were to (1) describe and investigate the geographical variation of GC incidence in Manitoba; (2) explore factors influencing the geographic change of GC incidence in Manitoba, and (3) investigate the geographical variation of GC incidence over time in Manitoba.

Methods: This research study adopted an ecological design. A spatial Poisson regression model was used to address research objectives (1) and (2), and a Spatio-temporal Poisson regression model was used to address research objective 3.

Results: SESI was significantly associated with cardia gastric cancer (CGC) and marginally associated with non-cardia gastric cancer (NCGC), while the Indigenous population proportion was marginally associated with CGC. In specific, 1 unit increase in SESI reduces the risk of CGC by 14% (IRR= 0.859; 95% CI: 0.780 - 0.947) and the risk of NCGC by approximately 10% (IRR = 0.898; 95% CI: 0.812 – 0.995); 1% increase in regional Indigenous population proportion reduces the risk of CGC by 1.4% (IRR = 0.986; 95% CI: 0.978 – 0.994). Also, 1 unit increase in SESI reduces the risk of CGC among women by 26.2% (IRR = 0.738; 95% CI: 0.618 – 0.879), and a 1% increase in Indigenous population proportion reduces the risk of CGC among women by 1.9% (IRR = 0.981; 95% CI: 0.966 – 0.996).

Conclusion: This study has demonstrated the existence of regional variation of GC incidence risk with temporal pattern in Manitoba.

Acknowledgments

To begin with, I want to thank the almighty God for the strength to persevere, wisdom to discern, and knowledge to complete this journey. My profound gratitude goes to my thesis advisor and mentor (Dr. Torabi). For his unquivering support right from the shore of my country since 2016. So with all my strength, I say a huge thank you for seeing this project through to completion.

I also appreciate the effort of my thesis committee members, Dr. Harminder Singh and Dr. Danielle Desautels. Their supportive and constructive criticism has helped to fine-tune this study through consistent provision of the medical and research expert advice.

I also want to acknowledge the financial support provided through Dr. Torabi's Natural Sciences and Engineering Research Council (NSERC) and Graduate Enhancement of Tri-Council stipend program (GETS) from the University of Manitoba.

I would also like to thank all staff of the George and Fay Yee Centre for Health Care Innovations (CHI) data science section. For the privilege given to me to serve and learn as a graduate research assistant intern during my first summer, an experience that exposed me to the world of health research.

I also want to extend my appreciation to the Research Data Center (RDC) university of Manitoba. For the privilege to carry out my research using their facilities and also awarding my research proposal, the RDC graduate student ward, I am grateful.

My appreciation also goes to all my Spatial Analytic Research group for their constant support during this project, especially Justin, who have been very supportive. And to my brother from another mother who has been very supportive right from my day one on this campus (Olawale) may you never walk alone, sir. To my parents here in Canada who took me in when I was homeless (pastor and Pastor Mrs. Adewumi), thanks for your support, especially spiritually.

Finally, my immeasurable gratitude goes to my family — especially my wife (Eyitayo), who has been supportive and understanding. For taking care of my angels (Anuoluwapo and Jesutofunmi) while daddy is far away in Canada studying, thanks dear, and I love you like *'kilode'*. Also, to the best mother in the world, I appreciate you, and I want to let you know that I am proud to be your son.

Dedication

This project is dedicated to the glory of almighty God, the author of our faith and the source of all knowledge and service to humanity.

Table of Contents

| | |
|--|------------|
| Abstract | i |
| Acknowledgments..... | iii |
| Dedication | v |
| List of Tables | ix |
| List of Figures | xi |
| List of Abbreviations | xv |
| | |
| Chapter 1: Introduction | 1 |
| 1.1: Background | 1 |
| 1.2: Research Aim and Objectives..... | 2 |
| 1.3: Research Contribution | 3 |
| 1.4: Ethics | 3 |
| | |
| Chapter 2: Literature Review | 4 |
| 2.1: Gastric Cancer: Risk and Protective Factors | 4 |
| 2.2: Environmental Risk Factors | 4 |
| 2.3: Genetic Risk Factors | 9 |
| | |
| Chapter 3: Methods | 10 |
| 3.1: Research Design | 10 |
| 3.2: Study Area | 10 |
| 3.3: Data | 12 |

| | |
|---|-----------|
| 3.3.1: Canada Cancer registry (CCR) | 12 |
| 3.3.2: Census Data | 13 |
| 3.3.3: Postal Code Conversion File (PCCF) | 17 |
| 3.3.4: Creation of Regional GC counts from the Canada Cancer Registry | 17 |
| 3.3.5: Creation of Regional Covariates from Census Data | 18 |
| 3.3.6: Merging of Regional GC and Regional Covariate dataset | 18 |
| 3.4: Statistical Analysis | 18 |
| 3.4.1: Poisson Regression Model | 18 |
| 3.4.2: Spatial Poisson Regression Model | 20 |
| 3.4.3: Spatio-temporal Poisson Regression Model | 23 |
| 3.4.4: Bayesian Inference Framework | 26 |
| 3.4.4.1: MCMC Convergence Diagnostics | 28 |
| 3.4.4.2: Prior Sensitivity Analysis | 30 |
| 3.5: Research Scope | 30 |
| Chapter 4: Results | 32 |
| 4.1: Descriptive Results | 32 |
| 4.2: Confirmatory results | 47 |
| 4.2.1: Poisson Regression Model | 47 |
| 4.2.2: Spatial Dependence Assessment | 52 |
| 4.2.3: Spatial Result: Ecological Regression Model | 53 |
| 4.2.4: Spatio-temporal Regression Result | 61 |
| 4.2.5: Prior Sensitivity Analysis Result | 69 |

| | |
|---|-----------|
| Chapter 5: Discussion and Conclusion | 72 |
| 5.1: Discussion | 72 |
| 5.1.1: Spatial Model | 72 |
| 5.1.2: Spatio-temporal Model | 73 |
| 5.2: Strengths and Limitations | 74 |
| 5.3: Future research | 75 |
| 5.4: Conclusion | 76 |
| References: | 77 |
| Appendices: | 88 |
| Appendix I: Deviance Information Criteria (DIC) for all models fitted | 88 |
| Appendix II: Deviance information criteria (DIC) for the Poisson and Spatial regression model..... | 90 |
| Appendix III: Map showing smoothed GC risk ratio for uniform and log gamma prior assumption | 91 |
| Appendix IV: Model diagnostics using Probability Integral transform (PIT) | 91 |
| Appendix V: list of identified districts with high-risk ratios of GC | 92 |
| Appendix VI: List of districts identified with a high risk of Men GC (22 districts) ... | 92 |
| Appendix VII: List of districts identified with a high risk of women GC (16 districts) ... | 93 |
| Appendix VIII: List of districts identified with a high risk of cardia GC (11 districts) ... | 94 |
| Appendix IX: List of districts identified with a high risk of non-cardia GC (27 districts) | 94 |
| Appendix IV: Code | 95 |

List of Tables

| | |
|---|----|
| Table 3.1: Description of all GC topographies using ICD-O3. | 12 |
| Table 4.1: Map of socioeconomic status score index across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96. | 47 |
| Table 4.2: Incidence risk ratio (IRR) and 95% credible interval for overall GC dataset using saturated Poisson regression model | 48 |
| Table 4.3: Incidence risk ratio (IRR) and 95% credible interval for men GC dataset using saturated Poisson regression model | 49 |
| Table 4.4: Incidence risk ratio (IRR) and 95% credible interval for female GC dataset using saturated Poisson regression model. | 50 |
| Table 4.5: Incidence risk ratio (IRR) and 95% credible interval for CGC dataset using saturated Poisson regression model | 50 |
| Table 4.6: Incidence risk ratio (IRR) and 95% credible interval for NCGC dataset using saturated Poisson regression model | 51 |
| Table 4.7: Incidence risk ratio (IRR) and 95% credible interval for overall, male and female GC dataset using spatial Poisson regression model | 54 |
| Table 4.8: Incidence risk ratio (IRR) and 95% credible interval for cardia and non-cardia GC dataset using spatial Poisson regression model | 54 |
| Table 4.9: Incidence risk ratio (IRR) and 95% credible interval for cardia GC dataset stratified by sex using spatial Poisson regression model. | 55 |
| Table 4.10: Incidence risk ratio (IRR) and 95% credible interval for non-cardia GC dataset stratified by sex using spatial Poisson regression model | 55 |

Table 4.11: Variance estimates, 95% credible interval, and percent of spatial variation explained for overall, male, female, cardia and non-cardia GC dataset using spatial Poisson regression 56

List of Figures

| | |
|--|----|
| Figure 3.1: Map of five Regional Health Authorities and their corresponding Regional Health Authority Districts in Manitoba | 11 |
| Figure 3.2: Schematic flowchart of data Partitioning | 20 |
| Figure 4.1: Plot of time trend of Gastric Cancer Crude rate (per 100,000 population) in Manitoba over 25 years (1992-2016) | 32 |
| Figure 4.2: Map of GC counts across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96 | 33 |
| Figure 4.3: Map of GC crude rate across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.. . . . | 34 |
| Figure 4.4: Map of age-sex adjusted incidence ratio of GC across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96. | 35 |
| Figure 4.5: Map of age-adjusted, sex-stratified incidence ratio of GC for males across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96. | 36 |
| Figure 4.6: Map of age-adjusted, sex-stratified incidence ratio of GC for females across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96 | 37 |
| Figure 4.7: Map of age-sex adjusted incidence ratio of cardia GC across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96...38 | |
| Figure 4.8: Map of age-sex adjusted incidence ratio of non-cardia GC across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96 ...39 | |

Figure 4.9: Map of median annual income across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96..... 40

Figure 4.10: Map of Indigenous People population across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96..... 42

Figure 4.11: Map of the unemployment rate across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96 43

Figure 4.12: Map of people with no post-secondary education across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96..... 44

Figure 4.13: Map of immigrants across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96..... 45

Figure 4.14: Map of socioeconomic status score index across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96..... 46

Figure 4.15: Map of unexplained variation in 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96 53

Figure 4.16: Map of standardized incidence ratio of overall GC for 96 RHADs in Manitoba and Winnipeg using spatial Poisson regression model; numbers in the map are the district number from 1 to 96 57

Figure 4.17: Map of standardized incidence ratio of male overall GC for 96 RHADs in Manitoba and Winnipeg using spatial Poisson regression model; numbers in the map are the district number from 1 to 96..... 58

Figure 4.18: Map of standardized incidence ratio of female overall GC for 96 RHADs in Manitoba and Winnipeg using spatial Poisson regression model; numbers in the map are the district number from 1 to 96 59

| | |
|---|----|
| Figure 4.19: Map of standardized incidence ratio of CGC for 96 RHADs in Manitoba and Winnipeg using spatial Poisson regression model; numbers in the map are the district number from 1 to 96 | 60 |
| Figure 4.20: Map of standardized incidence ratio of NCGC for 96 RHADs in Manitoba and Winnipeg using spatial Poisson regression model; numbers in the map are the district number from 1 to 96..... | 61 |
| Figure 4.21: Marginal temporal random effect and 95% credible interval plot for overall GC, male GC, female GC, NCGC, and CGC | 62 |
| Figure 4.22: Temporal random effect and 95% credible interval plot of overall GC for Porcupine Mountain, Pinawa, Lac Brochet, The Pas and Winnipeg Churchill districts | 64 |
| Figure 4.23: Maps of standardized incidence ratio of overall GC for 96 RHADs in Manitoba and Winnipeg for five time periods using spatio-temporal Poisson regression model; numbers in the map are the district number from 1 to 96..... | 65 |
| Figure 4.24: Maps of standardized incidence ratio of overall male GC for 96 RHADs in Manitoba and Winnipeg for five time periods using spatio-temporal Poisson regression model; numbers in the map are the district number from 1 to 96..... | 66 |
| Figure 4.25: Maps of standardized incidence ratio of overall female GC for 96 RHADs in Manitoba and Winnipeg for five time periods using spatio-temporal Poisson regression model; numbers in the map are the district number from 1 to 96 | 67 |
| Figure 4.26: Maps of standardized incidence ratio of CGC for 96 RHADs in Manitoba and Winnipeg for five time periods using spatio-temporal Poisson regression model; numbers in the map are the district number from 1 to 96..... | 68 |

Figure 4.27: Maps of standardized incidence ratio of NCGC for 96 RHADs in Manitoba and Winnipeg for five time periods using spatio-temporal Poisson regression model; numbers in the map are the district number from 1 to 96..... 69

Figure 4.28: Posterior density plot for fixed effect estimates and variance of random effect estimates 70

List of Abbreviations

ANOVA: Analysis of Variance

AR(1): Autoregressive Order one

BMI: Body Mass Index

BUGS: Bayesian Using Gibb's Sampling

CAR: Conditional Autoregressive

CCMB: Cancer Care Manitoba

CCR: Canadian Cancer Registry

CGC: Cardia Gastric Cancer

CORE: Course on Research Ethics

DIC: Deviance Information Criteria

EFA: Exploratory Factor Analysis

FA: Factor Analysis

GC: Gastric Cancer

GI: Gastrointestinal

ICD-03: International Classification of Disease for Oncology third edition

INLA: Integrated Nested Laplace Approximation

IRR: Incidence Risk Ratio

MCMC: Monte Carlo Markov Chain

NCGC: Non-Cardia Gastric Cancer

PCA: Principal Component Analysis

PCCF: Postal Code Conversion File

RDC: Research Data Center

RHA: Regional Health Authority

RHADs: Regional Health Authority Districts

RW (1): Random Walk Order one

RW (2): Random Walk Order two

SEER: Surveillance Epidemiology and End Results

SES: Socioeconomic Status

SESI: Socioeconomic status Score Index

SIR: Standardized Incidence Ratio

TCPS: Tri-Council Policy Statement

WHO: World Health Organization

WinBUGS: Window Bayesian Using Gibb's Sampling

WRHA: Winnipeg Regional Health Authority

CHAPTER 1: INTRODUCTION

1.1 Background

Cancer is one of the deadliest non-communicable diseases, formed as a result of uncontrollable cell division in the body. According to a recently published article by World Health Organization (WHO), 18.1 million new cancer cases were diagnosed in 2018, and 9.6 million deaths related to cancer were estimated worldwide (WHO, 2018). Similarly, the Canada Statistics sheet fact showed that cancer is the leading cause of death in Canada, accounting for approximately 30% of all deaths (Canadian Cancer Statistics, 2017). This finding was bolstered up by Canadian Cancer statistics, which also estimated 206,200 new cancer cases and 80,800 deaths in 2017 (Canadian Cancer Statistics, 2017). These estimates are expected to rise by 80% by the year 2030, despite the significant advancement in cancer care and treatment in Canada (Bray et al., 2018). The advances made in cancer care across Canada have contributed immensely to the country's steady decrease in both incidence and mortality rates over three decades. However, the occurrence of new cases has increased over the years, an increase that can be linked to the growing and aging population of Canada (Canada Cancer Statistics, 2018), which also varies by province.

One of the cancers that has consistently contributed to the global burden of cancer is Gastric Cancer (GC). It is the fifth most diagnosed cancer worldwide, which accounts for 6.1% of all cancer cases for both sexes globally. It stands as the fourth most diagnosed cancer in men accounting for 7.8% of whole cancer affecting men globally and the seventh most diagnosed cancer in women accounting for 4.3% of entire cancer affecting women globally (Bray et al., 2018).

The GC belongs to a family of gastrointestinal (GI) because it is one of the accessory organs of digestion. Almost half (43.5%) of all the GC cases are diagnosed at an advanced stage (stage IV), where cancer has metastasized to other parts of the body (Canada Cancer Statistics, 2018). Studies

in the United States (USA) have shown that approximately 36% of GC occurrence is diagnosed after it has spread to other parts of the body. And 33% of all GC diagnosed cases live beyond five years, which is one of the lowest cancer survival rates (American Cancer Society, 2019; SEER cancer statistics review, 2019).

In Manitoba, Cancer has been reported as the second leading cause of death as it accounts for 27.1% of all deaths in the province trailing after circulatory diseases (27.7%) (Annual Statistics, 2016). Statistics have shown that Manitoba has the second-highest age-standardized GC incidence risk for men (13.5 per 100,000) trailing after Newfoundland & Labrador (20.7 per 100,000), and the fifth-highest for women (5.6 per 100,000) with Newfoundland & Labrador retaining the first position (Canada Cancer Statistics, 2017). The occurrence of GC has been found to exhibit both geographical variation and temporal pattern over the years in some parts of the world (Mohebbi et al., 2011; Mahar et al., 2016), with little or no information about the geographical variation in Canada.

1.2 Research Aim and Objectives

The purpose of this study is to examine the effect of area-level risk factors on the incidence of GC, across the 96 regional health authority districts (RHADs) in Manitoba using twenty-five years (1992-2016) Canadian Cancer Registry (CCR) dataset. The following objectives will guide us in achieving the study aim:

1. Describe and investigate the geographical variation of GC incidence in Manitoba
2. Explore factors influencing the geographic difference of GC incidence in Manitoba
3. Investigate the geographic changes of GC incidence over time in Manitoba

1.3 Research Contribution

As at the time this project was conceived, the epidemiology of GC both in space and time in Manitoba was not well documented and has not been explicitly investigated. Hence, this study is the first of its kind to examine the geographical variation of GC in Manitoba. Map of the incidence (either standardized or crude) may not show sufficient information on its own about the latent risk factors driving the rate. However, combined with spatially varying factors could provide robust information about how the underlying factors influence the incidence in different locations. This study will, therefore, offer a comprehensive description of GC spatial epidemiology in the province, thereby serving as a means of education to the public. It will also enable public health agencies (especially those with a keen interest in cancer surveillance) such as CancerCare Manitoba (CCMB) and other public health agencies, including Manitoba Health to have a clear picture of the spatial variation of GC in Manitoba. An understanding that could serve as a tool in prioritizing intervention strategies/programs as a measure of eradicating GC.

1.4 Ethics

All required ethical conditions by the research data center (RDC), where my data was housed, were satisfied, and approval was granted. As part of the condition required in carrying out research using the RDC dataset, I have completed the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans Course on Research Ethics (TCPS 2: CORE).

Chapter 2: Literature Review

2.1 Gastric Cancer: Risk and Protective Factors

Most previous studies on the epidemiology of GC considered individuals as the unit of measurement (You et al., 2000; Go 2002; Parkin et al., 2002; Tsugane and Sasazuki 2007; Carrasco and Corvalan 2013; Davies 2013; Polinkevych and Pikas 2014; Watari et al., 2014). This consideration played a vital role in the discovery of factors that are associated with GC, which have been broadly categorized into environmental and genetic factors.

2.2 Environmental Risk Factors

Epidemiological studies of GC have suggested the existence of geographical variation in the incidence of GC across the continent. However, this geographic variation depends mainly on environmental factors (Chiu et al., 2011; Fitzmaurice et al., 2013; Liao et al., 2014), which is broadly categorized into a lifestyle (such as food) (Haenszel & Kurhara, 1960; Haenszel et al., 1972). Haenszel and Kurhara (1960), in their study of Japanese immigration from a region of high GC risk (Asia) to an area of low GC risk (United States), found that early subjection to environmental factors, such as lifestyle, play a vital role in GC incidence and mortality rate. A comparative study of three different Japanese populations, Japanese immigrants in Hawaii (USA), Sao Paulo (Brazil), and Japanese in Japan, revealed that lifestyle (especially diet) is associated with GC (Tsugane et al., 1990).

Significant documentation from various studies has shown that food is one of the major environmental factors associated with the etiology of GC globally (Kono & Hirohata 1996; Joossens et al., 1996; Shikata et al., 2006; Tsugane & Sasazuki 2007; Peleteiro et al., 2011). The influence of food on GC has been documented to exhibit a dyadic nature, causing it to either

function as a risk or protective factor. Evidence from ecological studies across the globe has identified and associated high intake of salt and salt-preserved food with increased risk of overall GC (Kneller et al., 1992; Joossens et al., 1996; Park et al., 2011; Watanabe et al., 2012; D'Elia et al., 2014; Umesawa et al., 2016). The high intake of salt in some parts of the world, especially developing countries, can be attributed to its use as a cheap means of food preservation (Kono & Hirohata 1996). This high intake has also been suggested to be responsible for the high incidence and mortality of GC in such regions (Kneller et al., 1992; Joossens et al., 1996; Tsugane & Sasazuki 2007).

Studies have also identified the consumption of vegetables, fruits and food containing high fiber content to be associated with overall GC (Jedrychowski et al., 1986; Riboli & Norat 2003; Larsson et al., 2006; Lunet 2007; Tsugane & Sasazuki 2007; Washington 2007; Liu and Russell 2008). Some of these studies showed a significant reduction in the risk of GC relative to fruit and vegetable consumption (Lunet et al., 2005; Gonzalez et al., 2006; Larsson et al., 2006; Tsugane & Sasazuki 2007). While some studies identified vegetables only as a significant protective factor against GC, some identified a non-significant association between GC and fruit intake (Larsson et al., 2006). More so, dietary fiber has been identified as a protective factor against overall GC, where studies have demonstrated that dietary fiber (e.g., cereal fiber) is significantly associated to a decrease in GC (Gonzalez et al., 2006; Ma et al., 2007; Zhang et al., 2013). An in vitro study by Moller et al. (1988) revealed the role of dietary fiber as a nitrate purifier. An attribute that offset the effect of a carcinogenic substance called N-nitroso compounds that can be obtained from the diet, tobacco smoking, and environmental sources (Moller et al., 1988).

Smoking is another environmental factor that has been associated with GC (Tredaniel et al., 1997; Smoke and Smoking, 2004; Ladeiras-Lopes, 2008). It has been shown to possess over 5000

chemical substances most of which are harmful to the body, and responsible for development of diverse kinds of cancer (Gonzalez et al., 2003; Munafò et al., 2012; Ferreccio et al., 2013; Aggarwal et al., 2014; Tang et al., 2014). Studies have associated smoking with an increase in the risk of GC, which was more prominent among men compared to women (Tredaniel et al., 1997; Nishino et al., 2006).

Studies have also shown that alcohol is directly associated with the overall GC (Barstad et al., 2005; Tramacere et al., 2012; Wang et al., 2018). With variability in individual alcohol consumption level and type of alcohol consumed, some researchers investigated the association between alcohol consumption level, type of liquor, and GC. The results of which show no association between total alcohol consumption and GC, a significant association between alcohol type and overall GC, and an association between daily use of wine and overall GC (Barstad et al., 2005). Another showed no association between moderate alcohol drinking and overall GC, while heavy alcohol drinkers are at higher risk (Tramacere et al., 2012; Wang et al., 2018). These studies, however, showed an inconsistent and unclear result about the association between alcohol and overall GC.

Obesity has also been documented to be associated with overall GC (Yang et al., 2009; Olsen et al., 2011; Turati et al., 2013; Lauby-Secretan, 2016) where the definition of obesity varies from a minimum body mass index of 25 kg/m^2 ($\text{BMI} \geq 25 \text{ kg/m}^2$) to a minimum body mass index of 30 kg/m^2 ($\text{BMI} \geq 30 \text{ kg/m}^2$), with equal likelihood of risk among both men and women. An increase in the strength of association between GC and obesity was also found to be related to increased BMI. However, there is insufficient information on the etiology of noncardia gastric cancer (NCGC), as reported by a group of IARC (Lauby-Secretan 2016). Socioeconomic status (SES) has also been linked with the etiology of overall GC, with an inconsistency in results

reported. Some of the studies identified a relationship between SES and GC (Nagel et al., 2007; Wu et al., 2014; Dong et al., 2017), while some did not identify any association (Loon et al., 1998). In a bid to further understand the epidemiology of GC and SES, researchers also partitioned GC into cardia and non-cardia GC. A typical study in this area was carried out in the United Kingdom (UK), which revealed an increasing incidence of Cardia GC among the professional class SES (Powell & McConkey 1990). A similar study in the USA partitioned their study population into SES group using two different criteria: average family income or housing rental and occupation. Their study recorded a higher incidence of GC among the low SES population for both sexes of the white race (Haenszel 1958).

One of the widely investigated and consistent factors associated with mostly overall GC is *Helicobacter Pylori* (*H. Pylori*) infection (Parkin et al., 2002; Carrasco and Corvalan 2013; Watari et al., 2014). While studies have shown that *H. Pylori* is associated with GC, it was also revealed that the impact of *H. Pylori* on the overall GC may be confounded by other factors such as salt/salty food consumption (Tsugane et al., 1994; Fox et al., 1999; Lee et al., 2003; Machida-Montani et al., 2004; Shikata et al., 2006; Polinkevych and Pikas 2014) and SES (Klein et al., 1991; Go 2002). This is also supported by statistics that show that 2% of the world's 50% population infected with *H. Pylori* end up having GC (Maev et al., 2014). The infection rate of *H. Pylori* has been documented to be more pronounced in developing countries while in most developed countries, the rate of *H. Pylori* is high among the Indigenous communities and immigrants from countries with high incidence rates such as Japan, Korea, and China (You et al., 2000; Go 2002).

The continuous enrichment of the Manitoba population with immigrants from countries with a high rate of *H. Pylori* could be partially responsible for the occurrence of GC in the regions where they are more dominant. Studies have investigated the existence of a relationship between immigration status and GC occurrence in different countries (Tsugane et al.,1990; Kamineni et al., 1999; Lee et al., 2007; Mousavi et al.,2011). The results of which are inconsistent, as some of the studies identified immigration status as a risk factor for GC, others showed no significant association. This may be due to other factors such as lifestyle and diet confounding the effect of immigration, as studies have shown that lifestyle and food are also factors associated with GC.

Another essential factor of interest that has not been extensively researched in the epidemiological study of GC in Canada is the Indigenous population (Roder & Currow 2009; Arnold et al., 2014; Moore et al., 2015; Coloquhoun et al., 2019). Most studies about the incidence and risk of overall GC among the Indigenous people across different countries suggest that Indigenous people are at more risk compared to their non-Indigenous counterparts (Roder & Currow 2009; Arnold et al., 2014). The term Indigenous refers to a group of people who are the original occupants of a particular area, region, or country. Most Indigenous populations across the globe share similar characteristics features, such as poor living lifestyles due to marginalization. A study combining a group of Indigenous people from Australia, New Zealand, USA, Chile, circumpolar region carried out by Melina et al., showed a higher increase of risk in GC among the Indigenous people compared to their non-Indigenous counterparts (Arnold et al., 2014). A similar study considering the Indigenous people in the Amazon of Ecuador revealed that the risk of GC among the Indigenous men and women is significantly lower than the non-Indigenous people (Sebastian et al., 2004).

Colquhoun et al. (2019), in their bid to bridge the paucity of information on overall GC in northern Canada, showed that the burden of overall GC in the Indigenous subpopulation is higher compared to non-Indigenous. Similarly, a study by Arnold M et al. (2013) revealed the presence of significance higher overall GC incidence in the Indigenous population relative to non-Indigenous. However, the classification of the overall GC by topographical subsite may provide more useful information regarding the influence of the Indigenous people on the risk of partitioned topographical subgroups.

2.3 Genetic Risk Factors

Similar to environmental factor, researchers have also shown that genetical factors such as blood group, familial predisposition and hereditary play essential roles in the etiology of overall GC (Aird et al., 1953; Hoskins et al., 1965; Langman 1998; Edgren et al., 2010; Oliveira et al., 2015; Van der Post et al., 2015). A population-based study carried out in Sweden using the Scandinavian donation and transfusion database identified a higher risk of overall GC among individuals with blood group A compared to other blood groups (Edgren et al., 2010). Similarly, studies have also shown that some types of cancer such as gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS), and familial intestinal gastric cancer are hereditary (Oliveira et al., 2015).

This study focuses more on environmental factors, though drawing a clear distinction between environmental and genetic factors may be almost impossible as they are interwoven.

Chapter 3: METHODS

3.1 Research Design

We adopt an ecological research design method, which enables us to quantify our measures (dependent and independent variables) of interest. For this study, we extracted all GC patients in Manitoba, diagnosed between 1992 - 2016 from the National Cancer Registry database.

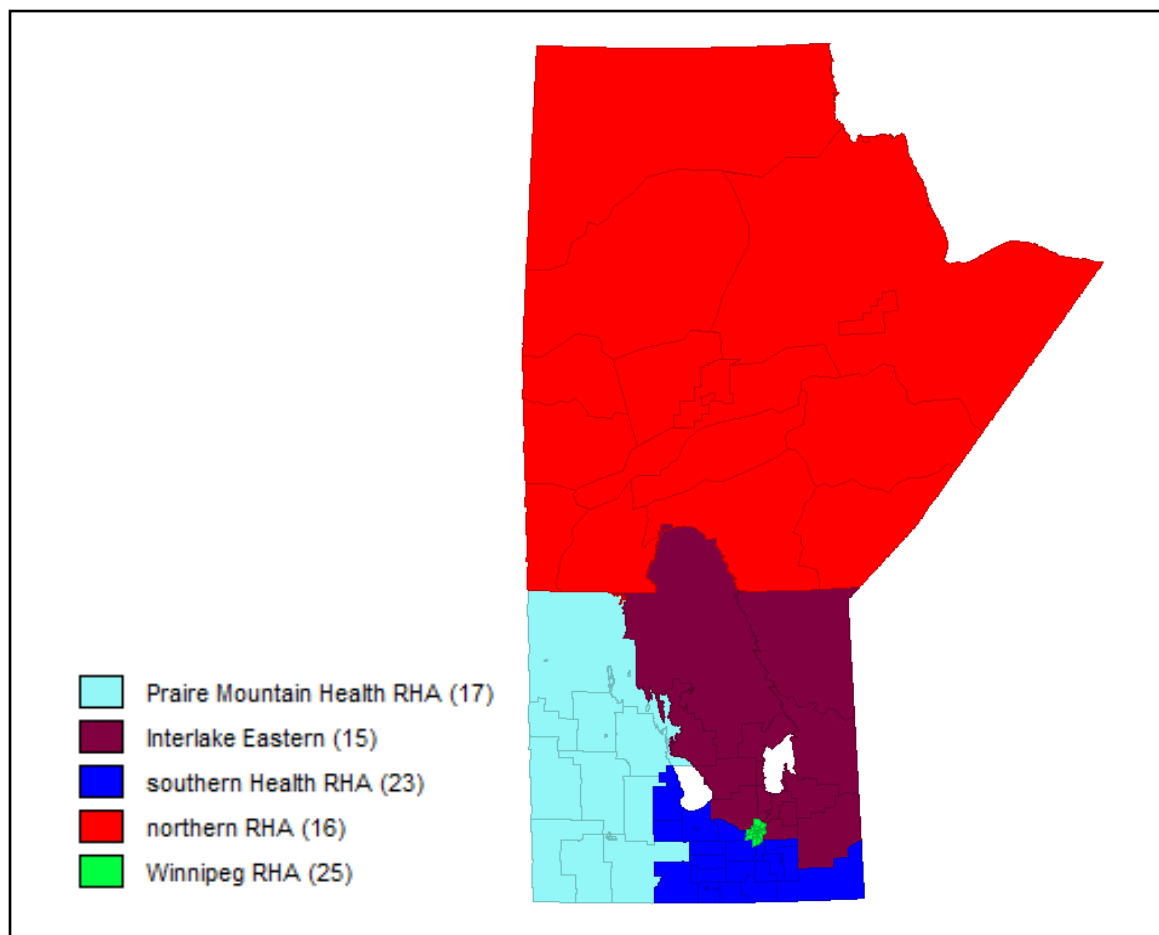
3.2 Study Area

The province of Manitoba is the study region in this research, and based on the 2016 population census, approximately 1,278,365 people are living in Manitoba, with 631,400 (49.39%) being male, and 646,965 (50.61%) being female. Winnipeg is the capital of the province which houses more than half (55.96%) of the entire Manitoba population. The province of Manitoba is bordered on the east by Ontario, on the west by Saskatchewan, on the north by Nunavut territory, northeast by Hudson Bay and south by the United States' Minnesota and North Dakota. It is seated on a 647,797 square km land with an approximate 107,966 square km (one-sixth) of inland water. The Province is enriched with immigrants from different countries, making it a diversified multi-cultural ethnic province and home to Indigenous peoples who have occupied the land for over a thousand years (Bumsted et al.,2019). The Province has five regional health authorities (RHAs) created from the initial eleven RHAs in 2012.

These RHAs across the Province are tasked with overseeing health services, both acute and community-based care. As a result of the requirement for improving the health service quality of individuals in each RHA, the RHAs were further sub-divided into districts, the Regional Health Authority Districts (RHADs), with the primary objective of fostering a heterogenous population via partitioning into homogenous populations (Fransoo et al., 2005). For more information

regarding the creation of the RHADs, readers can visit the Manitoba center for health policy (MCHP) website. The RHADs, as described in Figure 3.1 below, was adopted in this study as the area-level aggregate point. This is because it divides the provinces into smaller homogenous populations, and also oversees the health activities of communities under its jurisdiction. As at the time of this study, the province is divided into 96 RHADs, forming the unit analysis in this study.

Figure 3.1: Map of five Regional Health Authorities and their corresponding Regional Health Authority Districts in Manitoba



3.3 Data

All Manitoba population was considered in this study, while the cases of GC used in this study was from 1992 -2016 as there was no proper documentation of cancer diseases before 1992 (Statistics Canada, 2019). The universal disease classification code, as defined by the international classification of disease for oncology third edition (ICD-O3), is a hospital abstract data system for coding diseases and procedures (Fritz et al., 2000). The ICD-O3 assigns codes C160 – C169 to all topographical subsites of GC, which was used to extract all GC cases, as shown in Table 3.1 below.

Table 3.1: Description of all GC topographies using ICD-O3

| ICDO-O3 code | Topography subsite |
|--------------|-----------------------------------|
| C160 | Cardia |
| C161 | Fundus |
| C162 | Body |
| C163 | Antrum |
| C164 | Pyloric |
| C165 | Lesser curvature of the stomach |
| C166 | Greater Curvature of stomach |
| C168 | Overlapping lesion of the stomach |
| C169 | Stomach unspecified |

This information combined with three data files, Canadian Cancer Registry (CCR), Canada Census, and postal code conversion file (PCCF), was used to create an area-level dataset, which is the required first step in this study.

3.3.1 Canadian Cancer Registry (CCR)

The CCR consists of comprehensive information on people diagnosed with any form of cancer in the 13 provinces/territories in Canada. We accessed the file from the Research Data Centre (RDC) repository located at the University of Manitoba campuses (Bannatyne and Fort Gary). The CCR

is a census of respondents residing in any of the thirteen provinces/territories in Canada, either as a temporary or permanent resident who have been diagnosed with any form of cancer. The CCR dataset is a national collection of cancer patient records from multi-channel sources such as clinic files, radiotherapy, and hematology reports, documents from in-patient hospital stays, out-patient clinics, private hospitals, pathology, and other laboratory/autopsy reports, radiology and screening program reports, reports from physicians in private practice, medical billing and hospital discharge administrative databases, and reports on cancer deaths from Vital Statistics registrars as stated in the Canadian Cancer registry codebook. The ICD-O3, as described in Table 3.1 combined with the provincial code (46), and date of diagnosis was used to extract all GC patients who resided in Manitoba between 1992 to 2016 from the CCR. This extract represents our outcome (dependent variable) of interest in this study. Alongside the GC case extraction, we also extracted other variables of interest, e.g., demographic variables (the age when diagnosed with GC and sex). These demographic variables enable us to adjust for the influence of sex and age disparities across the RHADs to avoid spurious estimates.

3.3.2 Census Data

The Census data is the compiled census information of all people living in Canada in a particular Census year. The 2016 Census data was used in this study because it was the closest census to the upper limit time frame for the duration considered (1992 – 2016); we also assumed that the population of Manitoba was steady with no significant change (1.1 million – 1.2 million) within this period. This assumption also extends to the fixed covariates extracted from the Census data and the population size across the RHADs. The variables of interest relevant to this study obtained

from the Census data were household income, Indigenous status, immigrant status, post-secondary education, and employment status. The covariates were aggregated and geocoded into 96 RHADs.

We extracted a total of six covariates from the Census data described below:

Immigration status

The immigrant variable as used in this study is the population of individuals who identified themselves as an immigrant within the study period. The proportion of the people belonging to this group was used as a surrogate to represent area-level immigrants.

Indigenous status

The Indigenous variable represents the sub-population of Manitoba who identifies as First Nation, Metis, and Inuit in Manitoba

- *First Nation (North American Indian)*: this includes registered, status/non-status, treaty/non-treaty Indians, no Inuit. Presently, Manitoba has five First nation groups – Cree, Ojibway, Dakota, dene, and Oji-Cree.
- *Metis*: this includes Manitobans from a mixed-race, which are partly First nation and European or Canadian.
- *Inuit*: this consists of people from the arctic region in Manitoba. A typical town in Manitoba is Churchill, as it is located along the Hudson Bay at the peak of the Churchill River along the coast of the arctic ocean.

Socioeconomic status (SES)

As a result of collinearity between unemployment, income, and post-secondary school education, exploratory factor analysis (EFA) was carried out to create a linear combination of these three

variables using their shared variance to create a factor that was labeled socioeconomic score index (SESI). Two widely used methods in the creation of SES indices are factor analysis (FA) and Principal component analysis (PCA) (Pampalon and Raymond 2000; Salmond and Crampton 2002; Messer et al.,2006; Holt and Lo, 2008; Krishnan, 2010; Vincent and Jason 2013). The FA method as used by Abdul and Abd (Abdul and Abd, 2018) was adopted in creating our SESI; the procedure is described in the three steps below:

Step 1: Variable standardization (z_{ip})

Three variables that have been consistently used in the creation of SES indices creation are income, education, and unemployment (Salmond, Crampton and Sutton, 1998; Abdul and Abd, 2018). These three variables are also used in our creation of the SESI. The variables are standardized to have data ranged between 0 and 1. The standardization is as follows

$$z_{ip} = \frac{x_{ip} - \min(x_{ip})}{\max(x_{ip}) - \min(x_{ip})}$$

where

x_{ip} : actual covariate p at region i

z_{ip} : standardized covariate p at region i

$\min(x_{ip})$: the minimum value of actual covariate p at region i

$\max(x_{ip})$: the maximum value of actual covariate p at region i

Step 2: Weighted correlation matrix (w_{pf})

We obtained the weighted matrix by multiplying the correlation of the actual (unstandardized) variables with the loading of the variables on the factor extracted from the factor analysis. We computed the weighted correlation matrix as follows:

$$w_{pf} = R_{pp}^{-1} \times S_{pf}$$

where

R_{pp}^{-1} : inverse correlation matrix of the actual covariates

S_{pf} : loadings of the covariate on SESI factor extracted

Step 3: Socioeconomic score index (SESI)

We computed the SESI score for each RHAD by weighing the standardized covariates by the weighted matrix, i.e.

$$SESI_i = z_{ip} \times w_{pf}$$

The choice of the variable used is mostly dependent on the available information at the researchers' disposal and the need for creation. Hence there is no rule of thumb as to the number of variables required to create an SES score index. Researchers have measured SES using single variable such as education level (Pukala and Teppo 1986; Ferraroni et al., 1989; Van Loon et al., 1998), occupation (Palli et al., 1992; Faggiano et al., 1994; Van Loon et al., 1998) and income (Pukala and Teppo 1986; Brown et al., 1994) with a cautious interpretation during inferences.

