The Association Between Measles Cases and Migration/Settlement Patterns in Ontario

By Marcel Miron-Celis

A thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the
Master's of Science degree in Epidemiology

School of Epidemiology, Public Health and Preventative Medicine
Faculty of Medicine
University of Ottawa

My late grandfather, Bernardo, always kept high hopes that I would someday become a writer like himself. Here is my best attempt at being one. I dedicate this work to him as he continues to be a source of inspiration for me every day and to my family and close friends who supported me and believed in me as I embarked on this fruitful project.

Table of Contents

List of Tables and Figures

Abstract

Background

Measles is a serious infectious disease that contributes significantly to the burden of disease in many developing countries. In most developed nations, such as Canada, endemic transmission of measles has been declared eliminated thanks to rigorous vaccination programs, but isolated outbreaks of the disease continue to happen. Therefore, a thorough understanding of the factors contributing to these outbreaks is necessary.

Objectives

There were two main objectives of this thesis. The first objective was to assess the geospatial distribution of reported measles cases in Ontario with a goal of identifying clusters of reported measles. For this objective, the main hypothesis was that measles cases would not be randomly distributed across Ontario and instead would cluster in certain regions. The second objective was to explore some of the factors that may be associated with measles clusters. For this objective, the main hypothesis was that the proportion of immigrants, population density, low-income prevalence and education level would be associated with measles clusters.

Methods

The first objective was achieved through a thorough geospatial analysis using SaTScan and R. Individual forward sortation areas were used as the spatial unit of analysis. The analysis leveraged data from multiple sources: 2016 Census data, Ontario measles cases data from iPHIS from 2008 to 2019, a shapefile of all forward sortation areas in Canada from Statistics Canada

and centroid coordinates of forward sortation areas that were obtained using web scrapping techniques on the geolocation service of Natural Resources Canada. The maximal window size of the geospatial analysis was chosen using the maximum clustering heterogeneous set-proportion technique. The geospatial analysis was run with 99,999 Monte Carlo repetitions under a Poisson distribution using the purely spatial analysis. The Ontario population from the 2016 Census was used as the population at risk. Any cluster with a $p \leq 0.05$ was deemed statistically significant. The second objective was achieved through a case-control study: Forward sortation areas that were within statistically significant measles clusters were considered as cases and the rest of the forward sortation areas were considered as controls. Demographic data necessary to assess the factors of interest were extracted from the 2016 Census. A univariable logistic regression model was run to compute the odds ratio and test the association between the factors of interest and measles clusters. 95% confidence intervals were computed for each odds ratio. Data-curation techniques and data analysis were performed in R 4.0.4.

Results

From 2008 through 2019, 178 measles cases were identified. 82% of cases lacked necessary vaccination or vaccination records against measles, 35% of cases were linked to traveling outside of Ontario, 20% of cases reported being in contact with a known case, and 72% of cases were less than 5 years old or older than 21. Ten measles clusters were identified of which six were deemed statistically significant. These six significant clusters represented 7% of the population at risk but contained nearly 40% of all reported measles cases between 2008 and 2019. Measles clusters had a strong association with the proportion of immigrants living within them, population density and prevalence of low-income. No association was found between education level and measles clusters.

Conclusion

The results indicate that most measles cases in Ontario are unvaccinated or lack proof of vaccination; arise through secondary transmission within the province; arise from undetected transmission; and are adults or infants. Additionally, it is possible to see that the risk of reported measles cases is not randomly distributed across the province, but instead measles cases tend to cluster in certain regions. Such clusters tend to be characterized by specific population-level factors that may be contributing to the risk of reported measles. Targeted and equitable interventions are needed as we continue on the path to eradication.

Introduction: Why Infectious Diseases?

Most infectious disease epidemics can be traced back to human civilization and the complex cultural, social and economic practices that they entail. For instances, accounts of smallpox ravaging through emerging civilizations were observed as far back as the early Celestial Empire in Chinese scriptures and the ancient Egyptian civilization (1). There are also arguments to suggest that the fall of the Roman Empire was in part driven by widespread epidemics of malaria. Inadequate maintenance of drainage systems in newly conquered lands of Italy created ideal swampy conditions to maintain mosquito populations and promote the spread of malaria (2). In Europe of the Middle Ages, the plague was one of the most feared infectious diseases. The first major epidemic occurred when it was imported through the port of Genoa, Italy, in 1347. From there, it spread through continental Europe, killing nearly 25 million people—or a third of the known world population—in only five years. Contributing to the transcontinental spread of the plague were individuals who fled cities that had become epicenters of the disease, acting as intercity carriers of the plague (2). Another somber account of the complex relationship between human civilizations and infectious diseases is that of the colonization of America between the 15th and 16th century, when European colonizers introduced infectious diseases like the flu, varicella and measles, among others, to the native populations. As European colonizers and Native Americans were geographically isolated for thousands of years, the native populations had remained unblemished by the pathogens the Europeans had become accustomed to. The encounters between these two geographically isolated populations, and the epidemics that followed, killed millions, wiping out nearly 90% of the native population at the time (3). Most recently, the COVID-19 pandemic, which has given rise to over 200 million cases

and more than four million deaths worldwide at the time of writing, shows how quickly and effectively respiratory viruses can spread and kill people in a globalized setting (4).

Infectious diseases have been embedded in the fabric of civilization for millennia. Whenever humans congregate, communicable diseases tend to follow thereafter. This tendency is deeply rooted in survival mechanisms developed by bacteria and viruses through evolutionary processes that encompass three main factors. First, most pathological organisms require some form of reservoir to persist over long periods of time (5,6). A reservoir in the microbiological context entails living organisms or inanimate sites that provide the necessary resources for the basic survival of the pathogen (usually with limited reproduction). Secondly, some organisms also require a susceptible host that they can infect in order to reproduce in high quantities and advance their lineage (5,6). In some instances, the reservoir and the host can be the same for a given organism (5,6). Thirdly, viral and bacterial pathogens require some method of transmission to reach new susceptible individuals in order to reproduce in large quantities and advance their lineage (5). With multiple potential reservoirs in the likes of water supplies and livestock, along with millions of humans that can act as susceptible hosts and human activity to contribute to methods of transmission, metropoles are ideal environments for sustaining infectious pathogens.

It is by understanding these factors and interrupting the chain between them that humans have been able to reduce the burden of infectious diseases on global health. During the Soho Cholera Epidemic of 1854, John Snow pointed out that the increasing case counts were likely driven by the water from the Broad Street water pump and not the then-dominant miasma theory that pointed towards “bad air” as the main cause of cholera (7). Identifying contaminated water as the main reservoir of cholera through epidemiological surveys led to disabling the Broad Street water pump and ultimately helped halt the Cholera Outbreak in West London (7). Today,

Cholera outbreaks are practically nonexistent in developed countries thanks to water-chlorination programs and more effective sewage systems (8). Furthermore, in 1796, Edward Jenner introduced the world to the practice of vaccination. As Jenner realized that people inoculated with cowpox, which causes mild disease, would become immune to the more severe smallpox virus, he began to vaccinate volounteers with cowpox inoculum, and the practice spread through England and other European countries (9). This was the first step that was taken towards the eradication of smallpox. As humans are the only known reservoirs and susceptible hosts of the virus, smallpox could be eradicated through extensive immunization. This was finally achieved in 1980 through global vaccine programs (10). Finally, Alexander Fleming's serendipitous discovery of Penicillin in 1928 led to the first antibiotic to be mass produced and marked a turning point in a medical era overburdened by infectious diseases (11,12). Though discovered by chance, Alexander Fleming carefully studied what he called "mould juice" and pinpointed its antimicrobial properties against an array of gram-positive bacteria and some gram-negative organisms (13). Through their inhibiting action on the synthesis of the peptidoglycan layer of bacterial cell walls, Penicillin and other Beta Lactam antibiotics can effectively halt bacterial reproduction within susceptible hosts by compromising the structural integrity of pathogenic bacteria while leaving healthy human cells intact (14). Since then, antibiotics have saved millions of lives.

Whether they were a product of random chance or discovered through carefully considered epidemiological studies, these new insights pushed the boundaries of knowledge to develop new interventions that led developed countries to an epidemiological shift from an era dominated by infectious diseases to an era in which infectious diseases struggle to prevail and chronic lifestyle related diseases have become the new norm (15). As this epidemiological shift

continues, there remains multiple hurdles to overcome in order to effectively eliminate the burden of communicable diseases. Of particular concern is the measles virus and the deadly disease it causes. Though vaccine programs greatly contributed to the elimination of endemic measles in many developed countries, and a great reduction of cases worldwide from the 1990s to the early 2000s, cases and deaths due to the disease have begun to climb in recent years (16,17). In Canada, measles was declared eliminated in 1998, but isolated outbreaks continue to happen (18). As recent as 2011, Québec harbored the largest measles epidemic in North America in over a decade, totaling 725 cases, and Ontario has seen its fair share of measles cases as well, threatening Canada's measles-elimination status (19,20).

Therefore, it is necessary to understand the underlying factors that contribute to the persistence of said outbreaks to protect Canada's measles-elimination status. This thesis will focus on exploring such factors. The first chapter consists of a comprehensive literature review. This review aims to provide a thorough understanding of the origins of measles and its impacts through time, the natural history of measles, the epidemiology of measles and the status of measles in Canada. The second chapter consists of a geospatial analysis and a case-control study that investigates the association between immigrant communities in Ontario (among other factors) and measles outbreaks, building off the information presented in the first chapter. The third and final chapter contains a thorough discussion of the work presented in the context of existing literature and a general conclusion of the work.

Chapter 1: About Measles

1.1 Origins, Historical Context and Impacts Through Time

Measles morbillivirus (MeV) is a spherical, enveloped, non-segmented, single-stranded, negative-sense RNA virus that shares a long-lasting relationship with humans[18]. Its origin remains a topic of discussion, but the predominant theory links MeV to a spillover event that occurred from cattle to humans before the common era (BCE). The domestication of cattle by humans began approximately 10,000 years ago, as human societies shifted from a lifestyle of hunting and gathering towards a lifestyle of agriculture and settlements during the Neolithic revolution, marking the beginning of early civilizations (21). This newfound relationship increased the interaction between humans and cattle in a way that provided ample opportunity for a zoonotic event to take place (22).

There is an overwhelming body of evidence in support of this theory. First, MeV's closest relative is the *Rinderpest morbillivirus* (RPV), a cattle-infecting virus that is now eradicated (23). This supports the view that RPV and MeV share a commonality that is rooted in human–cattle interactions. In other words, RPV and MeV share a common ancestor that diverged and was ultimately transmitted from humans to cattle or from cattle to humans. Secondly, phylogeny of the *morbillivirus* genus provides evidence in support of zoonosis with a directionality from cattle to humans. Indeed, the *Plague des petits ruminants virus* (PPRV) is an older relative of both RPV and MeV that mainly infects sheep and goats (24). Since phylogenetic proximity between hosts facilitates cross-species transmission, it is more likely that the common ancestor of PPRV, RPV and MeV made the jump from small ruminants (e.g., sheep and goats) to cattle, which are both part of the *Bovidae* family, before eventually spilling over to humans

rather than jumping from small ruminants to humans and then cattle (25). Additionally, PPRV has been known to cause asymptomatic infection in cattle, further strengthening the evidence in support for the chain of transmission from small ruminants to cattle to humans (26,27). The question remains on when the divergence of MeV from its last common ancestor with RPV occurred. Previous studies have identified periods raging from the sixth century to the twelfth century CE (28,29). Most recently, advanced selection-aware bayesian molecular-clock modelling estimates show that MeV likely emerged in the sixth century BCE, concurrent with the rise of large metropoles in the Mediterranean region able to sustain measles transmission in the likes of Athens, Babylon, Alexandria and Rome (30–32).