3.3.3 Postal Code Conversion File (PCCF)

PCCF is the last piece of the data file that is linked with CCR and Census data to create an area-level dataset that we used in this study. The PCCF provides a correspondence between the Canadian six-alphanumeric postal code and Statistics Canada's standard geographic areas. The variables of interest in the PCCF consist of postal code, disseminated area code, RHADs name, district ID, and district code.

The area-level dataset (geocoding) which formed the basis of our study, is developed in three stages:

1. Creation of regional GC dataset from CCR
2. Creation of area-level covariates and population from Canadian Census data
3. Merging of area-level GC and covariate dataset

3.3.4 Creation of Regional GC Counts from the Canada Cancer Registry

The variables of interest at this stage are GC count, geographical variable (postal code), age of the patient, and year of diagnosis. These variables we extracted from CCR conditioned on correspondents living in Manitoba (using the provincial code 46). The correspondents were grouped into age categories using a class interval of four. We then linked the dataset to the PCCF which is also linked to RHADs and the disseminated area code, hence enabling us to connect it to the covariates. This linkage allows for us to geocode the GC cases and aggregate them into the 96 RHADs.

3.3.5 Creation of Regional Covariates from Census Data

The covariates that we extracted from the Census data are linked to the PCCF. This linkage allows us to aggregate over the 96 RHADs and relates it to the GC cases. We used the area-level population from the Census data to estimate the rate in each RHAD, and adjust for the disparities in age and sex distributions across the RHADs. The whole essence of this is to be able to estimate the expected count of GC assuming the age and sex distributions are uniform across the RHADs.

3.3.6 Merging of Regional GC and Regional Covariate Dataset

The merging is the stage where the regional GC dataset created from the CCR and area-level covariates created from the Census data are combined to form a single area-level dataset. The merged dataset is an overall summary dataset on which all other estimates and adjustments are carried out before modeling. The modeling aspect of the research is done using spatial and spatio-temporal models that are widely used to model areal-level data (Torabi and Rosychuk, 2011; Torabi, 2014; Torabi et al., 2014).

3.4 Statistical Analysis

3.4.1 Poisson Regression Model

As part of the preliminary exploratory analysis, we fitted a log-linear model to establish a relationship between the covariates and GC count across the 96 RHADs. This procedure is required to achieve a parsimonious model and also avoid type I or type II statistical error. We considered only the independent variables that were found to be significantly associated with overall GC (at the significance level of 0.05) and kept the variables for further analysis. This step enabled us to examine how each covariate influences the GC independently. We then proceed to fit a multivariable Poisson regression model to assess how the multiple variables affect GC and also

determine the model fit for the introduction of a spatial component. We evaluated the model goodness-of-fit via deviance information criteria (DIC).

Let $y_{ijk}; i = 1, \dots, 96; j = 1, 2; k = 1, \dots, 12$ represents the GC count in RHAD i , sex group j and age group $k = less\ than\ 35, 35 - 39, 40 - 44, \dots, 85 +$; y_{jk} is the GC count in sex j and age group k ; $y_i; I = 1, \dots, 96$ be the RHAD GC count; n_{ijk} be the population of people in RHAD i belonging to sex group j and age group k ; n_{jk} is the population of in sex j and age group k ; x_i be the RHAD covariates (SESI, the proportion of Indigenous, the proportion of immigrants), and assuming the total number of cases of GC in each RHADs follows a Poisson distribution, i.e. $Y_i \sim Poisson(\lambda_i E_i)$, where λ_i is risk and E_i is the expected number of GC count in RHAD i . The log-linear model is defined as

$$\log(\lambda_i) = \alpha + \beta x_i \quad (3.1)$$

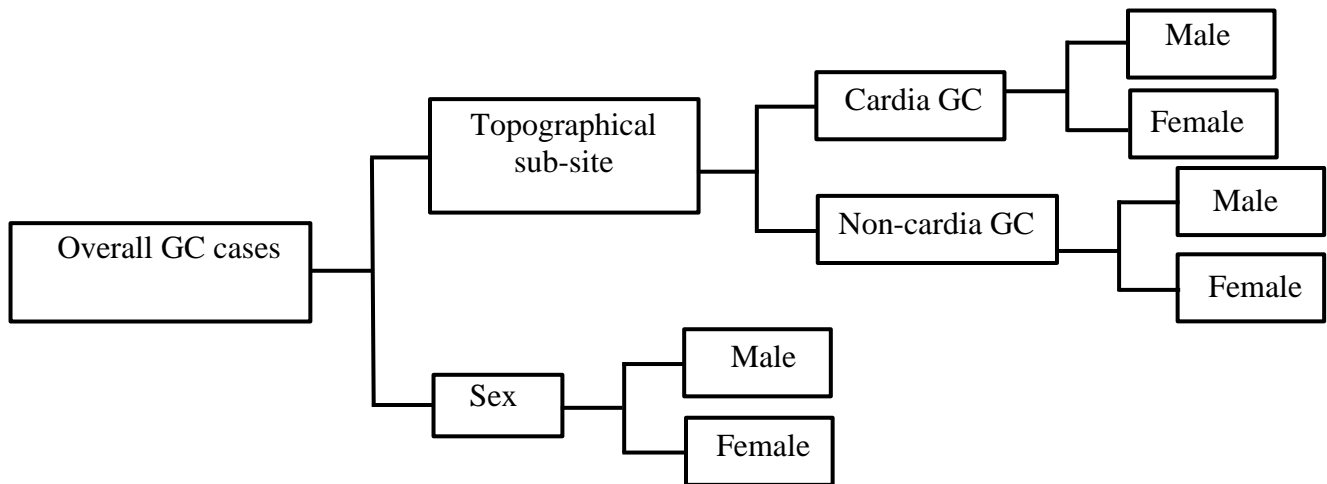
$$Y_i = \sum_{j=1}^2 \sum_{k=1}^{12} y_{ijk} \quad (3.2)$$

$$E_i = \sum_{j=1}^2 \sum_{k=1}^{12} \frac{y_{jk}}{n_{jk}} \times n_{ijk} \quad (3.3)$$

where α and β are mean ratio (intercept) and regression coefficients, respectively. The age and sex adjustment were essential as they enable us to eliminate the potential error of wrongly ascribing high/low risk to RHADs based on age/sex distribution, thereby concealing the real RHADs risk.

We partitioned the overall dataset by both sex and Topographical sub-sites of GC, as shown in Figure 3.1 below. To each of these sub-group datasets, we fitted both spatial and spatio-temporal models discussed in subsequent sections.

Figure 3.2: Schematic flowchart of data partitioning



3.4.2 Spatial Poisson Regression Model

The use of spatial regression model in epidemiological studies is to examine how a set of covariates affects the incidence of disease, while also adjusting the model to account for regional dependence in the dataset (Banerjee et al., 2014). This modeling approach is robust because most conventional regression models assume that observations are independent. This assumption is not often valid in spatially collected data. As a result of this assumption, spatial models draw their strength from relying on neighboring regions for a more reliable estimate. According to Tobler's first law of geography which states that *"Everything is related to everything else, but near things are more related than distant things"* (Tobler 1970; Sui 2004); regions that are close together get a higher weight compared to areas that are far apart. This ideology is often incorporated into the spatial regression model via the autoregressive model mechanism (Lawson 2013; Banerjee et al., 2014). Before using a spatial model in real-life applications, its suitability is often assessed and justified via both exploratory and confirmatory analysis.

The most applied confirmatory test used to investigate and ascertain the spatial dependency of an event in spatial studies is the Moran I statistics. The method determines spatial dependence by measuring the spatial auto-correlation of events and testing the statistical significance of the statistic (autocorrelation) estimated. The Moran's I statistic is given by

$$MI = \frac{I}{S_0} \left(\frac{\sum_{i=1}^I \sum_{j=1}^I w_{ij} (y_i - \bar{y}) (y_j - \bar{y})}{\sum_{i=1}^I (y_i - \bar{y})^2} \right)$$

where y_i as used in this study represents the region i count of GC cases; \bar{y} represents the average count of GC cases; w_{ij} represents the weight matrix obtained assuming a neighborhood structure where regions share a direct boundary; I is the number of RHADs ($I = 96$); $S_0 = \sum_{i=1}^I \sum_{j=1}^I w_{ij}$ is the aggregate weight. The range of MI is between -1 and 1, where -1 denotes a perfect clustering of different values (perfect dispersion), 0 denotes perfect randomness (no correlation), and 1 indicates an ideal clustering of similar values.

Using a similar notation described in sub-section 3.4.1 above, let λ_i be the relative GC risk observed for each i , assuming $Y_i \sim \text{Poisson}(\lambda_i E_i)$, then we define the modified Poisson regression model as:

$$\log(\lambda_i) = \alpha + \beta x_i + \phi_i \quad (3.4)$$

where $\lambda_i = E[Y_i | \beta_i, \theta_i, \phi_i] / E_i$ is the (conditional) risk, and ϕ_i is the heterogeneity in the dataset. In this study, the ϕ_i was partitioned into two random effects η_i and u_i , accounting for spatially correlated and uncorrelated heterogeneity, respectively (Besag et al., 1991; Knorr-Held & Richardson, 2003; Torabi, 2014; Torabi et al., 2014; Sharafi et al., 2018; Torabi et al., 2019). The

$u_i, u_i \sim N(0, \sigma_u^2)$, is the unstructured heterogeneity, “noise,” which materializes from our dataset that we cannot account for, while η_i is the structured heterogeneity described by a Gaussian conditional autoregressive (CAR) distribution. Given $\mathcal{N}(i), i = 1, \dots, 96$, to be the set of neighbors(s) to a specific RHAD which can be defined in different ways such as proximity or boundary sharing between regions (Waller and Gotway, 2004). The neighborhood as used in this study is defined as two regions sharing boundaries, and w_{ij} represents the relationship between any two regions (i, j) , which is defined as

$$w_{ij} = \begin{cases} 1, & i \text{ and } j \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases}$$

then, the conditional distribution of spatial random distribution $\eta_i | \eta_j, i \neq j$ is given as

$$\eta_i | \eta_j, i \neq j, \tau_\eta \sim N\left(\frac{1}{\mathcal{N}(i)} \sum_{j \sim i} w_{ij} \eta_j, \frac{\sigma_\eta^2}{\mathcal{N}(i)}\right), \quad (3.5)$$

where $\mathcal{N}(i)$ is the number of neighbors to region i and σ_η^2 is the spatial dispersion of region i . Considering the above modification to the conventional Poisson regression model, the spatial model considered in this study is defined as

$$\log(\lambda_i) = \alpha + \beta x_i + u_i + \eta_i \quad (3.6)$$

where α is the common intercept (mean ratio) for the entire region, and β is the effect of covariate x in region i . The use of CAR described in equation (3.5) provides the means of weighing the spatial random effect locally considering the neighbors to the specific region of interest. This mechanism allows the district to borrow strength locally (Besag et al., 1991), and the number of parameters and hyperparameters to be estimated are $\theta = (\alpha, \beta)$ and $\psi = (\tau_u, \tau_\eta)$ respectively, with $\tau = \frac{1}{\sigma^2}$ representing the precision of the estimated parameter.

3.4.3 Spatio-temporal Poisson Regression Model

In most epidemiological studies, a spatially-attributed disease is simultaneously attributed to a time point, which calls for the modification of the spatial model described in equation (3.6) to account for the time effect explicitly.

Let y_{it} be the count of GC cases aggregated over space and time which is assumed to follow a Poisson distribution $Y_i \sim \text{Poisson}(\lambda_{it}E_{it})$, then equation (3.6) can be extended to accommodate the temporal effect which is given as

$$\log(\lambda_{it}) = \alpha + \beta x_i + u_i + \eta_i + \gamma_t + \varphi_t \quad (3.7)$$

where the additional two components (γ_t, φ_t) represent the unstructured and structured temporal effects. The structured temporal effect, φ_t , is modeled by imposing a random walk of order one, RW (1), a distribution which is defined as a step function given as

$$\varphi_t | \varphi_{-t} \sim \begin{cases} \text{Normal}(\varphi_{t+1}, \tau_\varphi), & \text{for } t = 1 \\ \text{Normal}\left(\frac{\varphi_{t-1} + \varphi_{t+1}}{2}, \frac{\tau_\varphi}{2}\right), & \text{for } t = 2, \dots, T-1 \\ \text{Normal}(\varphi_{t-1}, \tau_\varphi), & \text{for } t = T \end{cases}$$

and an exchangeable prior $\gamma_t \sim (0, \tau_\gamma)$ is imposed on the unstructured temporal effect. The number of parameters and hyperparameters to be estimated are $\theta = (\alpha, \beta)$ and $\psi = (\tau_u, \tau_\eta, \tau_\gamma, \tau_\varphi)$ respectively. Extending equation (3.7) to allow accounting for spatio-temporal interaction effect leads to

$$\log(\lambda_{it}) = \alpha + \beta x_i + u_i + \eta_i + \gamma_t + \varphi_t + \delta_{it} \quad (3.8)$$

where $\delta_{it}, i = 1, \dots, I; t = 1, \dots, T$, is the spatio-temporal interaction which can be modeled in four different ways depending on the assumption imposed on the form of spatial and temporal component (structured or unstructured) interaction (Knorr-Held, 2000). The interaction δ_{it} is often defined by the precision matrix given as $\sigma_\delta^2 \mathbb{R}_\delta$ which under the assumption of separability can be separated into a product that is exclusively spatial and temporal components (i.e., $Cov(y_{ii'}, y_{tt'}) = Cov(y_{ii'})Cov(y_{tt'})$), and \mathbb{R}_δ is the structure matrix, representing the kind of time or space dependency between the elements of δ . The matrix of \mathbb{R}_δ is obtained as the Kronecker product between the space structure matrix and time structure matrix (Clayton, 1996; Knorr-Held, 2000). All possible forms of interaction described in the following section are considered in this study, where the one that best fits the model is judged using the DIC.

The interaction I: unstructured spatial and temporal interaction

This interaction assumption partitions the structure matrix \mathbb{R}_δ into unstructured spatial component \mathbb{R}_u and temporal component \mathbb{R}_γ with an exchangeable prior which when combined via Kronecker product yields an interaction effect with an identity structure matrix i.e.

$$\mathbb{R}_\delta = \mathbb{R}_u \otimes \mathbb{R}_\gamma = I$$

The main idea behind this interaction model is the assumption of no spatial or temporal structure which translates into a normal distribution i.e. $\delta_{it} \sim N(0, \sigma_\delta^2)$.

Interaction II: unstructured spatial and structured temporal interaction

Under this interaction assumption, a structured temporal effect \mathbb{R}_φ assuming a first-order, second-order random walk or autoregressive order one distribution is combined with an unstructured spatial effect \mathbb{R}_u with an exchangeable distribution, i.e.

$$\mathbb{R}_\delta = \mathbb{R}_u \otimes \mathbb{R}_\varphi$$

The resulting structure matrix \mathbb{R}_δ from the above, Kronecker product follows the assumed distribution choice (RW (1), RW (2), or AR (1)).

Interaction III: structured spatial and unstructured temporal interaction

This interaction assumption combines a structured spatial effect \mathbb{R}_s assuming a CAR distribution with an unstructured temporal effect \mathbb{R}_γ with an exchangeable prior, i.e.

$$\mathbb{R}_\delta = \mathbb{R}_s \otimes \mathbb{R}_\gamma$$

The resulting structure matrix \mathbb{R}_δ from the above, Kronecker product assumes a CAR distribution.

Interaction IV: structured spatial and temporal interaction

This interaction assumption combines a structured spatial effect \mathbb{R}_s assuming a CAR distribution with a structured temporal effect \mathbb{R}_φ assuming any choice of RW (1), RW (2) or AR (1) distribution, i.e.

$$\mathbb{R}_\delta = \mathbb{R}_s \otimes \mathbb{R}_\varphi$$

The resulting structure matrix \mathbb{R}_δ from the above, Kronecker product follows a complex distribution.

3.4.4 Bayesian Inference Framework

A Bayesian hierarchical model algorithm was used in this study. This approach was used because of its ability to handle a complex model like a spatial and spatio-temporal model with ease of computation and implementation via available statistical software like Bayesian Using Gibb's Sampling (BUGS) and or the window version WinBUGS. The Bayesian approach is a well-established method that has been increasingly used in spatial epidemiology over the last decade (Torabi et al., 2014; Torabi, 2014). The Bayesian model treats the parameter vector to be estimated as a random variable rather than a fixed variable and allows for its randomness via the assignment of a prior distribution (Carlin and Xia, 1999a; Lawson 2013).

Let \mathbf{y} represents the dataset which consists of y_i *the* response variable, which in our case is the spatial GC counts for 96 RHADs and \mathbf{X} is $n \times p$ matrix of covariates. Let $\boldsymbol{\theta}$ be a $p \times 1$ vector of the parameter to be estimated and $P(\boldsymbol{\theta})$ be the prior distribution of the parameter vector $\boldsymbol{\theta}$, and let

$P(\mathbf{y}|\boldsymbol{\theta})$ be the likelihood of obtaining data \mathbf{y} given the unknown parameters $\boldsymbol{\theta}$, then the joint distribution of the data \mathbf{y} and parameter $\boldsymbol{\theta}$ can be expressed as

$$P(\boldsymbol{\theta}, \mathbf{y}) = P(\mathbf{y}|\boldsymbol{\theta}) \times P(\boldsymbol{\theta}) \quad (3.9)$$

using Bayes' theorem, the posterior distribution can be expressed as

$$P(\boldsymbol{\theta}|\mathbf{y}) = \frac{P(\boldsymbol{\theta}, \mathbf{y})}{P(\mathbf{y})} = \frac{P(\mathbf{y}|\boldsymbol{\theta}) \times P(\boldsymbol{\theta})}{P(\mathbf{y})} \quad (3.10)$$

where

$$P(\mathbf{y}) = \int P(\mathbf{y}|\boldsymbol{\theta}) \times P(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (3.11)$$

Equation (3.11) is often considered as a constant since it is independent of the parameter of interest; hence the posterior is directly proportional to the product of the likelihood and the prior distribution, which is given as

$$P(\boldsymbol{\theta}|\mathbf{y}) \propto P(\mathbf{y}|\boldsymbol{\theta}) \times P(\boldsymbol{\theta}) \quad (3.12)$$

As a result of the complexity of equation (3.10) due to equation (3.11), which in our case requires integration over a parameter vector $\boldsymbol{\theta}$, for 96 RHADs, a closed solution may not be feasible. Hence a numerical integration approach is used to obtain the needed parameter estimates via Integrated Nested Laplace Approximation (INLA) using Monte Carlo Markov Chain (MCMC) numerical algorithm.

The MCMC is a numerical method used for calculating numerical approximations of multi-dimensional integrals. The main objective of using an MCMC is to draw samples from the probability distribution randomly. This method is integrated into Bayesian inference via the use of Gibbs sampling or Metropolis-Hasting scheme in WinBUGS software; some readily available packages created in R programming language was used during the model fitting and exploratory analysis.

3.4.4.1 MCMC Convergence Diagnostics

As a result of the mechanism involved in the parameter estimation, which is done via an MCMC sampling, we require a stationary point where further iteration via Gibbs sampling will not improve the parameter estimates. It is at this optimum point that we can draw reliable estimates that can be used for the inferential purpose about the influence of the predictors on the risk of GC incidences. Four methods – time series plot, Gelman Rubin statistic, auto-correlation plot, and the comparison between the process error and adjusted standard deviation (5% of the parameter estimates' standard deviation) – are used to assess the convergence of parameter estimates.

Time series plot

This exploratory analysis shows how the parameter estimates behave over time, and to determine if there is a wide inconsistent dispersion in the estimation procedure. A good practice is to use more than one chain and then look for disparity or variation in the chain pattern; a distinct contrast stipulates non-convergence of the Gibbs sampler, which leads to an unreliable and misleading inference.

Gelman Rubin statistic

This method was used to examine the convergence of multiple sequences simultaneously by comparing the within-sequence variance and between-sequence variance ratio (\hat{R}_c) (Gelman and Rubin 1992). The (\hat{R}_c) (potential scale reduction factor) is computed as follows

$$\hat{R}_c = \frac{d + 3 \hat{\sigma}_p^2}{d + 1 \hat{\sigma}_w^2}$$

$$\hat{\sigma}_p^2 = \frac{L - 1}{L} \hat{\sigma}_w^2 + \frac{m + 1}{ml} \hat{\sigma}_b^2$$

$$\hat{\sigma}_w^2 = \frac{1}{M} \sum_{m=1}^M \hat{\sigma}_m^2$$

$$\hat{\sigma}_b^2 = \frac{1}{M - 1} \sum_{m=1}^M (\hat{\theta}_m - \hat{\theta})^2$$

$$\hat{\theta} = \frac{1}{M} \sum_{m=1}^M \hat{\theta}_m$$

where d is the degree of freedom; $\hat{\sigma}_p^2$ is the pooled variance under the analysis of variance (ANOVA) assumption of the equality of variance; $\hat{\sigma}_w^2$ is the within-chain variance; $\hat{\sigma}_b^2$ is the between-chain variance; $\hat{\theta}$ is the overall sample posterior mean; $\hat{\sigma}_m^2$ is the sample variance of the m th chain; $\hat{\theta}_m$ is the sample mean for the m th chain; M and L are the numbers of chains and length of the chain respectively.

The rule of thumb according to Gelman and Brooks 1997 is that an (\hat{R}_c) < 1.2 is an indication of convergence that requires no further iteration.

Autocorrelation

The autocorrelation was used to show that the estimates obtained from a sampling process are independent. Since the process of getting the posterior distribution depends on iterative sampling via Gibbs sampler, the dependency of subsequent sampling estimate on the previous sampling estimate implies that the current parameter estimate is not different from the past. The recurring evaluation could be seen as the process been stocked in a repetitive loop without a new informative estimate. The general idea is to observe that as the process lag increases, the autocorrelation decreases.

Markov chain error

The idea here is to compare the Markov chain error (MC Error) with 5% of the standard deviation and ensure that the MC error is less than 5% standard error to support convergence.

3.4.4.2 Prior Sensitivity Analysis

Sensitivity analysis is often carried out in a Bayesian modeling approach due to the mechanism of the model construct. One primary concern in the Bayesian modeling approach is the influence of prior on the posterior estimates, which could lead to a spurious inference. In order to ascertain that the choice of our priors does not influence the parameter estimates, four variants of minimally informative priors were imposed on all models. The marginal plots of the distributions of the fixed effect parameter estimates and variance of each of the models under these four different prior assumptions are examined for any significant difference.

3.5 Research Scope

Before addressing the first research objective, an analysis was carried out to examine the existence of spatial dependency in GC risk ratios across the 96 RHADs. The assessment of spatial dependence was done by displaying the residual GC risk ratios graphically to find any systematic pattern or clusters, which was followed by a confirmatory significance test using Moran I statistic.

Upon the establishment of spatial dependency, a spatially smoothed GC incidence risk ratio was examined via a spatial regression model. The result of which is graphically explained via the aid of the map after adjusting for age and sex population difference across the 96 RHADs.

This procedure paved the way for addressing the second research objective, which was done by examining the effect of the covariates discussed above on the risk of GC.

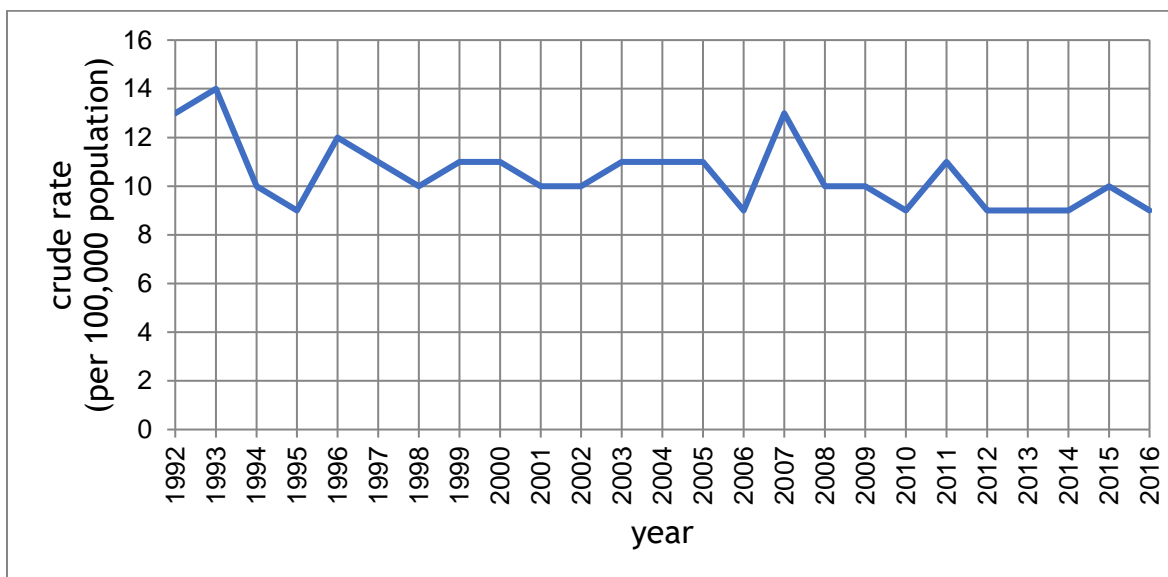
Lastly, the third research objective was addressed by incorporating a time component variable into the spatially smoothed model used in addressing the second research objective. The inclusion of the temporal component enables us to investigate both the impact of time on the risk of GC and change in GC risk ratio across the 96 RHADs. This was achieved by looking at the effect of the temporal component and the interactive impact of spatial component and time component.

Chapter 4: RESULTS

4.1 Descriptive Results

In this section, we explore the dataset using both graphical and numerical analysis to adjust for any inherent distortion present in the dataset. This procedure helps us to eliminate any recurrent errors that may affect our overall judgment. A total of 3172 cases of GC in Manitoba were extracted from the CCR, where 980 (approximately 31%) were cardia GC, and 2192 (approximately 69%) were non-cardia GC. The number of cases reported here spans over 25 years (1992 – 2016). To have a prior idea of the pattern of the distribution of the GC in the province to guide the analytical process, we categorized the GC cases into sub-groups (age-groups, sex, and year of diagnosis). The trend plot in Figure 4.1 shows an irregular pattern (which could be random or statistically significant) over the 25-years, with the two highest peak crude rates in 1993 and 2007, where the 1993 peak could be due to including prevalent cancer cases.

Figure 4.1: Plot of time trend of Gastric Cancer Crude rate (per 100,000 population) in Manitoba over 25 years (1992-2016)

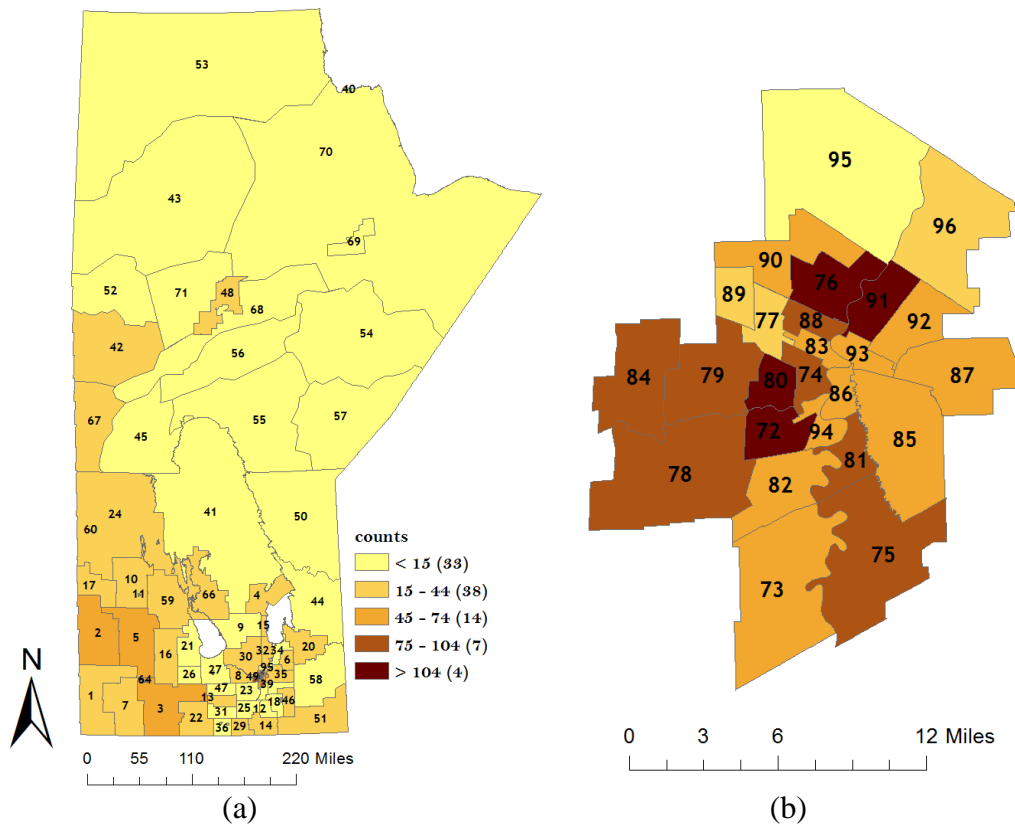


Overall, we observe a steady trend over time in the incidence of GC in the province.

Based on prior knowledge about the distribution of GC, which suggests that the main population at-risk are the adults and that the variation in population distribution could lead to a distorted risk estimate, an adjustment for age-group and sex was done to produce a standardized risk estimate.

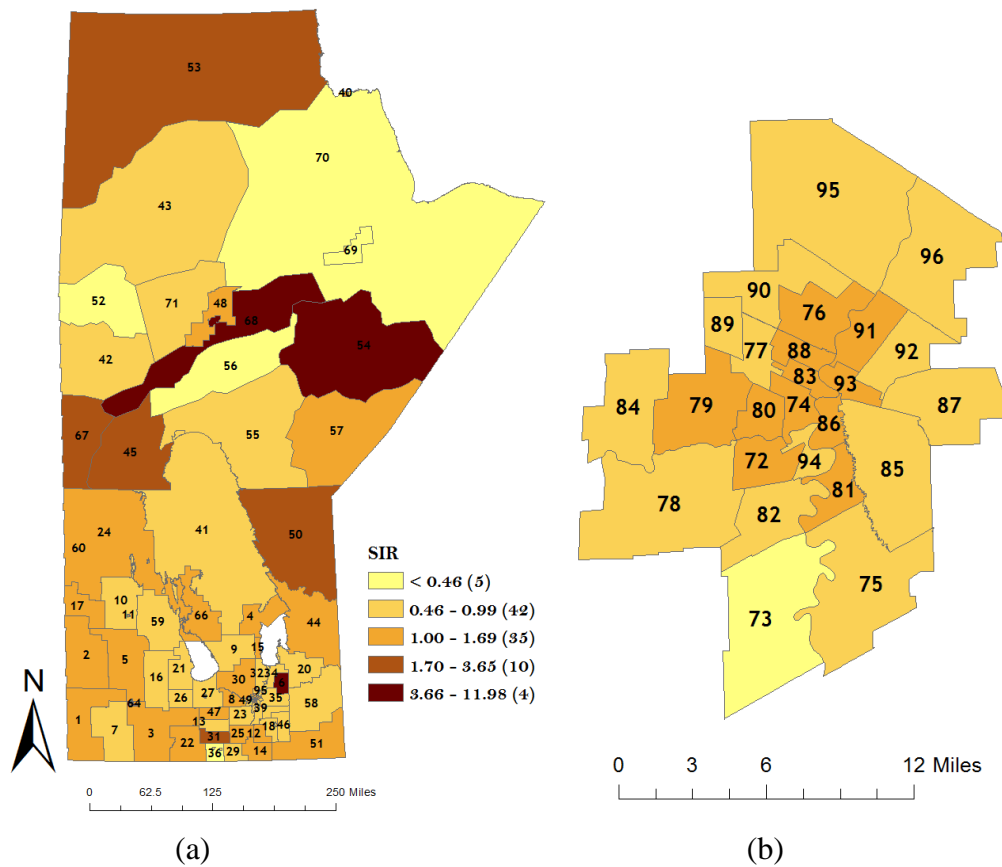
The standardization helps to reduce the distorting effect of the difference in age-sex population distribution on the estimated rate, as shown in Figures 4.2 - 4.4 below. Figure 4.2 (a-b) represents the spatial distribution of GC cases (counts). The map does not reveal much information in most regional health authorities (RHAs) (Figure 4.2a), except for WRHA (Figure 4.2b). The distributional pattern exhibited by WRHA can be attributed to its dense population compared to other RHAs.

Figure 4.2: Map of GC counts across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.



This effect of population variation was adjusted by computing the crude rate, which is the measurement of GC incidence relative to each district's population, as shown in Figure 4.3 (a-b). The map shows that some of the districts have a higher risk of GC compared to the inaccurate GC incidence risk displayed in Figure 4.2, which was not adjusted by the population size in each district.

Figure 4.3: Map of GC crude rate across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.



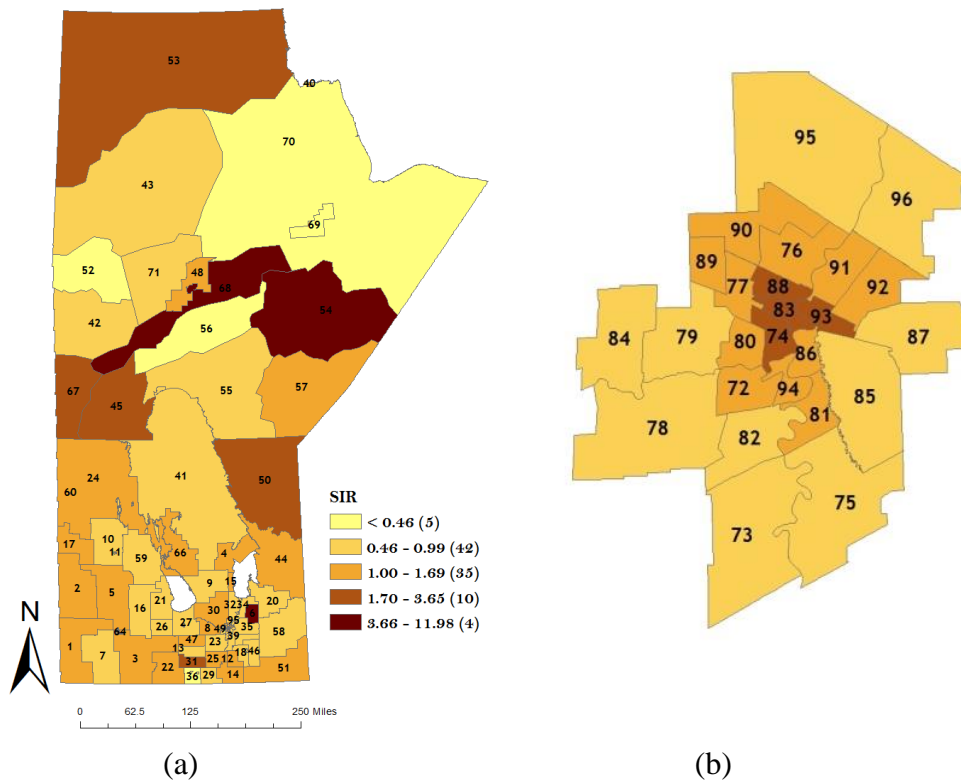
However, RHADs GC incidence risk ratio comparison based on the crude rate estimated in Figure 4.3 may still be misguided. This is because districts with a high population may have a high number

of cases, as shown in Figure 4.2b. This led to using the age and sex to adjusting for RHADs population difference with the result in Figure 4.4 (a - b). The adjusted risk ratio is defined as

$$SIR = \frac{y_i}{E_i} \quad (4.1)$$

where y_i is the observed GC cases and E_i is the expected GC cases as defined in Chapter 3.

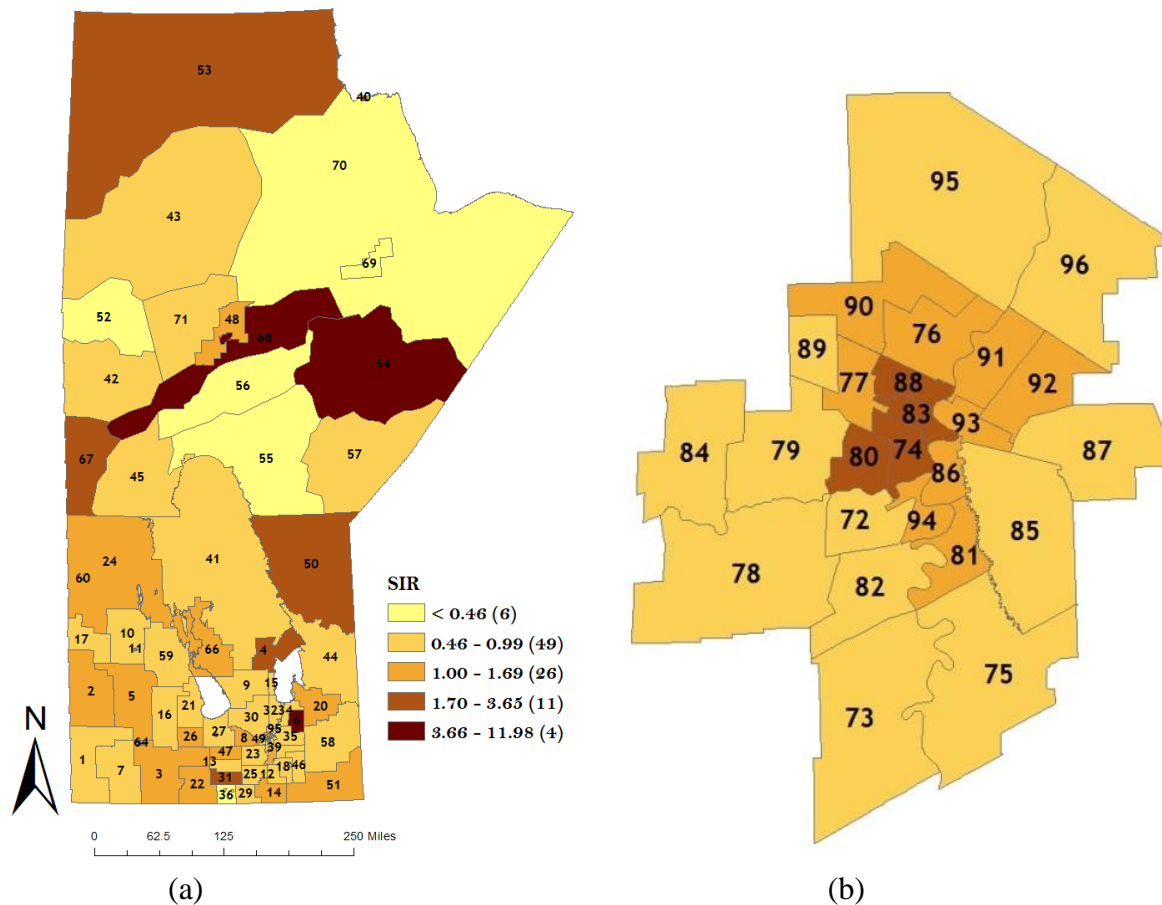
Figure 4.4: Map of age-sex adjusted incidence ratio of GC across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.



We also provided the maps of GC for age-adjusted, sex-stratified incidence ratio and age-sex adjusted incidence ratio for GC by topographical subsites (cardia and non-cardia) (Figures 4.5 - 4.12).

The result of the age-adjusted, sex-stratified GC risk ratio for men in Figure 4.5 below identified that men in four districts (Winnipeg Churchill in Northern RHA, Beausejour in Interlake Eastern RHA, Gods Lake Narrows in Northern RHA and Bay Line in northern RHA) had a risk value between 3.66 – 11.98. Also, men in eleven districts (Dauphin in Prairie mountain health RHA, Roland/Thompson in southern health RHA, Arborg in Interlake Eastern RHA, little grand rapids in Interlake Eastern RHA, The Pas in Northern RHA, Brandon East end in Prairie mountain RHA, Downtown East in Winnipeg RHA, Point Douglas South in Winnipeg RHA, Point Douglas North in Winnipeg RHA, River East South in Winnipeg RHA, and Lac Brochet in Northern RHA) were identified with GC risk between 1.70 – 3.65.

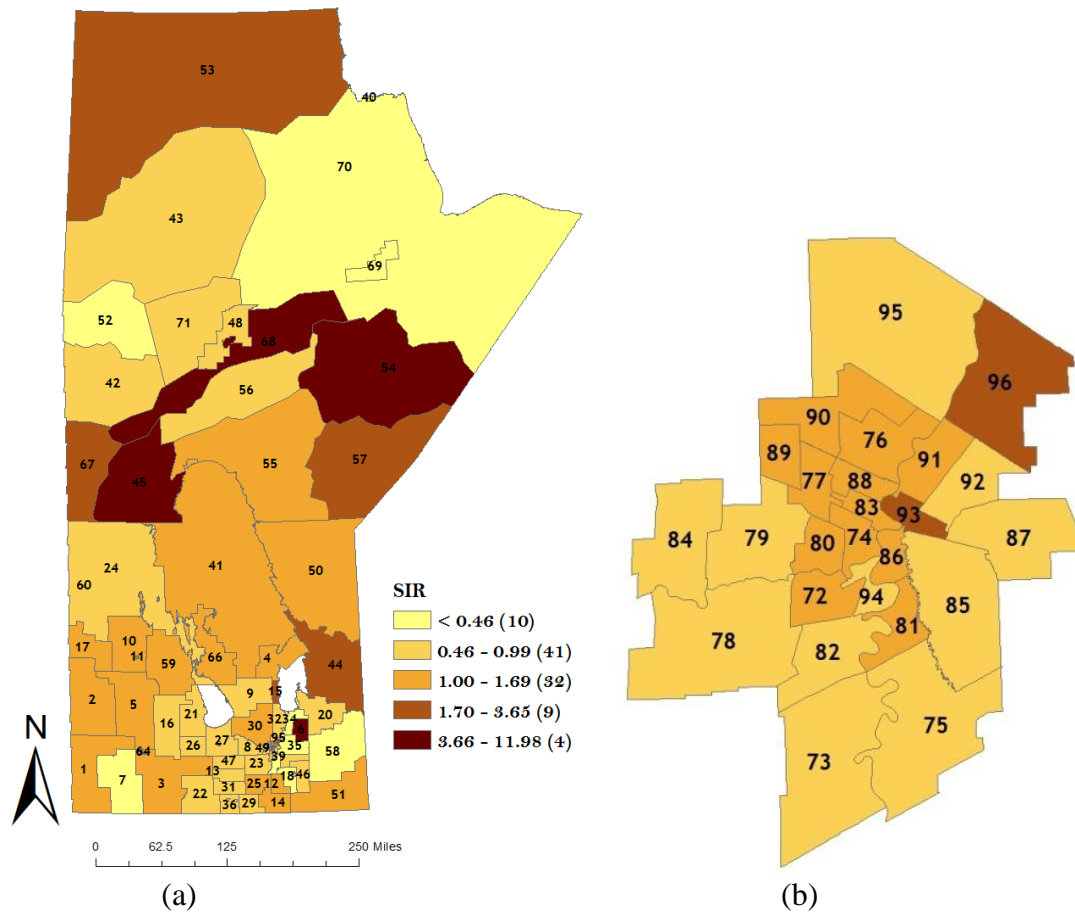
Figure 4.5: Map of age-adjusted, sex-stratified incidence ratio of GC for males across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.



From Figure 4.6, women in four districts (Beausejour in Interlake Eastern RHA, Grand Rapids in Northern RHA, Gods Lake Narrow in Northern RHA, and Bay Line in Northern RHA) were identified with GC risk between 3.66 –11.98.

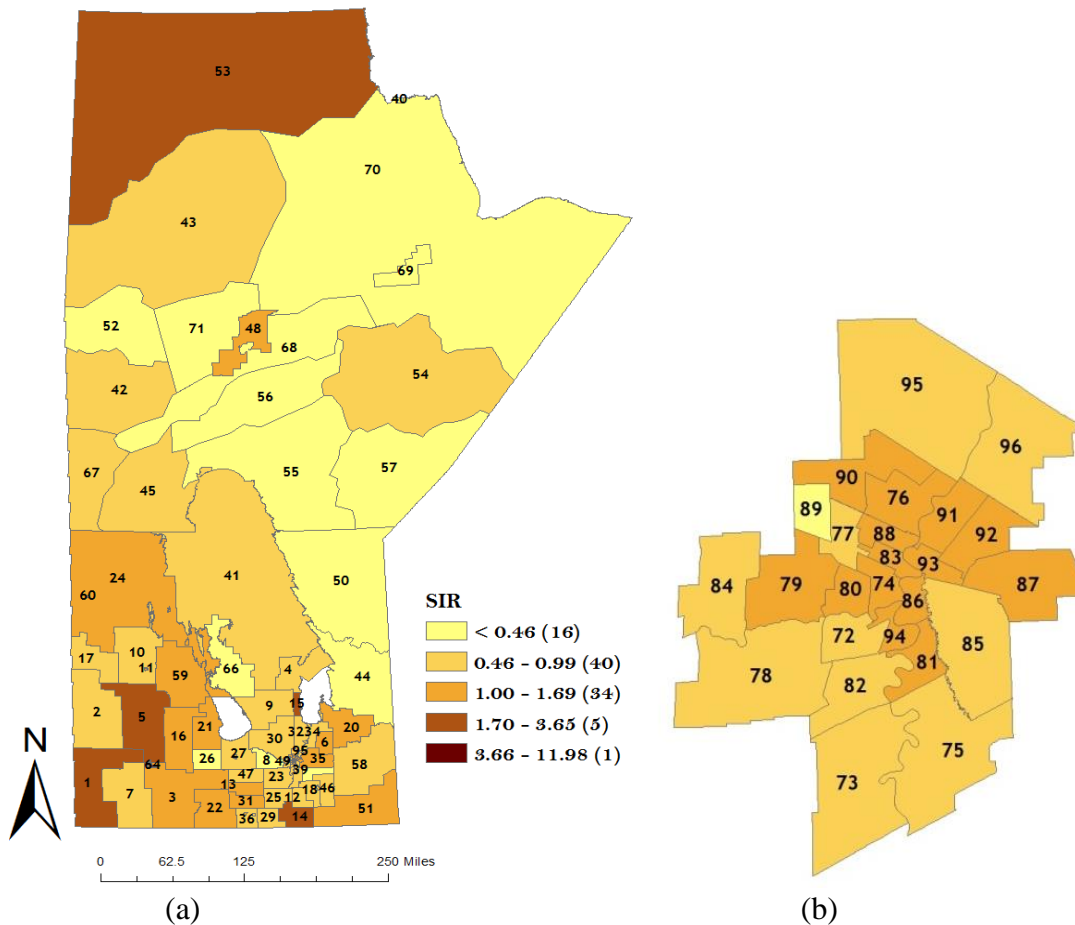
Also, women in 9 districts (Gimli in Interlake Eastern RHA, Winnipeg Churchill in Northern RHA, Powerview in Interlake Eastern RHA, Lac Brochet in Northern RHA, Island Lake in Northern RHA, Brandon East end in Prairie mountain health RHA, The Pas in Northern RHA, River East South in Winnipeg RHA, and River East North in Winnipeg RHA) were identified with a GC risk between 1.70 – 3.65

Figure 4.6: Map of age-adjusted, sex-stratified incidence ratio of GC for females across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.



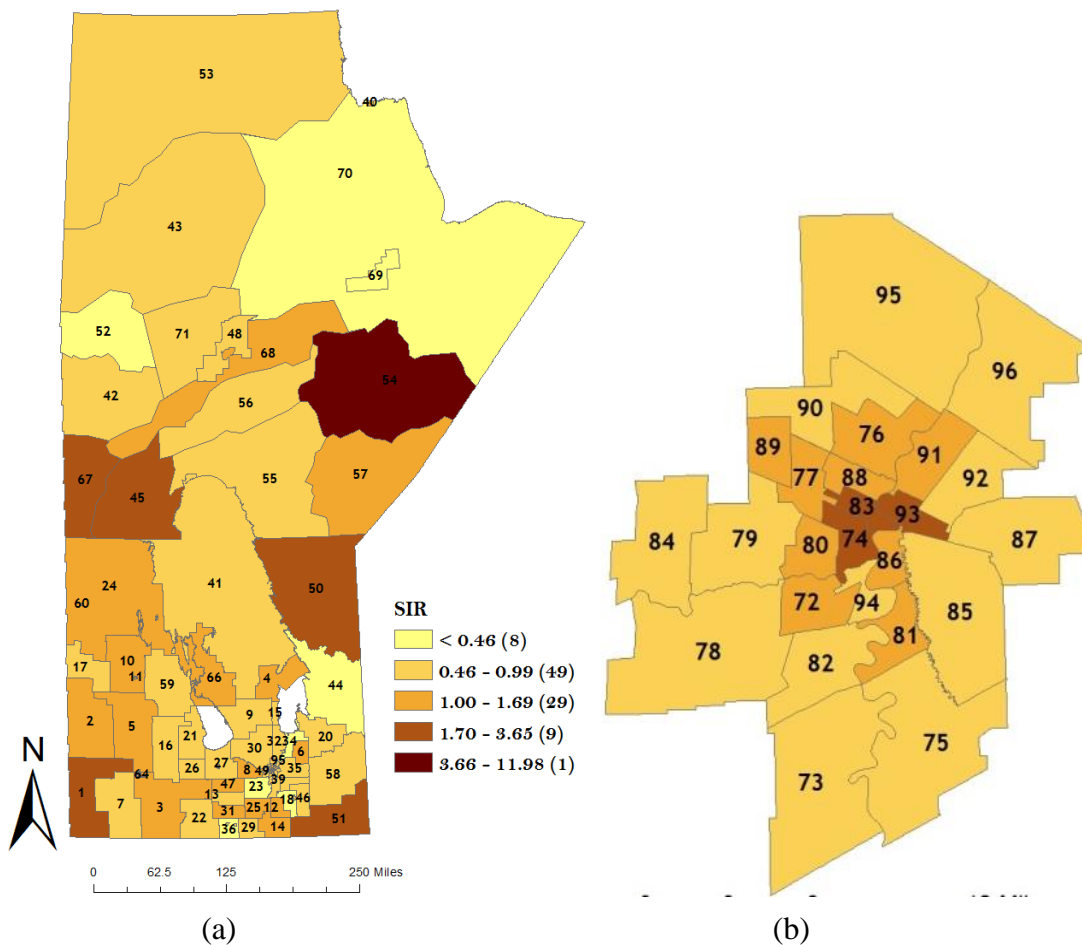
The result of CGC from Figure 4.7 below identified one district (Lac Brochet in Northern RHA) with risk of CGC between 3.66 – 11.98, and 5 districts (Souris River in Prairie mountain Health RHA, Little Saskatchewan in Prairie mountain health RHA, Red river south in southern health RHA, Gimli in Interlake eastern RHA, and Brandon east end in Prairie mountain health RHA) with cardia GC (CGC) risk between 1.70 – 3.65.

Figure 4.7: Map of age-sex adjusted incidence ratio of cardia GC across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.



The result of non-cardia GC (NCGC) in Figure 4.8 identified one district (Gods Lake Narrow in Northern RHA) with NCGC risk between 3.66 – 11.98 and nine districts (Souris River in Prairie Mountain Health RHA, Winnipeg Churchill in Northern RHA, Grand Rapids in Northern RHA, Little Grand Rapids in Interlake Eastern RHA, Rural East in Southern Health RHA, The Pas in Northern RHA, Downtown East in Winnipeg RHA, Point Douglas South in Winnipeg RHA and River East South in Winnipeg RHA) with NCGC risk between 1.70 – 3.65.

Figure 4.8: Map of age-sex adjusted incidence ratio of non-cardia GC across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.



These observed variations in risk for all data subsets across the province have been shown in the literature to be influenced by either environmental factors or genetic factors. As a result of this,

our study aims to investigate the etiology of GC relative to environmental factors, which also exhibit spatial variation, as shown in Figures 4.9 – 4.14 below.

One of the environmental variables of interest studied in this study used in the creation of the district level SES is the median household income, which has been aggregated into the 96 RHADs Figure 4.9 (a – b).

Figure 4.9: Map of median annual income across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.

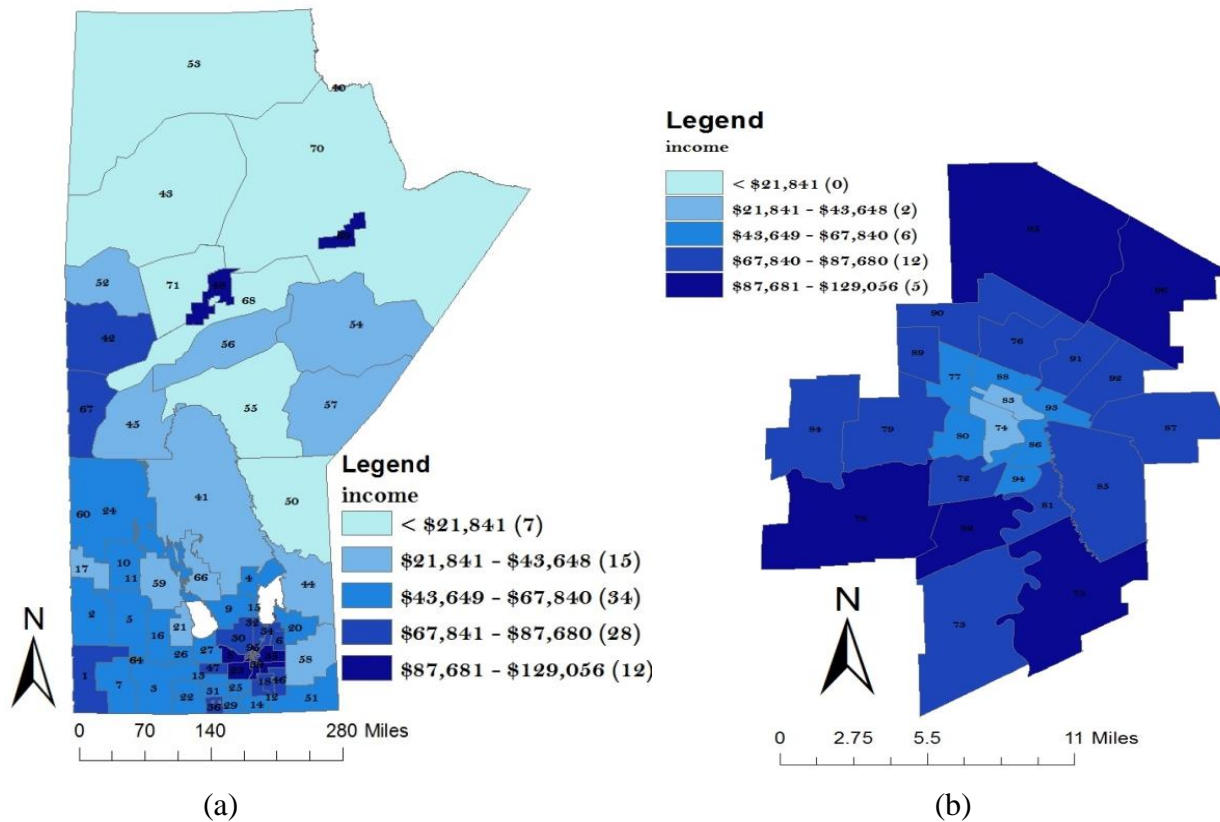
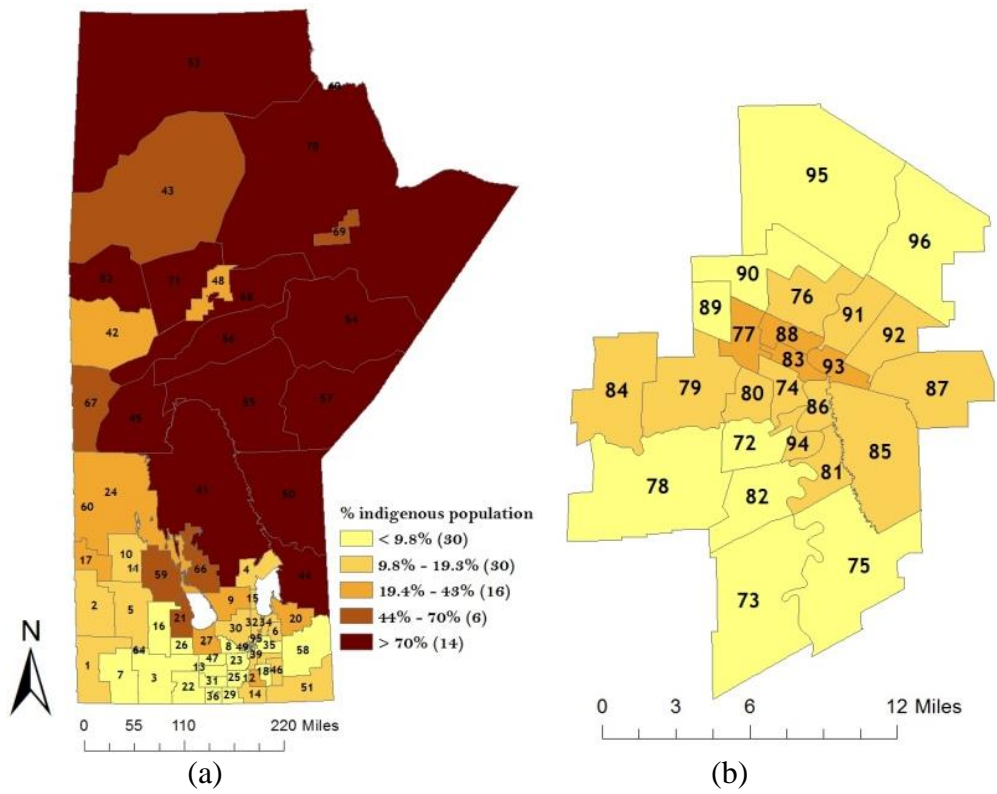


Figure 4.9 (a – b) represent the map of household income across the 96 RHADs in Manitoba and the 25 RHADs in Winnipeg RHA. We calculated the RHADs household income using the regional median as the measure of central tendency since we are dealing with income data, which tends to be skewed. The plot shows that more districts in the north have a median annual income less than

\$21,841 compared to other districts in the province, as 6 RHADs out of the 7 RHADs with a median yearly income less than \$21,841 were located in the northern region of the province. The result also shows that the minimum median income in Winnipeg RHA was between \$21,841 - \$43,648, and the maximum was between \$87,681 – \$129,056. Also, 5 RHADs out of the 12 RHADs (42%) with the highest income were located in WRHA, an indication that Winnipeg is among the highest-paid RHAs. This result shows the variation in income distribution in the province, which could partially be responsible for the difference in GC risk across the province.

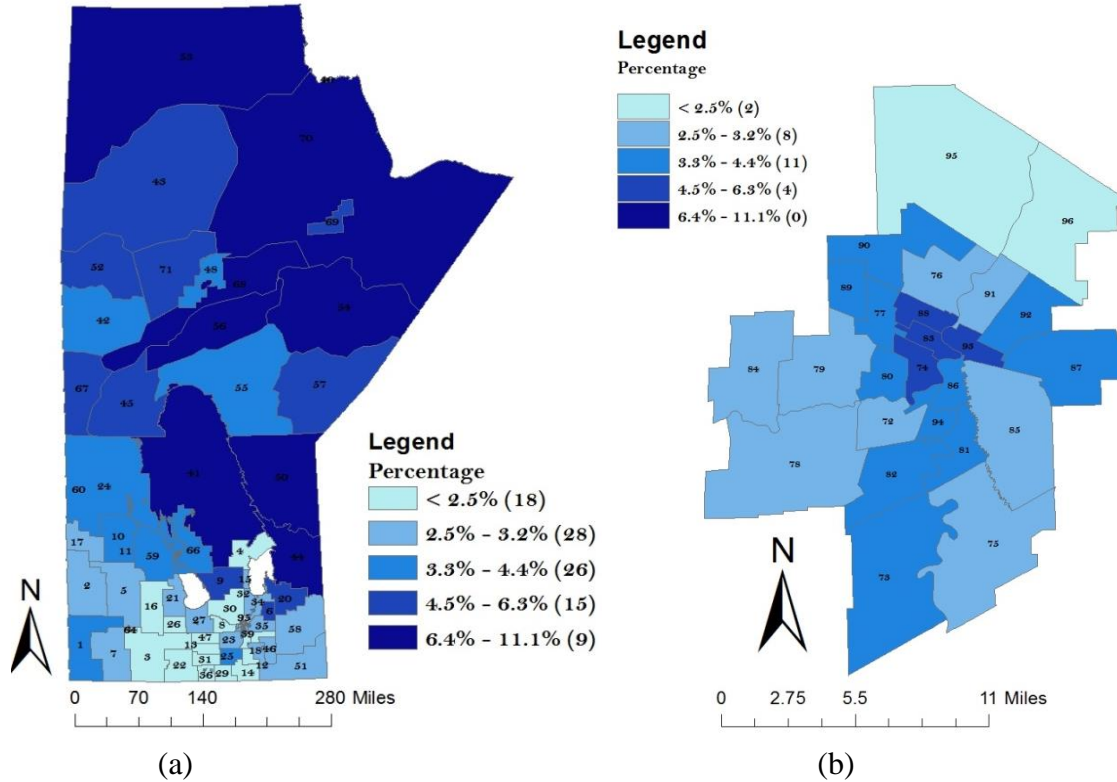
Another varying geographical factor that could also influence GC risk across the 96 RHADs is the Indigenous population. Indigenous Peoples are densely residents in the northern region of the province. Figure 4.10(a - b) shows that more than 50% of the majority of the districts in the north were made up of Indigenous Peoples. Note that 11 RHADs out of the 14 RHADs (78%) with 70.1% - 99% Indigenous population were located in the north, while very few resided in WRHA. The highest Indigenous population settlement in WRHA was between 19.4% - 43% located in 4 RHADs.

Figure 4.10: Map of Indigenous People population across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.



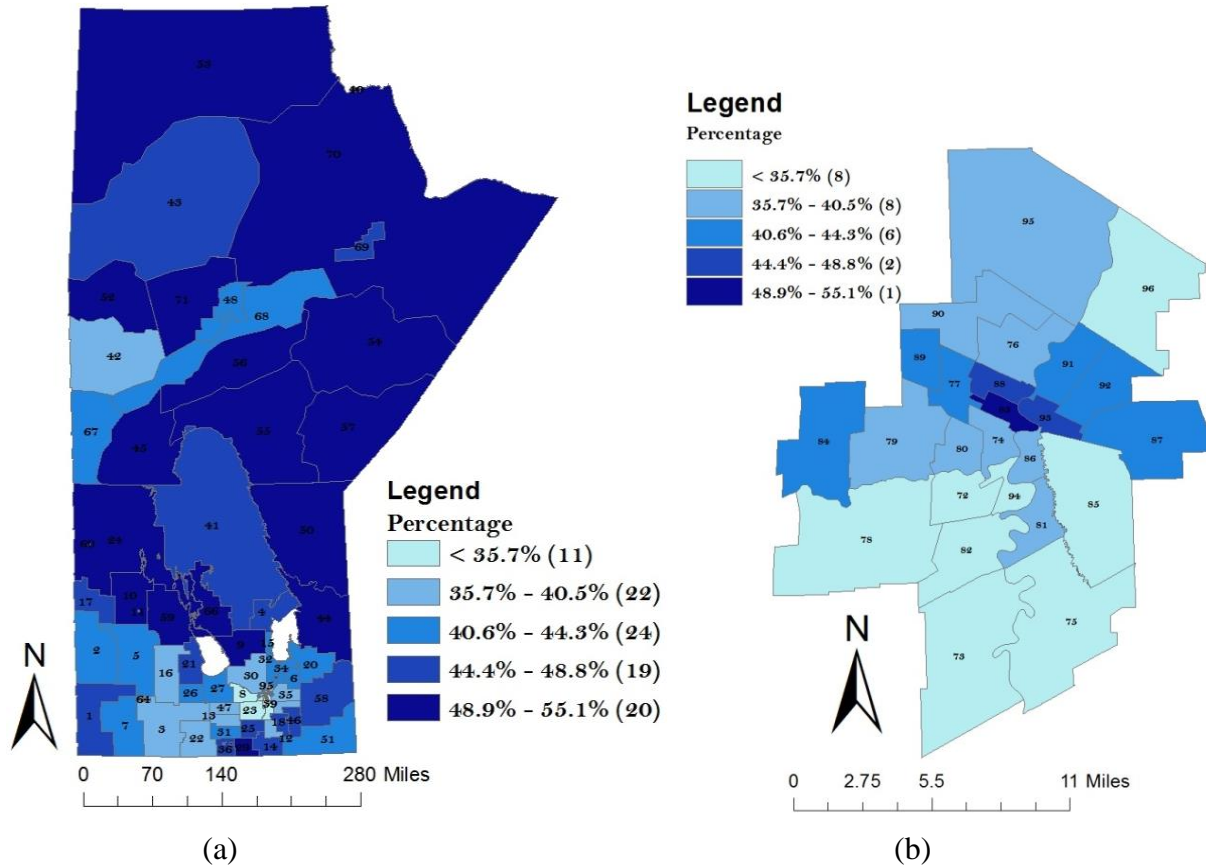
Unemployment is another variable that could influence the GC incidence rate. The highest unemployment rate was estimated to be 11.1%. The unemployment map presented in Figure 4.11 (a – b) reveals that there were more unemployed people in the northern region compared to the southern part of the province. Note that 6 RHADs out of the 9 RHADs (66.7%) with the highest unemployment population were located in the northern RHA. The highest unemployment rate in WRHA (Figure 4.12) was 6.3%, and 4 RHADs were found in this category (4.5% - 6.3%). Generally, more RHADs in WRHA had a lower unemployment rate (between 0% - 4.4%). The majority of the RHADs in northern RHA had a high percentage of the unemployment rate (6.34% - 11.12%) compared to other RHAs in southern Manitoba, where the unemployed percentage population ranged between 3.31% to 4.45%.

Figure 4.11: Map of the unemployment rate across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.



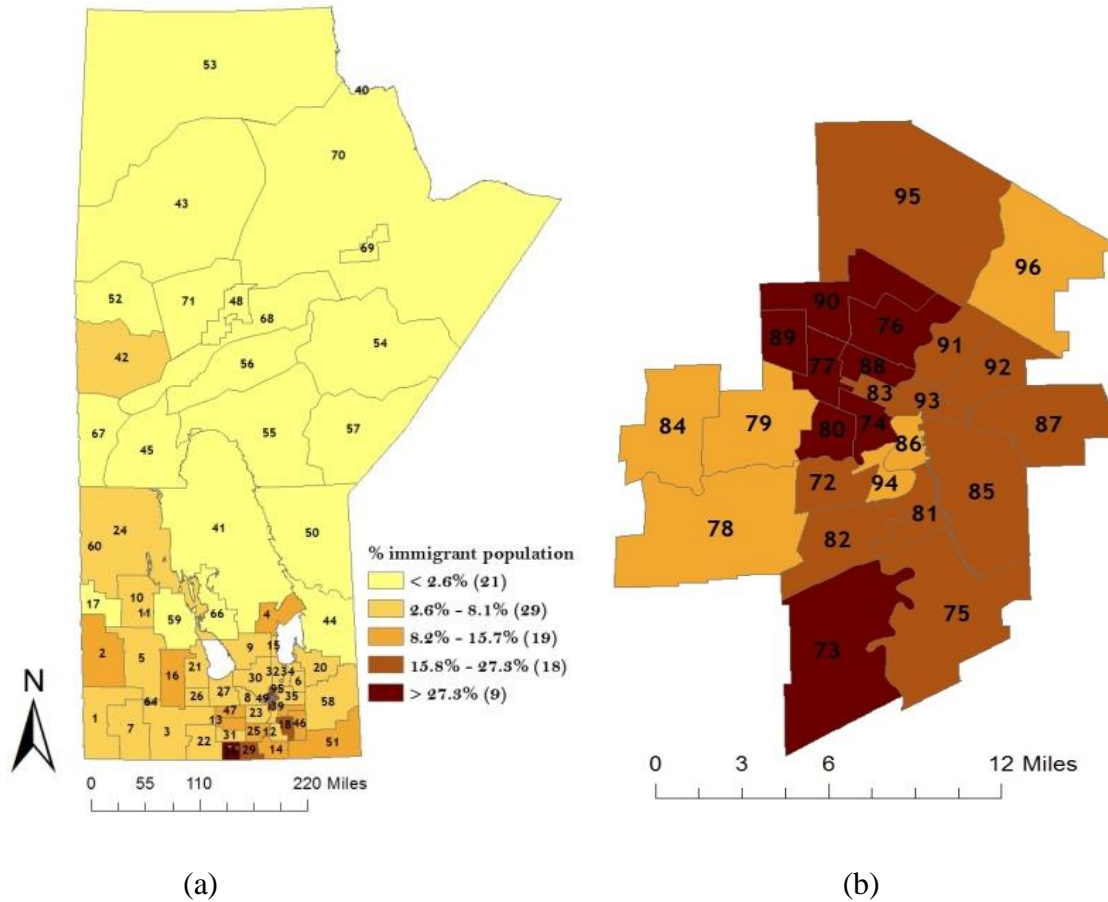
We also looked at the distribution of no post-secondary education in association with the GC incidence rate in Manitoba. This distribution, as shown in Figure 4.12 (a – b), showed that the majority of the northern and eastern regions did not have post-secondary education. In specific, 20 out of 96 RHADs had a percentage between 48.9 – 55.1 of their populations having no post-secondary school education. One RHAD (Point Douglas South) in WRHA was detected in the category with a high percentage of no post-secondary school education. Apart from Point Douglas south in WRHA, the majority of the districts in WRHA had a low rate of no post-secondary school education. Note that 22 RHADs out of 25 RHADs in WRHA had between 0 and 44.3% of no post-secondary education.

Figure 4.12: Map of people with no post-secondary education across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.



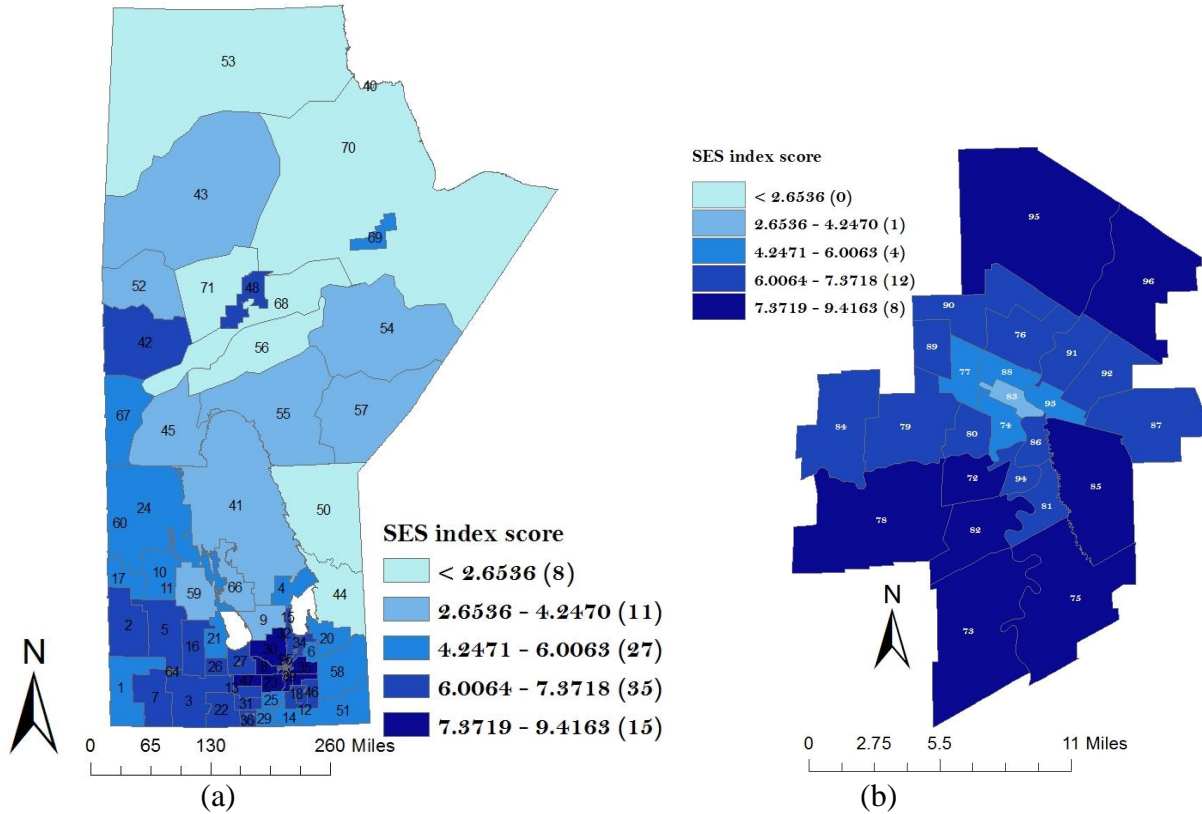
Similarly, we considered the distribution of immigrants across the 96 RHADs in Figure 4.13(a – b). We observed that the majority of the immigrant resided in the southern part of the province, where 9 RHADs had a high percentage between 27.4 – 53.9% of its population being immigrants. Only one RHAD (district 01) in the northern RHA had between 2.6 – 8.2% of its population to be immigrants, as shown in Figure 4.13a. WRHA had between 2.6 - 8.2% of its population to be immigrants, as shown in Figure 4.13b.

Figure 4.13: Map of immigrants across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.



Finally, we examined the distribution of socioeconomic score index (SESI) created by exploratory Factor analysis described in Chapter 3 using the most recent 2016 Census data. The result from the Factor analysis showed that the extracted factor (labeled SESI) accounts for 58% of the variation in the three variables. We also observed that the income variable was the most associated variable to SESI (highest loading value), and employment was the least associated variable to SESI. We also deduced that SESI captured 73 % of the variation in income, 60 % of the variation in education, and 41% of the variation in employment. The map of the SESI is represented in Figure 4.14 (a – b).

Figure 4.14: Map of socioeconomic status score index across the 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.



The SESI value ranged from 0.643397 to 9.4163, where a higher score implies a better socioeconomic status. Note that 15 RHADs had high SES score index; 8 RHADs (Fort Garry south, St. vital south, Assiniboine south, Fort Garry north, River height west, St. Boniface, River east north, and Seven Oaks north) were from Winnipeg RHA, 4 RHADs (MacDonald, Notre Dame/ St. Claude, Cartier/SFX, and Tache) from Southern health RHA, and 3 RHADs (District 02, Springfield, and District 03) from Interlake Eastern RHA. Most of these districts were located in the urban municipalities and majorly found in the southern part of the province.

4.2 Confirmatory Results

4.2.1 Poisson Regression Model

As part of the preliminary model, a Poisson regression model was fitted using the Bayesian approach, and the final summary is shown in Table 4.1. An initial univariate Poisson regression was fitted to determine the relationship between the independent variables (IV) and the GC risk. The result of this procedure, reported in Table 4.1, showed that all the 3 independent variables are significantly associated with the GC risk. This result suggested that each of the IVs was a potential candidate for inclusion in the final model. All MCMC convergence diagnostic conditions discussed in the method section (sub-section 3.4.4.) were satisfied.