It wasn't until 910 CE, when Rhazes documented the differential diagnosis of what he called *hasbah* that measles was formally described in historical literature, but it is possible that MeV was the culprit of epidemics that occurred much earlier (33). For instances, the great plague of Athens, which was famously described by Thucydides, is an epidemiological event that ambiguously points towards measles as the potential perpetrator. As it took place in the midst of the Peloponnesian War, at around 430 BCE, it is well established that the plague of Athens was a driving force in the defeat of the Athenians and ultimately led to the death of a quarter of the population in ancient Greece and the end of the Periclean golden age of Greek culture (34–36). Many infectious diseases have been considered to be at the root of this epidemic by historians and health experts alike, but the symptoms and progression of the disease described by Thucydides most closely align with those of measles or an ancient variant (37). Furthermore, the plague of Athens took place at approximately 430 BCE, which roughly concurs with the estimated emergence of MeV at 600 BCE. Measles may also provide an explanation for the deadly epidemics that struck the Roman Empire 70 years apart at the beginning of the common

era. Commonly known as the Antonine plague (165–180 CE) and the plague of Cyprian (251–270 CE), these epidemics dealt a significant blow to the Roman Empire through mass casualties, which amounted to a reduction in human activity and considerable economic contraction (38,39). As is the case with most accounts of early epidemics, equivocal descriptions by historic figures make it difficult to ascertain the root cause, but extrapolation from Cyprus's and Antonine's description make measles or smallpox seem likely candidates. Indeed, given that nearly 25% of the Roman population perished in both occurrences, it is commonly argued that the epidemics were manifestations of two different diseases to which the Romans were not immune, likely measles and smallpox, but the order in which they struck remains difficult to assert (40). Nevertheless, such accounts substantiate the cultural, economic and health impacts that measles has made through history.

Since MeV's divergence, globalization and increased interconnectivity between people have contributed to its success. Activities involving human mobilization, such as land exploration and world trading patterns, led to a globalized society through which measles could easily spread. European colonizers imported the disease to various areas of the continent of America, which produced outbreaks in many indigenous communities. For example, in 1529, a measles outbreak in what is now the island of Cuba killed two-thirds of the indigenous population. Merely two years later, a large portion of the population of present-day Honduras also perished due to measles as the disease swept through the Inca civilization (41). As early as the 19th century and even the 20th century, MeV was still being introduced to isolated communities that remained susceptible to its deadly effects. Indeed, nearly 20% of the native people of the Fiji islands fell victim to a measles epidemic when it was introduced by the most senior chief of Fiji following a diplomatic trip in 1875, while the people of Rotuma suffered a

similar fate in 1911 when a ship bearing two sick women introduced the virus onto the island (42). Measles continues to circulate globally with reoccurring outbreaks in all of the WHO regions around the globe, but some regions have eliminated endemic transmission of the disease (43).

1.2 Pathophysiology and Natural History of Measles

Given its historical impacts and continued persistence, MeV presents itself as one of the most successful pathogens known to humans. Therefore, understanding the pathophysiology of measles may help identify some of the factors that contribute to its persistence in the world and contribute to the development of an effective strategy to contain its spread in high-risk regions and eventually eradicate it all together.

MeV is primarily spread via infectious respiratory droplets and even small particle aerosols that can remain suspended in the air and infect the respiratory tract of susceptible individuals. However, the virus is viable for less than two hours at ambient temperatures on common surfaces and objects that come in contact with infectious droplets, whereas the aerosolized particles can only remain infectious for approximately 30 minutes, though some accounts of infectious aerosolized particles potentially circulating for hours have been reported (44,45). Once the virus reaches a susceptible individual, it leverages its functional proteins—which were specialized for human tropism through lengthy evolutionary processes—in order to establish an infection.

At the structural level, two envelope glycoproteins, the F (fusion) protein and the H (haemagglutinin) protein, play vital roles in the infection process and pathogenesis of the measles virus. Both the F protein and the H protein interact with each other in a way that

contributes to various stages of viral pathogenesis such as binding of the virus to cells, the fusion of the virus membrane and host-cell membranes, viral penetration and haemolysis (45,46). Once in the body, the measles virus utilizes its F and H proteins to bind to two human cell receptors known as CD46 and CD150. The CD46 receptors are expressed on all nucleated human cells, while the CD150 receptors are expressed mainly on activated T and B lymphocytes along with antigen presenting cells (47). Furthermore, MeV virions can incorporate the CypB protein to its structure to bind to the CD147 receptor located on most epithelial and neuronal cells (48). Other viral proteins important to establish a viral infection include the N (nucleoprotein), P (phosphoprotein), L (large polymerase protein), M (matrix protein), C and V proteins. The N protein plays a crucial role in the structural integrity of MeV virions, as it forms a helical nucleocapsid around the genomic RNA to create the protective ribonucleocapsid. The P and L proteins associate themselves with the ribonucleocapsid and play important roles in RNA synthesis. The M protein plays vital roles in RNA synthesis and in viral assembly by linking the ribonucleoprotein complex to the envelope glycoproteins of MeV (49). The C and V proteins are known to be important in the pathogenesis of MeV, as they interact with host cell proteins and regulate the response to infection (50–52).

Throughout its incubation period of approximately 12.5 days (from initial infection to the first onset of signs and symptoms), the measles virus replicates and spreads within its human host (53). The consensus model of measles pathogenesis pinpoints initial viral infection and replication in the epithelial cells of the upper respiratory tract, which mostly express the CD147 receptor. From there, the virus spreads to local lymphatic tissue where it replicates in necessary quantities to establish viremia and spreads to other organs such as the spleen, the lymphatic system, the lungs, the thymus, the liver, the kidney, the gastrointestinal tract, the skin and the

central nervous system, though wild-type measles preferentially infects and replicates within $CD150^{+}$ cells (i.e., activated T and B lymphocytes, antigen-presenting cells, etc.) (54,55).

By understanding the replication and spreading pattern of MeV through its human host, it is possible to make sense of the symptoms and health complications that often arise following infection. The first clinical signs of disease begin approximately 10 days after infection, with prodromal symptoms including fever, cough, conjunctivitis and coryza (54). The first prodromal symptoms are often accompanied by Koplik's spots (small white lesions on the buccal mucosa), while the distinctive generalized skin rash appears shortly afterwards, lasting 3–5 days (54). Complications tend to arise in up to 40% of patients and are often observed in unvaccinated individuals, children under the age of five and malnourished individuals (56,57). Given that MeV preferentially binds and replicates within immune cells with the CD150 receptor, the virus is also known for causing significant immunosuppression through lymphotropic properties, leading to much of its observed complications and mortality burden due to manifestations of secondary infections, such as pneumonia and diarrhoea (57). Complications within the central nervous system can also arise occasionally, likely linked to MeV's ability to bind and replicate in neuronal cells through CD147 receptors. For instances, MeV infection can cause blindness, memory loss, loss of motor function and seizures, which are a direct result from subacute sclerosing panencephalitis, a manifestation that often arises in children and young adults who were infected by the measles virus before the age of two (58,59). Additionally, measles can induce an auto-immune disorder that affects the central nervous system and leads to encephalomyelitis, mainly in older children and adults (60).

A measles infection takes significant time for the body to clear. Some reports show that MeV RNA can be detected in clinical samples approximately three months after the onset of the

rash, but individuals are mostly infective for a few days, when the viral loads are at their peak and the symptoms are most severe (61). Once the virus has been cleared, lifelong immunity against measles is typically attained through neutralising antibodies against the H protein, which is known for eliciting strong immune responses (62). Alternatively, infants can have immunity conferred through passively acquired maternal antibodies while in the womb of women with previously acquired immunity, but this form of immunity is usually temporary and fades away in early infancy (63). Therefore, the safest and most effective way to confer immunity against the measles virus is through the use of attenuated vaccines. The most widely used vaccine around the world is the Schwarz vaccine, which utilizes an attenuated version of the Edmonston strain of the measles virus and was licensed for use in the United States and other countries in 1965 (64). The vaccine is administered in two doses. The first dose is typically given between 9 and 15 months of age and usually generates protective concentrations of antibodies among 85–95% of the infants who are vaccinated(56). A second dose of the vaccine is usually given to ensure lasting immunity in individuals and achieve high levels of herd immunity (65).

1.3 The Epidemiology of Measles

Before vaccines were introduced in 1963, most children would become infected before the age of five, producing millions of cases worldwide and nearly 2.6 million deaths every year (66). Measles' health burden is further highlighted by historical accounts, in which measles has shown a mortality rate of 10–30% when introduced to completely susceptible populations (42,67). Mass-vaccination campaigns around the world have achieved considerable milestones in reducing the burden of measles at the global scale. Indeed, the Measles & Rubella Initiative, led by the WHO in partnership with other organisations, has contributed to the vaccination of over one billion children worldwide and ultimately helped reduce measles-related deaths by 73% from

536,000 in the year 2000 to 142,000 in 2018 (66,68). It is estimated that this mass-vaccination initiative has prevented nearly 21 million deaths worldwide since it first began its operations in 2001 (66). However, measles continues to hamper public health in the post-vaccine era, and there has been a recent increase in measles cases worldwide in the past few years, with a climb in deaths of nearly 50% from 2016 to 2019 (69).

Measles' mode of transmission is one of the most important driving forces for its continued success. Considering that respiratory droplets and small particle aerosols can come in contact with susceptible individuals so easily, it is not surprising that measles has a basic reproduction number (R_0) of 12–18, making it one of the most contagious diseases known to humans (70). Furthermore, individuals infected with the virus are often contagious several days before and after any distinct symptoms of measles, such as the generalized rash, which hinders quarantine control measures and makes it extremely difficult to effectively control the spread of the disease in a susceptible population (61). In fact, outbreaks can occur in populations in which fewer than 10% of people are susceptible and nearly 90% of susceptible individuals who come into close contact with a case can develop the disease (65,71).

Humans are the only known reservoirs of the measles virus. This has many implications from an epidemiological perspective. First, it implies that the measles virus requires a critical number of susceptible individuals who can sustain endemic transmission through an unbroken chain of susceptible hosts, known as the critical community size. This critical community size is estimated to be between 250,000 and 400,000 (72). Secondly, it explains the cyclical pattern of measles outbreaks. Indeed, when measles establishes itself as an endemic disease, outbreaks tend to occur in cyclical patterns. The cyclical nature of measles outbreaks in endemic regions is mainly due to an accumulation of susceptible individuals over sequential birth cohorts followed

by an ensuing decline of susceptible individuals in the aftermath of outbreaks. The typical temporal pattern is characterised by yearly seasonal epidemics within major epidemic cycles with a periodicity that can vary between two and five years (73). Finally, it also implies that measles can be eradicated if population immunity reaches sufficiently high levels. However, this can be challenging, since measles is among the most contagious infectious diseases. Indeed, given measles' R_0 value of 12–18, and its two-dose vaccine effectiveness of approximately 98%, vaccine campaigns would have to reach approximately 90–95% of the population to have a meaningful chance at eradicating measles (65). Some countries have made great progress in the global effort towards measles eradication, but one cannot confuse progress with success, especially since developed countries like Canada continue to see sporadic outbreaks.

1.4 Measles in Canada

Canada's battle against measles has been a lengthy one. Data going as far back as 1924 shows that Canada had from 10,000 to over 80,000 cases of measles or 7,700 cases per million population at its peak and nearly 1000 related deaths every year before the introduction of vaccines (74). Even after the measles vaccine was introduced and licensed for use in 1963, Canada was seeing thousands of cases of measles every year (75). Indeed, as early as the beginning of the 1990s, Canada was reporting over 6000 cases of measles in some years, and this was the situation even though one-dose vaccine coverage ranged from 95% to 100% in most regions (76–78). This encouraged Canada to introduce an intensive two-dose or catch-up vaccine campaign in 1996–1997 (78). Thanks to these comprehensive vaccine programs, measles was declared eliminated in Canada in 1998 (18). Within a fully functioning surveillance system, this classification means that Canada has effectively disrupted endemic transmission of measles

within its population so that any new outbreak dies out and that imported cases introduced into the population are not able to re-establish endemic transmission of the disease (18,65).

Despite its elimination status, the country continues to struggle with sporadic outbreaks of measles. Indeed, most outbreaks in Canada arise in under-vaccinated communities that have been in contact with index cases who travelled to international regions with continued endemic transmission of measles (79–86). Given the nature of said outbreaks, the annual incidence rate of measles varies considerably from year to year. For instances, the annual incidence rate of measles in Canada for 2013, 2014, 2015, 2016, 2017, 2018, and 2019 were 2.4, 11.8, 5.5, 0.3, 1.2, 0.8 and 3 cases per million population, respectively (79–85). Though these rates are considerably lower than those seen before the introduction of extensive vaccine programs, the outbreaks behind them continue to threaten Canada's elimination status. Consequently, the study presented in the next chapter focuses on studying the characteristics of measles cases and related outbreaks in Ontario. Specifically, we examine the geospatial distribution of said cases and the factors associated with them in an attempt to gain new insights regarding the status of measles in Ontario and inform public health policy.