Table 4.1: Incidence risk ratio (IRR) and 95% credible interval for overall GC dataset using univariate Poisson regression model

| Parameter | Incidence Risk Ratio | (95% credible interval) |
|------------|----------------------|-------------------------|
| Immigrant | 1.65 | (1.2289, 2.2155) |
| Indigenous | 1.62 | (1.3134, 1.9991) |
| SESI | 0.90 | (0.8742, 0.9226) |

This result was followed by a manual forward stepwise regression model procedure to determine the most parsimonious log-linear model. The initial non-spatial model derived from this procedure is given as:

$$\log(\lambda_i) = \alpha + sesi_i * \beta_1 + immigrant_i * \beta_2 + indigenous_i * \beta_3 \quad (4.21)$$

The model was fitted using R-INLA default prior settings assuming a less informative prior for the parameters to be estimated, to give the model more power to rely on the data. In specific, we

assumed a uniform prior distribution with a broad range (i.e., $\alpha \sim U(-\infty, +\infty)$) for intercept and a normal prior distribution with mean zero and a large variance (i.e., $\beta \sim N(0, 1000)$ where $\sigma^2 = \tau^{-1}$). The results are presented in Table 4.2 below.

Table 4.2: Incidence risk ratio (IRR) and 95% credible interval for overall GC dataset using saturated Poisson regression model

| Parameter | Incidence Risk Ratio | 95% Credible Interval (Lower, Upper) |
|-------------------|-----------------------------|---|
| SESI | 0.9039 | (0.8694, 0.9399) |
| Immigrant | 1.0101 | (1.007, 1.0131) |
| Indigenous | 1.003 | (1.000, 1.006) |

The result presented in Table 4.2 is a preliminary procedure, which is one of the exploratory analyses. The interpretation herein is for exploratory purposes only which differs from our final confirmatory analysis in later sections. The result showed a significant association between socioeconomic status, immigrant and risk of overall GC as the 95% credible interval does not contain the null hypothesis and a merely significant association between Indigenous population and risk of overall GC

($H_0: \text{Incidence Risk Ratio} = 1$).

It can be deduced from the result summary in Table 4.3 that a unit increase in RHADs socioeconomic score index reduces GC risk by 9% while adjusting for immigrant and Indigenous variables. Also, a 1% percent increase in the proportion of immigrants in a specific RHAD

increases the RHAD GC risk by approximately 1% while adjusting for the effect of socioeconomic score index and Indigenous effect.

We obtained a similar result for the male GC population (Table 4.3), where the socioeconomic score index and the immigrant population were significantly associated with the risk of GC in males. In specific, a unit increase in socioeconomic score index decrease the risk of GC in men population by approximately 10%. 1% increase in the proportion of immigrants reduces the risk of GC in the men by less than 1% while the Indigenous proportion was not significantly associated with the risk of GC in the men population.

Table 4.3: Incidence risk ratio (IRR) and 95% credible interval for men GC dataset using saturated Poisson regression model

| Parameter | Incidence Risk Ratio | 95% Credible Interval (Lower, Upper) |
|-------------------|-----------------------------|---|
| SESI | 0.8994 | (0.8573, 0.9436) |
| Immigrant | 0.991 | (0.955, 0.997) |
| Indigenous | 0.998 | (0.994, 1.006) |

Table 4.4: Incidence risk ratio (IRR) and 95% credible interval for female GC dataset using saturated Poisson regression model

| Parameter | Incidence Risk Ratio | 95% Credible Interval (Lower, Upper) |
|-------------------|-----------------------------|---|
| SESI | 0.9021 | (0.8403, 0.9646) |
| Immigrant | 1.008 | (1.003, 1.1503) |
| Indigenous | 1.005 | (0.999, 1.0111) |

Considering the risk of GC in the women population (Table 4.4), we observed a similar result where only the SESI and the immigrant population proportion were significantly associated with the risk of GC in the women population. Table 4.4 shows that a unit increase in SESI reduces the risk of GC in the women population by approximately 10%. A 1% increase in the proportion of immigrants increases the risk of GC in the women population by approximately 1%.

Table 4.5: Incidence risk ratio (IRR) and 95% credible interval for CGC dataset using saturated Poisson regression model

| Parameter | Incidence Risk Ratio | 95% Credible Interval (Lower, Upper) |
|-------------------|-----------------------------|---|
| SESI | 0.8564 | (0.7993, 0.9185) |
| Immigrant | 1.005 | (0.99, 1.0101) |
| Indigenous | 1.0131 | (0.9812, 1.0182) |

For CGC (Table 4.5), only the SESI index showed a significant relationship with the risk of CGC as a unit increase in the socioeconomic score index decreases the risk of CGC by 14%. We also obtained a similar result for NCGC (Table 4.6), where the SESI reduced the risk of NCGC by 10%.

Table 4.6: Incidence risk ratio (IRR) and 95% credible interval for NCGC dataset using saturated Poisson regression model

| Parameter | Incidence Risk Ratio | 95% Credible Interval (Lower, Upper) |
|-------------------|-----------------------------|---|
| SESI | 0.8994 | (0.8741, 0.9436) |
| Immigrant | 1.009 | (1.006, 1.0131) |
| Indigenous | 1.002 | (0.998, 1.005) |

However, the estimated parameters may be overestimated. This issue was addressed by introducing a spatial component effect in the model, which among other benefits, adjusts the variance via the introduction of the spatial dependence variance-covariance matrix discussed in the method section (Chapter 3).

The next phase of this study was the incorporation of an extra-Poisson component into the Poisson regression model (equation 4.21) to describe the structured spatial effect on GC risk. To do this, we examined the distribution of the unexplained variation across the RHADs and also determined the spatial dependence of GC occurrence using the Moran's I statistic.

4.2.2 Spatial Dependence Assessment

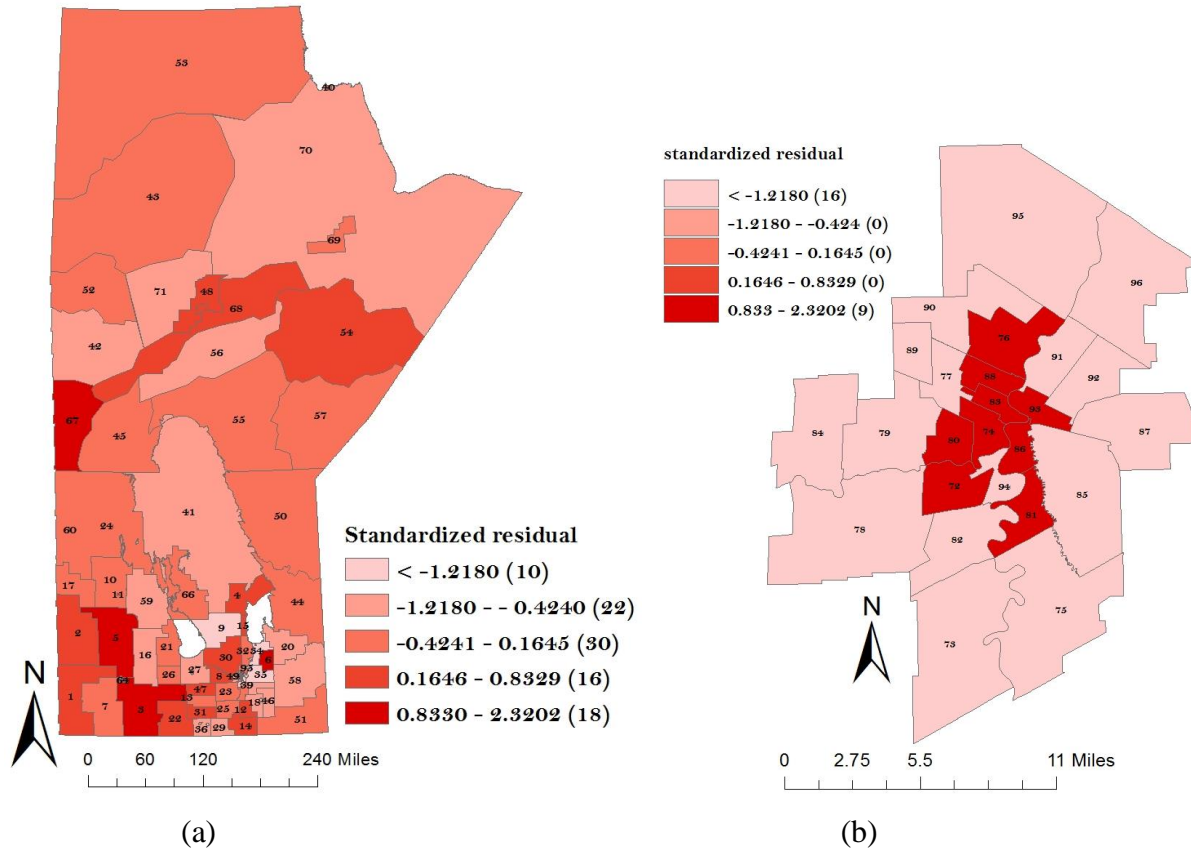
In this sub-section, apart from the fact that our response of interest is spatially aggregated which suggests the use of the spatial model as the appropriate analytic tool, we also justify the reason for including a spatial component into our model. We adopted both exploratory and confirmatory analyses to support our decision.

Firstly, we plot the unexplained variation (residuals) in GC risk to show that the unexplained variation differs across the 96 RHADs (Figure 4.15). The map shows the existence of unaccounted-for spatial variation by the Poisson regression model. This result supports our decision to include an extra-Poisson component into the model to describe this variation.

Also, we discovered that some RHADs had similar unexplained variation (clusters), which suggested similarity in GC risk among these districts.

Also, a Moran's I value of **0.1563 (p-value = 0.007)** was obtained, indicating the presence of spatial dependence (clusters). These tests confirmed the presence of spatial correlation in GC cases and are addressed in the subsequent sections.

Figure 4.15: Map of unexplained variation in 96 RHADs in Manitoba and Winnipeg; numbers in the map are the district number from 1 to 96.



4.2.3 Spatial Result: Ecological Regression Model

In order to address our second research objective, an ecological regression model was fitted using R-INLA for all data subset extracted from the parent data and the results were shown in Table 4.7. We imposed an ICAR on the spatial effect via Besag York Mollie (BYM), and minimally informative priors were imposed on all parameters as follows. A uniform prior distribution was assumed for the intercept $\alpha \sim U(-\infty, +\infty)$, a normal prior distribution for the slopes $\beta \sim N(0, 1000)$, and a log Gamma prior for both spatially structured and unstructured precision $\log \tau \sim \log \text{Gamma}(0.1, 0.01)$.

Table 4.7: Incidence risk ratio (IRR) and 95% credible interval for overall, male and female GC dataset using spatial Poisson regression model

| Parameter | Overall population | | Male population | | Female population | |
|-------------------|--------------------|--------------------------------------|-----------------|--------------------------------------|-------------------|--------------------------------------|
| | IRR | 95% Credible Interval (Lower, Upper) | IRR | 95% Credible Interval (Lower, Upper) | IRR | 95% Credible Interval (Lower, Upper) |
| SESI | 0.9213 | (0.836, 1.013) | 0.916 | (0.833, 1.008) | 0.906 | (0.814, 1.010) |
| Immigrant | 1.003 | (0.992, 1.014) | 1.004 | (0.993, 1.014) | 1.003 | (0.992, 1.014) |
| Indigenous | 1.001 | (0.994, 1.008) | 1.000 | (0.992, 1.007) | 1.004 | (0.961, 1.012) |

From Table 4.7, it can be seen that the covariates did not significantly predict the risk of GC for the overall GC dataset and the sex-stratified GC dataset.

Table 4.8: Incidence risk ratio (IRR) and 95% credible interval for cardia and non-cardia GC dataset using spatial Poisson regression model

| Parameter | Cardia GC | | Non-cardia GC | |
|-------------------|-----------|--------------------------------------|---------------|--------------------------------------|
| | IRR | 95% Credible Interval (Lower, Upper) | IRR | 95% Credible Interval (Lower, Upper) |
| SESI | 0.859 | (0.780, 0.947) | 0.898 | (0.812, 0.995) |
| Immigrant | 0.994 | (0.983, 1.003) | 1.002 | (0.990, 1.014) |
| Indigenous | 0.986 | (0.978, 0.994) | 1.000 | (0.990, 1.007) |

Table 4.8 represents the CGC and NCGC stratified dataset. Unlike the overall result, SESI was a significant factor associated with CGC. A unit increase in SES score decreases the risk of CGC by 14%, and the risk of NCGC by 10%. The risk of CGC among the Indigenous People, which was

marginally significant, is reduced by 1.4% compared to the rest of the population, while that of NCGC among the Indigenous was not significant.

Table 4.9: Incidence risk ratio (IRR) and 95% credible interval for cardia GC dataset stratified by sex using spatial Poisson regression model

| Parameter | Male | | Female | |
|-------------------|-------|---|--------|---|
| | IRR | 95% Credible Interval (Lower, Upper) | IRR | 95% Credible Interval (Lower, Upper) |
| SESI | 0.930 | (0.827, 1.012) | 0.738 | (0.618, 0.879) |
| Immigrant | 1.003 | (0.986, 1.006) | 0.994 | (0.979, 1.009) |
| Indigenous | 1.010 | (0.982, 1.049) | 0.981 | (0.966, 0.996) |

Further stratification of CGC by sex, as shown in Table 4.9, revealed no significant impact of SESI, immigrant and Indigenous variables on the risk of CGC for the men population. A different result for the women population was observed, where a unit increase in district SES score decreases the risk of CGC among the women population by a significant 26.2%, and the risk of CGC among Indigenous women population is reduced by 1.9% compared to the general population. The result of the non-cardia GC sex-stratified spatial model displayed in Table 4.10 showed no significant impact of the covariates on the risk of NCGC.

Table 4.10: Incidence risk ratio (IRR) and 95% credible interval for non-cardia GC dataset stratified by sex using spatial Poisson regression model

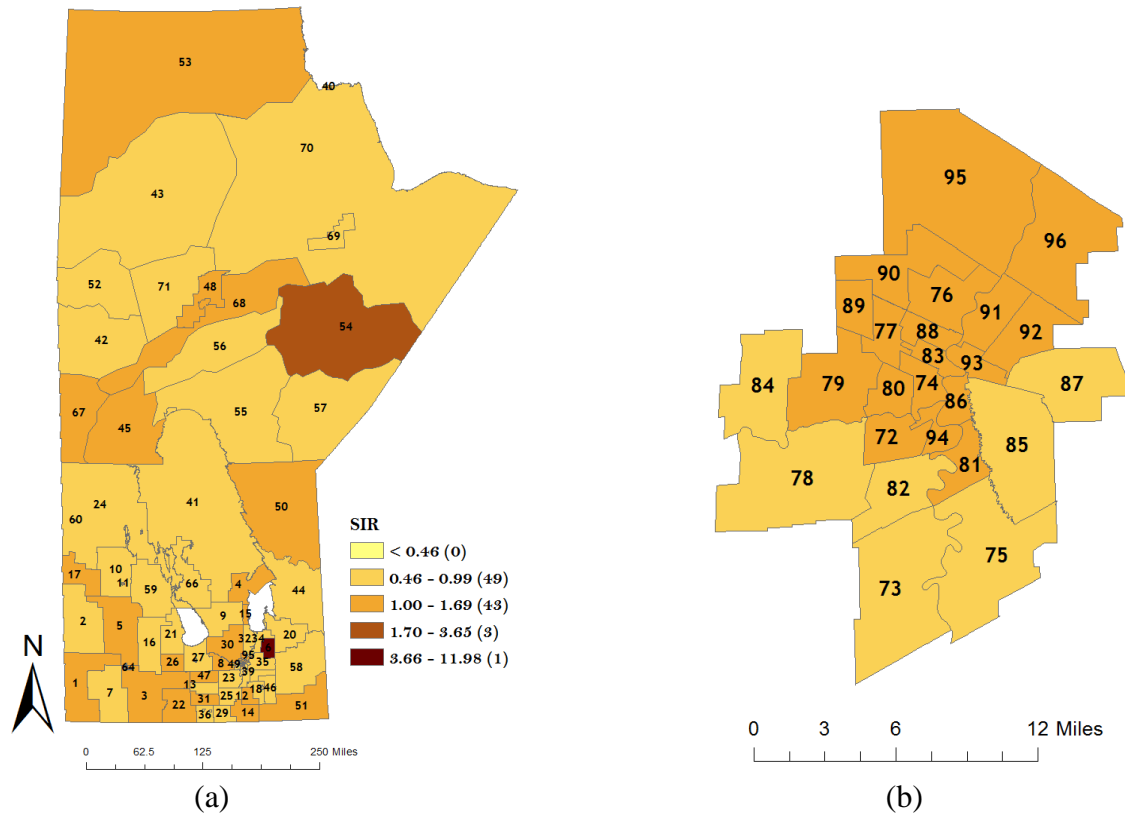
| Parameter | Male | | Female | |
|-------------------|-------|---|--------|---|
| | IRR | 95% Credible Interval (Lower, Upper) | IRR | 95% Credible Interval (Lower, Upper) |
| SESI | 0.910 | (0.811, 1.022) | 0.922 | (0.819, 1.040) |
| Immigrant | 1.010 | (0.997, 1.022) | 1.001 | (0.987, 1.014) |
| Indigenous | 1.001 | (0.992, 1.010) | 1.004 | (0.995, 1.013) |

Table 4.11: Variance estimates, 95% credible interval, and percent of spatial variation explained for overall, male, female, cardia and non-cardia GC dataset using spatial Poisson regression model

| Model | Spatial variance (95% Credible Interval) | Non-spatial variance (95% Credible Interval) | Spatial variation explained (%) |
|---------------------------|---|---|--|
| Overall population | 0.013 (0.004, 0.090) | 0.106 (0.062, 0.1937) | 11 |
| Male population | 0.015 (0.004, 0.102) | 0.079 (0.039, 0.170) | 16 |
| Female population | 0.014 (0.004, 0.102) | 0.060 (0.023, 0.181) | 19 |
| Cardia | 0.028 (0.003, 0.084) | 0.014 (2.99E-05, 0.020) | 67 |
| Non-cardia | 0.042 (0.014, 0.188) | 0.074 (0.035, 0.201) | 36 |

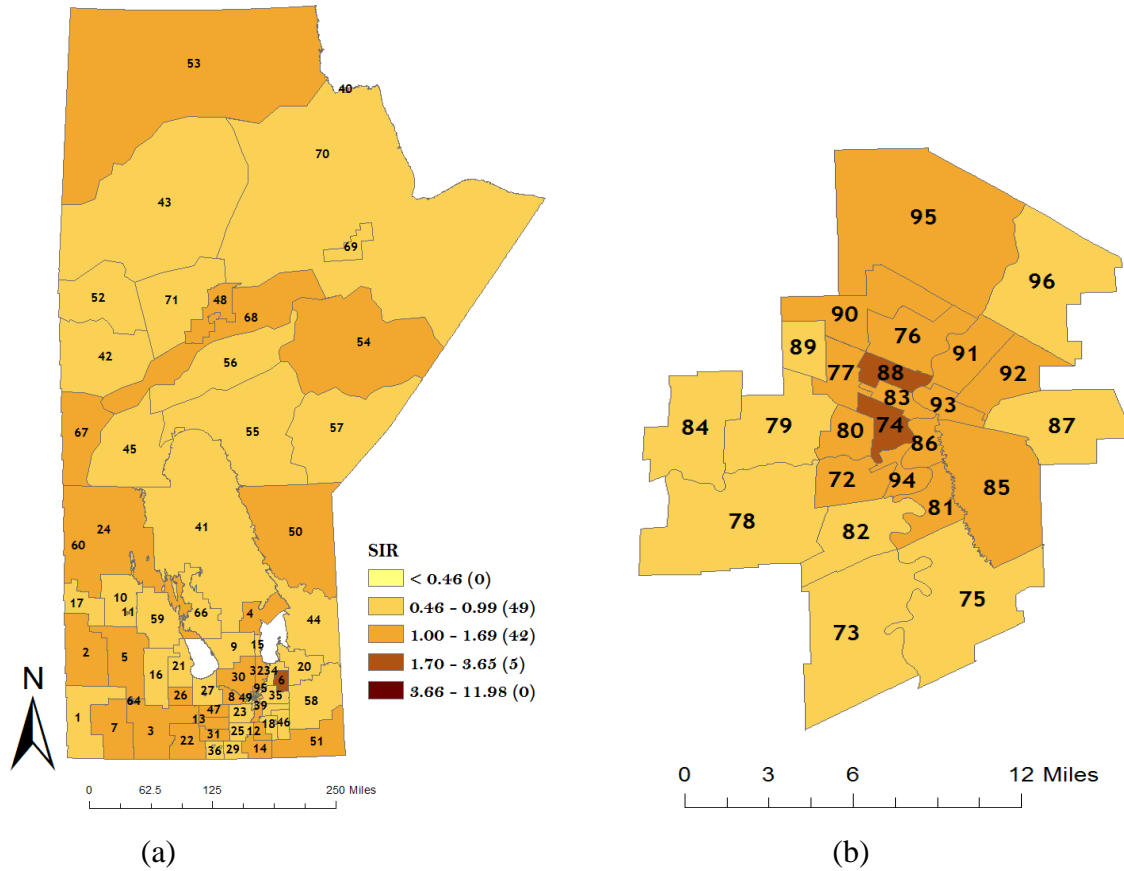
Table 4.11 shows that 11% of the variation in overall GC dataset was explained by spatial component, 16% of the variation in men stratified GC dataset was explained by spatial component, 19% of the variation in women stratified GC dataset was explained by spatial component, 67% of the variation in cardia CG was explained by spatial component, and 37% of the variation in Non-cardia CG was explained by spatial component. To further understand spatial variation of GC, CGC, and NCGC across the 96 RHADs, the corresponding smoothed spatial maps were plotted as shown in Figures 4.16 – 4.20.

Figure 4.16: Map of standardized incidence ratio of overall GC for 96 RHADs in Manitoba and Winnipeg using spatial Poisson regression model; numbers in the map are the district number from 1 to 96.



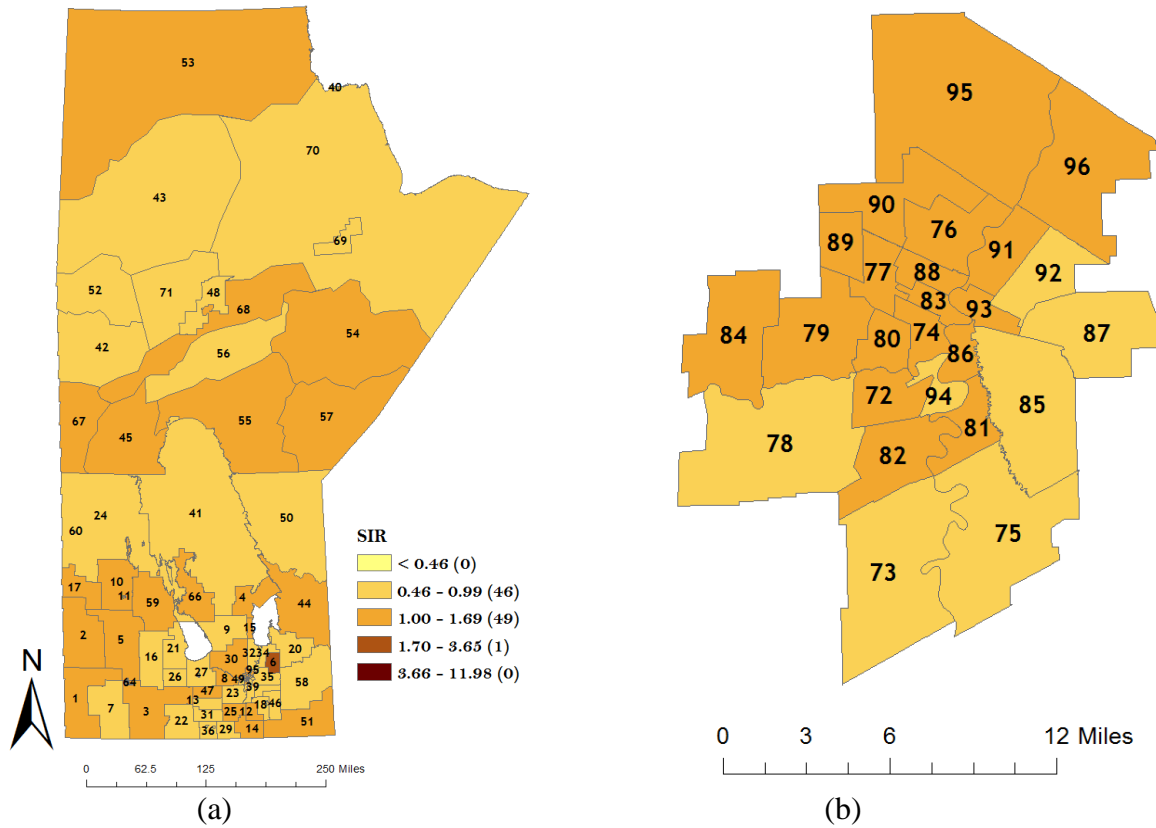
The spatial map of SIR of overall GC after adjusting for the effect of SES, Indigenous population, and immigrant population displayed in Figure 4.16 (a – b) identified a total of 25 districts with a higher overall GC risk compared to the rest of the population. Note that 11 (44%) out of the 25 districts with high overall GC were located in WRHA central. One district (District 06) in Interlake eastern RHA had the highest risk compared to the rest of the population (3.66 – 11.98). Three districts (Winnipeg Churchill in Northern RHA, district 12 in northern RHA, and Brandon east end in Prairie mountain RHA) had a higher risk than the rest of the population (1.70 – 3.65).

Figure 4.17: Map of standardized incidence ratio of male overall GC for 96 RHADs in Manitoba and Winnipeg using spatial Poisson regression model; numbers in the map are the district number from 1 to 96.



The result of the spatial variation of GC stratified by sex for men population displayed in Figure 4.17 (a – b) identified 22 districts with GC incidence risk ratio greater than the rest of the population. Note that 7 RHADs out of the 22 RHADs (i.e., 32%) had a high GC incidence risk ratio relative to the rest the population which were located in WRHA. Five districts (Winnipeg Churchill in the northern RHA, Brandon East end and Dauphin in Prairie Mountain RHA, and downtown East & Point Douglas North in Winnipeg RHA) had the highest risk ratios between 1.70 – 3.65.

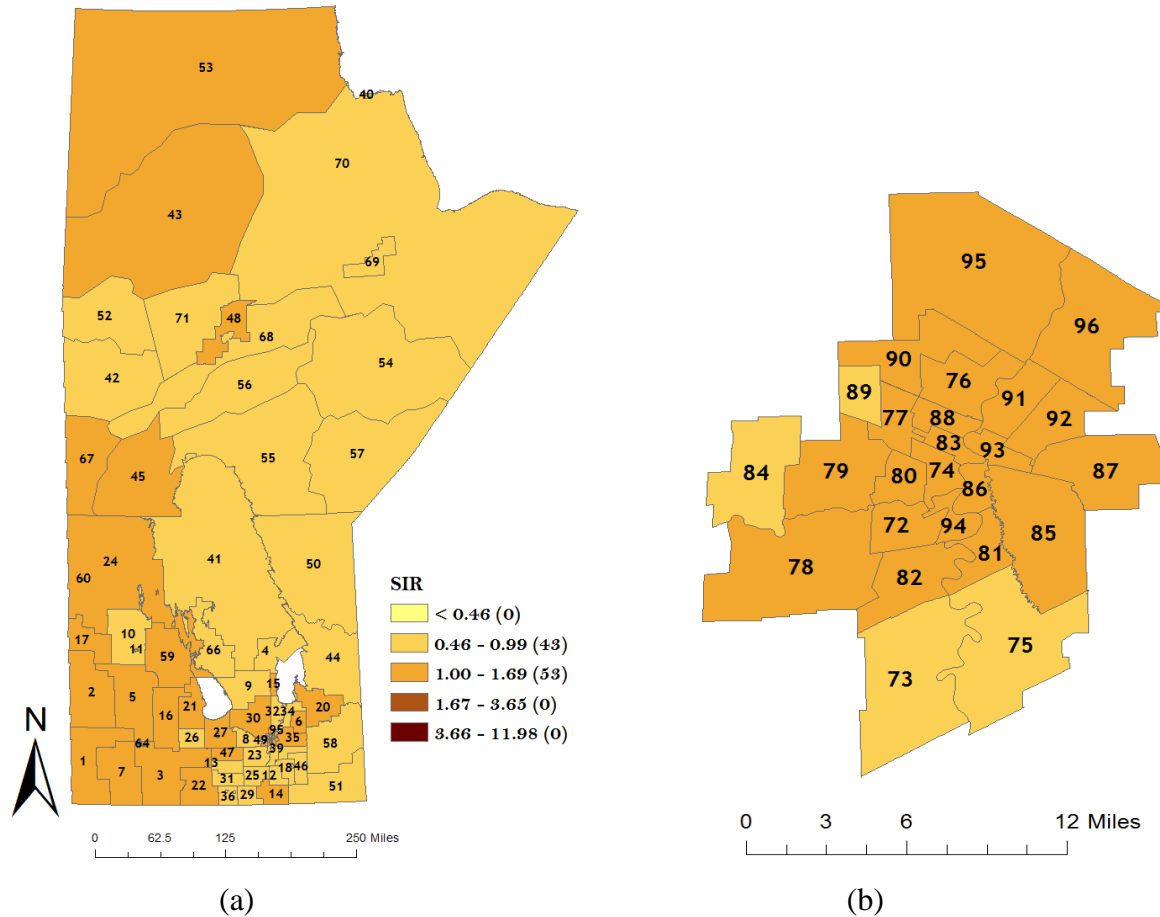
Figure 4.18: Map of standardized incidence ratio of female overall GC for 96 RHADs in Manitoba and Winnipeg using spatial Poisson regression model; numbers in the map are the district number from 1 to 96.



For the women GC sub-population, a total of 16 RHADs were identified with a high GC incidence risk ratio relative to the rest of the population. Note that 10 out of the 16 RHADs (i.e., 63%) were located in WRHA. No district was found in the highest risk range, one RHAD (district 06 in Interlake RHA) was identified with the highest GC incidence risk ratios (Figure 4.18).

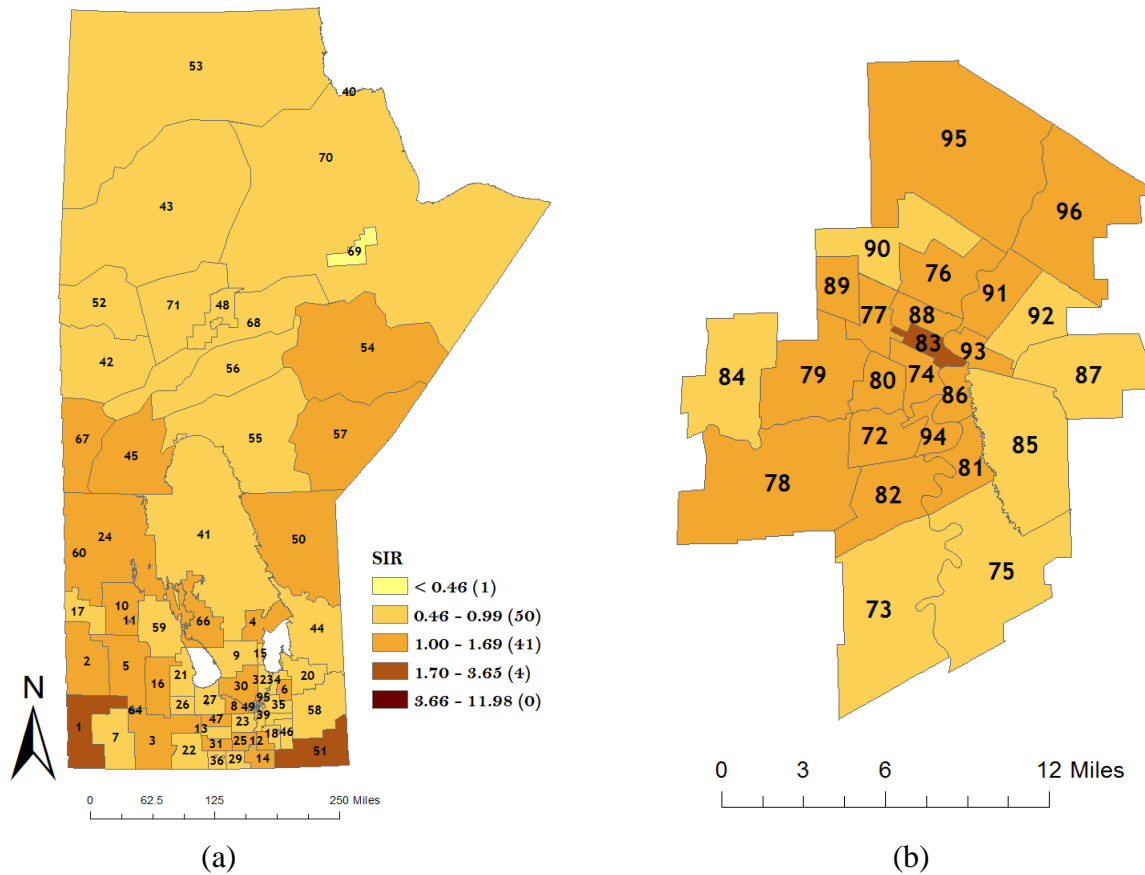
The result of CGC spatial variation in Figure. 4.19 identified 11 RHADs with a high CGC incidence risk ratio relative to the rest of the population. Note that 6 out of the 11 RHADs (i.e., 55%) were located in WRHA.

Figure 4.19: Map of standardized incidence ratio of CGC for 96 RHADs in Manitoba and Winnipeg using spatial Poisson regression model; numbers in the map are the district number from 1 to 96.



The result of the NCGC in Figure 4.20 identified 27 RHADs with a higher risk of NCGC relative to the rest of the population. Note that 11 out of the 27 RHADs (i.e., 41%) with a high NCGC incidence risk ratio were located in WRHA. Similar to CGC, no RHAD was found in the highest risk class interval where 4 RHADs (Souris River in Prairie Mountain RHA, Winnipeg Churchill in northern RHA, Rural East in Southern Health RHA, and Point Douglas South in Winnipeg RHA) were identified with the highest NCGC incidence risk ratios.

Figure 4.20: Map of standardized incidence ratio of NCGC for 96 RHADs in Manitoba and Winnipeg using spatial Poisson regression model; numbers in the map are the district number from 1 to 96.



4.2.4 Spatio-temporal Regression Result

In order to address the third research question, the ecological regression model was modified by adding temporal random effect and interaction effect between temporal random and spatial random effects. Due to the rareness of GC, the twenty-five-time frame was compressed into five time periods as follows:

Period 1: summation of GC cases from 1992 – 1996

Period 2: summation of GC cases from 1997 – 2001

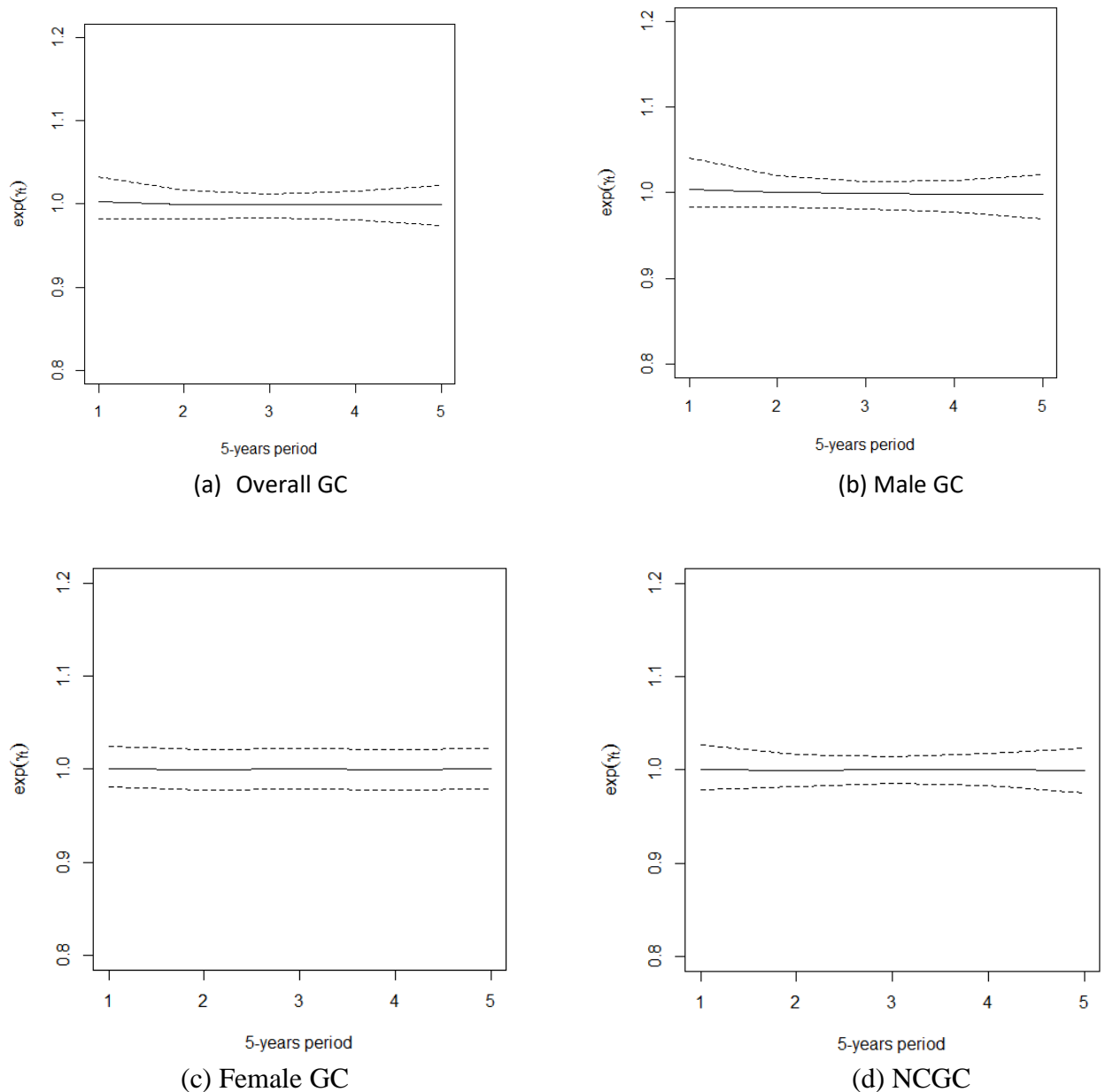
Period 3: summation of GC cases from 2002 – 2006

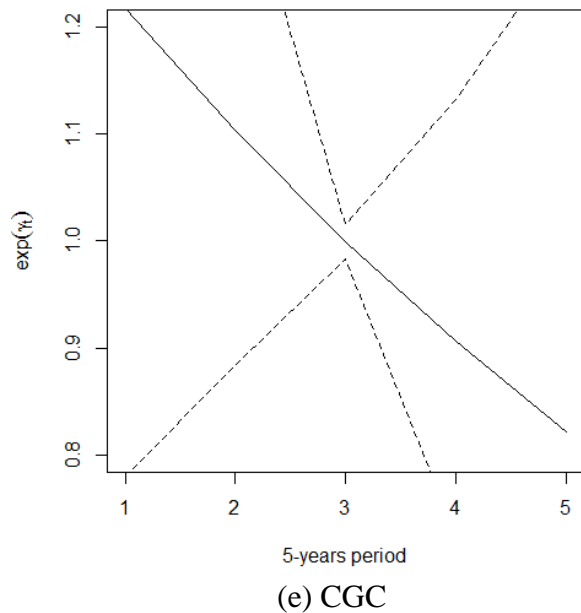
Period 4: summation of GC cases from 2007 – 2011

Period 5: summation of GC cases from 2012 – 2016

A combination of three smoothers (RW(1), RW(2) and AR(1)) for the temporal random effect was applied on the four types of spatio-temporal interaction discussed in the method section (Chapter 3) using R-INLA. The model that best fits the data was selected using the DIC (Appendix I).

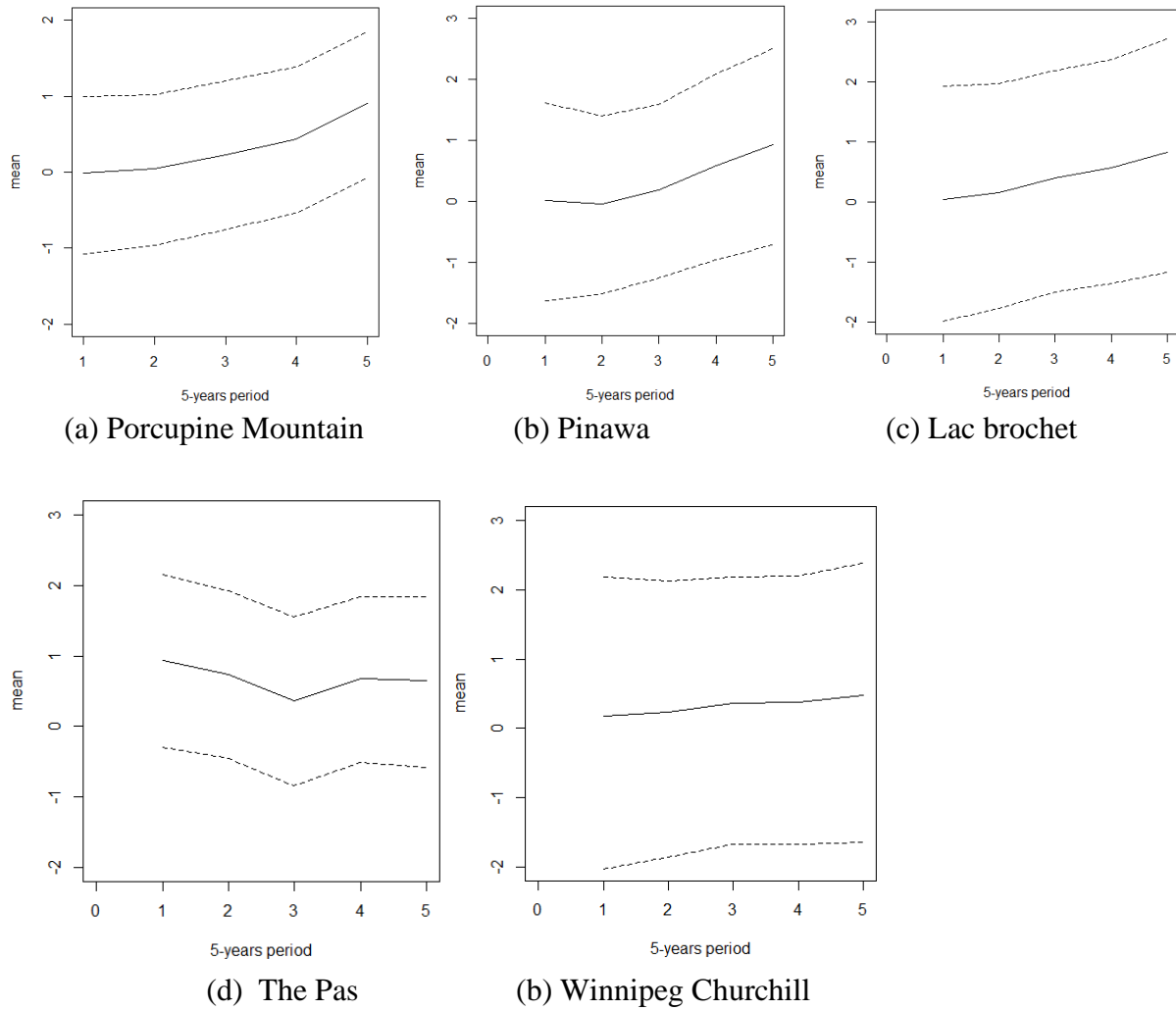
Figure 4.21: Marginal temporal random effect and 95% credible interval plot for overall GC, male GC, female GC, NCGC, and CGC





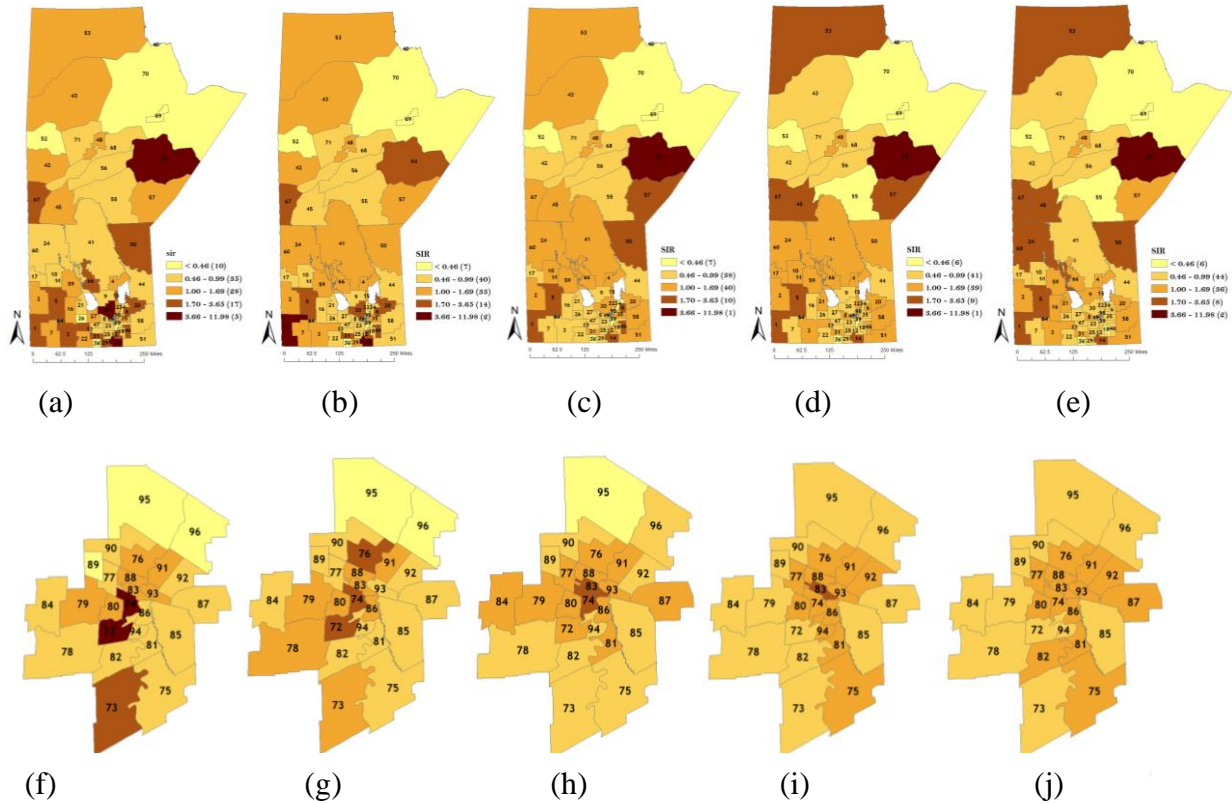
The results of marginal temporal effect plots, reported in Figure 4.21 (a – e), showed a steady non increasing or decreasing trend for overall GC (a), male GC (b), female GC (c) and NCGC (d), while a decreasing trend is observed for CGC (e). However, the steadiness of the overall GC, male GC, female GC, and NCGC did not hold for all districts as some districts like Porcupine Mountain in Prairie Mountain Health RHA, Pinawa in Northern RHA, Lac Brochet in Northern RHA, and Winnipeg Churchill in Northern RHA exhibited an increased risk of overall GC while The Pas from northern RHA exhibited a decrease in overall GC risk compared to the rest of the population (Figure 4.22).

Figure 4.22: Temporal random effect and 95% credible interval plot of overall GC for Porcupine Mountain, Pinawa, Lac Brochet, The Pas and Winnipeg Churchill district



Lastly, we examined the spatio-temporal random effect for overall GC, male GC, female GC, CGC, and NCGC, which showed spatial variation in risk over time.

Figure 4.23: Maps of standardized incidence ratio of overall GC for 96 RHADs in Manitoba and Winnipeg for five time periods using spatio-temporal Poisson regression model; numbers in the map are the district number from 1 to 96.

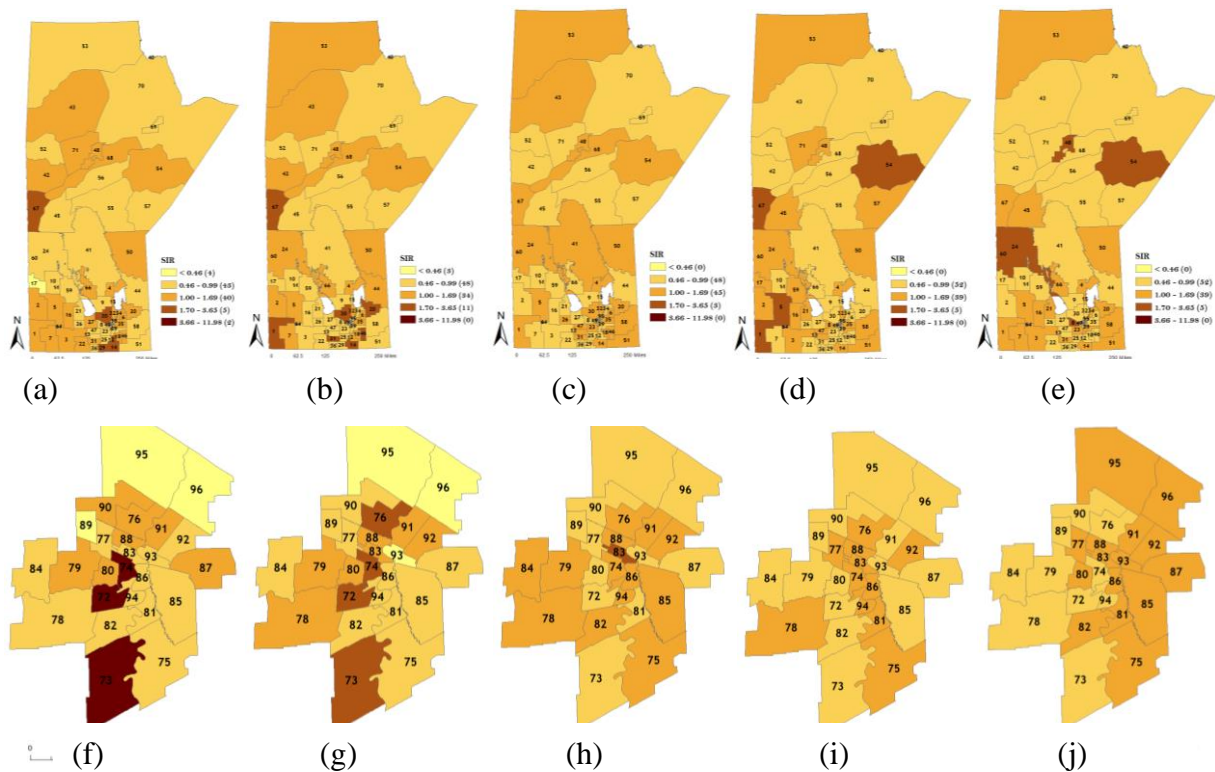


We considered several spatio-temporal models with different temporal smoothers and various interactions between space and time, as described in Chapter 3. The model with RW (1) smoother for the structured temporal random effect and RW (1) for the structured temporal effect and unstructured spatial random effect in the space-time interaction effect were adjudged the best with the DIC value of 2202.78 (Appendix 1). Figure 4.23(a - j) above, demonstrated the presence of geographical variation of overall GC incidence risk across the 96 RHADs over time.

Similarly, for the male strata, the spatio-temporal model with RW (1) smoother for the temporal random effect, and AR (1) for the structured temporal random effect in the spatio-temporal interaction effect and the unstructured spatial random effect were adjusted the best model based

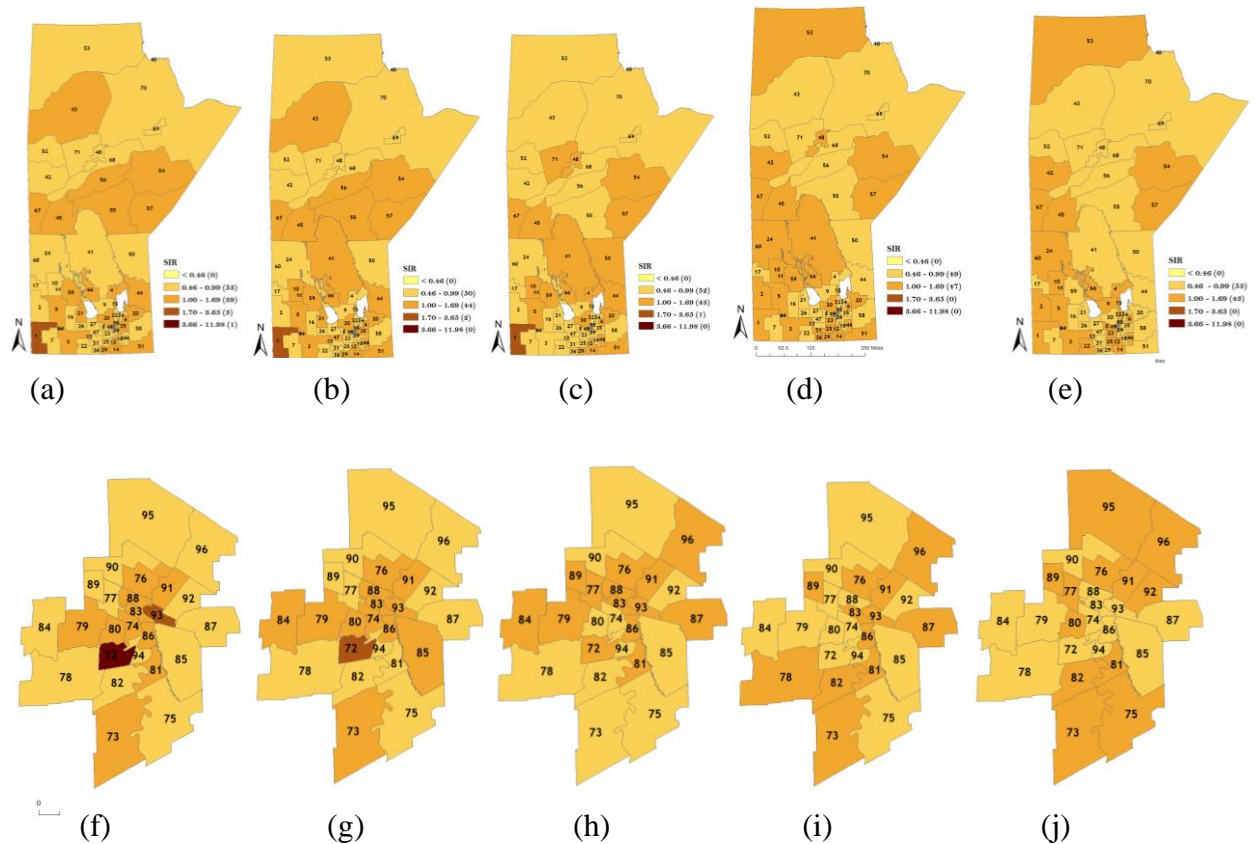
on the DIC value 1992.44 (Appendix 1). An increasing trend over time was observed across most of the RHADs, especially in the northern region of the province.

Figure 4.24: Maps of standardized incidence ratio of overall male GC for 96 RHADs in Manitoba and Winnipeg for five time periods using spatio-temporal Poisson regression model; numbers in the map are the district number from 1 to 96



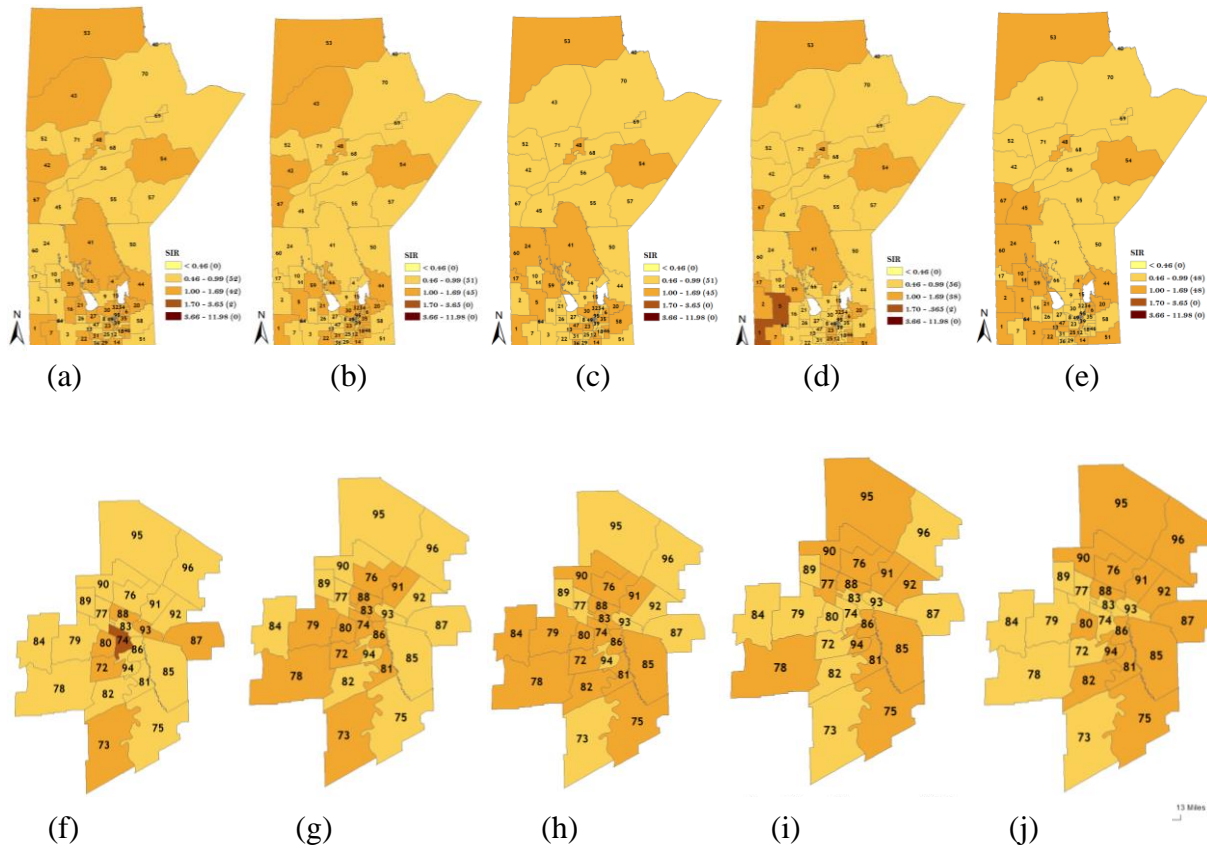
The result presented in Figure 4.25 showed the spatio-temporal model for female GC strata. The spatio-temporal model with AR (1) smoother for the temporal effect and AR (1) for the structured temporal random effect in the spatio-temporal interaction and the unstructured spatial random effect were adjudged the best model, having the smallest DIC value 1559.08 (Appendix 1). We observed a steady risk of GC for women across most RHADs.

Figure 4.25: Maps of standardized incidence ratio of overall female GC for 96 RHADs in Manitoba and Winnipeg for five time periods using spatio-temporal Poisson regression model; numbers in the map are the district number from 1 to 96



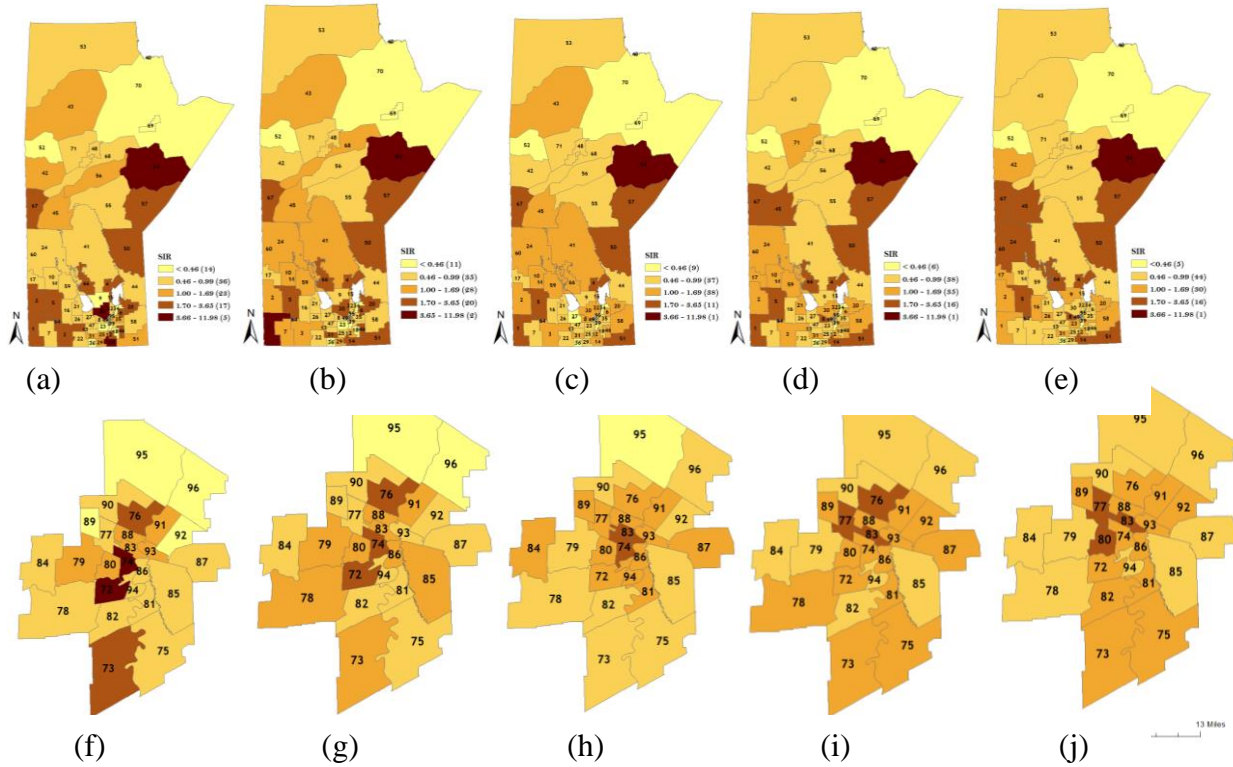
The spatio-temporal result for CGC was based on the spatio-temporal model with RW (2) smoother for the temporal random effect, and an AR (1) smoother for the structured temporal effect in the spatio-temporal interaction and the unstructured spatial random effect (Appendix 1). Figure 4.26 shares a very close pattern with that of the women GC with almost steady risk over time and a little variation in Prairie Mountain and Winnipeg RHA.

Figure 4.26: Maps of standardized incidence ratio of CGC for 96 RHADs in Manitoba and Winnipeg for five time periods using spatio-temporal Poisson regression model; numbers in the map are the district number from 1 to 96



Lastly for the NCGC, the spatio-temporal model with RW (1) smoother imposed on the temporal random effect, and RW (1) smoother imposed on the structured temporal effect in the spatio-temporal interaction and unstructured spatial random effect were adjudged the best model with the smallest DIC value of 1987.02 (Appendix 1). The result revealed a significant variation in the risk NCGC across the RHADs in Manitoba over time, as almost all the RHA showed some variations over time, as shown in Figure 4.27.

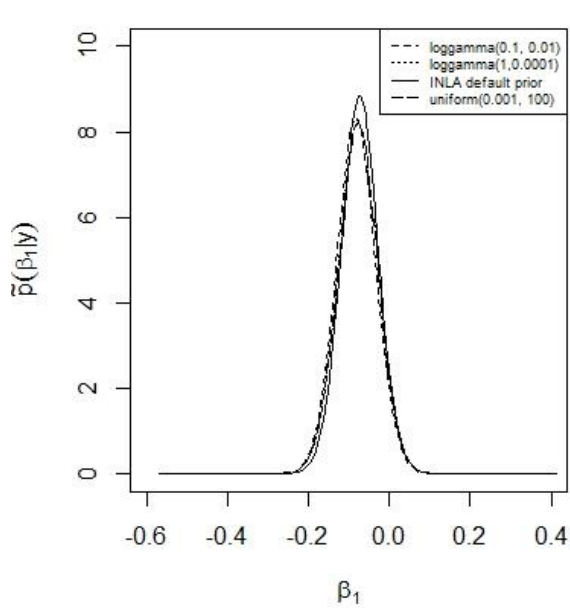
Figure 4.27: Maps of standardized incidence ratio of NCGC for 96 RHADs in Manitoba and Winnipeg for five time periods using spatio-temporal Poisson regression model; numbers in the map are the district number from 1 to 96



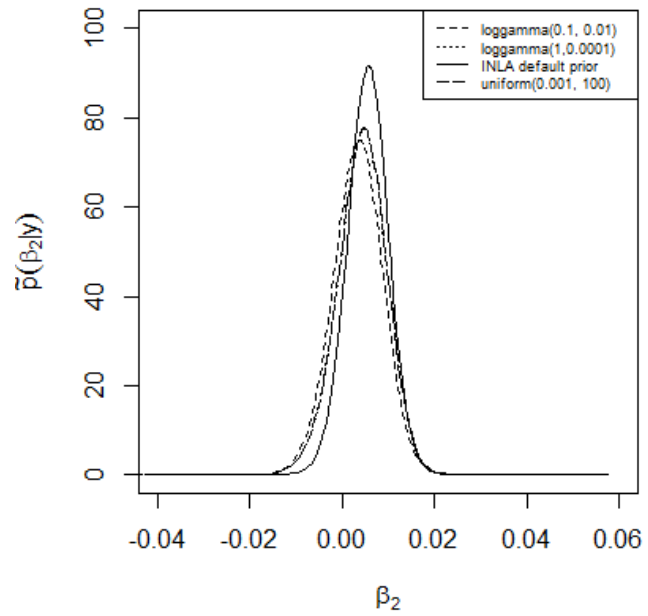
4.2.5 Prior Sensitivity Analysis Result

A sensitivity analysis was done using four types of minimally informative prior: log gamma (0.1, 0.01), log gamma (1, 0.0001), INLA default prior log gamma (1, 0.0005), and Uniform prior (0.001, 100). The result presented in Figure 4.28 (a – e) representing the sensitivity analysis for the overall model showed a consistent distribution of marginal posterior for both fixed-effect parameters and variance components. The analysis was done for all models and similar results were obtained.

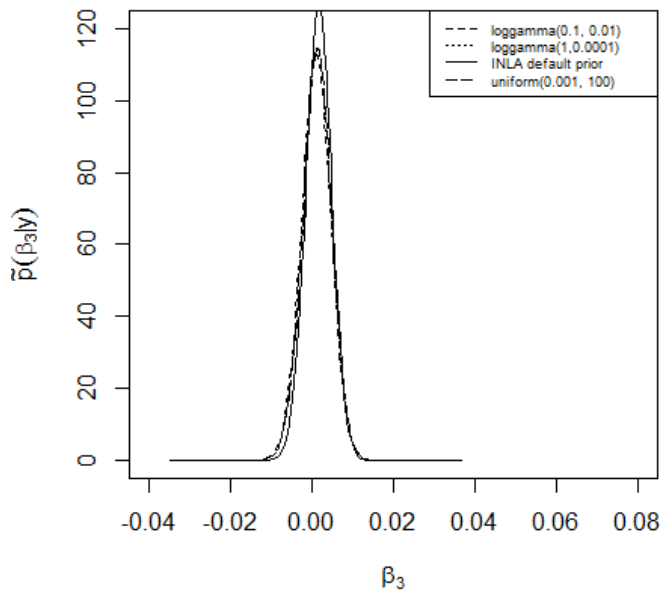
Figure 4.28: Posterior density plot for fixed effect estimates and variance of random effect estimates



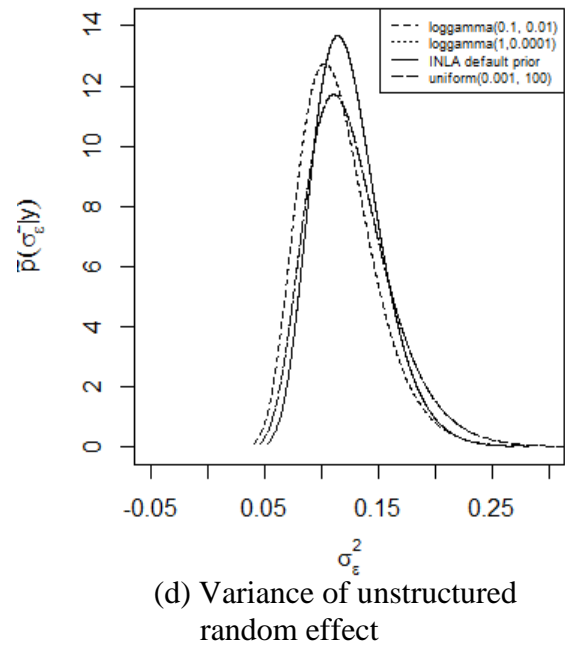
(a) SESI



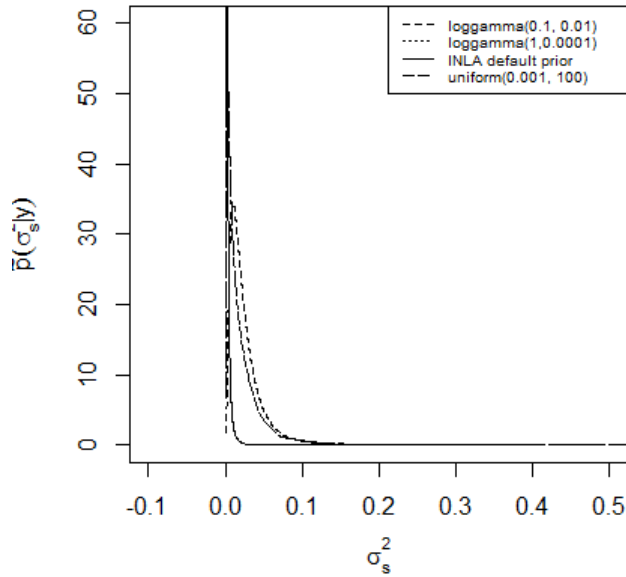
(b) Immigrant



(c) Indigenous



(d) Variance of unstructured random effect



(e) Variance of structured random effect

The result showed consistency in marginal posterior estimates as there was no significant differences in the parameter estimates generated via these four priors. Also, the marginal posterior of the spatial random effect under these priors were plotted in a map where an identical result was obtained as shown in Appendix 2.

Chapter 5: Discussion and Conclusions

5.1 Discussion

We used the Bayesian hierarchical spatial and spatio-temporal models, which were implemented in R-INLA, to describe the risk of GC in 96 RHADs in Manitoba. The result of which was broadly categorized into a spatial and spatio-temporal model is discussed below.

5.1.1 Spatial Model

Using the ecological regression model, we observed no significant relationship between SES, immigrant population proportion, and Indigenous population proportion and overall GC even when we stratified by sex. However, a marginal difference in the association of the covariates with GC risk was noticed when the overall GC was stratified by the topographical subset. A significant association between SES and CGC, and a marginally significant association between Indigenous population proportion and CGC was observed, where they both covariates decreased the risk of CGC. This result is partly supported by literature as some studies also suggest a decrease in the risk of GC among both sexes of the Indigenous people (Sebastian et al., 2004). A possible suggestion regarding this could be confounding of the Indigenous variable by the lifestyle and dietary pattern, which has been documented to also be a risk factor of GC (Tsugane et al., 1990). We observed a marginally significant association between SES and NCGS. Further partitioning of CGC by sex showed that the effect of SES on women was more than that of men, and the risk of CGC among Indigenous women was reduced by 1.9% compared to the rest of the population.

Despite all the identified RHADs with a higher risk of GC relative to the rest of the population, across all data partitions, five RHADs were identified in all data sub-groups. Little Saskatchewan and Brandon East End in Prairie Mountain RHA and Point Douglas North, Downtown East, and

St. Boniface West in WRHA all exhibited a high risk of overall GC, GC for both sexes, for CGC, and NCGC. Combining all cases of GC, the districts identified with a high risk of GC shared similar environmental characteristics such as low median annual income, a high proportion of Indigenous People, low to a medium proportion of immigrants, and a high percentage of people with no post-secondary school education. The two districts in the North, Winnipeg Churchill and district 12, had the highest risk of GC and were predominated by Indigenous people (the Chipewyan and Cree natives). These above-identified RHADs were also among the RHADs with the highest risk of GC for the men population. They were also associated with high percentage of no post-secondary school education people (35.7% - 48.8%), unemployment (6.4% - 11.1%), a low income (\$21,841 - \$43, 648), and moderate proportion of immigrants (27.3%). District 06 was identified with a high risk of overall GC for both sexes. The district had a high income (\$67,841 - \$87,680), high unemployment (4.5% - 6.3%), high uneducated people (35.7% - 40.5%), but had an average SES score index (4.2471 – 6.0063). Consistent identification of districts in Northern RHA, which were predominated by Indigenous People with a high risk of GC, is supported in the literature (Salmond et al., 1998; Sebastian & Hurtig, 2004).

5.1.2 Spatio-temporal Model

We also observed a provincial overall steady risk of GC, male GC, female GC, CGC, and NCGC over time from our smoothed temporal random effect. However, this trend did not hold for most of the RHADs in the province. The spatio-temporal result presented in Figures 4.23 – 4.27 demonstrated an apparent change in the color gradient for each map replicated over time, suggesting changes in the risk of GC across the RHADs over time. This finding enabled us to identify RHADs with an increase in GC incidence risk over time, a useful tool for disease

surveillance and or monitoring. Few of these identified RHADs (Porcupine Mountain, Pinawa, Lac Brochet, The Pas, and Winnipeg Churchill) were extracted, and the risk of GC for those RHADs over time was plotted. The plot (Figure 4.22) revealed an increasing trend as against the provincial steady risk rate. This information can be used by any organization interested in Cancer care (e.g., CancerCare Manitoba) to identify RHADs in need of urgent intervention programs.

5.2 Strengths and Limitations

One of the main strengths of this study is that it is based on population data and therefore not subject to selection bias. This study has demonstrated the use of a more confiscated method in the estimation of reliable and stable risk for area-level data by including a geographical variation of data into the model. In addition to obtaining a reliable estimate, the method simultaneously identified districts within the province with significantly different risks compared to the rest of the population.

The inclusion of time random effect helped us to acknowledge and account for time effect, thereby adjusting the variance-covariance matrix to avoid over-estimation which could lead to a wrong judgment regarding the significance of the parameter estimate and study biasedness. The time effect also helped us to investigate the trend of the disease over time which is a vital tool in disease surveillance and intervention as a recent report released by Canadian Cancer Statistics 2019 has estimated that approximately 50% of the incidence of GC can be prevented given the right tools. Though the absolute risk of GC was small, a consideration to put in place screening facility to enable early detection of GC may be considered after a thorough review of the financial implication as such a project may require a considerable budget. A possible initial solution may be the establishment of such a screening facility in districts which a consistent increase in the risk

of GC over time. Programs that can improve socioeconomic status in districts with a high risk of GC should be considered.

However, this study is limited to the fact that we were unable to adjust for several other factors such as smoking, obesity, lifestyle, *Helicobacter pylori*, and food, which are important risk factors associated with the etiology of GC. Some of these factors that were not accounted for may be confounders to some of the factors considered. Also, we used data from the 2016 census to determine a set of fixed covariates used in the spatio-temporal model, whereby a time changing covariate might explain how changes in the covariates over time influence the variation in GC over time. Another significant limitation of this study involves the grouping of unspecified GC with the non-cardia GC, which may also affect the findings relative to non-cardia GC.

5.3 Future Research

Based on the limitation of this study, we suggest further study on the investigation of risk of GC at the districts by adjusting for more factors such as lifestyle, diet and *helicobacter pylori*. These factors may serve as confounders for some of the factors considered in this study. Also, we suggest the use of time changing covariates for the spatio-temporal Poisson regression model, which could lead to a more insightful finding of how changes in factors over time influence changes in risk of GC over time. Similarly, we recommend the separation of unspecified GC from noncardia GC for a more accurate result about the variation of noncardia GC. Finally, we suggest further study in the districts that have been found to show a high risk of GC, especially those that had a consistent increase over time to identify underlying causes which are responsible for the rise in those districts.