Chapter 2: Geospatial Analysis of Measles risk in Ontario: A Case-Control Study

2.1 Introduction

Although Canada lagged behind other developed nations in the fight against measles eradication in the 1980s and 1990s, Canada has redeemed itself as one the world leaders in the area and posited itself as one of the few countries with a measles-elimination status (78). However, the information provided above highlights the need for continued efforts in order to contain the spread of measles in Canada and maintain the country's elimination status. Ontario is the most populous province in Canada, containing approximately 40% of the country's population (87). Surprisingly, only a small number of measles cases have been reported in Ontario. Indeed, only 10% of all measles cases reported in Canada from 2013 to 2019 resided in Ontario (20,79–85). Official public-health sources state that most measles cases in Ontario and the rest of Canada tend to arise from sporadic travel cases and secondary cases that occur when a susceptible individual encounters a person with travel-acquired measles (20,79–85). Although this may be true, such statements give the impression that most measles cases are random in nature and have no specific epidemiological patterns, which suggests that there is little to be done to address the issue. However, further exploration of the issue brings to light some factors that may be contributing to the occurrence of measles in Ontario, which merit a thorough investigation.

Ontario's vaccine-surveillance practices are a good starting point to identify areas of improvement. Ontario houses immunization records in the Digital Health Immunization Repository (DHIR), which utilizes a reporting a system known as Panorama to collect and report vaccine-coverage data for various vaccine-preventable diseases, including measles, across all the

public-health units (88). Parents and guardians are required to submit immunization records of their child at the time of school enrolment and/or when assessment activities are carried out by their corresponding public health-unit. Through this system, Ontario conducts yearly immunization coverage assessments to ensure that Ontario's population maintains sufficiently high vaccination rates. The most recent report, published in August 2020, showed that two-dose measles vaccine coverage among Ontario's school pupils aged seven and seventeen years old for the 2018–2019 school year reached 86.6% and 95.5%, respectively (89). Although these seem like encouraging figures, two elements of concern standout from this report. First, as the title of the report and its corresponding methods suggest, vaccine-coverage surveillance is only performed for school-aged children in Ontario. This means that individuals who are not attending school or did not attend school during their childhood in Ontario are systematically missed from vaccine surveillance practices. Second, although the average vaccination proportion remains relatively high among school pupils in Ontario, the report shows that some public-health units are reporting two-dose vaccination rates as low as 50% for seven-year-olds and below 90% for seventeen-year-olds. This insinuates that there may be clusters of under-immunized individuals at a higher risk for contracting measles.

Owing to the fact that individuals unexposed to the provincial school system are being systematically missed in measles-vaccination surveillance, there is some merit in conducting a thorough analysis to identify particular groups that may require additional public-health interventions and support. A particular population subgroup of concern are migrant communities. Immigrants arriving to Ontario before the age of seven or past the age of seventeen are likely being missed by Ontario's vaccine-surveillance practices. As nearly 40% of immigrants to Canada tend to settle in Ontario, this likely leads to tens of thousands of individuals not being

captured in measles-vaccination programs and can produce pockets of under-immunized migrant communities, which puts them at higher risk for measles outbreaks (90). Additionally, newcomers tend to be unfamiliar with the Canadian healthcare system, so there are issues pertaining to health literacy and the use of health services among these communities in Canada. Indeed, it is well known that existing language, cultural and economic barriers translate into a lack of knowledge and even a sense of hesitancy regarding available health services and information, such as vaccination programs (91,92).

Furthermore, given the fact that some public-health units are showing subpar immunization rates among school pupils while the province in general maintains relatively high immunization rates, there is a rationale to explore the possibility of geospatial differences for the risk of measles in the province not only for migrant communities, but also for vaccine hesitant communities and densely populated areas. As lack of vaccination or immunity against measles is the most prominent risk factor for developing the disease, factors associated with vaccine hesitancy and how they vary throughout the region should be explored. For instance, parents with a lower income level tend to have greater concern about the safety and necessity of vaccines compared to those with a higher income level, despite the overwhelming body of evidence in support of vaccines (93–96). Education level has also shown to be associated with vaccine hesitancy, with individuals with a lower education level typically expressing more concerns about vaccines than those with a higher education level (93,94,96,97). Finally, higher population density has been found to be associated with an increased risk of measles outbreaks in previous studies, which justifies further exploration in the Ontarian context (98).

In light of this information, this study seeks to identify high-risk areas for sporadic measles outbreaks in Ontario through a geospatial analysis. Then, using the results from the

geospatial analysis, the study will focus on identifying factors that may be associated with a higher risk for measles outbreaks through a case-control study. The first hypothesis is that the risk for measles will not be randomly distributed across the province; instead, measles cases will appear to cluster in certain areas, creating regions with higher risk for measles cases. The second hypothesis is that a higher proportion of migrants, higher population density, lower income and lower education level will be associated with an increased risk of measles clusters.

2.2 Methods

Some deviations from what was originally planned are to noted before providing a thorough description of the study methods. The original plan was to determine measles-importation areas in addition to the geospatial analysis and case-control study. This required linking genotype data from the integrated Public Health Information System (iPHIS) of Ontario to the Measles Nucleotide Surveillance (MeaNS) database from the World Health Organization. Although steps were taken to gain access to both datasets in order to achieve this objective, it was only possible to obtain data from iPHIS, as the administrators from the MeaNS database did not respond to inquiries regarding data access. Apart from this deviation, the study was carried out as planned.

2.2.1 Data sources and Variables Used

This study leveraged data from multiple sources to achieve its objectives. Data from the integrated Public Health Information System (iPHIS) of Ontario was used to conduct the geospatial analysis of high-risk areas for sporadic measles outbreaks in Ontario. This database collects data pertaining to cases of multiple diseases of public-health significance and vaccine-preventable diseases for surveillance purposes, including all measles cases reported in Ontario

along with demographics and potential risk factors associated with each case (99). Data are usually entered into the system by a qualified case manager from the corresponding public-health unit of the case. The data can be accessed by interested stakeholders through a formal data request to Data and Information Management Services at Public Health Ontario (100). From this dataset, the variables that were used include the unique case number of each measles case that was reported between 2008 and 2019, the year the case was reported, the age group of each case, their respective forward sortation area (FSA), the genotype of the MeV virus involved with the case and aggregated risk factors such as immunization status, close contact with a case and travel outside Ontario. FSAs designate a geographical unit within Canada and are represented by the first three characters of a postal code. For example, all postal codes that start with K1G fall under the same FSA. A measles case within the iPHIS database was defined as an individual with a laboratory-confirmed infection of MeV with clinically compatible signs and symptoms or an individual with clinically compatible signs and symptoms with a known epidemiologic link to a laboratory-confirmed case of measles (101).

Data from the 2016 Canadian Census was also necessary to conduct the geospatial analysis as well as the case-control study. This dataset is easily accessible by any member of the public online (102). The Canadian Census is a large population-based survey carried out by Statistics Canada every five years. The survey is required, by law, to be completed by each household in Canada and essentially covers the bulk of the Canadian population, with a few exceptions. As a result, the response rate for the 2016 Census was 98.4% for the short-form questionnaire and 97.8% for the long-form questionnaire (103). The short-form questionnaire is distributed to all Canadian dwellings and captures general information pertaining to the number of people living in a given dwelling, the language spoken, their age, their sex and their marital

status (104). The long-form questionnaire was distributed with a one-quarter sampling fraction using a systematic approach where a random dwelling is picked along with every fourth dwelling thereon from the coverage-unit list. Then, weighting techniques were used by Statistics Canada to extrapolate the results from the sampled dwellings to all dwellings in the prescribed coverage unit (105). It includes the same dimensions as the short-form questionnaire, but it goes a step further in collecting information on activities of daily living, sociocultural information, mobility, place of birth of parents, education and labour market activities (106). Specific variables that were sought out from this dataset captured population counts, immigrant status, low-income prevalence and education level. These variables were grouped at the FSA level by Statistics Canada. Immigrants were defined as individuals who were not born in Canada and arrived through the immigration system at any specified time period. Non-permanent residents were also included in the immigrant category for the purpose of this study. The proportion of immigrants was computed by dividing the number of immigrants by the sum of total immigrants and non-immigrants within a given FSA. The prevalence of low-income earners was ascertained by Statistics Canada by computing the proportion of indivduals that are below the Low-income cut-off threshold within a given FSA. The proportion of post-secondary educated individuals was ascertained by dividing the number of indviduals aged 25 years or older who completed any form of post-secondary education by the total number of individuals aged 25 years or older within a given FSA.

Two additional data sources were necessary to complement the iPHIS dataset and the 2016 Census data. A cartographic boundary file (i.e., a map) of all the FSAs in Canada, created by Statistics Canada with permission from Canada Post Corporation, was used for the geospatial analysis. The cartographic boundary file can be accessed by anyone interested through Statistics

Canada's website (107). This file depicts the full extent of FSAs and their geographic boundaries along with the shoreline of the major land mass of Canada and its coastal islands, which provides the necessary framework for mapping and geospatial analysis on statistical software. The cartographic boundary file was also necessary to estimate the land mass area in squared kilometers of each FSA. The geolocation service from Natural Resources Canada was also used to obtain centroid coordinates of all the FSAs in Ontario which were expressed as longitude/latitude in decimal degrees. The geolocation service is an open-source tool available for use by anyone through their website (108). This provided the necessary location indicators to conduct the geospatial analysis.

2.2.2 Data Curation

Significant data curation was required in preparation for the statistical analyses. Data curation techniques were executed in R-4.0.3 using the following packages: tidyverse, readxl, rvest, spdep, and raster (109–114). The full extent of the R script that was written to complete this study can be found in Appendix A.

In preparation for the geospatial analysis, data from multiple sources had to be collected, cleaned and merged into a working dataset. First, each unique case number from the iPHIS dataset was used to create measles case counts, and case counts were then grouped by FSA. Any measles cases that were not assigned to an FSA were excluded. Second, the 2016 Census dataset was manipulated to extract all data for FSAs within Ontario to facilitate the data-curation process. This was done by selecting all FSAs that started with either K, L, M, N or P, which are the first letters of all FSAs in Ontario (115). Then, the FSAs and their corresponding population total for 2016 were extracted from the 2016 Census dataset and merged with the measles case-

count data by the FSA variable, which was the observational unit. Furthermore, web scraping techniques were used to extract the latitude and longitude coordinates of FSA centroids using the link listed in the geolocation service website from Natural Resources Canada. In simple terms, each FSA from the previously merged dataset was autonomously and systematically entered into the geolocation service link, which produced individual web pages from which the corresponding coordinates were autonomously extracted by selecting specific HTML nodes. The scraped coordinates were then cleaned using string-pattern matching and merged by FSA into the existing dataset containing measles counts and population counts. Any FSAs that had no coordinates associated to them were excluded. This constituted the final working dataset for the geospatial analysis. A cartographic boundary file of Ontario FSAs was also necessary for the geospatial analysis. Using the cartographic boundary file of FSAs in Canada, all the FSAs in Ontario were selected to obtain a shapefile or a map of Ontario FSAs. The shapefile of Ontario FSAs was then used to create a list indicating contiguous relationships between all the FSAs in Ontario. Any FSAs from the final geospatial analysis dataset that were not in the Ontario shapefile were excluded.

Similar approaches were used to prepare the working dataset for the case-control study with some minor deviations. First, all data for Ontario was extracted from the 2016 Census dataset. Then, relevant variables such as immigrant status, total population, low-income prevalence and education level were extracted and merged into a working dataset where the unit of observation was each individual FSA. Output from the geospatial analysis was also used to create a binary variable to identify case FSAs as 1 and control FSAs as 0, where cases were FSAs within measles clusters and controls were FSAs outside measles clusters. The geometry list of the Ontario shapefile, which contains the polygon shape corresponding to each FSA, was also

merged into the case-control dataset. The geometry list created by Statistics Canada used the NAD83 coordinates reference system, which were changed to the WGS84 coordinates reference system to streamline the analysis. The geometry list was used to estimate the land mass of each FSA, determine the population density of each FSA and map the general areas of the identified clusters along with variables of interest. Finally, any FSAs that were not in the Ontario shapefile were excluded from the case-control dataset.

Finally, the iPHIS dataset also had to be manipulated to create histograms of each identified cluster. The output from SaTScan, which provided all FSAs within each cluster, was used to assign a cluster to each measles case in the iPHIS dataset. A categorical variable indicating the cluster that each case belongs to was created. All cases that had an FSA that was found to be within a given cluster were assigned a number corresponding to that cluster. Cases that did not belong to any identified cluster were assigned a 0.

2.2.3 Statistical Analysis

The statistical analysis was also done in R-4.0.3 using the rsatscan, gmodels and cowplot packages (109,116–118). As rsatscan runs in conjunction with SaTScan, SaTScan was also installed (119). Additionally, a draft package called ParSatscan was used. This package is yet to be published in the R ecosystem and was kindly provided by Wei Wang from the Western China School of Public Health at Sichuan University. The R script containing the full extent of the data analysis can also be found in Appendix A.