5.4 Conclusions

In conclusion, our study has demonstrated that the incidence risk of GC varied across the province and that the overall incidence risk of GC at the provincial level was steady. However, this is not valid for all districts within the province and as such, measures should be taken to address the risk in these identified districts. The identified districts with increasing high risk of GC over time can be further studied for possible prevention. Also, we showed that area-level SES affected GC stratified by topographical sub-site. We were also able to identify 25 districts with a high risk of overall GC, 22 districts with a high risk of overall GC for men, 16 districts with a high risk of overall GC for women, 11 districts with a high risk of cardia GC, and 27 districts with a high risk of non-cardia GC. We also observed that Brandon East End district in Prairie Mountain RHA, and Point Douglas North, River East South, Downtown East, and St. Boniface West districts in Winnipeg RHA all exhibited a consistent high risk across all data subgroups. Finally, we demonstrated that spatial and spatio-temporal regression models were robust for the analysis of datasets collected over space and time, and their usages in addressing more complex data health problems are encouraged.

References

1. Abdul Rahman, N., & Abd Naeem, N. S. (2018). Construction of a Socio-Economic Status (SES) Index in Peninsular Malaysia Using the Factor Analysis Approach. *Pertanika Journal of Social Sciences & Humanities*, 26(3).
2. Aggarwal, B. B., Sung, B., & Gupta, S. C. (Eds.). (2014). *Inflammation and cancer* (Vol. 816). Springer.
3. Aird, I., Bentall, H. H., & Roberts, J. F. (1953). Relationship between cancer of stomach and the ABO blood groups. *British medical journal*, 1(4814), 799.
4. American Cancer Society: Cancer Facts and Figures 2019. Atlanta, Ga: American Cancer Society, 2019. Available online at <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2019/cancer-facts-and-figures-2019.pdf>. Accessed May 05, 2019.
5. Annual Statistics 2015 – 2016: Manitoba Health, Seniors and Active Living information management and analytics Available online at: <https://www.gov.mb.ca/health/annstats/as1617.pdf>. Accessed May 08, 2019
6. Arnold, M., Moore, S. P., Hassler, S., Ellison-Loschmann, L., Forman, D., & Bray, F. (2014). The burden of stomach cancer in indigenous populations: a systematic review and global assessment. *Gut*, 63(1), 64-71.
7. Banerjee S, Carlin BP, Gelfand AE. 2014. Hierarchical Modeling and Analysis for Spatial Data, Second Edition, Monographs on Statistics and Applied Probability (2nd ed.), Chapman and Hall/CRC, ISBN 9781439819173.
8. Barstad, B., Sørensen, T. I. A., Tjønneland, A., Johansen, D., Becker, U., Andersen, I. B., & Grønbaek, M. (2005). Intake of wine, beer and spirits and risk of gastric cancer. *European journal of cancer prevention*, 14(3), 239-243.
9. Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43(1), 1-20.
10. Blair, V., Kahokehr, A., & Sammour, T. (2013). Cancer in Māori: lessons from prostate, colorectal and gastric cancer and progress in hereditary stomach cancer in New Zealand. *ANZ journal of surgery*, 83(1-2), 42-48.
11. Boccolini, P. D. M. M., Asmus, C. I. R. F., Chrisman, J. D. R., Câmara, V. D. M., Markowitz, S. B., & Meyer, A. (2014). Stomach cancer mortality among agricultural workers: results from a death certificate-based case-control study. *Cadernos Saúde Coletiva*, 22(1), 86-92.
12. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, in press. The online GLOBOCAN 2018 database is accessible at <http://gco.iarc.fr/>
13. Brooks, S. P., and A. Gelman. (1997). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics* 7: 434–455.

-
14. Brown LM, Silverman DT, Pottern LM et al (1994). Adenocarcinoma of the esophagus and esophagogastric junction in white men in the United States: alcohol, tobacco, and socioeconomic factors. *Cancer Causes Control* 5:333–40.
 15. Brownell, M., Lix, L., Ekuma, O., Derksen, S., De Haney, S., et. al., *Why is the Health Status of Some Manitobans Not Improving? The Widening Gap in the Health Status of Manitobans*. Winnipeg, Manitoba: Manitoba Centre for Health Policy, 2003.
 16. Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics 2017. Toronto, ON: Canadian Cancer Society; 2017. Available at: <http://www.cancer.ca/~media/cancer.ca/CW/cancer%20information/cancer%20101/Canadian%20cancer%20statistics/Canadian-Cancer-Statistics-2017-EN.pdf> Accessed May 08, 2019
 17. Canadian Cancer Statistics Advisory Committee. Canadian Cancer Statistics 2018. Toronto, ON: Canadian Cancer Society; 2018. Available at: [cancer.ca/Canadian-Cancer-Statistics-2018-EN](http://www.cancer.ca/Canadian-Cancer-Statistics-2018-EN) (accessed December 23, 2018).
 18. Carlin BP, Xia H, 1999a. Assessing environmental justice using Bayesian hierarchical models: two case studies. *Journal of Exposure Analysis & Environmental Epidemiology*, 9:66-78.
 19. Carrasco, G., & Corvalan, A. H. (2013). Helicobacter pylori-induced chronic gastritis and assessing risks for gastric cancer. *Gastroenterology research and practice*, 2013.
 20. Chiu, H. F., Tsai, S. S., Chen, P. S., Liao, Y. H., Liou, S. H., Wu, T. N., & Yang, C. Y. (2011). Traffic air pollution and risk of death from gastric cancer in Taiwan: petrol station density as an indicator of air pollutant exposure. *Journal of Toxicology and Environmental Health, Part A*, 74(18), 1215-1224.
 21. Clayton, D. G. (1996). Generalized linear mixed models. *Markov chain Monte Carlo in practice*, 1, 275-302.
 22. Colquhoun, A., Hannah, H., Corriveau, A., Hanley, B., Yuan, Y., Goodman, K. J., & CANHelp Working Group. (2019). Gastric Cancer in Northern Canadian Populations: A Focus on Cardia and Non-Cardia Subsites. *Cancers*, 11(4), 534.
 23. D'Elia, L., Galletti, F., & Strazzullo, P. (2014). Dietary salt intake and risk of gastric cancer. In *Advances in Nutrition and Cancer* (pp. 83-95). Springer, Berlin, Heidelberg.
 24. Davies, I. J. T. (2013). *The clinical significance of the essential biological metals*. Butterworth-Heinemann.
 25. Dong, E., Duan, L., & Wu, B. U. (2017). Racial and ethnic minorities at increased risk for gastric cancer in a regional US population study. *Clinical Gastroenterology and Hepatology*, 15(4), 511-517.
 26. Edgren, G., Hjalgrim, H., Rostgaard, K., Norda, R., Wikman, A., Melbye, M., & Nyrén, O. (2010). Risk of gastric cancer and peptic ulcers in relation to ABO blood type: a cohort study. *American journal of epidemiology*, 172(11), 1280-1285.

-
27. Faggiano F, Zanetti R, Costa G (1994). Cancer risk and social inequalities in Italy. *J Epidemiol Community Health* 48:447–52.
 28. Ferraroni M, Negri E, La Vecchia C, D'Avanzo B, Franceschi S (1989). Socioeconomic indicators, tobacco and alcohol in the aetiology of digestive tract neoplasms. *Int J Epidemiol* 18:556–62.
 29. Ferreccio, C., Yuan, Y., Calle, J., Benítez, H., Parra, R. L., Acevedo, J., ... & Steinmaus, C. (2013). Arsenic, tobacco smoke, and occupation: associations of multiple agents with lung and bladder cancer. *Epidemiology (Cambridge, Mass.)*, 24(6), 898.
 30. Fitzmaurice, C., Dicker, D., Pain, A., Hamavid, H., Moradi-Lakeh, M., MacIntyre, M. F., ... & Hamadeh, R. R. (2015). The global burden of cancer 2013. *JAMA oncology*, 1(4), 505-527.
 31. Fox, J. G., Dangler, C. A., Taylor, N. S., King, A., Koh, T. J., & Wang, T. C. (1999). High-salt diet induces gastric epithelial hyperplasia and parietal cell loss, and enhances *Helicobacter pylori* colonization in C57BL/6 mice. *Cancer research*, 59(19), 4823-4828.
 32. Fransoo, R., Martens, P., The Need to Know Team, Burland, E, Prior, H., Burchill, C., Chateau, D., Walld, R., Sex Differences in Health Status, Health Care Use, and Quality of Care: A Population-Based Analysis for Manitoba's Regional Health Authorities. Winnipeg, Manitoba Centre for Health Policy, November, 2005.
 33. Fritz P, Percy C, Jack A, Shanmugaratnuers K, Solin L, Parkin D: International classification of diseases for oncology. 3 edition. Geneva: World Health Organization; 2000.
 34. Gelman, Andrew, and Donald B. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 457-472
 35. Go, M. F. (2002). Review article: natural history and epidemiology of. *Helicobacter pylori*, 3-15.
 36. González, C. A., & Agudo, A. (2012). Carcinogenesis, prevention and early detection of gastric cancer: where we are and where we should go. *International journal of cancer*, 130(4), 745-753.
 37. González, C. A., Pera, G., Agudo, A., Bueno-de-Mesquita, H. B., Ceroti, M., Boeing, H., ... & Berrino, F. (2006). Fruit and vegetable intake and the risk of stomach and oesophagus adenocarcinoma in the European Prospective Investigation into Cancer and Nutrition (EPIC–EURGAST). *International journal of cancer*, 118(10), 2559-2566.
 38. González, C. A., Pera, G., Agudo, A., Palli, D., Krogh, V., Vineis, P., ... & Nyrén, O. (2003). Smoking and the risk of gastric cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *International journal of cancer*, 107(4), 629-634.
 39. Haenszel, W. (1958). Variation in incidence of and mortality from stomach cancer, with particular reference to the United States. *Journal of the National Cancer Institute*, 21(2), 213-262.

-
40. Haenszel, W., & Kurihara, M. (1968). Studies of Japanese migrants. I. Mortality from cancer and other diseases among Japanese in the United States. *Journal of the National Cancer Institute*, 40(1), 43-68.
 41. Haenszel, W., Kurihara, M., Segi, M., & Lee, R. K. (1972). Stomach cancer among Japanese in Hawaii. *Journal of the National Cancer Institute*, 49(4), 969-988.
 42. Holt J, and Lo C (2008). The geography of mortality in the Atlanta metropolitan area. *Computers, Environment and Urban Systems*, 32(2), 149 – 164.
 43. Hoskins, L. C., Loux, H. A., Britten, A., & Zamcheck, N. (1965). Distribution of ABO blood groups in patients with pernicious anemia, gastric carcinoma and gastric carcinoma associated with pernicious anemia. *New England Journal of Medicine*, 273(12), 633-637.
 44. Indigenous inclusion directorate. Indigenous identity declaration: A guide for parents and guardians available at <https://www.edu.gov.mb.ca/aed/aid/brochure.pdf> Accessed on June 19,2019
 45. International Agency for Research on Cancer: world Health Organization. *Cancer today*. Available at: <http://gco.iarc.fr/today/online-analysis-table?> (Accessed March 4, 2019)
 46. Jedrychowski, W., Wahrendorf, J., Popiela, T., & Rachtan, J. (1986). A case-control study of dietary factors and stomach cancer risk in Poland. *International journal of cancer*, 37(6), 837-842.
 47. John M. Bumsted, Peter McLintock and Others: Manitoba. *Encyclopedia Britannica, inc*, 2019. Available at <https://www.britannica.com/place/Manitoba> Accessed on May 13, 2019.
 48. Joossens, J. V., Hill, M. J., Elliott, P., Stamler, R., Stamler, J., Lesaffre, E., ... & EUROPEAN CANCER PREVENTION (ECP) AND THE INTERSALT COOPERATIVE RESEARCH GROUP. (1996). Dietary salt, nitrate and stomach cancer mortality in 24 countries. *International journal of epidemiology*, 25(3), 494-504.
 49. Kamineni, A., Williams, M. A., Schwartz, S. M., Cook, L. S., & Weiss, N. S. (1999). The incidence of gastric carcinoma in Asian migrants to the United States and their descendants. *Cancer causes & control*, 10(1), 77-83.
 50. Klein, P. D., Opekun, A. R., Smith, E. O., Graham, D. Y., Gaillour, A., & Gastrointestinal Physiology Working Group. (1991). Water source as risk factor for Helicobacter pylori infection in Peruvian children. *The Lancet*, 337(8756), 1503-1506.
 51. Kneller, R. W., Guo, W. D., Hsing, A. W., Chen, J. S., Blot, W. J., Li, J. Y., ... & Fraumeni, J. F. (1992). Risk factors for stomach cancer in sixty-five Chinese counties. *Cancer Epidemiology and Prevention Biomarkers*, 1(2), 113-118.
 52. Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in medicine*, 19(17-18), 2555-2567.
 53. Knorr-Held, L., & Richardson, S. (2003). A hierarchical model for space–time surveillance data on meningococcal disease incidence. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(2), 169-183.

-
54. Kono, S., & Hirohata, T. (1996). Nutrition and stomach cancer. *Cancer Causes & Control*, 7(1), 41-55.
 55. Krishnan V (2010). Constructing an area-based socioeconomic status index: A principal component analysis approach. *Early Childhood Council Annual Conference 2010, 2010*, May 7, Christchurch, New Zealand.
 56. Ladeiras-Lopes, R., Pereira, A. K., Nogueira, A., Pinheiro-Torres, T., Pinto, I., Santos-Pereira, R., & Lunet, N. (2008). Smoking and gastric cancer: systematic review and meta-analysis of cohort studies. *Cancer causes & control*, 19(7), 689-701.
 57. Larsson, S. C., Bergkvist, L., & Wolk, A. (2006). Fruit and vegetable consumption and incidence of gastric cancer: a prospective study. *Cancer Epidemiology and Prevention Biomarkers*, 15(10), 1998-2001.
 58. Lauby-Secretan, B., Scoccianti, C., Loomis, D., Grosse, Y., Bianchini, F., & Straif, K. (2016). Body fatness and cancer—viewpoint of the IARC Working Group. *New England Journal of Medicine*, 375(8), 794-798.
 59. Lawson AB. 2013. Bayesian disease mapping: hierarchical modeling in spatial epidemiology. Chapman and Hall/CRC press.
 60. Lee, J., Demissie, K., Lu, S. E., & Rhoads, G. G. (2007). Cancer incidence among Korean-American immigrants in the United States and native Koreans in South Korea. *Cancer Control*, 14(1), 78-85.
 61. Lee, S. A., Kang, D., Shim, K., Choe, J., Hong, W., & Choi, H. (2003). Original article effect of diet and Helicobacter pylori infection to the risk of early gastric cancer. *Journal of epidemiology*, 13(3), 162-168.
 62. Liao, L. M., Hofmann, J. N., Kamangar, F., Strickland, P. T., Ji, B. T., Yang, G., ... & Gao, Y. T. (2014). Polycyclic aromatic hydrocarbons and risk of gastric cancer in the Shanghai Women's Health Study. *International journal of molecular epidemiology and genetics*, 5(3), 140.
 63. Liu, C., & Russell, R. M. (2008). Nutrition and gastric cancer risk: an update. *Nutrition reviews*, 66(5), 237-249.
 64. Lunet, N., Lacerda-Vieira, A., & Barros, H. (2005). Fruit and vegetables consumption and gastric cancer: a systematic review and meta-analysis of cohort studies. *Nutrition and cancer*, 53(1), 1-10.
 65. Lunet, N., Valbuena, C., Vieira, A. L., Lopes, C., Lopes, C., David, L., ... & Barros, H. (2007). Fruit and vegetable consumption and gastric cancer by location and histological type: case-control and meta-analysis. *European Journal of Cancer Prevention*, 16(4), 312-327.
 66. MA, M., Pera, G., Agudo, A., Bueno-de-Mesquita, H. B., Palli, D., Boeing, H., ... & Panico, S. (2007). Cereal fiber intake may reduce risk of gastric adenocarcinomas: The EPIC-EURGAST study. *International journal of cancer*, 121(7), 1618-1623.

-
67. Machida-Montani, A., Sasazuki, S., Inoue, M., Natsukawa, S., Shaura, K., Koizumi, Y., ... & Tsugane, S. (2004). Association of *Helicobacter pylori* infection and environmental factors in non-cardia gastric cancer in Japan. *Gastric Cancer*, 7(1), 46-53.
 68. Maev, I. V., Kucheriavyĭ, I., Andreev, D. N., & Barkalova, E. V. (2014). Eradication therapy for *Helicobacter pylori* infection: review of world trends. *Terapevticheskii arkhiv*, 86(3), 94-99.
 69. Mahar, A. L., Coburn, N. G., Kagedan, D. J., Viola, R., & Johnson, A. P. (2016). Regional variation in the management of metastatic gastric cancer in Ontario. *Current Oncology*, 23(4), 250.
 70. Martens, P., Fransoo, R., et. al., The Manitoba RHA Indicators Atlas: Population-Based Comparisons of Health and Health Care Use. Winnipeg, Manitoba: Manitoba Centre for Health Policy, 2003.
 71. Messer, L., Laraia, B., Kaufman, J., Eyser, J., Holzman, C., Culhane, J., Elo, I., Burke, J., and O'Campo, P. (2006). The development of a standardized neighborhood deprivation index. *Journal of Urban Health*, Volume 83, number 6 (1041-1062).
 72. Mills, P. K., & Yang, R. C. (2007). Agricultural exposures and gastric cancer risk in Hispanic farm workers in California. *Environmental research*, 104(2), 282-289.
 73. Mohebbi, M., Wolfe, R., Jolley, D., Forbes, A. B., Mahmoodi, M., & Burton, R. C. (2011). The spatial distribution of esophageal and gastric cancer in Caspian region of Iran: An ecological analysis of diet and socio-economic influences. *International journal of health geographics*, 10(1), 13.
 74. Møller, M. E., Dahl, R., & Bøckman, O. C. (1988). A possible role of the dietary fibre product, wheat bran, as a nitrite scavenger. *Food and chemical toxicology*, 26(10), 841-845.
 75. Moore, S. P., Antoni, S., Colquhoun, A., Healy, B., Ellison-Loschmann, L., Potter, J. D., ... & Bray, F. (2015). Cancer incidence in indigenous people in Australia, New Zealand, Canada, and the USA: a comparative population-based study. *The Lancet Oncology*, 16(15), 1483-1492.
 76. Mousavi, S. M., Sundquist, J., & Hemminki, K. (2011). Does immigration play a role in the risk of gastric cancer by site and by histological type? A study of first-generation immigrants in Sweden. *Gastric cancer*, 14(3), 285-289.
 77. Munafò, M. R., Timofeeva, M. N., Morris, R. W., Prieto-Merino, D., Sattar, N., Brennan, P., ... & Whincup, P. H. (2012). Association between genetic variants on chromosome 15q25 locus and objective measures of tobacco exposure. *Journal of the National Cancer Institute*, 104(10), 740-748.
 78. Nagel, G., Linseisen, J., Boshuizen, H. C., Pera, G., Del Giudice, G., Westert, G. P., ... & Peeters, P. H. (2007). Socioeconomic position and the risk of gastric and oesophageal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC-EURGAST). *International journal of epidemiology*, 36(1), 66-76.

-
79. Nishino, Y., Inoue, M., Tsuji, I., Wakai, K., Nagata, C., Mizoue, T., ... & Tsugane, S. (2006). Tobacco smoking and gastric cancer risk: an evaluation based on a systematic review of epidemiologic evidence among the Japanese population. *Japanese journal of clinical oncology*, 36(12), 800-807.
 80. Oiso, T. (1975). Incidence of stomach cancer and its relation to dietary habits and nutrition in Japan between 1900 and 1975. *Cancer research*, 35(11 Part 2), 3254-3258.
 81. Oliveira, C., Pinheiro, H., Figueiredo, J., Seruca, R., & Carneiro, F. (2015). Familial gastric cancer: genetic susceptibility, pathology, and implications for management. *The Lancet Oncology*, 16(2), e60-e70.
 82. Olsen, C. M., Pandeya, N., Green, A. C., Webb, P. M., Whiteman, D. C., & Australian Cancer Study. (2011). Population attributable fractions of adenocarcinoma of the esophagus and gastroesophageal junction. *American journal of epidemiology*, 174(5), 582-590.
 83. Palli D, Bianchi S, Decarli A et al (1992). A case-control study of cancers of the gastric cardia in Italy. *Br J Cancer* 65:263–66.
 84. Park, B., Shin, A., Park, S. K., Ko, K. P., Ma, S. H., Lee, E. H., ... & Yoo, K. Y. (2011). Ecological study for refrigerator use, salt, vegetable, and fruit intakes, and gastric cancer. *Cancer Causes & Control*, 22(11), 1497.
 85. Parkin, D. M., Whelan, S. L., Ferlay, J., Teppo, L., & Thomas, D. B. (2002). Cancer Incidence in Five Continents Vol VIII: IARC Scientific Publication No.
 86. Pearce, N., Susser, M., & Boffetta, P. (1997). Social inequalities and cancer.
 87. Polinkevych, B. S., & Pikas, P. B. (2014). The role of Helicobacter pylori in occurrence of the digestive system pathologic processes. *Klinichna khirurgiia*, (6), 66.
 88. Pukkala E, Teppo L (1986). Socioeconomic status and education as risk determinants of gastrointestinal cancer. *Prev Med* 15:127–38.
 89. Raj, A., Mayberry, J. F., & Podas, T. (2003). Occupation and gastric cancer. *Postgraduate Medical Journal*, 79(931), 252-258.
 90. Ren, J. S., Kamangar, F., Forman, D., & Islami, F. (2012). Pickled food and risk of gastric cancer—a systematic review and meta-analysis of English and Chinese literature. *Cancer Epidemiology and Prevention Biomarkers*, 21(6), 905-915.
 91. Riboli, E., & Norat, T. (2003). Epidemiologic evidence of the protective effect of fruit and vegetables on cancer risk. *The American journal of clinical nutrition*, 78(3), 559S-569S.
 92. Roder, D., & Currow, D. (2009). Cancer in aboriginal and Torres Strait Islander people of Australia. *Asian Pac J Cancer Prev*, 10(5), 729-733.
 93. Roos, N., Shapiro, E., Bond, R., Black, C., Finlayson, G., et. al., Changes in Health and Health Care Use of Manitobans: 1985-1998. Winnipeg, Manitoba: Manitoba Centre for Health Policy, 2001.

-
94. Salmond C, Crampton P, and Sutton F (1998). NZDep91: A New Zealand index of deprivation. *Australia and New Zealand Journal of Public Health*, 22(7), 825 – 837.
 95. Salvador, I., Mercado, A., Bravo, G. L., Baldeón, M., & Fornasini, M. (2015). Risk and protective factors for gastric metaplasia and cancer: a hospital-based case-control study in Ecuador. *Nutrición hospitalaria*, 32(3), 1193-1199.
 96. Sebastián, M. S., & Hurtig, A. K. (2004). Cancer among indigenous people in the Amazon Basin of Ecuador, 1985-2000. *Revista Panamericana de Salud Pública*, 16, 328-333.
 97. SEER Cancer statistics review: Stomach Cancer , available at <https://seer.cancer.gov/statfacts/html/stomach.html> assessed on July, 28 2019.
 98. Sharafi, Z., Asmarian, N., Hoorang, S., & Mousavi, A. (2018). Bayesian spatio-temporal analysis of stomach cancer incidence in Iran, 2003–2010. *Stochastic environmental research and risk assessment*, 32(10), 2943-2950.
 99. Shikata, K., Kiyohara, Y., Kubo, M., Yonemoto, K., Ninomiya, T., Shiota, T., ... & Matsumoto, T. (2006). A prospective study of dietary salt intake and gastric cancer incidence in a defined Japanese population: the Hisayama study. *International journal of cancer*, 119(1), 196-201.
 100. Statistics Canada. 2019. *Canadian Cancer registry*. <https://www23-statcan-gc-ca.uml.idm.oclc.org/imdb/p2SV.pl?Function=getSurvey&Id=1211637> (accessed on December 18).
 101. Sui, D. Z. (2004). Tobler's first law of geography: A big idea for a small world? *Annals of the Association of American Geographers*, 94(2), 269-277.
 102. Tang, H., Wei, P., Duell, E. J., Risch, H. A., Olson, S. H., Bueno-de-Mesquita, H. B., ... & McWilliams, R. R. (2014). Axonal guidance signaling pathway interacting with smoking in modifying the risk of pancreatic cancer: a gene-and pathway-based interaction analysis of GWAS data. *Carcinogenesis*, 35(5), 1039-1045.
 103. Tobler WR. 1970. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* 46: 234 -240.
 104. Torabi, M (2014). Bowel disorders and its spatial trend in Manitoba, Canada. *BMC Public Health*. 14:285
 105. Torabi, M, Green, C., Nugent, Z, et. al., (2014). Geographical variation and factors associated with colorectal cancer mortality in a universal health care system. *Can J Gastroenterol Hepatol*. 28(4), 191–197.
 106. Torabi, M., Bernstein, C. N., Yu, B. N., Wickramasinghe, L., Blanchard, J. F., & Singh, H. (2019). Geographical variation and factors associated with inflammatory bowel disease in a central Canadian province. *Inflammatory Bowel Diseases*.

-
107. Tramacere, I., Pelucchi, C., Bagnardi, V., Rota, M., Scotti, L., Islami, F., ... & Negri, E. (2011). A meta-analysis on alcohol drinking and esophageal and gastric cardia adenocarcinoma risk. *Annals of oncology*, 23(2), 287-297.
 108. Trédaniel, J., Boffetta, P., Buiatti, E., Saracci, R., & Hirsch, A. (1997). Tobacco smoking and gastric cancer: review and meta-analysis. *International journal of cancer*, 72(4), 565-573.
 109. Trott, P. A. (1977). International classification of diseases for oncology. *Journal of clinical pathology*, 30(8), 782.
 110. Tsugane, S., & Sasazuki, S. (2007). Diet and the risk of gastric cancer: review of epidemiological evidence. *Gastric cancer*, 10(2), 75-83.
 111. Tsugane, S., de Souza, J. M., Costa, M. L., Mirra, A. P., Gotlieb, S. L., Laurenti, R., & Watanabe, S. (1990). Cancer incidence rates among Japanese immigrants in the city of Sao Paulo, Brazil, 1969–78. *Cancer Causes & Control*, 1(2), 189-193.
 112. Tsugane, S., de Souza, J. M., Costa, M. L., Mirra, A. P., Gotlieb, S. L., Laurenti, R., & Watanabe, S. (1990). Cancer incidence rates among Japanese immigrants in the city of Sao Paulo, Brazil, 1969–78. *Cancer Causes & Control*, 1(2), 189-193.
 113. Tsugane, S., Tei, Y., Takahashi, T., Watanabe, S., & Sugano, K. (1994). Salty food intake and risk of Helicobacter pylori infection. *Japanese journal of cancer research*, 85(5), 474-478.
 114. Turati, F., Tramacere, I., La Vecchia, C., & Negri, E. (2012). A meta-analysis of body mass index and esophageal and gastric cardia adenocarcinoma. *Annals of oncology*, 24(3), 609-617.
 115. Umesawa, M., Iso, H., Fujino, Y., Kikuchi, S., & Tamakoshi, A. (2016). Salty food preference and intake and risk of gastric cancer: the JACC study. *Journal of epidemiology*, 26(2), 92-97.
 116. van der Post, R. S., Vogelaar, I. P., Carneiro, F., Guilford, P., Huntsman, D., Hoogerbrugge, N., ... & Bardram, L. (2015). Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. *Journal of medical genetics*, 52(6), 361-374.
 117. van Loon AJ, Goldbohm RA, van den Brandt PA (1998). Socioeconomic status and stomach cancer incidence in men: results from The Netherlands Cohort Study. *J Epidemiol Community Health* 52:166–71.
 118. Vincent, Kyle, and Jason M. Sutherland. "A review of methods for deriving an index for socioeconomic status in British Columbia." *Vancouver, Canada: University of British Columbia, Centre for Health Services and Policy Research. Available at: <http://healthcarefunding.ca/files/2013/04/Review-of-Methods-for-SES-Index-for-BC.pdf>* (2013). Pampalon, R., and Raymon, G. (2000). A deprivation index for health and welfare planning in Quebec. *Chronic Diseases in Canada*, Volume 21, number 3 (104-113).

-
119. Wang, S., Freedman, N. D., Loftfield, E., Hua, X., & Abnet, C. C. (2018). Alcohol consumption and risk of gastric cardia adenocarcinoma and gastric noncardia adenocarcinoma: A 16-year prospective analysis from the NIH-AARP diet and health cohort. *International journal of cancer*, *143*(11), 2749-2757.
 120. Wang, X. Q., Yan, H., Terry, P. D., Wang, J. S., Cheng, L., Wu, W. A., & Hu, S. K. (2012). Interaction between dietary factors and Helicobacter pylori infection in noncardia gastric cancer: a population-based case-control study in China. *Journal of the American College of Nutrition*, *31*(5), 375-384.
 121. Washington, D. C. (2007). World Cancer Research Fund/American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. AICR.
 122. Watanabe, M., Kato, J., Inoue, I., Yoshimura, N., Yoshida, T., Mukoubayashi, C., ... & Iguchi, M. (2012). Development of gastric cancer in nonatrophic stomach with highly active inflammation identified by serum levels of pepsinogen and Helicobacter pylori antibody together with endoscopic rugal hyperplastic gastritis. *International journal of cancer*, *131*(11), 2632-2642.
 123. Watari, J., Chen, N., Amenta, P. S., Fukui, H., Oshima, T., Tomita, T., ... & Das, K. M. (2014). Helicobacter pylori associated chronic gastritis, clinical syndromes, precancerous lesions, and pathogenesis of gastric cancer development. *World Journal of Gastroenterology: WJG*, *20*(18), 5461.
 124. Welling, R., Beaumont, J. J., Petersen, S. J., Alexeeff, G. V., & Steinmaus, C. (2015). Chromium VI and stomach cancer: a meta-analysis of the current epidemiological evidence. *Occup Environ Med*, *72*(2), 151-159.
 125. World Health Organization. International Agency for Research on Cancer. Available at https://www.iarc.fr/wp-content/uploads/2018/09/pr263_E.pdf. Accessed: march 03, 2019.
 126. Wu, C. C., Hsu, T. W., Chang, C. M., Yu, C. H., Wang, Y. F., & Lee, C. C. (2014). The effect of individual and neighborhood socioeconomic status on gastric cancer survival. *PloS one*, *9*(2), e89655.
 127. Wu, Y., Fan, Y., Jiang, Y., Wang, Y., Liu, H., & Wei, M. (2013). Analysis of risk factors associated with precancerous lesion of gastric cancer in patients from eastern China: a comparative study. *Journal of cancer research and therapeutics*, *9*(2), 205.
 128. Yang, P., Zhou, Y., Chen, B., Wan, H. W., Jia, G. Q., Bai, H. L., & Wu, X. T. (2009). Overweight, obesity and gastric cancer risk: results from a meta-analysis of cohort studies. *European journal of cancer*, *45*(16), 2867-2873.
 129. You, W. C., Zhang, L., Gail, M. H., Chang, Y. S., Liu, W. D., Ma, J. L., ... & Blaser, M. J. (2000). Gastric dysplasia and gastric cancer: Helicobacter pylori, serum vitamin C, and other risk factors. *Journal of the National Cancer Institute*, *92*(19), 1607-1612.

-
130. Zamani, N., Hajifaraji, M., Malekshah, A. T., Keshtkar, A. A., Esmailzadeh, A., & Malekzadeh, R. (2013). A case-control study of the relationship between gastric cancer and meat consumption in Iran. *Archives of Iranian medicine*, *16*(6), 324-329.
 131. Zhang, L., Kim, S., Ding, W., Tong, Y., Zhang, X., Pan, M., & Chen, S. (2015). Arsenic sulfide inhibits cell migration and invasion of gastric cancer in vitro and in vivo. *Drug design, development and therapy*, *9*, 5579.
 132. Zhang, Z., Xu, G., Ma, M., Yang, J., & Liu, X. (2013). Dietary fiber intake reduces risk for gastric cancer: a meta-analysis. *Gastroenterology*, *145*(1), 113-120

Appendices

Appendix I: Deviance information criteria (DIC) for the all fitted models

| Model code | Overall model | Male strata | Female strata | Cardia strata | Non-cardia strata |
|---------------|----------------|----------------|----------------|----------------|-------------------|
| 1. Temporal 1 | 2531.33 | 2244.43 | 1607.8 | 2531.33 | 2251.36 |
| 2. Temporal 2 | 2530.23 | 2530.23 | 1608.8 | 2530.23 | 2251.20 |
| 3. Temporal 3 | 2531.85 | 2531.85 | 1607.95 | 2531.85 | 2251.72 |
| 4. Mod.st2a | 2233.31 | 2008.84 | 1575.58 | 1557.06 | 2005.52 |
| 5. Mod.st2b | 2234.52 | 2008.76 | 1576.12 | 1555.25 | 2006.16 |
| 6. Mod.st2c | 2233.30 | 2008.59 | 1574.96 | 1556.27 | 2004.96 |
| 7. Mod.st3a | 2202.78 | 1993.01 | 1568.74 | 1565.07 | 1987.02 |
| 8. Mod.st3b | 2203.22 | 1993.42 | 1569.37 | 1564.61 | 1987.63 |
| 9. Mod.st3c | 2202.86 | 1993.07 | 1568.9 | 1565.19 | 1987.12 |
| 10. Mod.st4a | 2233.95 | 2032.83 | 1640.68 | 1632.37 | 2031.43 |
| 11. Mod.st4b | 2233.94 | 2032.87 | 1640.7 | 1632.18 | 2031.43 |
| 12. Mod.st4c | 2234.09 | 2032.95 | 1641.46 | 1632.56 | 2031.57 |
| 13. Mod.st5a | 2208.57 | 1992.44 | 1559.34 | 1547.31 | 1988.16 |
| 14. Mod.st5b | 2208.04 | 1992.59 | 1560.27 | 1546.99 | 1989.03 |
| 15. Mod.st5c | 2208.38 | 1992.74 | 1559.08 | 1548.13 | 1988.12 |
| 16. Mod.st6a | 2257.42 | 2045.57 | 1589.76 | 1553.0 | 2033.56 |
| 17. Mod.st6b | 2255.22 | 2046.90 | 1589.76 | 1553.59 | 2032.51 |
| 18. Mod.st6c | 2256.34 | 2046.49 | 1590.23 | 1554.39 | 2033.18 |

- **Temporal 1:** RW (1) smoother for structured temporal effect with no spatio-temporal interaction
- **Temporal 2:** RW (2) smoother for structured temporal effect with no spatio-temporal interaction
- **Temporal 3:** AR (1) smoother for structured temporal effect with no spatio-temporal interaction
- **Mod.st2a:** Unstructured spatial and temporal random interaction effects with RW (1) for temporal structured effect
- **Mod.st2b:** Unstructured spatial and temporal random interaction effects with RW (2) for temporal structured effect
- **Mod.st2c:** Unstructured spatial and temporal random interaction effects with AR (1) for temporal structured effect

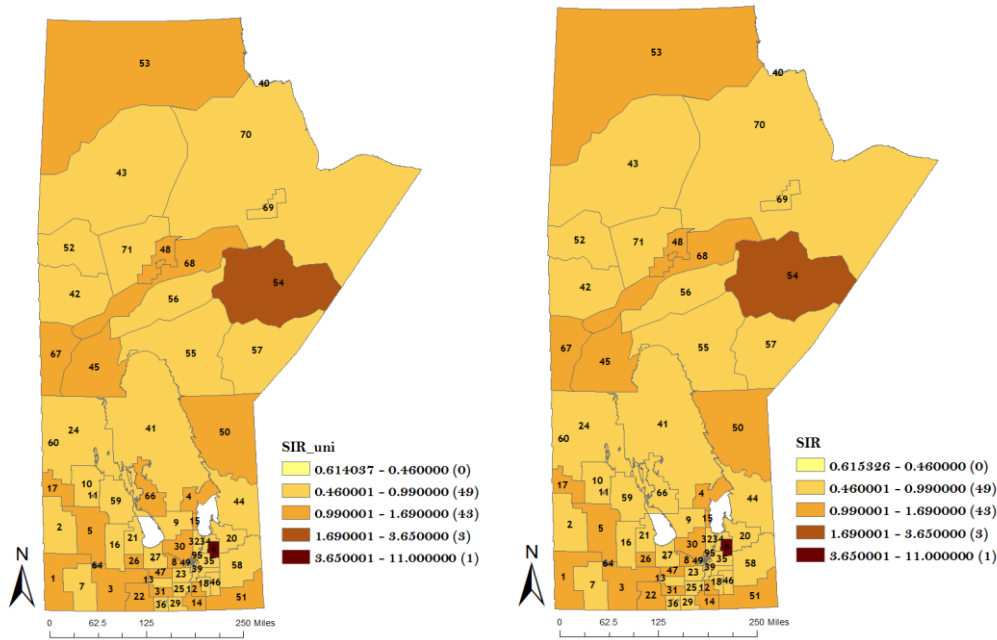
-
- **Mod.st3a:** RW (1) temporal random and unstructured spatial random effects in the interaction effect with RW (1) for temporal structured effect
 - **Mod.st3b:** RW (1) temporal random and unstructured spatial random effects in the interaction effect with RW (2) for temporal structured effect
 - **Mod.st3c:** RW (1) temporal random and unstructured spatial random effects in the interaction effect with AR (1) for temporal structured effect
 - **Mod.st4a:** RW (2) temporal random and unstructured spatial random effects in the interaction effect with RW (1) for temporal structured effect
 - **Mod.st4b:** RW (2) temporal random and unstructured spatial random effects in the interaction effect with RW (2) for temporal structured effect
 - **Mod.st4a:** RW (2) temporal random and unstructured spatial random effects in the interaction effect with AR (1) for temporal structured effect
 - **Mod.st5a:** AR (1) temporal random and unstructured spatial random effects in the interaction effect with RW (1) for temporal structured effect
 - **Mod.st5b:** AR (1) temporal random and unstructured spatial random effects in the interaction effect with RW (2) for temporal structured effect
 - **Mod.st5c:** AR (1) temporal random and unstructured spatial random effects in the interaction effect with AR (1) for temporal structured effect
 - **Mod.st6a:** Unstructured temporal random and conditional autoregressive (CAR) spatial random effect in the interaction effect with RW (1) for temporal structured effect
 - **Mod.st6b:** Unstructured temporal random and conditional autoregressive (CAR) spatial random effect in the interaction effect with RW (2) for temporal structured effect

- **Mod.st6c:** Unstructured temporal random and conditional autoregressive (CAR) spatial random effect in the interaction effect with AR (1) for temporal structured effect

Appendix II: Deviance information criteria (DIC) for Poisson and spatial regression model

| Model | Poisson model | Spatial model |
|------------------|---------------|---------------|
| Overall GC | 817.35 | 645.48 |
| Male GC | 698.92 | 597.90 |
| Female GC | 554.14 | 519.36 |
| Cardia GC | 492.84 | 478.06 |
| Noncardia GC | 748.77 | 602.59 |
| Cardia male | 465.67 | 452.10 |
| Cardia female | 265.15 | 261.39 |
| Noncardia male | 622.39 | 538.97 |
| Noncardia female | 534.37 | 491.91 |

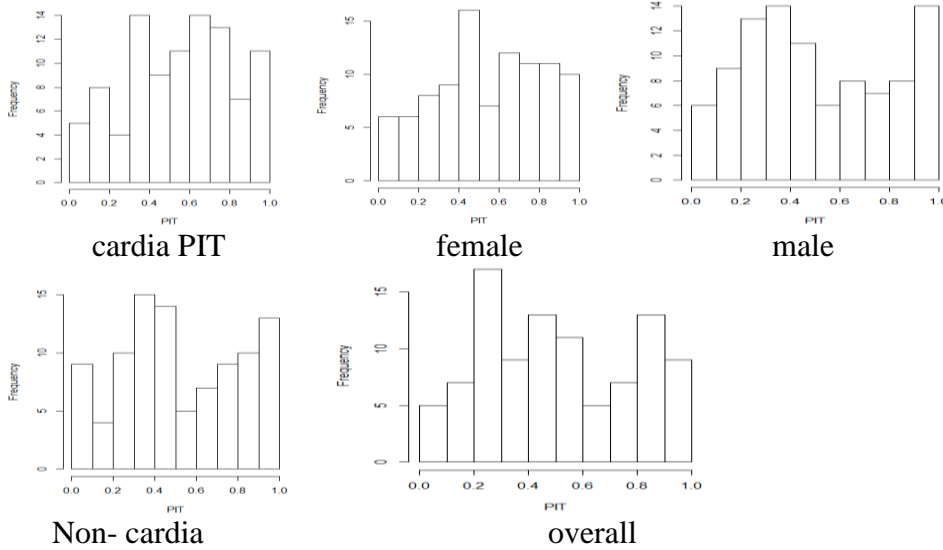
Appendix III: Map of the GC risk ratio for uniform and log gamma prior distributions for variance components



(a) Uniform prior

(b) log gamma prior

Appendix IV: Model diagnostics using probability integral transform (PIT)



Appendix V: list of identified districts with high risk ratio of GC

- **Northern regional health authority**

Gods Lake Narrows
Winnipeg Churchill

- **Interlake regional health authority**

Beausejour
Lac Brochet
Nelson House
The Pas
Stonewall/Teulon

- **Prairie mountain regional health authority**

Riding Mountain
Dauphin
Spruce wood
Brandon east end

- **Southern health regional health authority**

Carman
Notre dame
Red river south

- **Winnipeg regional health authority**

Seven oaks east
Point douglas north
River east west
Point douglas south
River east south
Downtown east
St. Boniface west
Downtown west
River height east
River height west
St. vital north

Appendix VI: List of districts identified with high risk of Men GC (22 districts)

- **Northern regional health authority**

Gods Lake Narrows
Winnipeg Churchill
The pas

- **Interlake regional health authority**

Thompson/Mystery Lake

The pas

Nelson House

Gillam/Fox Lake

- **Prairie mountain regional health authority**

Asessippi

Brandon east end

Riding mountain

dauphin

- **Southern health regional health authority**

Louise

Carman

cartier

Niverville/Richot

- **Winnipeg regional health authority**

Point douglas south

River east south

Downtown east

St. Boniface west

Downtown west

River height east

Point douglas north

Appendix VII: List of districts identified with high risk of women GC (16 districts)

- **Interlake regional health authority**

Stonewall/Teulon

Lac brochet

Gillam/Fox Lake

- **Prairie mountain regional health authority**

Souris

Brandon east end

Riding mountain

- **Winnipeg regional health authority**

River east north

Seven oaks east

River east west

Point douglas north

River east south

Inkster west

Downtown east
St. Boniface west
River height west
St. vital north

Appendix VIII: List of districts identified with high risk of cardia GC (11 districts)

- **Northern regional health authority**

Winnipeg churchill

- **Prairie mountain regional health authority**

Souris

Brandon east end

Riding mountain

Dauphin

- **Winnipeg regional health authority**

Seven oaks west

Point douglas north

Downtown west

Downtown east

St. Boniface west

St. vital north

Appendix IX: List of districts identified with high risk of non-cardia GC (27 districts)

- **Northern regional health authority**

Grand rapids

- **Interlake regional health authority**

The pas

Eriksdale/Ashen

Nelson House

- **Prairie mountain regional health authority**

Souris

Asessippi

Riding mountain

Brandon east end

Spruce wood

Dauphin

Swan river

Porcupine mountain

- **Southern health regional health authority**

Carman

cartier

Red river south

Rural east

- **Winnipeg regional health authority**

Seven oaks west
River east west
Inkster east
Point douglas north
Point douglas south
River east south
Downtown west
Downtown east
St. Boniface west
River height west
St. vital north

Appendix X: Code

SAS Code for extracting gastric cancer cases from Cancer registry

```
/* code for extracting manitoba dataset */
libname gbenga
'z:\CCR_RCC\CCR_RCC_IARC_1992_2016_v1\data_donnees\data\sas_en';
options nofmterr;
/* subset all interested variables*/
data predictors;
set gbenga.ccr_iarc_1992_2016_incid_f1_v1 (keep = PERSON_ID PSEX
PDATBIR PDCCRNBRMRS TREPPROV TPIN TPLACRES TPOSTCOD TDATDIAG TICD_O2T
TDCCRAGEDIAG TDCCRAGEGRP);
if TREPPROV = 46;
run;
proc export data = work.predictors outfile = 'P:\Fakanye_5646\misc\new-
datasets\gc_2016' dbms = xlsx replace;
sheet = 'gc16';
newfile = yes;
run;
```

R code for data cleaning, partition and analysis

```
#----- changing my path
mypaths <-
.libPaths("P:/AA_RESEARCHER_TRAINING_INFORMATION/R_packages/June
_2018")
mypaths <-
.libPaths("P:/AA_RESEARCHER_TRAINING_INFORMATION/R_packages/July
2019")
# getting working directory and setting work directory
```



```

getwd()
setwd("P:/Fakanye_5646/New folder/R code")
#####
#####
# loading 2016 CCR dataset
rdata <- read.csv("P:\\Fakanye_5646\\New folder\\rawdata.csv",
sep=",")

#extrcating stomach cancer cases
data.gc <- subset(rdata, ccr$TICD_O2T == 'C160'| ccr$TICD_O2T ==
'C161'| ccr$TICD_O2T =='C162'|ccr$TICD_O2T =='C163'|ccr$TICD_O2T
=='C164'|ccr$TICD_O2T =='C165'|ccr$TICD_O2T
=='C166'|ccr$TICD_O2T =='C167'|ccr$TICD_O2T
=='C168'|ccr$TICD_O2T =='C169')
write.csv(data.gc, file = "data-gc16.csv")
#renaming variables
names(data.gc) <- c('id',
'pr','tpin','location','postcode','dod','year','gc_code','stage_
cat1','stage_cat2','stage_cat3','age','agegrp','number','sex','d
ob')

# subsetting datasets by sex, cancer type
data.male <- subset(data.gc, sex == 1)
data.female <- subset(data.gc, sex == 2)
data.cad <- subset(data.gc, gc_code == "C160")
data.ncad <- subset(data.gc, gc_code == "C161"| gc_code ==
"C162" | gc_code == "C163" | gc_code == "C164" | gc_code ==
"C165" | gc_code == "C166"|
gc_code=="C167"|gc_code=="C168"|gc_code=="C169")
#-----
-----
#extracting required variables
data.cad <- data.cad[,c(4,5,7,12,13,15)]
data.ncad <- data.ncad[,c(4,5,7,12,13,15)]
data.male <- data.male[,c(4,5,6,7,12)]
data.female <- data.female[,c(4,5,6,7,12)]
data.gc2 <- data.gc[,c(4,5,6,7,12)]

#stratification of cardia and non-cardia by sex
m_cardia <- subset(data.cad, sex ==1)
m_ncardia <- subset(data.ncad, sex ==1)
f_cardia <- subset(data.cad, sex ==2)
f_ncardia <- subset(data.ncad, sex ==2)
#-----
-----

```

```

# extracting cases by year and age group and aggregating by
postcode
library(tidyverse)
#1. overall data
data.gc2_pc <- summarise(group_by(data.gc2, postcode), cardia =
sum(gc_code == "C160"), Noncardia = sum(gc_code != "C160"),x1992
= sum(year == "1992"),x1993 = sum(year == "1993"),x1994 =
sum(year == "1994"),x1995 = sum(year == "1995"),x1996 = sum(year
== "1996"),x1997 = sum(year == "1997"),x1998 = sum(year ==
"1998"),x1999 = sum(year == "1999"),x2000 = sum(year ==
"2000"),x2001 = sum(year == "2001"),x2002 = sum(year ==
"2002"),x2003 = sum(year == "2003"),x2004 = sum(year ==
"2004"),x2005 = sum(year == "2005"),x2006 = sum(year ==
"2006"),x2007 = sum(year == "2007"),x2008 = sum(year ==
"2008"),x2009 = sum(year == "2009"),x2010 = sum(year ==
"2010"),x2011 = sum(year == "2011"),x2012 = sum(year ==
"2012"),x2013 = sum(year == "2013"),x2014 = sum(year ==
"2014"),x2015 = sum(year == "2015"),x2016 = sum(year == "2016"),
"below 30" = sum(age.grp == 1|age.grp ==2|age.grp ==3|age.grp
==4|age.grp ==5|age.grp ==6), "30-34" = sum(age.grp == 7),"35-
39" = sum(age.grp == 8), "40-44" = sum(age.grp == 9),"45-49" =
sum(age.grp == 10), "50-54" = sum(age.grp == 11), "55-59" =
sum(age.grp == 12),"60-64" = sum(age.grp == 13),"65-69" =
sum(age.grp == 14),"70-74" = sum(age.grp == 15),"75-79" =
sum(age.grp == 16),"80-84" = sum(age.grp == 17),"85-89" =
sum(age.grp == 18),"90 +" = sum(age.grp == 19 |age.grp==20))

```

```

#2. male
data.male_pc <- summarise(group_by(data.male, postcode), cardia
= sum(gc_code == "C160"), Noncardia = sum(gc_code !=
"C160"),x1992 = sum(year == "1992"),x1993 = sum(year ==
"1993"),x1994 = sum(year == "1994"),x1995 = sum(year ==
"1995"),x1996 = sum(year == "1996"),x1997 = sum(year ==
"1997"),x1998 = sum(year == "1998"),x1999 = sum(year ==
"1999"),x2000 = sum(year == "2000"),x2001 = sum(year ==
"2001"),x2002 = sum(year == "2002"),x2003 = sum(year ==
"2003"),x2004 = sum(year == "2004"),x2005 = sum(year ==
"2005"),x2006 = sum(year == "2006"),x2007 = sum(year ==
"2007"),x2008 = sum(year == "2008"),x2009 = sum(year ==
"2009"),x2010 = sum(year == "2010"),x2011 = sum(year ==
"2011"),x2012 = sum(year == "2012"),x2013 = sum(year ==
"2013"),x2014 = sum(year == "2014"),x2015 = sum(year ==
"2015"),x2016 = sum(year == "2016"), "below 30" = sum(age.grp ==
1|age.grp ==2|age.grp ==3|age.grp ==4|age.grp ==5|age.grp ==6),
"30-34" = sum(age.grp == 7),"35-39" = sum(age.grp == 8), "40-44"
= sum(age.grp == 9),"45-49" = sum(age.grp == 10), "50-54" =
sum(age.grp == 11), "55-59" = sum(age.grp == 12),"60-64" =

```

```
sum(age.grp == 13), "65-69" = sum(age.grp == 14), "70-74" =
sum(age.grp == 15), "75-79" = sum(age.grp == 16), "80-84" =
sum(age.grp == 17), "85-89" = sum(age.grp == 18), "90 +" =
sum(age.grp == 19 |age.grp==20))
```

#3. female

```
data.female_pc <- summarise(group_by(data.female, postcode),
cardia = sum(gc_code == "C160"), Noncardia = sum(gc_code !=
"C160"), x1992 = sum(year == "1992"), x1993 = sum(year ==
"1993"), x1994 = sum(year == "1994"), x1995 = sum(year ==
"1995"), x1996 = sum(year == "1996"), x1997 = sum(year ==
"1997"), x1998 = sum(year == "1998"), x1999 = sum(year ==
"1999"), x2000 = sum(year == "2000"), x2001 = sum(year ==
"2001"), x2002 = sum(year == "2002"), x2003 = sum(year ==
"2003"), x2004 = sum(year == "2004"), x2005 = sum(year ==
"2005"), x2006 = sum(year == "2006"), x2007 = sum(year ==
"2007"), x2008 = sum(year == "2008"), x2009 = sum(year ==
"2009"), x2010 = sum(year == "2010"), x2011 = sum(year ==
"2011"), x2012 = sum(year == "2012"), x2013 = sum(year ==
"2013"), x2014 = sum(year == "2014"), x2015 = sum(year ==
"2015"), x2016 = sum(year == "2016"), "below 30" = sum(age.grp ==
1|age.grp ==2|age.grp ==3|age.grp ==4|age.grp ==5|age.grp ==6),
"30-34" = sum(age.grp == 7), "35-39" = sum(age.grp == 8), "40-44"
= sum(age.grp == 9), "45-49" = sum(age.grp == 10), "50-54" =
sum(age.grp == 11), "55-59" = sum(age.grp == 12), "60-64" =
sum(age.grp == 13), "65-69" = sum(age.grp == 14), "70-74" =
sum(age.grp == 15), "75-79" = sum(age.grp == 16), "80-84" =
sum(age.grp == 17), "85-89" = sum(age.grp == 18), "90 +" =
sum(age.grp == 19 |age.grp==20))
```

#4. Cardia

```
data.cad_pc <- summarise(group_by(data.cad, postcode), x1992 =
sum(year == "1992"), x1993 = sum(year == "1993"), x1994 = sum(year
== "1994"), x1995 = sum(year == "1995"), x1996 = sum(year ==
"1996"), x1997 = sum(year == "1997"), x1998 = sum(year ==
"1998"), x1999 = sum(year == "1999"), x2000 = sum(year ==
"2000"), x2001 = sum(year == "2001"), x2002 = sum(year ==
"2002"), x2003 = sum(year == "2003"), x2004 = sum(year ==
"2004"), x2005 = sum(year == "2005"), x2006 = sum(year ==
"2006"), x2007 = sum(year == "2007"), x2008 = sum(year ==
"2008"), x2009 = sum(year == "2009"), x2010 = sum(year ==
"2010"), x2011 = sum(year == "2011"), x2012 = sum(year ==
"2012"), x2013 = sum(year == "2013"), x2014 = sum(year ==
"2014"), x2015 = sum(year == "2015"), x2016 = sum(year == "2016"),
"below 30" = sum(age.grp == 1|age.grp ==2|age.grp ==3|age.grp
==4|age.grp ==5|age.grp ==6), "30-34" = sum(age.grp == 7), "35-
39" = sum(age.grp == 8), "40-44" = sum(age.grp == 9), "45-49" =
```

```
sum(age.grp == 10), "50-54" = sum(age.grp == 11), "55-59" =  
sum(age.grp == 12), "60-64" = sum(age.grp == 13), "65-69" =  
sum(age.grp == 14), "70-74" = sum(age.grp == 15), "75-79" =  
sum(age.grp == 16), "80-84" = sum(age.grp == 17), "85-89" =  
sum(age.grp == 18), "90 +" = sum(age.grp == 19 |age.grp==20))
```

#5. Non-Cardia

```
data.ncad_pc <- summarise(group_by(data.ncad, postcode), x1992 =  
sum(year == "1992"), x1993 = sum(year == "1993"), x1994 = sum(year  
== "1994"), x1995 = sum(year == "1995"), x1996 = sum(year ==  
"1996"), x1997 = sum(year == "1997"), x1998 = sum(year ==  
"1998"), x1999 = sum(year == "1999"), x2000 = sum(year ==  
"2000"), x2001 = sum(year == "2001"), x2002 = sum(year ==  
"2002"), x2003 = sum(year == "2003"), x2004 = sum(year ==  
"2004"), x2005 = sum(year == "2005"), x2006 = sum(year ==  
"2006"), x2007 = sum(year == "2007"), x2008 = sum(year ==  
"2008"), x2009 = sum(year == "2009"), x2010 = sum(year ==  
"2010"), x2011 = sum(year == "2011"), x2012 = sum(year ==  
"2012"), x2013 = sum(year == "2013"), x2014 = sum(year ==  
"2014"), x2015 = sum(year == "2015"), x2016 = sum(year == "2016"),  
"below 30" = sum(age.grp == 1|age.grp ==2|age.grp ==3|age.grp  
==4|age.grp ==5|age.grp ==6), "30-34" = sum(age.grp == 7), "35-  
39" = sum(age.grp == 8), "40-44" = sum(age.grp == 9), "45-49" =  
sum(age.grp == 10), "50-54" = sum(age.grp == 11), "55-59" =  
sum(age.grp == 12), "60-64" = sum(age.grp == 13), "65-69" =  
sum(age.grp == 14), "70-74" = sum(age.grp == 15), "75-79" =  
sum(age.grp == 16), "80-84" = sum(age.grp == 17), "85-89" =  
sum(age.grp == 18), "90 +" = sum(age.grp == 19 |age.grp==20))
```

#8. male cardia

```
mcardia_pc <- summarise(group_by(m_cardia, postcode), "1992" =  
sum(year == "1992"), "1993" = sum(year == "1993"), "1994" =  
sum(year == "1994"), "1995" = sum(year == "1995"), "1996" =  
sum(year == "1996"), "1997" = sum(year == "1997"), "1998" =  
sum(year == "1998"), "1999" = sum(year == "1999"), "2000" =  
sum(year == "2000"), "2001" = sum(year == "2001"), "2002" =  
sum(year == "2002"), "2003" = sum(year == "2003"), "2004" =  
sum(year == "2004"), "2005" = sum(year == "2005"), "2006" =  
sum(year == "2006"), "2007" = sum(year == "2007"), "2008" =  
sum(year == "2008"), "2009" = sum(year == "2009"), "2010" =  
sum(year == "2010"), "2011" = sum(year == "2011"), "2012" =  
sum(year == "2012"), "2013" = sum(year == "2013"), "2014" =  
sum(year == "2014"), "2015" = sum(year == "2015"), "2016" =  
sum(year == "2016"), "below 35" = sum(agegrp == 1|agegrp
```

```
==2|agegrp ==3|agegrp ==4|agegrp ==5|agegrp ==6|agegrp==7), "35-39" = sum(agegrp == 8), "40-44" = sum(agegrp == 9), "45-49" = sum(agegrp == 10), "50-54" = sum(agegrp == 11), "55-59" = sum(agegrp == 12), "60-64" = sum(agegrp == 13), "65-69" = sum(agegrp == 14), "70-74" = sum(agegrp == 15), "75-79" = sum(agegrp == 16), "80-84" = sum(agegrp == 17), "85 +" = sum(agegrp == 18| agegrp == 19 |agegrp ==20))
```

```
#9. female cardia
```

```
fcardia_pc <- summarise(group_by(f_cardia, postcode), "1992" = sum(year == "1992"), "1993" = sum(year == "1993"), "1994" = sum(year == "1994"), "1995" = sum(year == "1995"), "1996" = sum(year == "1996"), "1997" = sum(year == "1997"), "1998" = sum(year == "1998"), "1999" = sum(year == "1999"), "2000" = sum(year == "2000"), "2001" = sum(year == "2001"), "2002" = sum(year == "2002"), "2003" = sum(year == "2003"), "2004" = sum(year == "2004"), "2005" = sum(year == "2005"), "2006" = sum(year == "2006"), "2007" = sum(year == "2007"), "2008" = sum(year == "2008"), "2009" = sum(year == "2009"), "2010" = sum(year == "2010"), "2011" = sum(year == "2011"), "2012" = sum(year == "2012"), "2013" = sum(year == "2013"), "2014" = sum(year == "2014"), "2015" = sum(year == "2015"), "2016" = sum(year == "2016"), "below 35" = sum(agegrp == 1|agegrp ==2|agegrp ==3|agegrp ==4|agegrp ==5|agegrp ==6|agegrp==7), "35-39" = sum(agegrp == 8), "40-44" = sum(agegrp == 9), "45-49" = sum(agegrp == 10), "50-54" = sum(agegrp == 11), "55-59" = sum(agegrp == 12), "60-64" = sum(agegrp == 13), "65-69" = sum(agegrp == 14), "70-74" = sum(agegrp == 15), "75-79" = sum(agegrp == 16), "80-84" = sum(agegrp == 17), "85 +" = sum(agegrp == 18| agegrp == 19 |agegrp ==20))
```

```
#10. male non-cardia
```

```
mncardia_pc <- summarise(group_by(m_ncardia, postcode), "1992" = sum(year == "1992"), "1993" = sum(year == "1993"), "1994" = sum(year == "1994"), "1995" = sum(year == "1995"), "1996" = sum(year == "1996"), "1997" = sum(year == "1997"), "1998" = sum(year == "1998"), "1999" = sum(year == "1999"), "2000" = sum(year == "2000"), "2001" = sum(year == "2001"), "2002" = sum(year == "2002"), "2003" = sum(year == "2003"), "2004" = sum(year == "2004"), "2005" = sum(year == "2005"), "2006" = sum(year == "2006"), "2007" = sum(year == "2007"), "2008" = sum(year == "2008"), "2009" = sum(year == "2009"), "2010" = sum(year == "2010"), "2011" = sum(year == "2011"), "2012" = sum(year == "2012"), "2013" = sum(year == "2013"), "2014" = sum(year == "2014"), "2015" = sum(year == "2015"), "2016" =
```

```
sum(year == "2016"), "below 35" = sum(agegrp == 1|agegrp
==2|agegrp ==3|agegrp ==4|agegrp ==5|agegrp ==6|agegrp==7), "35-
39" = sum(agegrp == 8), "40-44" = sum(agegrp == 9), "45-49" =
sum(agegrp == 10), "50-54" = sum(agegrp == 11), "55-59" =
sum(agegrp == 12), "60-64" = sum(agegrp == 13), "65-69" =
sum(agegrp == 14), "70-74" = sum(agegrp == 15), "75-79" =
sum(agegrp == 16), "80-84" = sum(agegrp == 17), "85 +" =
sum(agegrp == 18| agegrp == 19 |agegrp ==20))
```

```
#11. female non-cardia
```

```
fncardia_pc <- summarise(group_by(f_ncardia, postcode), "1992" =
sum(year == "1992"), "1993" = sum(year == "1993"), "1994" =
sum(year == "1994"), "1995" = sum(year == "1995"), "1996" =
sum(year == "1996"), "1997" = sum(year == "1997"), "1998" =
sum(year == "1998"), "1999" = sum(year == "1999"), "2000" =
sum(year == "2000"), "2001" = sum(year == "2001"), "2002" =
sum(year == "2002"), "2003" = sum(year == "2003"), "2004" =
sum(year == "2004"), "2005" = sum(year == "2005"), "2006" =
sum(year == "2006"), "2007" = sum(year == "2007"), "2008" =
sum(year == "2008"), "2009" = sum(year == "2009"), "2010" =
sum(year == "2010"), "2011" = sum(year == "2011"), "2012" =
sum(year == "2012"), "2013" = sum(year == "2013"), "2014" =
sum(year == "2014"), "2015" = sum(year == "2015"), "2016" =
sum(year == "2016"), "below 35" = sum(agegrp == 1|agegrp
==2|agegrp ==3|agegrp ==4|agegrp ==5|agegrp ==6|agegrp==7), "35-
39" = sum(agegrp == 8), "40-44" = sum(agegrp == 9), "45-49" =
sum(agegrp == 10), "50-54" = sum(agegrp == 11), "55-59" =
sum(agegrp == 12), "60-64" = sum(agegrp == 13), "65-69" =
sum(agegrp == 14), "70-74" = sum(agegrp == 15), "75-79" =
sum(agegrp == 16), "80-84" = sum(agegrp == 17), "85 +" =
sum(agegrp == 18| agegrp == 19 |agegrp ==20))
```

```
# merging cases (gc_by_pcode) with RHDA (96) using postal code
rhda <- read.csv("P:\\Fakanye_5646\\New folder\\R
code\\rhda.csv", sep=",")
data.gc_rh <- merge(rhda, data.gc2_pc, by.x = "pcode", by.y =
"postcode", all.y = TRUE)
data.male_rh <- merge(rhda, data.male_pc, by.x = "pcode", by.y =
"postcode", all.y = TRUE)
data.female_rh <- merge(rhda, data.female_pc, by.x = "pcode",
by.y = "postcode", all.y = TRUE)
data.cad_rh <- merge(rhda, data.cad_pc, by.x = "pcode", by.y =
"postcode", all.y = TRUE)
data.ncad_rh <- merge(rhda, data.ncad_pc, by.x = "pcode", by.y =
"postcode", all.y = TRUE)
```

```

mcardia_rh <- merge(rhda, mcardia_pc, by.x = "pcode", by.y =
"postcode", all.y = TRUE)
fcardia_rh <- merge(rhda, fcardia_pc, by.x = "pcode", by.y =
"postcode", all.y = TRUE)
mncardia_rh <- merge(rhda, mncardia_pc, by.x = "pcode", by.y =
"postcode", all.y = TRUE)
fncardia_rh <- merge(rhda, fncardia_pc, by.x = "pcode", by.y =
"postcode", all.y = TRUE)

#saving in .csv file, the rhad datat; so we can manually adjust
for regions that are not matched
write.csv(data.gc_rh, file = "data.gc_rh.csv")
write.csv(data.male_rh, file = "data.male_rh.csv")
write.csv(data.female_rh, file = "data.female_rh.csv")
write.csv(data.cad_rh, file = "data.cad_rh.csv")
write.csv(data.ncad_rh, file = "data.ncad_rh.csv")
write.csv(mcardia_rh, file = "mcardia_rh.csv")
write.csv(fcardia_rh, file = "fcardia_rh.csv")
write.csv(mncardia_rh, file = "mncardia_rh.csv")
write.csv(fncardia_rh, file = "fncardia_rh.csv")
#-----
-----
# Aggregating the merged gc cases data by RHDA
#loading manually edited aggregated data by postcode
o.rhad <- summarise(group_by(data.gc_rh, id, District), cardia =
sum(cardia), Noncardia = sum(Noncardia), "1992" =
sum(x1992), "1993" = sum(x1993), "1994" = sum(x1994), "1995" =
sum(x1995), "1996" = sum(x1996), "1997" = sum(x1997), "1998" =
sum(x1998), "1999" = sum(x1999), "2000" = sum(x2000), "2001" =
sum(x2001), "2002" = sum(x2002), "2003" = sum(x2003), "2004" =
sum(x2004), "2005" = sum(x2005), "2006" = sum(x2006), "2007" =
sum(x2007), "2008" = sum(x2008), "2009" = sum(x2009), "2010" =
sum(x2010), "2011" = sum(x2011), "2012" = sum(x2012), "2013" =
sum(x2013), "2014" = sum(x2014), "2015" = sum(x2015), "2016" =
sum(x2016), "below 30" = sum(below.30), "30-34" =
sum(X30.34), "35-39" = sum(X35.39), "40-44" = sum(X40.44), "45-49"
= sum(X45.49), "50-54" = sum(X50.54), "55-59" = sum(X55.59), "60-
64" = sum(X60.64), "65-69" = sum(X65.69), "70-74" =
sum(X70.74), "75-79" = sum(X75.79), "80-84" = sum(X80.84), "85-89"
= sum(X85.89), "90+" = sum(X90..))

male.rhad <- summarise(group_by(data.male_rh, id, District),
cardia = sum(cardia), Noncardia = sum(Noncardia), "1992" =
sum(x1992), "1993" = sum(x1993), "1994" = sum(x1994), "1995" =
sum(x1995), "1996" = sum(x1996), "1997" = sum(x1997), "1998" =
sum(x1998), "1999" = sum(x1999), "2000" = sum(x2000), "2001" =

```

```
sum(x2001), "2002" = sum(x2002), "2003" = sum(x2003), "2004" =
sum(x2004), "2005" = sum(x2005), "2006" = sum(x2006), "2007" =
sum(x2007), "2008" = sum(x2008), "2009" = sum(x2009), "2010" =
sum(x2010), "2011" = sum(x2011), "2012" = sum(x2012), "2013" =
sum(x2013), "2014" = sum(x2014), "2015" = sum(x2015), "2016" =
sum(x2016), "below 30" = sum(below.30), "30-34" =
sum(X30.34), "35-39" = sum(X35.39), "40-44" = sum(X40.44), "45-49"
= sum(X45.49), "50-54" = sum(X50.54), "55-59" = sum(X55.59), "60-
64" = sum(X60.64), "65-69" = sum(X65.69), "70-74" =
sum(X70.74), "75-79" = sum(X75.79), "80-84" = sum(X80.84), "85-89"
= sum(X85.89), "90+" = sum(X90..))
```

```
female.rhad <- summarise(group_by(data.female_rh, id, District),
cardia = sum(cardia), Noncardia = sum(Noncardia), "1992" =
sum(x1992), "1993" = sum(x1993), "1994" = sum(x1994), "1995" =
sum(x1995), "1996" = sum(x1996), "1997" = sum(x1997), "1998" =
sum(x1998), "1999" = sum(x1999), "2000" = sum(x2000), "2001" =
sum(x2001), "2002" = sum(x2002), "2003" = sum(x2003), "2004" =
sum(x2004), "2005" = sum(x2005), "2006" = sum(x2006), "2007" =
sum(x2007), "2008" = sum(x2008), "2009" = sum(x2009), "2010" =
sum(x2010), "2011" = sum(x2011), "2012" = sum(x2012), "2013" =
sum(x2013), "2014" = sum(x2014), "2015" = sum(x2015), "2016" =
sum(x2016), "below 30" = sum(below.30), "30-34" =
sum(X30.34), "35-39" = sum(X35.39), "40-44" = sum(X40.44), "45-49"
= sum(X45.49), "50-54" = sum(X50.54), "55-59" = sum(X55.59), "60-
64" = sum(X60.64), "65-69" = sum(X65.69), "70-74" =
sum(X70.74), "75-79" = sum(X75.79), "80-84" = sum(X80.84), "85-89"
= sum(X85.89), "90+" = sum(X90..))
```

```
cad.rhad <- summarise(group_by(data.cad_rh, id, District),
"1992" = sum(x1992), "1993" = sum(x1993), "1994" =
sum(x1994), "1995" = sum(x1995), "1996" = sum(x1996), "1997" =
sum(x1997), "1998" = sum(x1998), "1999" = sum(x1999), "2000" =
sum(x2000), "2001" = sum(x2001), "2002" = sum(x2002), "2003" =
sum(x2003), "2004" = sum(x2004), "2005" = sum(x2005), "2006" =
sum(x2006), "2007" = sum(x2007), "2008" = sum(x2008), "2009" =
sum(x2009), "2010" = sum(x2010), "2011" = sum(x2011), "2012" =
sum(x2012), "2013" = sum(x2013), "2014" = sum(x2014), "2015" =
sum(x2015), "2016" = sum(x2016), "below 30" = sum(below.30), "30-
34" = sum(X30.34), "35-39" = sum(X35.39), "40-44" =
sum(X40.44), "45-49" = sum(X45.49), "50-54" = sum(X50.54), "55-
59" = sum(X55.59), "60-64" = sum(X60.64), "65-69" =
sum(X65.69), "70-74" = sum(X70.74), "75-79" = sum(X75.79), "80-84"
= sum(X80.84), "85-89" = sum(X85.89), "90+" = sum(X90..))
```

```
ncad.rhad <- summarise(group_by(data.ncad_rh, id, District),
"1992" = sum(x1992), "1993" = sum(x1993), "1994" =
```

```
sum(x1994), "1995" = sum(x1995), "1996" = sum(x1996), "1997" =
sum(x1997), "1998" = sum(x1998), "1999" = sum(x1999), "2000" =
sum(x2000), "2001" = sum(x2001), "2002" = sum(x2002), "2003" =
sum(x2003), "2004" = sum(x2004), "2005" = sum(x2005), "2006" =
sum(x2006), "2007" = sum(x2007), "2008" = sum(x2008), "2009" =
sum(x2009), "2010" = sum(x2010), "2011" = sum(x2011), "2012" =
sum(x2012), "2013" = sum(x2013), "2014" = sum(x2014), "2015" =
sum(x2015), "2016" = sum(x2016), "below 30" = sum(below.30), "30-
34" = sum(X30.34), "35-39" = sum(X35.39), "40-44" =
sum(X40.44), "45-49" = sum(X45.49), "50-54" = sum(X50.54), "55-
59" = sum(X55.59), "60-64" = sum(X60.64), "65-69" =
sum(X65.69), "70-74" = sum(X70.74), "75-79" = sum(X75.79), "80-84"
= sum(X80.84), "85-89" = sum(X85.89), "90+" = sum(X90..))
```

```
mcardia.rhad <- summarise(group_by(mcardia_rh, id, District),
"1992" = sum(X1992), "1993" = sum(X1993), "1994" =
sum(X1994), "1995" = sum(X1995), "1996" = sum(X1996), "1997" =
sum(X1997), "1998" = sum(X1998), "1999" = sum(X1999), "2000" =
sum(X2000), "2001" = sum(X2001), "2002" = sum(X2002), "2003" =
sum(X2003), "2004" = sum(X2004), "2005" = sum(X2005), "2006" =
sum(X2006), "2007" = sum(X2007), "2008" = sum(X2008), "2009" =
sum(X2009), "2010" = sum(X2010), "2011" = sum(X2011), "2012" =
sum(X2012), "2013" = sum(X2013), "2014" = sum(X2014), "2015" =
sum(X2015), "2016" = sum(X2016), "below 35" = sum(below.35), "35-
39" = sum(X35.39), "40-44" = sum(X40.44), "45-49" = sum(X45.49),
"50-54" = sum(X50.54), "55-59" = sum(X55.59), "60-64" =
sum(X60.64), "65-69" = sum(X65.69), "70-74" = sum(X70.74), "75-79"
= sum(X75.79), "80-84" = sum(X80.84), "85 +" = sum(X85..))
```

```
fcardia.rhad <- summarise(group_by(fcardia_rh, id, District),
"1992" = sum(X1992), "1993" = sum(X1993), "1994" =
sum(X1994), "1995" = sum(X1995), "1996" = sum(X1996), "1997" =
sum(X1997), "1998" = sum(X1998), "1999" = sum(X1999), "2000" =
sum(X2000), "2001" = sum(X2001), "2002" = sum(X2002), "2003" =
sum(X2003), "2004" = sum(X2004), "2005" = sum(X2005), "2006" =
sum(X2006), "2007" = sum(X2007), "2008" = sum(X2008), "2009" =
sum(X2009), "2010" = sum(X2010), "2011" = sum(X2011), "2012" =
sum(X2012), "2013" = sum(X2013), "2014" = sum(X2014), "2015" =
sum(X2015), "2016" = sum(X2016), "below 35" = sum(below.35), "35-
39" = sum(X35.39), "40-44" = sum(X40.44), "45-49" = sum(X45.49),
"50-54" = sum(X50.54), "55-59" = sum(X55.59), "60-64" =
sum(X60.64), "65-69" = sum(X65.69), "70-74" = sum(X70.74), "75-79"
= sum(X75.79), "80-84" = sum(X80.84), "85 +" = sum(X85..))
```

```
fncardia.rhad <- summarise(group_by(fncardia_rh, id, District),
"1992" = sum(X1992), "1993" = sum(X1993), "1994" =
sum(X1994), "1995" = sum(X1995), "1996" = sum(X1996), "1997" =
```

```

sum(X1997), "1998" = sum(X1998), "1999" = sum(X1999), "2000" =
sum(X2000), "2001" = sum(X2001), "2002" = sum(X2002), "2003" =
sum(X2003), "2004" = sum(X2004), "2005" = sum(X2005), "2006" =
sum(X2006), "2007" = sum(X2007), "2008" = sum(X2008), "2009" =
sum(X2009), "2010" = sum(X2010), "2011" = sum(X2011), "2012" =
sum(X2012), "2013" = sum(X2013), "2014" = sum(X2014), "2015" =
sum(X2015), "2016" = sum(X2016), "below 35" = sum(below.35), "35-
39" = sum(X35.39), "40-44" = sum(X40.44), "45-49" = sum(X45.49),
"50-54" = sum(X50.54), "55-59" = sum(X55.59), "60-64" =
sum(X60.64), "65-69" = sum(X65.69), "70-74" = sum(X70.74), "75-79"
= sum(X75.79), "80-84" = sum(X80.84), "85 +" = sum(X85..))

```

```

mncardia.rhad <- summarise(group_by(mncardia_rh, id, District),
"1992" = sum(X1992), "1993" = sum(X1993), "1994" =
sum(X1994), "1995" = sum(X1995), "1996" = sum(X1996), "1997" =
sum(X1997), "1998" = sum(X1998), "1999" = sum(X1999), "2000" =
sum(X2000), "2001" = sum(X2001), "2002" = sum(X2002), "2003" =
sum(X2003), "2004" = sum(X2004), "2005" = sum(X2005), "2006" =
sum(X2006), "2007" = sum(X2007), "2008" = sum(X2008), "2009" =
sum(X2009), "2010" = sum(X2010), "2011" = sum(X2011), "2012" =
sum(X2012), "2013" = sum(X2013), "2014" = sum(X2014), "2015" =
sum(X2015), "2016" = sum(X2016), "below 35" = sum(below.35), "35-
39" = sum(X35.39), "40-44" = sum(X40.44), "45-49" = sum(X45.49),
"50-54" = sum(X50.54), "55-59" = sum(X55.59), "60-64" =
sum(X60.64), "65-69" = sum(X65.69), "70-74" = sum(X70.74), "75-79"
= sum(X75.79), "80-84" = sum(X80.84), "85 +" = sum(X85..))