Although the statistical computations for the geospatial analysis are calculated using purposely made software, it is important to understand the theoretical framework behind such techniques to acknowledge their validity. SaTScan utilizes a spatial scan statistic, which was first

described by Kulldorff, to detect the presence and location of geographic clusters of a given outcome within spatial datasets similar to the one described above (120). The spatial scan statistic is used to identify the maximum likelihood clusters (i.e., likely clusters) using a scanning window z, which contains a number of spatial units, to reject the null hypothesis that the outcome is randomly/equally distributed in a given study area (i.e., no clusters exist) and accept the alternative hypothesis that the outcome is not randomly distributed in a given study area (i.e., there are clusters). In this case, the spatial units are individual FSAs, the study area is Ontario and the outcome is measles cases. The logarithm of the likelihood ratio of a scanning window in question, $LLR(z)$, is computed under a Poisson distribution using the following formula:

$$LLR(Z) = ln\left\{\left(\frac{c_z}{n_z}\right)^{c_z}\left(\frac{C-c_z}{C-n_z}\right)^{C-c_z}\right\},$$

where C represents the total number of observed cases in the study area (i.e., Ontario) while n_z and c_z are the expected cases in z under the null hypothesis and the observed cases in z, respectively. Monte Carlo simulation is used to scan the study area, which produces a distribution of the scan statistic and a critical value under the null hypothesis that can be used to compare to the test statistic of a given scanning window. The critical value is determined by the desired α level of the researcher, which determines the rate of Type 1 error. If the goal is to identify high-risk areas with an α level of 0.05, then a test statistic for a given scanning window that is in the top 5% of the distribution is considered statistically significant. . If the goal was to identify low-risk areas, then the criterion would be inversed. The number of Monte Carlo simulations, among other factors, influences the statistical power (i.e., β or Type 2 error rate), reflecting the number of scans done over the study area. In SatScan, the number of replications

must be at least 999 to ensure excellent power for all types of data sets while 9,999 replications are recommended for small to medium size data sets.

Most of the input parameters for the scan statistic are pre-determined. Indeed, the expected number of cases and the observed number of cases are determined by the case counts and population counts in the input data set. However, the size of the scanning window, known as the maximum window size (MWS), has to be determined by the researcher and is a critical parameter that has a major influence on the detected clusters. The MWS is described by a proportion of the total population at risk (i.e., the total population in the study area). For instance, if the MWS is set at 15%, then the sum of the population from each spatial unit (i.e., FSA) included in a given scanning window would be between 0% and 15% of the total population at risk. This means that different MWSs will produce different sets of scanning windows (z) and ultimately different detected clusters, which brings up the issue of selecting the optimal MWS to maximize the detection of true clusters (i.e., true positives) while limiting the detection of false clusters (i.e., false positives).

An array of methods and indicators have been used to decide on the optimal MWS. The SaTScan user manual gives vague guidelines on the selection of MWS while Kulldorff has reported that a MWS of 50% can generally be used to limit false-negative clusters (121,122). However, such a window size can lead to very large clusters that falsely cover regions between several small real clusters, therefore increasing the false-positive rate. Given this issue, investigators have proposed and employed multiple indicators in an attempt to select the optimal MWS. For instance, classic performance measures such as sensitivity, specificity, positive predictive value and Youden's index have been used in an attempt to capture the misclassification of clusters (123,124). Though these methods are relatively robust in theory,

they are impractical for use in real-world datasets like the one used for this study. Indeed, without any prior knowledge of true clusters within the study area, one cannot compute these classic performance measures to identify the optimal MWS. Another indicator that has been more commonly adopted, thanks to its implementation in the SaTScan software, is the Gini coefficient. The Gini coefficient essentially uses the reported clusters obtained from multiple window sizes smaller than or equal to the chosen MWS to identify the optimal Maximum Reported Window Size (MRWS). Using this method, the MRWS with the largest Gini coefficient is deemed the most optimal, and its corresponding clusters are considered to be the most accurate (125). Although the Gini coefficient generally provides better accuracy of detected clusters over the generic 50% MWS method, it is still likely to report overly large clusters that lead to some false positives; this problem is exacerbated with complex datasets that have heterogenous clusters (i.e., clusters with significantly different relative risks) (126).

More recently, the maximum clustering heterogeneous set-proportion (MCHS-P) was proposed by Wang et al. to select the maximum window size for the spatial scan statistic. In essence, this approach conducts a sensitivity analysis of a range of window sizes that is chosen by the researcher to select the optimal MWS that produces the clusters with the highest relative risks. This technique was shown to be the most effective at selecting the optimal window size to avoid detecting overly large clusters, which improves the accuracy of the results in practical datasets, even when heterogenous clusters are present. Indeed, the MCHS-P has shown an overall sensitivity and specificity of 82% and 99%, respectively (126). This represents a modest improvement over the Gini coefficient, which had sensitivity and specificity of 81% and 96%, respectively (126). However, it is important to note that the MCHS-P far outperforms the Gini coefficient in complex scenarios where there are heterogenous clusters close to each

other.Therefore, when using datasets where there is no prior knowledge of true clusters, such as the one used for this study, the MCHS-P is the most appropriate method to use to select the MWS.

As the MCHS-P tends to prevail over other methods, this technique was used to select the MWS during the geospatial analysis of measles cases in Ontario. To do so, the ParSatscan package described earlier was used. The MCHS-P requires multiple input parameters in order to compute the optimal MWS. Namely, the FSA code was used as the location ID for each spatial unit, the measles case counts along with population counts by FSA were used and the data list indicating contiguous relationships between FSAs was also imputed. The latitude and longitude coordinates of each spatial unit was also necessary. Finally, the maximum window sizes tested ranged from 0.1% to 50% with incremental increases of 0.1%. Each window size was tested using 999 Monte Carlo repetitions for a total of 499,500 repetitions. The analysis type was purely spatial, and a Poisson model was used.

Then, using the output MWS from the MCHS-P, a formal geospatial analysis was performed to identify clusters of measles in Ontario. The input parameters were similar to those used for the MCHS-P. Since this thesis focused on identifying regions with higher risk and not time periods of higher risk for reported measles cases, a purely spatial analysis was performed. The MWS was set to the optimal MWS identified by the MCHS-P technique. As only a single MWS was used for the formal geospatial analysis, more Monte Carlo repetitions could be run to increase statistical power while remaining within computational constraints. Therefore, a total of 99,999 Monte Carlo repetitions were done for the geospatial analysis. Potential case clusters were selected using a significance level of 0.05 or lower.

Using the results from the geospatial analysis, a case-control study was constructed where the cases were all FSAs within high-risk areas for measles (i.e., clusters) and the controls were all FSAs not within high-risk areas for measles. Factors such as immigrant proportion, population density, prevalence of low-income earners and prevalence of post-secondary education were considered to test their association with measles clusters. As these variables did not respect the linearity assumption they were assessed as categorical variables. For the immigrant-status variable, exposure for a given FSA was defined as having a proportion of immigrants larger than the proportion of immigrants in the general Canadian population of 21.9%; FSAs with a proportion of immigrants equal to or less than the proportion of immigrants in the general Canadian population were considered unexposed (127). Exposure for the population-density variable was defined as having a population density greater than 1000 people/km^2; an unexposed FSA would have a population density equal to or less than 1000 people/km^2. This threshold for the population-density was chosen based on a previous study that showed that a population density greater than 1000 people/km^2 tends to increase the risk for measles outbreaks (98). The exposure categories for educational attainment were determined by proportion of individuals that completed post-secondary education within each respective FSA. FSAs with a proportion of completed post-secondary education among 25 to 64 year olds lower than the Canadian average of 54% would be considered exposed while those with a proportion equal to or greater than 54% would be considered unexposed (128). Finally, the income exposure was also divided into two groups. FSAs with a prevalence of low income (based on the low-income cutoffs) higher than the Canadian average of 8.8% were categorized as exposed, whereas FSAs with a prevalence of low income equal to or smaller than the Canadian average were categorized as unexposed. Due to sample size limitations and small cell counts for some variable

categories, it was not possible to assess the independent effects of each exposure of interest using a multivariable logistic regression model. Therefore, to measure the association between these variables and measles hotspots, the odds ratio of each exposure and their respective 95% confidence intervals between case and control FSAs were computed using a univariable logistic regression model.

Finally, multiple tables and figures were created to complement the results. Using aggregated data from the iPHIS dataset, summary statistics were computed to help characterize the risk factors of reported measles cases within Ontario, such as immunization status, age group, travelling outside Ontario or being in close contact with a case. Given that data for risk factors were aggregated by Public Health Ontairo, it was not possible to exclude the same cases that were removed using the eligibility criteria mentioned earlier, so the data should be interpreted with this in mind. In addition to the characterization of risk factors for measles cases, histograms were created to help characterize each cluster, and mapping techniques were used to create a map of the general areas of measles clusters in Ontario along with the variables of interest from the case-control study (i.e., population density, immigrant status, low-income prevalence and education level). For the histograms, the measles case count from each cluster was plotted by year and grouped by MeV genotype. Any cases that were missing information regarding the MeV genotype involved in the infection were labeled as “NA”. To map the clusters, an elliptical shape was plotted around coordinates that were known to be within each cluster. The elliptical shapes were purposely made larger than the true area of each cluster to include additional FSAs and prevent the direct identification of case FSAs. Nevertheless, they provide a useful visual representation of measles clusters. The variables of interest from the case-control study were

individually added as an additional layer on the map to provide a visual representation of the geographical distribution of the variables and their relationship with the measles clusters.

2.2.4 Ethical Considerations and Ethics Approval

The intentions of this project are to identify areas in Ontario with a higher risk for reported measles cases with the goal of providing the information necessary for targeted interventions. Although the intentions of this project are to help improve the overall health of Canadian the population, it is important to recognize that the results of this research project could be used to stigmatize the very communities that it intends to help. To mitigate this issue, the three letter codes of control and case FSAs were not disclosed throughout this dissertation nor were their geolocation. Additionally, when mapping the clusters, the cluster area was purposely made larger to avoid the direct identification of case FSAs or their specific location. Ethics approval was obtained from the Office of Research Ethics and Integrity of the University of Ottawa. A letter indicating such approval can be found in Appendix B.

2.3 Results

2.3.1 Geospatial Analysis

From 2008 through to 2019, a total of 178 measles cases were identified through the iPHIS surveillance system. Of these 178 cases, 28 did not have an FSA associated with them, which led to their exclusion from the geospatial analysis. The 2016 Census had a total of 523 FSAs containing 13,448,492 people. One FSA was excluded from the analysis due to missing coordinates, but no measles cases were reported in this FSA, and it had a small population of 723. An additional nine FSAs were excluded due to the fact that they were not present in the cartographic boundary file. These FSAs appeared to be non-residential areas (e.g., government

institutions, airports, hospitals, etc.). Additionally, no measles cases were reported in these FSAs, and their individual population was 589 or lower. Following these exclusions, a total of 150 measles cases (16%excluded due to missing FSAs) were used for the geospatial analysis along with 513 FSAs containing 13,447,125 people (<0.01% excluded), which served as the population at risk

Table 1 provides a characterization of the risk factors associated with the 178 measles cases that were identified in Ontario. One can see from this table that the majority of measles cases in Ontario lacked vaccination against measles or proof of vaccination, as no vaccination records were available in iPHIS. Only 35% of cases reported travelling outside of Ontario prior to symptom onset, while merely 20% of cases were reportedly in close contact with known cases of measles. Furthermore, nearly half of measles cases were 21 years of age or older.

Table 1: General characteristics of measles cases in Ontario by year as reported in Ontario's iPHIS system

Characteristics	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total (%)
Immunization Status													
≥ 1 dose	8	3	2	0	1	3	6	5	1	0	1	2	32 (18%)
Unimmunized	1	0	1	6	2	8	13	9	5	2	5	8	60 (34%)
Not recorded	49	4	6	2	0	4	3	6	1	5	2	4	86 (48%)
Travel Outside Ontario													
Yes	3	1	4	6	2	6	11	1	6	5	7	11	63 (35%)
No	55	6	5	2	1	9	11	19	1	2	1	3	115 (65%)
Contact with a Case													
Yes	0	0	0	4	0	8	12	5	1	3	1	2	36 (20%)
No	58	7	9	4	3	7	10	15	6	4	7	12	142 (80%)
Age Group													
≤5	9	1	2	1	2	9	6	4	6	1	3	3	47 (26%)
6-20	19	5	4	5	1	2	4	5	0	1	1	2	49 (28%)
≥21	30	1	3	2	0	4	12	11	1	5	4	9	82 (46%)

The optimal MWS for the geospatial analysis, as determined by the MCHS-P technique, was 4.1% of the population at risk. Using the optimal MWS and the input parameters described in the methods, 10 different measles clusters were identified with SaTScan, of which six were deemed statistically significant. The identified clusters are described in detail in Table 2. Briefly, the first cluster had a rate of measles that was 4.98 times higher ($p<0.001$) than outside the cluster. The second cluster had a rate of measles that was 21.43 times higher ($p<0.001$) than outside the cluster. The third cluster had a rate of measles that was 26.15 times higher ($p=0.02$) than outside the cluster. The fourth cluster had a rate of measles that was 25.75 times higher ($p=0.02$) than outside the cluster. The fifth cluster had a risk for observed measles cases that was 14.43 times higher ($p=0.03$) than outside the cluster. The sixth cluster had a risk for observed measles cases that was 4.02 times higher ($p=0.04$) relative to the risk outside the cluster. The seventh, eighth, ninth and tenth clusters had relative risks for observed measles cases of 10.74 ($p=0.08$), 17.28 ($p=0.35$), 9.73 ($p=0.41$) and 2.63 ($p=0.99$), respectively. Overall, the regions that were within the ten measles clusters contained 1,282,177 people, which represents approximately 9.5% of the study population. With only 9.5% of the study population, these regions collectively reported 77 cases of measles from 2008 through 2019, which is 51% of all measles cases included in the study. The regions within the six statistically significant clusters contained 940,126 people, which represents approximately 7% of the study population, while 58 measles cases—or 40% of all cases included in the study—were reported within these six significant clusters.

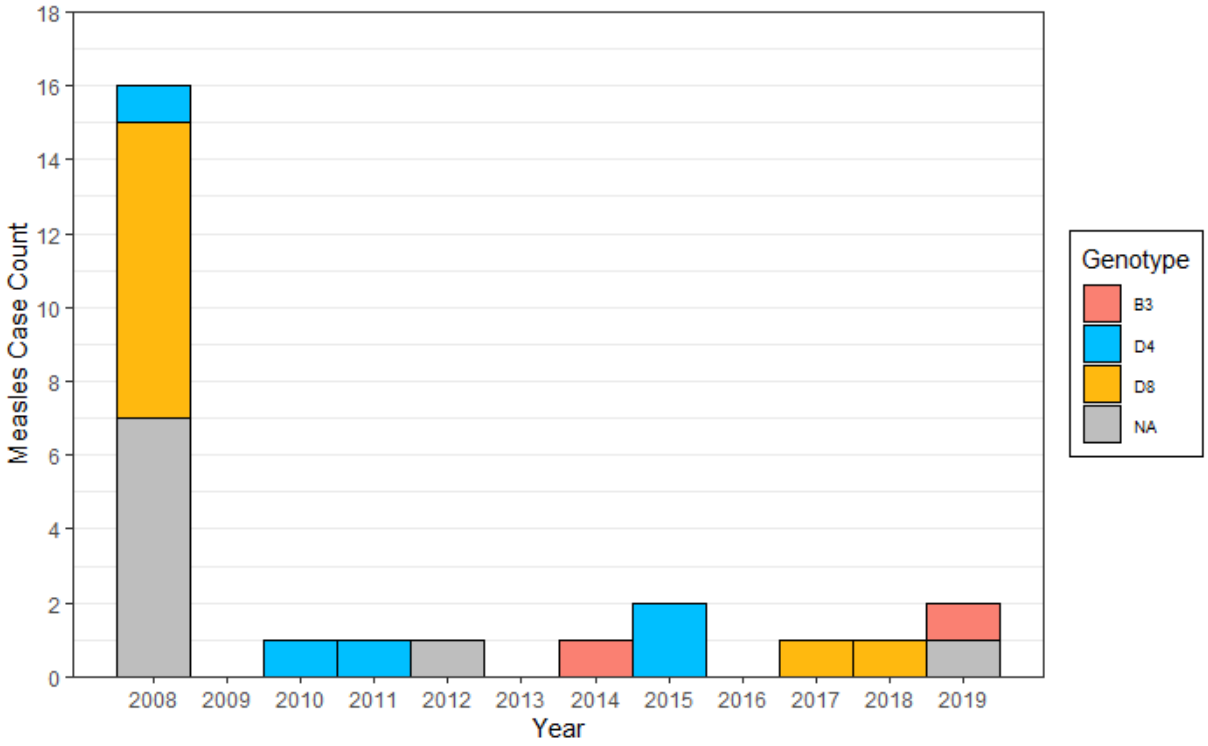
Table 2: Cluster characteristics from geospatial analysis

Region	FSAs	Population N = 13,447,125	Measles Cases N = 150	Relative Risk*	*p* value
Ontario	513	13,447,125	150	NA	NA
Cluster 1	23	543,204	26	4.98	<0.001
Cluster 2	1	26,096	6	26.15	0.001
Cluster 3	3	14,074	4	26.15	0.02
Cluster 4	1	14,311	4	25.72	0.02
Cluster 5	1	32,061	5	14.43	0.03
Cluster 6	12	310,380	13	4.02	0.04
Cluster 7	1	43,047	5	10.74	0.10
Cluster 8	1	15,866	3	17.28	0.37
Cluster 9	1	37,764	4	9.73	0.43
Cluster 10	6	245,374	7	2.63	0.99

*Relative risks of reported measles cases are computed by SatScan by dividing the estimated risk within the cluster by the estimated risk outside the cluster

Figures 1 through 6 depict the measles case counts in each of the six significant clusters plotted by year and grouped by the genotype of the measles virus responsible for each case. For Cluster 1, one can see that the majority of the cases occurred in 2008 (*n=16/26*); for that year, 50% of the cases were caused by MeV D8 genotype. Additional cases were recorded for 2010, 2011, 2012, 2014, 2015, 2017, 2018 and 2019, but these years saw no more than two cases each that were caused by varying genotypes of MeV such as B3, D4 and D8.

Figure 1: Count of reported measles cases within Cluster 1 by year and grouped by the genotype of the virus



In Cluster 2, all of the cases (*n=6/6*) were reported in 2015, and 83% of those cases were caused by MeV D4. In Cluster 3, half of the cases (*n=2/4*) occurred in 2008, whereas the other half (*n=2/4*) occurred in 2015. Both cases in 2008 were caused by the D8 genotype, whereas the cases in 2015 were caused by the D4 genotype.

Figure 2: Count of reported measles cases within Cluster 2 by year and grouped by the genotype of the virus

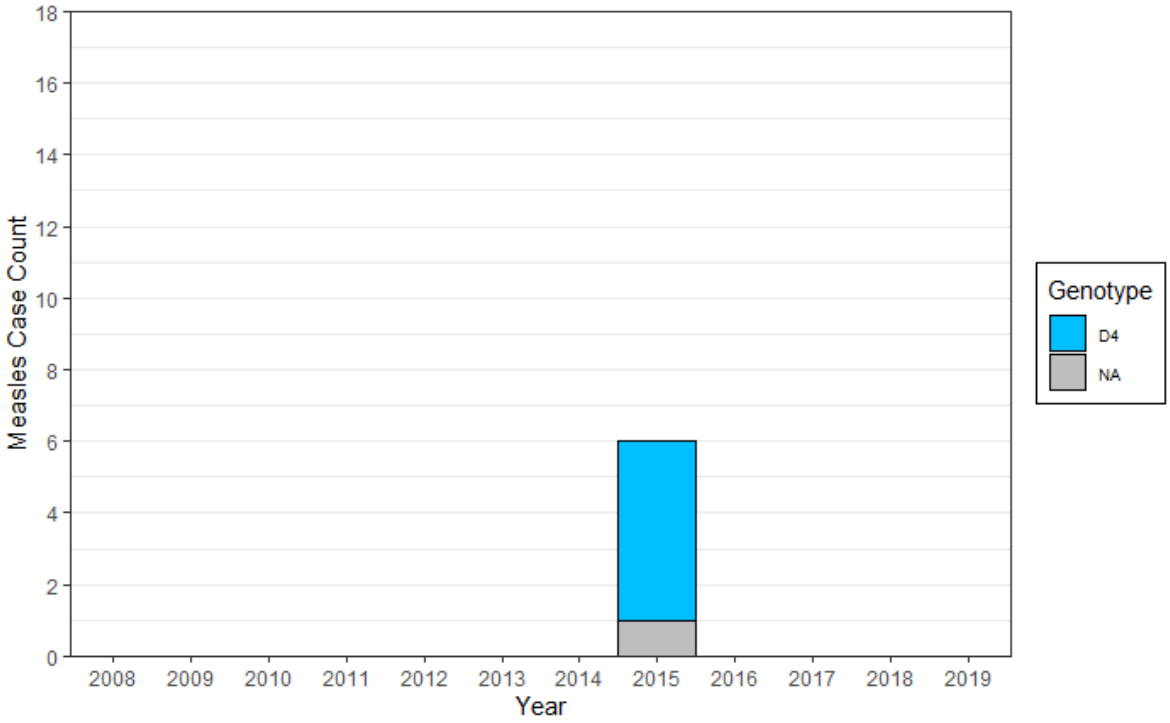
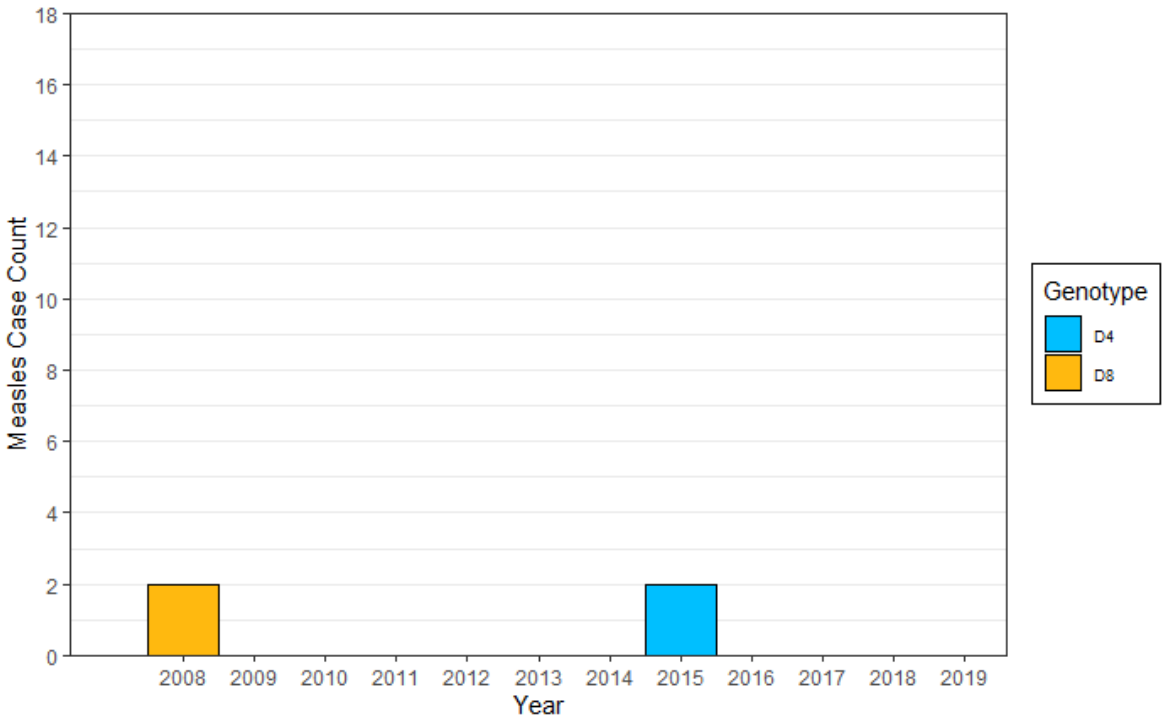


Figure 3: Count of reported measles cases within Cluster 3 by year and grouped by the genotype of the virus



In Cluster 4, all cases (*n=4/4*) were reported in 2008, and 50% of those were known to be caused by the D8 genotype of MeV. In Cluster 5, four cases (*n=4/5*) were reported in 2014, whereas one case (*n=1/5*) was reported in 2019. 75% of the cases reported in 2014 within Cluster 5 were caused by the B3 genotype of MeV, whereas the lone case in 2019 was caused by the D8 genotype. Finally, in Cluster 6, the majority of the cases were reported in 2013 (*n=5/13*) and 2015 (*n=4/13*), whereas lone cases were reported in 2008, 2014, 2018 and 2019. All of the cases reported in 2013 were caused by the B3 genotype of MeV, whereas all of the cases in 2015 were caused by the D4 genotype.