```

```

#-----
# merging the gc cases with covariates
cov.rhad<- read.csv("P:\\Fakanye_5646\\New
folder\\cov.rhad.csv", sep=",")
#1. Overall
overall <- merge(cov.rhad, o.rhad, by.x = "OBJECTID", by.y =
"id", all.x = TRUE)
#2. male
male <- merge(cov.rhad, male.rhad, by.x = "OBJECTID", by.y =
"id", all.x = TRUE)
#3. female
female <- merge(cov.rhad, female.rhad, by.x = "OBJECTID", by.y =
"id", all.x = TRUE)
#4. cardia
cardia <- merge(cov.rhad, cad.rhad, by.x = "OBJECTID", by.y =
"id", all.x = TRUE)
#5. Noncardia
noncardia <- merge(cov.rhad, ncad.rhad, by.x = "OBJECTID", by.y
= "id", all.x = TRUE)

```

```

#8. male cardia
mcard.data <- merge(cov.rhad, mcardia.rhad, by.x = "OBJECTID",
by.y = "id", all.x = TRUE)

#9. female cardia
fcad.data <- merge(cov.rhad, fcardia.rhad, by.x = "OBJECTID",
by.y = "id", all.x = TRUE)

#10. male non-cardia
mncad.data <- merge(cov.rhad, mncardia.rhad, by.x = "OBJECTID",
by.y = "id", all.x = TRUE)

#11. female non-cardia
fncad.data <- merge(cov.rhad, fncardia.rhad, by.x = "OBJECTID",
by.y = "id", all.x = TRUE)

#-----
-----

#final output data
write.csv(overall, file = "overall.csv")
write.csv(male, file = "male.csv")
write.csv(female, file = "female.csv")
write.csv(cardia, file = "cardia.csv")
write.csv(noncardia, file = "noncardia.csv")
write.csv(mcard.data, file = "male_cardia.csv")
write.csv(fcad.data, file = "female_cardia.csv")
write.csv(mncad.data, file = "male_ncardia.csv")
write.csv(fncad.data, file = "female_ncardia.csv")

#####
#           temporal dataset extraction           #
#####
male.st1 <- read.csv("P:\\Fakanye_5646\\New folder\\inla-
folder\\male-st1.csv", sep=",")
female.st1 <- read.csv("P:\\Fakanye_5646\\New folder\\inla-
folder\\female-st1.csv", sep=",")

#cecking for linearity of the continuous covariates and the log
of response
#1. imigrant
imigrant <- overall[,c(6,10,11)]
imigrant$y <- imigrant$cardia + imigrant$Noncardia
quantile(imigrant$imigp, na.rm = T) # min = 0, Q1 = 3.375539, Q2
= 7.415129, Q3 = 17.482804, max = 53.918346
imigrant$q1 <- 0

```

```

imigrant$q2 <- 0
imigrant$q3 <- 0
for(i in 1:nrow(imigrant)){
  ifelse(imigrant$imigp[i] <= 3.375539, imigrant$q1[i] <- 1,
imigrant$q1[i] <- 0)
  ifelse(imigrant$imigp[i] > 3.375539 & imigrant$imigp[i] <=
7.415129, imigrant$q2[i] <- 1, imigrant$q2[i] <- 0)
  ifelse(imigrant$imigp[i] > 7.415129 & imigrant$imigp[i] <=
17.482804, imigrant$q3[i] <- 1, imigrant$q3[i] <- 0)
}

#poison egression
mod <- glm(y ~ q1 + q2 + q3, data = imigrant, family =
poisson(link = log))

#2. indigenous
indi <- overall[,c(7,10,11)]
indi$y <- indi$cardia + indi$Noncardia
quantile(indi$indip, na.rm = T) # min = 1.171, Q1 = 9.3, Q2 =
15.11, Q3 = 31.19, max = 98
indi$q1 <- 0
indi$q2 <- 0
indi$q3 <- 0
for(i in 1:nrow(indi)){
  ifelse(indi$indip[i] <= 9.3, indi$q1[i] <- 1, indi$q1[i] <- 0)
  ifelse(indi$indip[i] > 9.3 & indi$indip[i] <= 15.11,
indi$q2[i] <- 1, indi$q2[i] <- 0)
  ifelse(indi$indip[i] > 15.11 & indi$indip[i] <= 31.19,
indi$q3[i] <- 1, indi$q3[i] <- 0)
}

#poison egression
mod <- glm(y ~ q1 + q2 + q3, data = indi, family = poisson(link
= log))

#-----Exploratory Analysis-----
-#
#a. Income
plot(mydata$income, mydata$sir,xlab = "Median Income", ylab =
"GC incicende ratio", col='blue')
lines(lowess(mydata$income, mydata$sir), col = "red")

#-----
-----
# model fitting
# POISSON MODEL
#1. empty model

```

```

mod1 <- glm(y ~ 1, data = mydata, family = poisson(link = log))

#2. checking association between the independent variable and
the dependent variable
#Univariate model
mod1.1 <- glm(y ~ offset(log(E)) + income, data = mydata, family
= poisson(link = log))
mod1.2 <- glm(y ~ offset(log(E)) + npsedup, data = mydata,
family = poisson(link = log))
mod1.3 <- glm(y ~ offset(log(E)) + aborp, data = mydata, family
= poisson(link = log))
mod1.4 <- glm(y ~ offset(log(E)) + imigp, data = mydata, family
= poisson(link = log))
mod1.5 <- glm(y ~ offset(log(E)) + vismp, data = mydata, family
= poisson(link = log))
mod1.6 <- glm(y ~ offset(log(E)) + unempp, data = mydata, family
= poisson(link = log))
mod1.7 <- glm(y ~ offset(log(E)) + agricp, data = mydata, family
= poisson(link = log))
mod1.8 <- glm(y ~ offset(log(E)) + mingp, data = mydata, family
= poisson(link = log))
#-----
-----#
#-----CREATING SES INDEX SCORE-----
-----#
mydata1 <- read.csv("P:\\Fakanye_5646\\R code\\data_2016.csv",
sep=",")
# recoding variable and standardization
mydata <- mydata1[, -c(1:17,26:50)]
mydata$popp <- mydata1$tpop/sum(mydata1$tpop)
mydata$Edu <- 1 - mydata$npsedup
mydata$Empl <- 1 - mydata$unempp
#=====

# standardizing the predictors
mydata$simig <- (mydata$imigp-
min(mydata$imigp))/(max(mydata$imigp) - min(mydata$imigp))
mydata$sabor <- (mydata$aborp-
min(mydata$aborp))/(max(mydata$aborp) - min(mydata$aborp))
mydata$svism <- (mydata$vismp-
min(mydata$vismp))/(max(mydata$vismp) - min(mydata$vismp))
mydata$sunemp <- (mydata$unempp-
min(mydata$unempp))/(max(mydata$unempp) - min(mydata$unempp))
mydata$sedu <- (mydata$npsedup-
min(mydata$npsedup))/(max(mydata$npsedup) - min(mydata$npsedup))
mydata$sming <- (mydata$mingp-
min(mydata$mingp))/(max(mydata$mingp) - min(mydata$mingp))

```

```

mydata$sfarm <- (mydata$agricp-
min(mydata$agricp))/(max(mydata$agricp) - min(mydata$agricp))
mydata$sincome <- (mydata$income-
min(mydata$income))/(max(mydata$income) - min(mydata$income))
mydata$spopp <- (mydata$poppp-
min(mydata$poppp))/(max(mydata$poppp) - min(mydata$poppp))
mydata$sEmpl <- (mydata$Empl-
min(mydata$Empl))/(max(mydata$Empl) - min(mydata$Empl))
mydata$sEdu <- (mydata$Edu- min(mydata$Edu))/(max(mydata$Edu) -
min(mydata$Edu))

#----- Factor Analysis -----#
#EFA with income, education and employment
sdata2 <- sdata[,c(11,13,14)]# standardized data
udata2<- mydata[,c(1,9,10)]# unstandadrized data

# correlation matrix
udata_cor <- cor(udata2)
inv_cor <- 1/udata_cor
fa <- factanal(sdata2, factors =1, rotation = "varimax", scores
= "regression")

# factor loadings as matrix data
skf <- as.matrix.data.frame(fa$loadings) # rotated
component/loadings
skfmat <- as.matrix(skf)

# inverse correlation matrix of unstandardized data
cor <- cor(udata2)
rkk <- as.matrix(1/cor)

# weighted correlation matrix
wkf <- as.matrix(round(rkk%*%skfmat, 4))

#=====
# index computation
zik <- as.matrix(sdata2)
ses <- zik%*%wkf
mydata1$sesi <- ses
write.csv(mydata1, file = "data_sesi.csv")

#----- investigating spatial dependency using
Moran's I -----#
library('spdep')
library('rgdal')

```

```

shp <- readOGR(dsn =
"P:\\Fakanye_5646\\dat\\geoshape3\\geodashp3.shp", layer =
"geodashp3")
queen.nb <- read.gal("queen.gal", region.id = shp$OBJECTID)
moran.test(as.numeric(shp$field_3), nb2listw(queen.nb),
length(shp$OBJECTID)) # GC count
#moran.test(as.numeric(shp$field_4), nb2listw(queen.nb),
length(shp$OBJECTID)) # crude rate

#=====
# multicollinearity assessment
#=====
x <- cbind(data$medcome,data$imigp, data$aborp, data$vismp,
data$unempp, data$agricp, data$mingp, data$npsedup)
xcor <- cor(x)
xcor2 <- cor.test(x)
write.csv(xcor, file = "cor.csv")
#=====
=====

#----- computing of fixed-effect model Residual-----
-----#

#=====#
# log(mu) = log(E)+ beta0 + beta1*sesi + beta2*immigration +
beta3*indigenous
#log(mu) = log(E) + 0.9658 + (-0.1161)*sesi + 0.691*immigration
+ 0.02497*indigenous
data <- read.csv("P:\\Fakanye_5646\\R code\\data_sesi.csv",
sep=",")
lge <- log(data$E)

b1x1 <- data$sesi*(-0.1161)
b2x2 <- data$imigp*0.691
b3x3 <- data$aborp*0.02497

xbtotal <- 0.6418 + b1x1 + b2x2 + b3x3# total
lambda <- exp(xbtotal) # rate
mu <- lambda*data$E # predicted count
#residual
data$residual <- data$y - mu

#standardised residual
data$sresid <- scale(data$residual, center = TRUE, scale = TRUE)
data$smooth_SIR <- mu/data$E

```

```

write.csv(data, file= "data-wit-residual.csv")

#1. residual plots
data2 <- read.csv("P:\\Fakanye_5646\\R code\\data-wit-
residual.csv", sep=",")
#a. SESI
plot(data2$smooth_SIR, data2$sesi,xlab = "smoothed risk", ylab =
" Socioeconomic score index", col='blue')
lines(lowess(data2$smooth_SIR, data2$sesi), col = "red")

#c. indigenous proportion
plot(data2$smooth_SIR, data2$aborp,xlab = "smoothed risk", ylab
= " Indigenous proportion", col='blue')
lines(lowess(data2$smooth_SIR, data2$aborp), col = 'red')

#e. Imigrant proportion
plot(data2$smooth_SIR, data2$imigp, xlab = "smoothed risk", ylab
= "immigrant Proportion", col='blue')
lines(lowess(data2$smooth_SIR, data2$imigp), col='red')

#----- Model fitting -----#
setwd("P:\\Fakanye_5646\\New folder\\inla-folder")
mypaths <-
.libPaths("P:/AA_RESEARCHER_TRAINING_INFORMATION/R_packages/Ju
ne_2018")
mypaths <- .libPaths("P:/Fakanye_5646/INLA_19.09.03")
#--- load required packages -----#
library(sp)
library(maptools)
library(Matrix)
library(spdep)
library(rgdal)
library(parallel)
library(ggplot2)
library(INLA)
#-- Prepare the map --#
#Import the data
data <- read.csv("overall-ST.csv")
mb <- readOGR("shp96.shp")
data.mb = attr(mb, "data")

#Create the graph for adjacencies in INLA
z2 <- poly2nb(mb)
nb2INLA("mb.graph", z2) # output spatial neighbours for INLA
(mb.graph is file where adjacency matrix will be stored)
#this create a file called "mb-INLA.adj" with the graph for
INLA

```

```

mb.adj <- paste(getwd(), "/mb.graph", sep="")
#--Transform the data to be in the right format for INLA--#
y.vector <- as.vector(as.matrix(data[,11:15]))#by column
E.vector <- as.vector(as.matrix(data[,16:20]))#by column
x1 <- data$ses
x2 <- data$imigp
x3 <- data$indip
year <- numeric(0)
for(i in 1:5){
  year <- append(year, rep(i, dim(data)[1]))
}
rhad <- as.factor(rep(data[,1], 5))

data2<- data.frame(y = y.vector, E= E.vector, x1 = x1, x2 =
x2, x3 = x3, ID.area1 = as.numeric(data$OBJECTID), ID.area2 =
as.numeric(data$OBJECTID), year = year,
                  ID.year1 = year, ID.year2 = year,
ID.area.year = seq(1, length(rhad)))

# Poisson model without spatial component      #
# model = alpha + x1 + x2 + x3                #
data.s <- read.csv("S-overall.csv")
data.st <- read.csv("data2.csv")
data.ns<- data.frame(y = data.s$y, E= data.s$E, x1 =
data.s$ses, x2 = data.s$imigp, x3 = data.s$indip)
formula.ns <- y ~ 1 + x1 + x2 + x3
mod.ns <- inla(formula.ns, family = "poisson", data = data.ns,
E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))

#spatial model
# model = alpha + x1 + x2+ x3 + u_i + s_i
data.s2 <- data.frame(y = data.s$y, E= data.s$E, x1 =
data.s$ses, x2 = data.s$imigp, x3 = data.s$indip, ID.area1 =
as.numeric(data.s$OBJECTID), ID.area2 =
as.numeric(data.s$OBJECTID))
formula.s1 <- y ~ 1 + x1 + x2 + x3 + f(ID.area1, model="bym",
graph = mb.adj, scale.model = TRUE,
hyper = list(prec.unstruct = list(prior = "loggamma",
param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01)))
mod.s1 <- inla(formula.s1, family = "poisson", data = data.s2,
E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))

# less informative prior for precision parameter

```

```

formula.s2 <- y ~ 1 + x1 + x2 + x3 + f(ID.areal, model="bym",
graph = mb.adj, scale.model = TRUE,
hyper = list(prec.unstruct = list(prior = "loggamma",
param = c(1, 0.001)),
prec.spatial = list(prior = "loggamma", param=c(1,0.001)))
mod.s2 <- inla(formula.s2, family = "poisson", data = data.s2,
E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))

```

```

formula.s3 <- y ~ 1 + x1 + x2 + x3 + f(ID.areal, model="bym",
graph = mb.adj, scale.model = TRUE)
mod.s3 <- inla(formula.s3, family = "poisson", data = data.s2,
E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))

```

```

#spatial and temporal effect model without interaction      #
# model = alpha + x1 + x2+ x3 + u_i + s_i + phi_t + psi_t    #
data.st<- data.frame(y = y.vector, E= E.vector, x1 = x1, x2 =
x2, x3 = x3, ID.areal = as.numeric(data$OBJECTID), ID.area2 =
as.numeric(data$OBJECTID),ID.year1 = year, ID.year2 = year)

```

#1. RW(1) prior for temporal effect

```

formula.t1 <- y ~ 1 + x1 + x2 + x3 + f(ID.areal, model="bym",
graph = mb.adj, scale.model = TRUE,
hyper = list(prec.unstruct = list(prior = "loggamma",
param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01))) +
f(ID.year1, model = "rw1") + f(ID.year2, model = "iid")
mod.t1 <- inla(formula.t1, family = "poisson", data = data.st, E
= E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))

```

#2. Rw(2) prior for temporal effect

```

formula.t2 <- y ~ 1 + x1 + x2 + x3 + f(ID.areal, model="bym",
graph = mb.adj, scale.model = TRUE,
hyper = list(prec.unstruct = list(prior = "loggamma", param =
c(0.1, 0.01)),prec.spatial = list(prior = "loggamma",
param=c(0.1,0.01))) +
f(ID.year1, model = "rw2") + f(ID.year2, model = "iid")
mod.t2 <- inla(formula.t2, family = "poisson", data = data.st, E
= E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))

```

#3. AR(1) prior for temporal effect

```

formula.t3 <- y ~ 1 + x1 + x2 + x3 + f(ID.areal, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),

```

```

prec.spatial = list(prior = "loggamma", param=c(0.1,0.01))) +
f(ID.year1, model = "ar1") + f(ID.year2, model = "iid")
mod.t3 <- inla(formula.t3, family = "poisson", data = data.st, E
= E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))

```

```

# model including spatial and temporal interaction #
#Interaction I: excahangeable prior
#model = alpha + x1 + x2 + x3 + u_i + s_i + phi_t + psi_t +
delta_it
#2a rw(1) prior for temporal effect

```

```

formula.st2a <- y ~ 1 + x1 + x2 + x3 + f(ID.areal, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01))) +
f(ID.year1, model = "rw1") + f(ID.year2, model = "iid") +
f(ID.area.year, model = "iid")
mod.st2a <- inla(formula.st2a, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))

```

```

#2b rw(2) prior for temporal effect

```

```

formula.st2b <- y ~ 1 + x1 + x2 + x3 + f(ID.areal, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01))) +
f(ID.year1, model = "rw2") + f(ID.year2, model = "iid") +
f(ID.area.year, model = "iid")

```

```

mod.st2b <- inla(formula.st2b, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))

```

```

#2c ar(1) prior for temporal effect

```

```

formula.st2c <- y ~ 1 + x1 + x2 + x3 + f(ID.areal, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01))) +
f(ID.year1, model = "ar1") + f(ID.year2, model = "iid") +
f(ID.area.year, model = "iid")
mod.st2c <- inla(formula.st2c, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))

```

```

#3. Interaction II: RW1 prior

```

```

# Model = alpha + x1 + x2 + x3 + u_i + s_i + phi_t + psi_t + delta_it
ID.area.int <- data.st2$ID.area1
ID.year.int <- data.st2$ID.year1

#3a. rw1 prior for temporal effect and rw(1) for interaction effect
formula.st3a <- y ~ 1 + x1 + x2 + x3 + f(ID.area1, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1, 0.01)))) +
f(ID.year1, model = "rw1") + f(ID.year2, model = "iid") +
f(ID.area.int, model = "iid", group = ID.year.int, control.group
= list(model="rw1"))
mod.st3a <- inla(formula.st3a, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))

#3b. rw2 prior for temporal effect and rw(1) for interaction effect
formula.st3b <- y ~ 1 + x1 + x2 + x3 + f(ID.area1, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01)))) +
f(ID.year1, model = "rw2") + f(ID.year2, model = "iid") +
f(ID.area.int, model = "iid", group = ID.year.int, control.group
= list(model="rw1"))
mod.st3b <- inla(formula.st3b, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))

#3c. ar(1) prior for temporal effect and rw(1) for interaction effect
formula.st3c <- y ~ 1 + x1 + x2 + x3 + f(ID.area1, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01)))) +
f(ID.year1, model = "ar1") + f(ID.year2, model = "iid") +
f(ID.area.int, model = "iid", group = ID.year.int, control.group
= list(model="rw1"))
mod.st3c <- inla(formula.st3c, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))

#4. Interaction II: RW2 prior
# Model = alpha + x1 + x2 + x3 + u_i + s_i + phi_t + psi_t + delta_it

```

```
ID.area.int <- data.st2$ID.area1
ID.year.int <- data.st2$ID.year1
```

#4a. rw1 prior for temporal effect and rw(2) for interaction effect

```
formula.st4a <- y ~ 1 + x1 + x2 + x3 + f(ID.area1, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01)))) +
f(ID.year1, model = "rw1") + f(ID.year2, model = "iid") +
f(ID.area.int, model = "iid", group = ID.year.int, control.group
= list(model="rw2"))
mod.st4a <- inla(formula.st4a, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))
```

#4b. rw2 prior for temporal effect and rw(2) for interaction effect

```
formula.st4b <- y ~ 1 + x1 + x2 + x3 + f(ID.area1, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01)))) +
f(ID.year1, model = "rw2") + f(ID.year2, model = "iid") +
f(ID.area.int, model = "iid", group = ID.year.int, control.group
= list(model="rw2"))
mod.st4b <- inla(formula.st4b, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))
```

#4c. ar(1) prior for temporal effect and rw(2) for interaction effect

```
formula.st4c <- y ~ 1 + x1 + x2 + x3 + f(ID.area1, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01)))) +
f(ID.year1, model = "ar1") + f(ID.year2, model = "iid") +
f(ID.area.int, model = "iid", group = ID.year.int, control.group
= list(model="rw2"))
mod.st4c <- inla(formula.st4c, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))
```

#5. Interaction II: AR1 prior

```
# Model = alpha + x1 + x2 + x3 + u_i + s_i + phi_t + psi_t +
delta_it
```

```
ID.area.int <- data.st2$ID.area1
ID.year.int <- data.st2$ID.year1
```

#5a. rw1 prior for temporal effect and ar(1) for interaction effect

```
formula.st5a <- y ~ 1 + x1 + x2 + x3 + f(ID.area1, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01)))) +
f(ID.year1, model = "rw1") + f(ID.year2, model = "iid") +
f(ID.area.int, model = "iid", group = ID.year.int, control.group
= list(model="ar1"))
mod.st5a <- inla(formula.st5a, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))
```

#5b. rw2 prior for temporal effect and ar(1) for interaction effect

```
formula.st5b <- y ~ 1 + x1 + x2 + x3 + f(ID.area1, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01)))) +
f(ID.year1, model = "rw2") + f(ID.year2, model = "iid") +
f(ID.area.int, model = "iid", group = ID.year.int, control.group
= list(model="ar1"))
mod.st5b <- inla(formula.st5b, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))
```

#5c. ar(1) prior for temporal effect and ar(1) for interaction effect

```
formula.st5c <- y ~ 1 + x1 + x2 + x3 + f(ID.area1, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01)))) +
f(ID.year1, model = "ar1") + f(ID.year2, model = "iid") +
f(ID.area.int, model = "iid", group = ID.year.int, control.group
= list(model="ar1"))
mod.st5c <- inla(formula.st5c, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))
```

#6. Interaction III: CAR prior

```
# Model =  $\alpha + x_1 + x_2 + x_3 + u_i + s_i + \phi_t + \psi_t + \delta_{it}$ 
ID.area.int <- data2$ID.area1
ID.year.int <- data2$ID.year1
```

#6a. CAR for interaction and rw1 for temporal

```
formula.st6a <- y ~ 1 + x1 + x2 + x3 + f(ID.areal, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01)))) +
f(ID.year1, model = "rw1") + f(ID.year2, model = "iid") +
f(ID.year.int, model = "iid", group = ID.area.int, control.group
= list(model="besag", graph = mb.adj))
mod.st6a <- inla(formula.st6a, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))
```

#6b. CAR for interaction and rw2 for temporal

```
formula.st6b <- y ~ 1 + x1 + x2 + x3 +f(ID.areal, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0001)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01)))) +
f(ID.year1, model = "rw2") + f(ID.year2, model = "iid") +
f(ID.year.int, model = "iid", group = ID.area.int, control.group
= list(model="besag", graph = mb.adj))
mod.st6b <- inla(formula.st6b, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))
```

#6c. CAR for interaction and ar1 for temporal

```
formula.st6c <- y ~ 1 + x1 + x2 + x3 + f(ID.areal, model="bym",
graph = mb.adj, scale.model = TRUE, hyper = list(prec.unstruct =
list(prior = "loggamma", param = c(0.1, 0.01)),
prec.spatial = list(prior = "loggamma", param=c(0.1,0.01)))) +
f(ID.year1, model = "ar1") + f(ID.year2, model = "iid") +
f(ID.year.int, model = "iid", group = ID.area.int, control.group
= list(model="besag", graph = mb.adj))
mod.st6c <- inla(formula.st6c, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))
```

Interaction IV: RW1 distribution δ_{it} ($\alpha + u_i + s_i + \phi_t + \psi_t + \delta_{it}$)

```
ID.area.int <- data2$ID.areal
ID.year.int <- data2$ID.year1
```

#7a. rw1 for temporal structured interaction

```

formula.st7a <- y ~ 1+ x1 + x2 + x3 + f(ID.areal, model="bym",
graph = mb.adj) + f(ID.year1, model = "rw1") +
  f(ID.year2, model = "iid") +
  f(ID.year.int, model = "rw1", group = ID.area.int,
control.group=list(model = "besag", graph = mb.adj))

```

```

mod.st7a <- inla(formula.st7a, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo =
TRUE))

```

#7b. rw2 for temporal structured interaction

```

formula.st7b <- y ~ 1+ x1 + x2 + x3 + f(ID.areal, model="bym",
graph = mb.adj) +
  f(ID.year1, model = "rw2") +
  f(ID.year2, model = "iid") +
  f(ID.year.int, model = "rw1", group = ID.area.int,
control.group=list(model = "besag", graph = mb.adj))

```

```

mod.st7b <- inla(formula.st7b, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo =
TRUE))

```

#7c. ar1 for temporal structured interaction

```

formula.st7c <- y ~ 1+ x1 + x2 + x3 + f(ID.areal, model="bym",
graph = mb.adj) +
  f(ID.year1, model = "ar1") +
  f(ID.year2, model = "iid") +
  f(ID.year.int, model = "rw1", group = ID.area.int,
control.group=list(model = "besag", graph = mb.adj))

```

```

mod.st7c <- inla(formula.st7c, family = "poisson", data =
data.st2, E = E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo =
TRUE))

```

#Put the temporal effect (gamma_j+phi_j) on the natural scale

```

temporal<-lapply(model.ST1$marginals.lincomb.derived,
function(X){
  marg <- inla.marginal.transform(function(x) exp(x), X)
  inla.emarginal(mean, marg)
})

```

extracting the spatial component from the model with best fit #1. spatial model overall (mod.s2)


```

zeta.o <- lapply(mod.sl$marginals.random$ID.area1[1:96],
function(x) inla.emarginal(exp,x))
zeta.m <- lapply(mod.ms$marginals.random$ID.area1[1:96],
function(x) inla.emarginal(exp,x))
zeta.f <- lapply(mod.fs$marginals.random$ID.area1[1:96],
function(x) inla.emarginal(exp,x))
zeta.c <- lapply(mod.cs$marginals.random$ID.area1[1:96],
function(x) inla.emarginal(exp,x))
zeta.nc <- lapply(mod.ncs$marginals.random$ID.area1[1:96],
function(x) inla.emarginal(exp,x))

#-- transforming the derived temporal effect into natural scale
---#
#----- overall temporal effect: oTemp1 = structured -----#
oTemp1 <- lapply(mod.st3a$marginals.random$ID.year1,
function(X){marg <- inla.tmarginal(function(x)
exp(x), X)
inla.emarginal(mean, marg)
})
# male temporal effect
mTemp1 <- lapply(mod.mst5a$marginals.random$ID.year1,
function(X){marg <- inla.tmarginal(function(x)
exp(x), X)
inla.emarginal(mean, marg)
})
#----- female temporal effect -----#
fTemp1 <- lapply(mod.fst5c$marginals.random$ID.year1,
function(X){marg <- inla.tmarginal(function(x)
exp(x), X)
inla.emarginal(mean, marg)
})
#----- Cardia temporal effect -----#
cTemp1 <- lapply(mod.cst5b$marginals.random$ID.year1,
function(X){marg <- inla.tmarginal(function(x)
exp(x), X)
inla.emarginal(mean, marg)
})
#----- non-cardia temporal effect -----#
#
ncTemp1 <- lapply(mod.ncst3a$marginals.random$ID.year1,
function(X){marg <- inla.tmarginal(function(x)
exp(x), X)
inla.emarginal(mean, marg)
})
#----- extraction of temporal effect -----#
otemp <- mod.st3a$summary.random$ID.year1

```

```

mtemp <- mod.mst5a$summary.random$ID.year1
ftemp <- mod.fst5c$summary.random$ID.year1
ctemp <- mod.cst5b$summary.random$ID.year1
nctemp <- mod.ncst3a$summary.random$ID.year1

```

```

#----- plot of temporal effect with confidence interval -----

```

```

#

```

```

tempo_o <- read.csv("temporal_O.csv")
plot(seq(1,5),seq(0.8,1.2,length=5),type="n",xlab="5-years
period",ylab=expression(exp(gamma[t])))
lines(tempo_o$sir, col = 1, lty = 1)
lines(tempo_o$lbound, col = 1, lty = 2)
lines(tempo_o$Ubound, col = 1, lty = 2)

```

```

tempo_M <- read.csv("temporal_M.csv")
plot(seq(1,5),seq(0.8,1.2,length=5),type="n",xlab="5-years
period",ylab=expression(exp(gamma[t])))
lines(tempo_M$sir, col = 1, lty = 1)
lines(tempo_M$lb, col = 1, lty = 2)
lines(tempo_M$ub, col = 1, lty = 2)

```

```

tempo_f <- read.csv("temporal_F.csv")
plot(seq(1,5),seq(0.8,1.2,length=5),type="n",xlab="5-years
period",ylab=expression(exp(gamma[t])))
lines(tempo_f$sir, col = 1, lty = 1)
lines(tempo_f$lb, col = 1, lty = 2)
lines(tempo_f$ub, col = 1, lty = 2)

```

```

tempo_c <- read.csv("temporal_C.csv")
plot(seq(1,5),seq(0.4,2.0,length=5),type="n",xlab="5-years
period",ylab=expression(exp(gamma[t])))
lines(tempo_c$sir, col = 1, lty = 1)
lines(tempo_c$lb, col = 1, lty = 2)
lines(tempo_c$ub, col = 1, lty = 2)

```

```

tempo_nc <- read.csv("temporal_NC.csv")
plot(seq(1,5),seq(0.8,1.2,length=5),type="n",xlab="5-years
period",ylab=expression(exp(gamma[t])))
lines(tempo_nc$sir, col = 1, lty = 1)
lines(tempo_nc$lb, col = 1, lty = 2)
lines(tempo_nc$ub, col = 1, lty = 2)

```

```

#separate plot of temporal effect

```

```

#1. overall

```

```

plot(seq(1,5),seq(0.8,1.2,length=5),type="n",xlab="5-years
period",ylab=expression(exp(gamma[t])))
lines(unlist(oTemp1), col = 4, lty = 3)

```

```
abline(h=1, lty = 1)
```

#2. male

```
plot(seq(1,5),seq(0.8,1.2,length=5),type="n",xlab="5-years
period",ylab=expression(exp(gamma[t])))
lines(unlist(mTemp1), col = 2, lty = 3)
abline(h=1, lty = 1)
```

#3. female

```
plot(seq(1,5),seq(0.8,1.2,length=5),type="n",xlab="5-years
period",ylab=expression(exp(gamma[t])))
lines(unlist(fTemp1), col = 3, lty = 3)
abline(h=1, lty = 1)
```

#4. cardia

```
plot(seq(1,5),seq(0.8,1.2,length=5),type="n",xlab="5-years
period",ylab=expression(exp(gamma[t])))
lines(unlist(cTemp1), col = 6, lty = 3)
abline(h=1, lty = 1)
```

#5. noncardia

```
plot(seq(1,5),seq(0.8,2.0,length=5),type="n",xlab="5-years
period",ylab=expression(exp(gamma[t])))
lines(unlist(ncTemp1), col = 8, lty = 3)
abline(h=1, lty = 1)
```

#----- space-Time interaction -----#

#1. extracting the results for each data subset

```
o_int <-
data.frame(round(mod.st3a$summary.random$ID.area.int[,1:6],4))
# overall population
m_int <-
data.frame(round(mod.mst5a$summary.random$ID.area.int[,1:6],4)
) # male population
f_int <-
data.frame(round(mod.fst5c$summary.random$ID.area.int[,1:6],4)
) # female population
nc_int <-
data.frame(round(mod.ncst3a$summary.random$ID.area.int[,1:6],4)
) # non-cardia population
c_int <-
data.frame(round(mod.cst5b$summary.random$ID.area.int[,1:6],4)
) # cardia population
```

#----- extracting the spatial effect over time -----#

```

delta <- data.frame(year=ID.year1, ID.area =ID.area.int,
overall = o_int[,2], male=m_int[,2],
female=f_int[,2],ncardia=nc_int[,2], cardia=c_int[,2])
#----- improving the CPO and PIT estimates -----
--#

# sensitivity analysis
UN.prior1 <- "expression:
a = 0.001;
b = 100;
dens = 0-log(sigma*sigma) - log(b-a);
return(dens);"
formula.s4 <- y ~ 1 + x1 + x2 + x3 + f(ID.area1, model="bym",
graph = mb.adj, scale.model = TRUE,
hyper = list(prec.unstruct = list(prior = UN.prior1),
prec.spatial = list(prior = UN.prior1))
mod.s4 <- inla(formula.s4, family = "poisson", data = data.s2, E
= E, control.predictor = list(compute = TRUE),
control.compute = list(dic = TRUE, cpo = TRUE))
#-----
----
#posterior of spatial random effect variance marginal plot
marg.s1 <- inla.tmarginal(function(x) 1/x,
mod.s1$marginals.hyperpar$`Precision for ID.area1 (spatial
component)` )
marg.s2 <- inla.tmarginal(function(x) 1/x,
mod.s2$marginals.hyperpar$`Precision for ID.area1 (spatial
component)` )
marg.s3 <- inla.tmarginal(function(x) 1/x,
mod.s3$marginals.hyperpar$`Precision for ID.area1 (spatial
component)` )
marg.s4 <- inla.tmarginal(function(x) 1/x,
mod.s4$marginals.hyperpar$`Precision for ID.area1 (spatial
component)` )
plot(seq(-0.1, .5, 0.1), seq(0, 60, 10), type="n", xlab=
expression(sigma[s]^2), ylab = expression(tilde(p)
(paste(sigma[s]^2, "|", y))))
lines(marg.s1, lty=2, col= 1)
lines(marg.s2, lty = 3, col=2)
lines(marg.s3, lty = 1, col=3)
lines(marg.s4, lty = 5, col=4)
legend("topright", c("loggamma(0.1, 0.01)",
"loggamma(1,0.0001)", "INLA default prior", "uniform(0.001,
100)"), cex = 0.6, lty = c(2,3,1,5))

```

```

#-----
----
# posterior of non-spatial random effect marginal plot
marg.ns1 <- inla.tmarginal(function(x) 1/x,
mod.s1$marginals.hyperpar$`Precision for ID.area1 (iid
component)` )
marg.ns2 <- inla.tmarginal(function(x) 1/x,
mod.s2$marginals.hyperpar$`Precision for ID.area1 (iid
component)` )
marg.ns3 <- inla.tmarginal(function(x) 1/x,
mod.s3$marginals.hyperpar$`Precision for ID.area1 (iid
component)` )
marg.ns4 <- inla.tmarginal(function(x) 1/x,
mod.s4$marginals.hyperpar$`Precision for ID.area1 (iid
component)` )
  plot(seq(-0.05,0.30, 0.05),seq(0,14,2),type="n",xlab=
  expression(sigma[epsilon]^2), ylab = expression(tilde(p)
  (paste(sigma[epsilon]^2,"|",y))))
  lines(marg.ns1, lty=2)
  lines(marg.ns2, lty = 3)
  lines(marg.ns3, lty = 1)
  lines(marg.ns4, lty = 5)
  legend("topright", c("loggamma(0.1, 0.01)",
  "loggamma(1,0.0001)", "INLA default prior", "uniform(0.001,
  100)"), cex = 0.58, lty = c(2,3,1,5))
#-----
----
#1. marginal posterior plot of SES fixed effects
  marg.ses1 <- mod.s1$marginals.fixed$x1
  marg.ses2 <- mod.s2$marginals.fixed$x1
  marg.ses3 <- mod.s3$marginals.fixed$x1
  marg.ses4 <- mod.s4$marginals.fixed$x1
plot(seq(-0.6,0.4, 0.2),seq(0,10,2),type="n",xlab=
expression(beta[1]), ylab=
expression(tilde(p) (paste(beta[1],"|",y))))
lines(marg.ses1, col =1, lty = 2)
lines(marg.ses2, col = 1, lty = 3)
lines(marg.ses3, col = 1, lty = 1)
lines(marg.ses4, col = 1, lty = 5)
legend("topright", c("loggamma(0.1, 0.01)",
"loggamma(1,0.0001)", "INLA default prior", "uniform(0.001,
100)"), cex = 0.58, lty = c(2,3,1,5))

#2. marginal posterior plot of immigrant fixed effects
  marg.imig1 <- mod.s1$marginals.fixed$x2
  marg.imig2 <- mod.s2$marginals.fixed$x2
  marg.imig3 <- mod.s3$marginals.fixed$x2

```

```

  marg.imig4 <- mod.s4$marginals.fixed$x2
plot(marg.imig1, col =1, lty = 2)
plot(seq(-0.04,0.06, 0.02),seq(0,100,20),type="n",xlab=
expression(beta[2]), ylab=
expression(tilde(p) (paste(beta[2],"|",y))))
lines(marg.imig1, col =1, lty = 2)
lines(marg.imig2, col = 1, lty = 3)
lines(marg.imig3, col = 1, lty = 1)
lines(marg.imig4, col = 1, lty = 5)
legend("topright", c("loggamma(0.1, 0.01)",
"loggamma(1,0.0001)", "INLA default prior", "uniform(0.001,
100)"), cex = 0.58, lty = c(2,3,1,5))

```

#3. marginal posterior plot of indigenous fixed effects

```

marg.indi1 <- mod.s1$marginals.fixed$x3
marg.indi2 <- mod.s2$marginals.fixed$x3
marg.indi3 <- mod.s3$marginals.fixed$x3
marg.indi4 <- mod.s4$marginals.fixed$x3
plot(marg.indi1, col =1, lty = 2)
plot(seq(-0.04,0.08, 0.02),seq(0,120,20),type="n",xlab=
expression(beta[3]), ylab=
expression(tilde(p) (paste(beta[3],"|",y))))
lines(marg.indi1, col =1, lty = 2)
lines(marg.indi2, col = 1, lty = 3)
lines(marg.indi3, col = 1, lty = 1)
lines(marg.indi4, col = 1, lty = 5)
legend("topright", c("loggamma(0.1, 0.01)",
"loggamma(1,0.0001)", "INLA default prior", "uniform(0.001,
100)"), cex = 0.58, lty = c(2,3,1,5))

```