Figure 4: Count of reported measles cases within Cluster 4 by year and grouped by the genotype of the virus

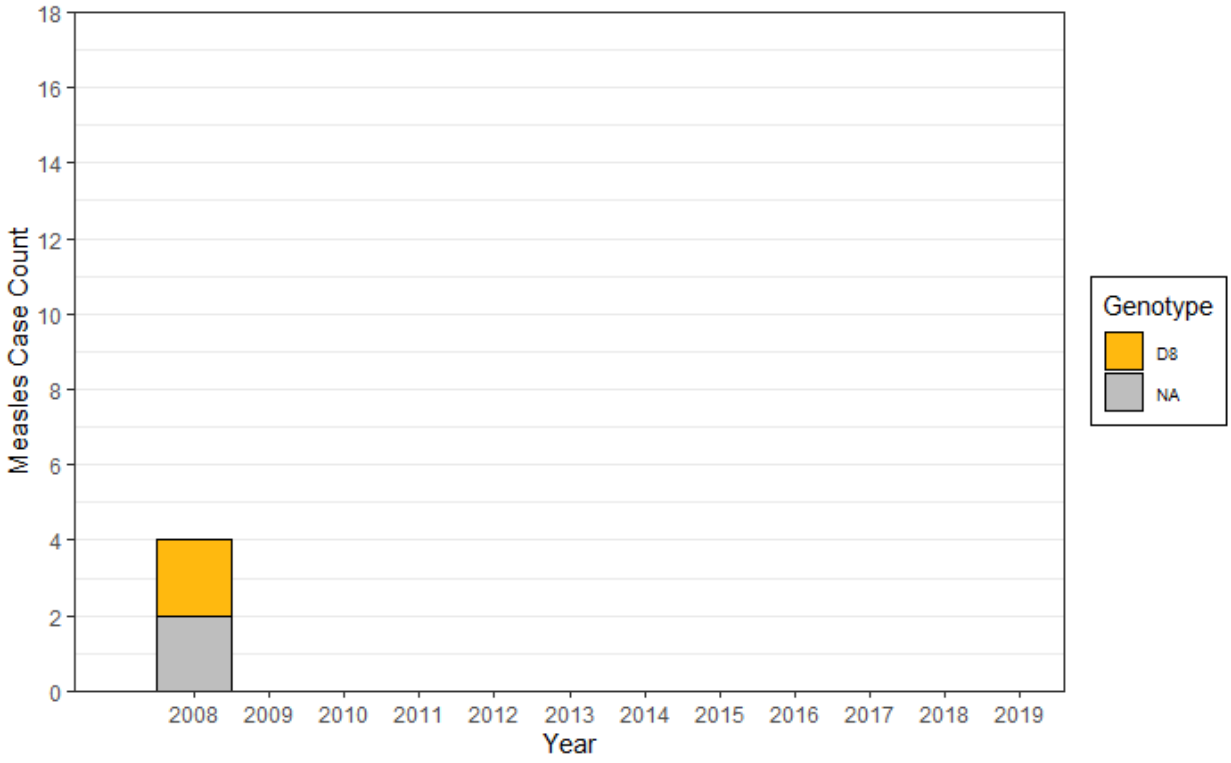


Figure 5: Count of reported measles cases within Cluster 5 by year and grouped by the genotype of the virus

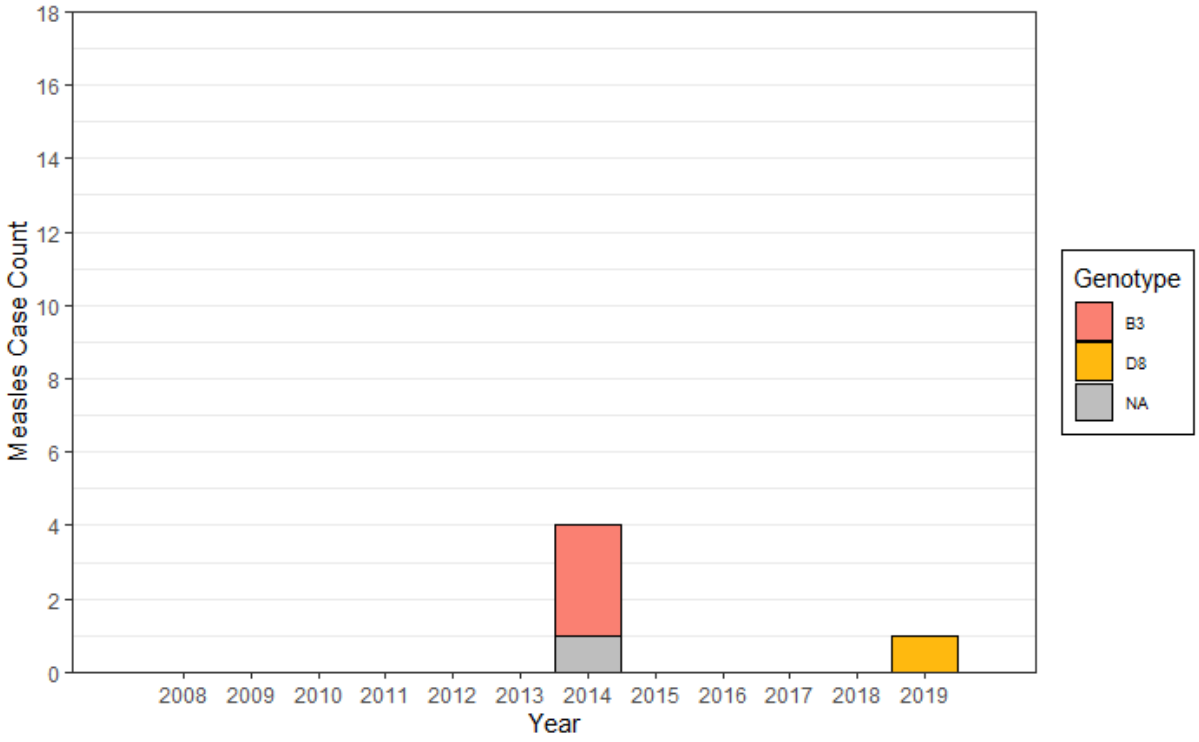


Figure 6: Count of reported measles cases within Cluster 6 by year and grouped by the genotype of the virus

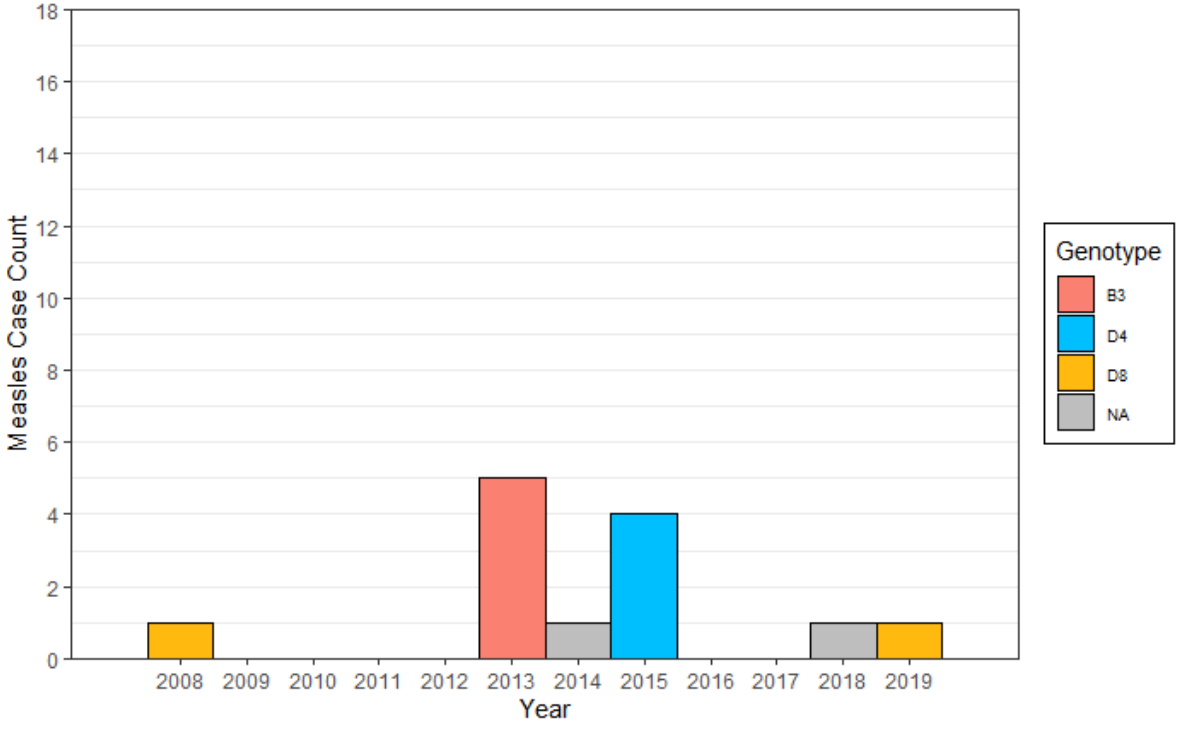
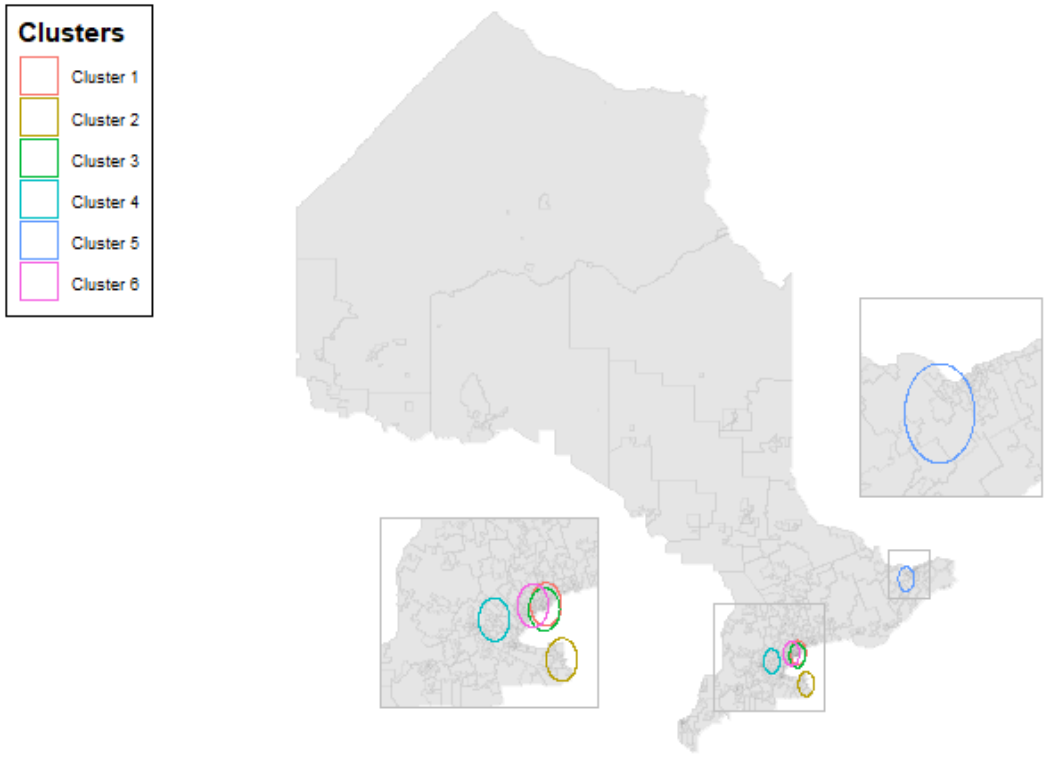


Figure 7 allows for further appreciation of the six significant clusters, as it provides a visual representation of the geographical location of the clusters. It is possible to see that Clusters 1,3,4 and 6 were located in the Greater Toronto Area and its surrounding regions, whereas Cluster 5 was in the Ottawa region and Cluster 2 appears to be in the Niagara region.

Figure 7: General location* of the six significant clusters within Ontario



***The area of clusters was plotted larger than the true area to avoid the direct identification of case FSAs.

2.3.2 Case-Control Study

The same exclusion criteria from the geospatial analysis were used for the case-control study, which led to the exclusion of the same ten FSAs described above. Additionally, any FSAs that were missing data for the variables of interest (i.e., low-income earners prevalence, immigrant counts, or education level) were excluded from the study. This led to the exclusion of two more FSAs. These FSAs had low population counts of 5 and 30, which explains their lack of

data for the variables of interest. These exclusion criteria produced a total study sample of 511 FSAs. Using the six significant clusters determined through the geospatial analysis, a total of 40 case FSAs were identified. The remaining 471 FSAs were used as control FSAs.

Table 3 summarizes the characteristics of case and control FSAs. It is possible to see that a larger proportion of case FSAs had an immigrant population greater than 21.9% of its total population compared to control FSAs (*p<0.0001*). Additionally, a larger proportion of case FSAs had a population density greater than 1000 people/km^2 relative to control FSAs (*p<0.0001*). A similar tendency was noted for the prevalence of low-income earners, as a larger proportion of case FSAs had a prevalence of low income greater than 8.8% compared to control FSAs (*p<0.0001*). Finally, a slightly larger proportion of control FSAs had post-secondary education level less than 54% among their population compared to case FSAs, but this difference was not statistically significant (*p=0.44*).

Table 3: Difference in characteristics of case and control FSAs

Characteristics	**Control FSAs**		**Case FSAs**		***p* value**
	N = 471	**Proportion (%)**	**N = 40**	**Proportion (%)**	
Immigrant Status					
≤ 21.9% immigrant	247	52.4	4	10.0	<0.0001
> 21.9% immigrant	224	47.6	36	90.0	
Population Density					
≤ 1000 people/km^2	247	52.4	3	7.5	<0.0001
> 1000 people/km^2	224	47.6	37	92.5	
Low Income Prevalence					
≤ 8.8% low-income earners	298	63.3	6	15.0	<0.0001
> 8.8% low-income earners	173	36.7	34	85.0	
Education Level					
≥ 54% post-secondary education	403	85.6	36	90.0	0.44
< 54% post-secondary education	68	14.4	4	10.0	

Table 4 provides the measures of association and their respective 95% confidence intervals for each exposure of interest. It is possible to see that immigrant status, population density and income level are all significantly associated with measles hotspots, whereas education level did not have any significant association. Indeed, FSAs that are within measles hotspots (i.e., cases) have 9.92 (95% CI: 3.48–28.32) times the odds of having a population that is more than 21.9% immigrant relative to FSAs that are not within measles hotspots (i.e., controls). The same is true for population density as FSAs that are within measles hotspots have 13.60 (95% CI: 4.14–44.72) times the odds of having a population density greater than 1000 people/km^2 relative to FSAs that are not within measles hotspots. Likewise, FSAs that are within measles hotspots have 9.76 (95% CI: 4.02–23.72) times the odds of having a proportion of low-income earners greater than 8.8% relative to FSAs that are not within measles hotspots. Regarding education, FSAs that are within measles hotspots have 0.67 (95% CI: 0.23–1.91) times the odds of having a proportion of post-secondary-education individuals that is lower than the Canadian average (i.e., 54%).

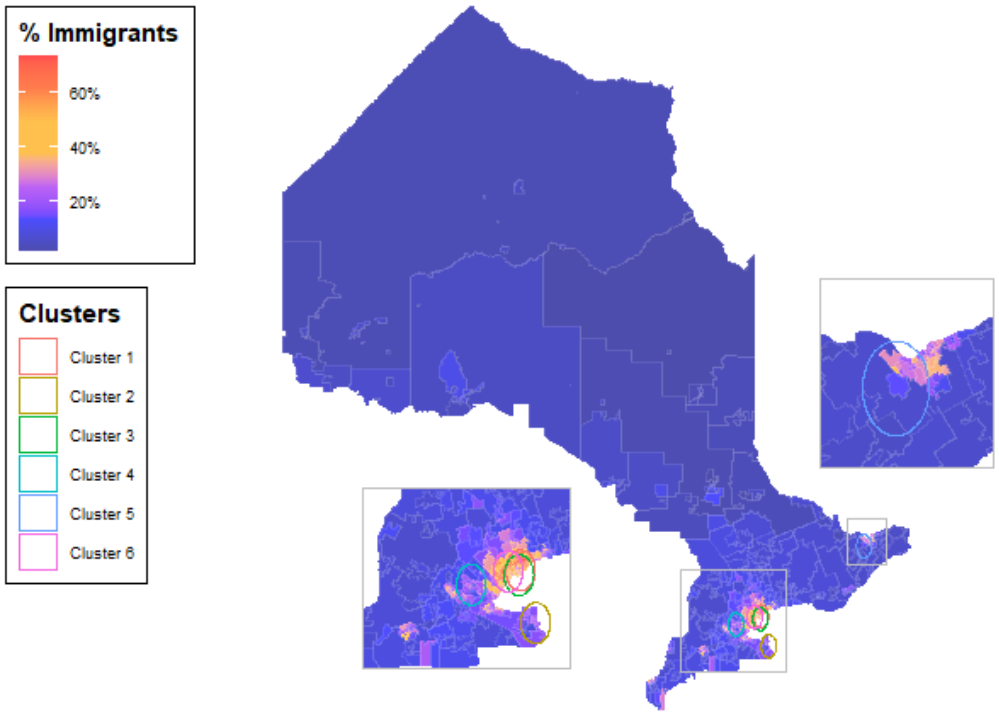
Table 4: Unadjusted odds ratio of case FSAs having an exposure of interest relative to control FSAs (with 95% confidence intervals)

Exposure	**Odds Ratio (95% CI)**
> 21.9% immigrant	9.92 (3.48–28.32)
> 1000 people/km^2	13.60 (4.14–44.72)
> 8.8% low-income earners	9.76 (4.02–23.72)
< 54% post-secondary education	0.67 (0.23–1.91)

Figures 8 through 11 allow for greater appreciation of the relationship between measles clusters and the proportion of immigrants, the prevalence of low-income, population density and

educational attainment, as they provide a visual representation of the geographical location of the measles clusters along with the geographical distribution of all the variables of interest. In Figure 8, one can clearly see that all identified measles clusters were in or near areas with a relatively high proportion of immigrants.

Figure 8: Proportion of immigrants within each FSA and the general location* of the six significant clusters within Ontario

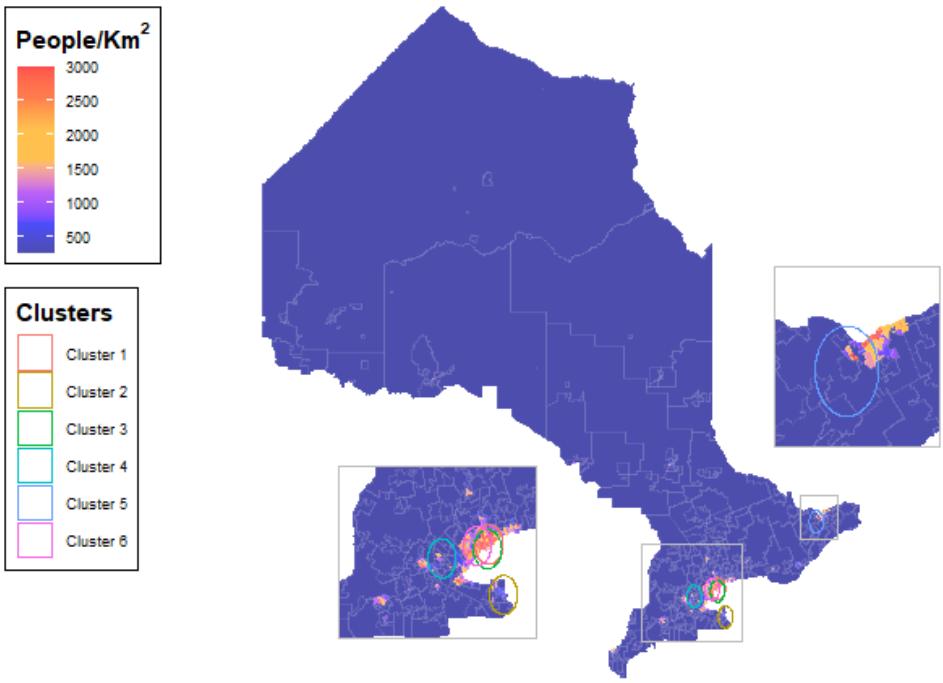


***The area of clusters was plotted larger than the true area to avoid the direct identification of case FSAs.

Similarly, Figure 9 shows that higher population density and measles clusters are geographically linked to each other, as most identified clusters were in close proximity of areas with higher population density. The same can be said for low-income prevalence, as Figure 10 shows that the prevalence of low-income tends to be higher near or in measles clusters. As for Figure 11, one can see that the proportion of individuals who completed post-secondary

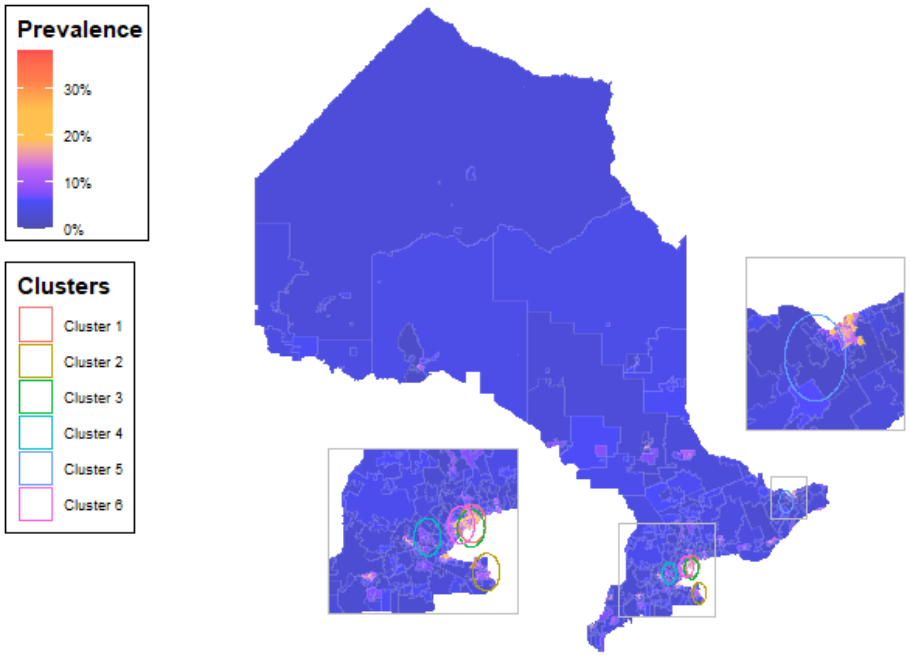
education is relatively high across most geographic regions, and it does not appear to differ significantly in proximity of measles clusters.

Figure 9: Population density (people/km^2) within each FSA and the general location* of the six significant clusters within Ontario



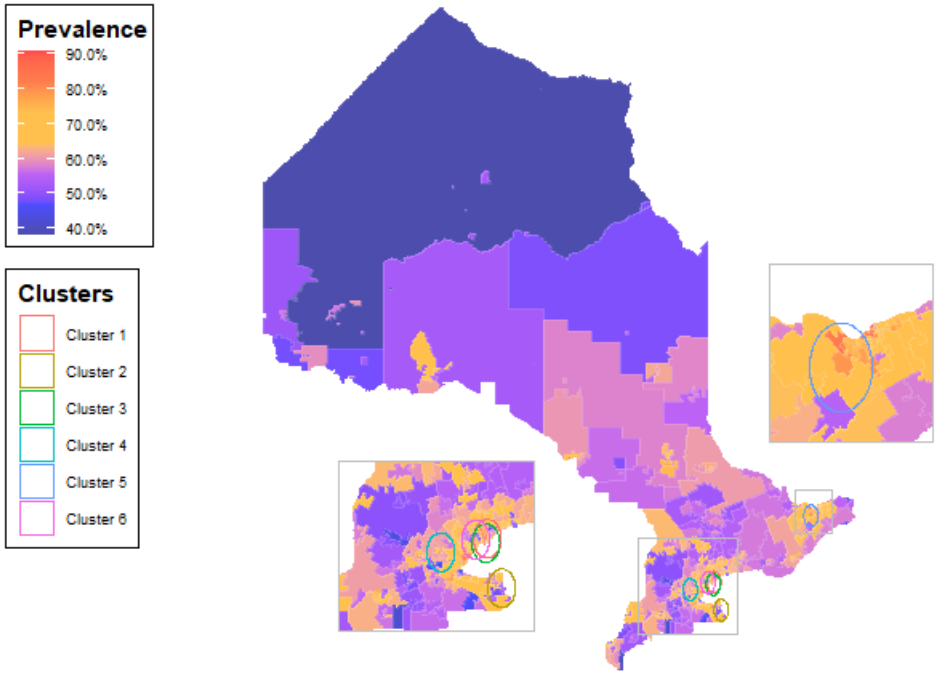
*The area of clusters was plotted larger than the true area to avoid the direct identification of case FSAs.

Figure 10: Prevalence (%) of low-income earners within each FSA and the general location* of the six significant clusters within Ontario



*The area of clusters was plotted larger than the true area to avoid the direct identification of case FSAs.

Figure 11: Prevalence (%) of post-secondary education within each FSA and the general location* of the six significant clusters within Ontario



*The area of clusters was plotted larger than the true area to avoid the direct identification of case FSAs

Chapter 3: Discussion

3.1 Interpretation of Results

Table 1 provides some general but useful characteristics of all the measles cases that were reported in Ontario from 2008 to 2019. It shows that only 18% of measles cases were known to be vaccinated with at least one dose of the measles vaccine, whereas 34% were unimmunized and 48% of cases did not have any immunization records available in iPHIS. This implies that the majority of measles cases in Ontario are driven by unimmunized individuals and individuals that are missed by vaccine-surveillance practices. Furthermore, only 35% of cases reported travelling outside of Ontario prior to symptom onset, and only 20% reported being in direct contact with a case, which implies that the majority of cases may have occurred through undetected secondary infections following importations. Therefore, even though endemic transmission of measles has been eliminated in Ontario and throughout Canada, imported index cases are still able to produce a limited amount of secondary cases. Finally, only 26% of cases were five years of age or younger, whereas 46% were 21 years or older. This goes against the typical epidemiological patterns of measles observed in regions with ongoing endemic transmission and is more in line with the epidemiological patterns observed in an elimination setting (129). Indeed, in an elimination setting with robust vaccination programs, the probability of contracting measles at a young age is lower, as vaccination usually occurs at a young age. Instead, cases usually tend to arise in older individuals due to waining immunity or, as may be the case in Ontario, due to older indviduals who may be unvaccinated and are missed from vaccine-surveillance practices.

The geospatial analysis of measles cases in Ontario found 10 measles clusters, of which six were deemed statistically significant. These six clusters contained only 7% of the study population but reported over 40% of all measles cases included in the study. These results indicate that measles cases are not randomly distributed across Ontario, but instead tend to congregate in certain regions within the province, even if most cases result from imported index cases. In other words, the risk of reported measles varies significantly across geographic regions within the province, and one cannot assume that one region will have the same risk to see a measles case (imported or not) as another region. Figures 1 through 6 allow for better appreciation of these six significant clusters, as they provide some clues of the epidemiological mechanisms behind such clusters. Indeed, one can see that the majority of cases within each cluster tend to arise within a year or two. This is especially true for Cluster 1, where one can see that over half of the cases occurred in 2008. A similar trend can be observed for Cluster 2, as all of the cases were reported in 2015; in Cluster 3, half of the cases were reported in 2008 and the other half was reported in 2015. All of the cases in Cluster 4 were reported in 2008, whereas nearly 80% of the cases in Cluster 5 were reported in 2014.

Finally, for Cluster 6 one can see that nearly 70% of cases were reported in 2013 and 2015. Additionally, one can see that most cases that occurred within one or two years in each cluster tended to be caused by the same genotype of MeV. These epidemiological patterns are suggestive of outbreaks rather than singular importation events. Indeed, the fact that most cases were closely related in time, rather than spread out as lone cases through different years, indicates that imported index cases likely produced limited outbreaks within their community or even just their household in some instances where the case counts are low (e.g., ≤ 5). Furthermore, the fact that the bulk of the cases within a given year were usually caused by the

same MeV genotype provides further evidence of measles transmission. In fact, the mutation rate of measles is very slow, which implies that individuals that were infected by the same MeV genotype typically belong to the same chain of transmission (130).

The primary goal of this thesis was to establish whether reported measles cases tend to be sporadic in their geographical distribution or whether some patterns could be noted regarding the geographical distribution of measles in Ontario. The results of the geospatial analysis showed that the risk of reported measles cases was variable across the province, as some regions were more likely to report measles cases that may have caused short-lived outbreaks. This is concurrent with the primary hypothesis of the study, which proposed that reported measles cases would appear to cluster in specific regions of the province, but the question remains as to what factors are behind such a tendency. The hypothesis that was set out at the beginning of the study was founded on known social psychology theories about human behavioural patterns, which stipulate that people with similar characteristics, values, goals or beliefs tend to congregate in social groups (131,132). With this in mind, it can be extrapolated that such a mechanism may be driving the occurrence of measles cases in Ontario. Although other explanations may exist for the results of the geospatial analysis, it is possible that people with characteristics and beliefs that increase their risk for measles may be congregating together and creating measles clusters within in the province.

This segues into the second objective of the thesis, which was to explore the population characteristics that may be contributing to these measles clusters. The case-control study provided a deeper insight into the identified measles clusters by exploring some factors that may be associated with these clusters. The results showed a statistically significant and very strong association between measles clusters and their proportion of immigrants, low-income prevalence

and population density. Indeed, measles clusters tended to have a higher-than-average proportion of immigrants living in them, a higher-than-average prevalence of low income and a higher population density. Meanwhile, no statistically significant association was found between measles hotspots and education level.

Figures 7 through 11 allow for further appreciation of the six significant clusters, as they provide a visual representation of general location of the measles clusters while also showing the geographical distribution of the variables of interest. In Figure 7, one can note the general locations of the measles clusters and see that they are located in and around the Ottawa, Toronto and Niagara regions, which are some of Ontario's largest metropoles. In concordance with the clusters, one can also note a higher proportion of immigrants, a higher population density and a higher prevalence of low-income earners in those regions through Figure 8, 9 and 10, respectively. However, as shown in Figure 11, education level appears to be more equally distributed across the province, with little contrast between the cluster areas and other regions. Such patterns are not surprising, since metropoles are, by definition, regions with higher population density. Additionally, given the larger size of metropoles, there is often more human activity, which creates greater economic opportunities that would incentivize more economic immigrants to settle down in such regions. Concurrently, economic immigrants seeking better opportunities would likely be from lower economic status, leading to areas with a higher prevalence of low-income earners.

The main factor that was proposed as a driving force behind potential clustering of measles cases was immigrant communities, and the results showed a very strong association between the proportion of immigrants in the population and measles hotspots. There may be multiple reasons behind this association, but the most likely mechanism from a public-health

perspective may be Ontario's vaccine-surveillance practices. As mentioned in the previous chapter, Ontario only conducts routine surveillance for measles vaccination among 7- to 17-year-old school pupils, which leaves out many individuals that did not live within Ontario when they were within this age stratum, such as immigrants. This is problematic, because new immigrants to Canada do not require proof of immunization when entering the country (133). Therefore, immigrants who are coming from countries that lack the infrastructure to provide effective vaccination programs may be arriving in Canada completely unimmunized or may have suboptimal immunity to measles due to subpar immunization schedules. In today's globalized setting, pockets of unimmunized communities around the world can transmit infectious diseases between each other and sustain epidemics (134,135). This means that even though countries like Canada may have eliminated endemic transmission of certain diseases within their borders, transmission of diseases may continue to happen across various countries through networks of unimmunized communities. This may very well be the case with immigrant communities in Canada. As they arrive to Canada with subpar immunization, they could be travelling back to their home countries where endemic transmission of measles is still prevalent and coming back infected, leading to short-lived outbreaks and observed measles clusters.

Other factors were also explored for their association with measles clusters. Indeed, a prevalence of low-income earners higher than the Canadian average was found to be strongly associated with measles clusters as well. This factor was initially explored due to its potential association with vaccine hesitancy (93–96). Therefore, the clustering of measles cases may also be driven by congregated individuals with low income that may be hesitant to get vaccinated against measles. However, immigrants also tend to have higher rates of low-income earners than the general Canadian population (136,137). Therefore, the association between low-income

prevalence and measles clusters may be confounded by the proportion of immigrants living within an FSA or vice-versa. Education level was also explored as a potential factor due to its association with vaccine hesitancy (93,94,96,97). No significant association was found between the proportion of individuals with post-secondary education within FSAs and measles clusters. Education level and low income may share a close relationship when it comes to their association with vaccine hesitancy. Indeed, people with low income tend to have a lower level of education, and people with a lower level of education may lack the critical-thinking skills required to assess the risks and benefits of vaccines with an objective point of view, leading to vaccine hesitancy (94,138). Therefore, education level may be a confounder in the relationship between income level and measles clusters. However, given that no association was found between education level and measles clusters, the association between low-income prevalence and measles clusters can be likely attributed to the proportion of immigrants living within FSAs. In other words, measles clusters are likely driven by suboptimal vaccine-surveillance practices that are disproportionately afflicting immigrant communities rather than low-income earners or individual choices driven by a lack of formal education. Unfortunately, it was not possible to assess the confounding effects of these variables due to a fixed sample size that was too small (especially for case FSAs) to introduce more than one variable at a time to the logistic regression model.

Finally, population density was also explored as a contributing factor to measles clusters and was found to have a significant association with measles clusters. This association was explored based on recent literature that showed that urban regions with population density greater that 1000 people/km^2 are at increased risk for measles outbreaks (98). This tendency is also logical from an epidemiological perspective. As discussed in Chapter 1, measles is a

respiratory virus that spreads from one person to another through respiratory droplets and small particle aerosols (44,45). Given this mechanism of spread, it becomes obvious that higher population density increases the likelihood of contact between an infectious case (or their respiratory droplets) and a susceptible case and therefore increases the probability of observing numerous measles outbreaks that could be larger in size. The question remains whether measles clusters are mostly driven by migrant communities or by population density. The most likely scenario is that these two factors interact with each other in a way that increases the risk for measles clusters to a level that is more than the sum of the risks attributed to migrant communities and population density alone. Indeed, if one assumes that migrant communities are under-immunized to measles due to poor vaccination programs in their countries of origin and inadequate vaccine surveillance in Ontario and these communities congregate in densely populated areas where there is a higher likelihood of susceptible individuals encountering an infected case, than the risk for measles outbreaks becomes a function of these two factors. Unfortunately, it was not viable to test for interaction between immigrant status of FSAs and population density, given the small sample size of the study.

3.2 Implications of Results

The results presented in this thesis bring forward many implications of public-health relevance. For instance, the results of the geospatial analysis, which postulate that the risk of measles outbreaks is concentrated in certain regions of the province, brings forward a new insight that was not previously considered. Indeed, previous public-health reports stated that the bulk of measles cases in Ontario arise from sporadic importations that caused short-lived outbreaks (20,79–85). These interpretations suggest that there was little to be done other than responding to the outbreaks and conducting thorough contact tracing. However, the results from

the geospatial analysis brings forth the possibility for targeted public-health interventions. As measles cases tend to cluster in specific regions of the province, public-health officials must consider taking a proactive approach to measles surveillance and prevention in Ontario. This is especially true considering that responding to measles outbreaks and conducting contact tracing is not deemed cost-effective in an elimination setting like Ontario (139). Indeed, a recent study showed that conducting contact tracing following a measles outbreak in Ontario in 2016 prevented an estimated 16 associated measles cases, which amounts to an estimated incremental cost-effectiveness ratio of $739,063 CAD per QALY gained (139). This is far beyond the currently accepted consensus of $50,000 per QALY outlined by the Canadian Agency for Drugs and Technologies in Health (CADTH) in the guidelines for the economic evaluation of health technologies in Canada (140). In other words, given a cost per QALY gained that is nearly 15 times higher the currently accepted threshold, responding to measles outbreaks is not cost-effective. Therefore, proactive and targeted interventions that are cost-effective are needed to substitute the currently reactive approach of outbreak response and contact tracing.

Targeted and proactive interventions that may help alleviate the issue are plentiful. To begin, a thorough assessment of measles immunity throughout the province should be carried out. Population immunity to measles has been recently assessed (141). This study conducted a sero-epidemiology survey using an enzyme immunoassay to test the sera of 1,199 individuals for measles immunity in Ontario. This study found that 86.3% of samples tested were above the threshold for measles protection, 5.8% were equivocal and 7.8% were below the measles-protection threshold, with some variations among age groups. For instance, only 79.4% of 12–19 year olds had the levels of immunity required for measles protection, despite two-dose vaccination coverage ranging from 90.4% to 96.7% (89,142,143). This may be due to the

sampling method used in the study or could be indicative of waning immunity due to lack of exposure to measles antigens in an elimination setting (144,145). Although this study provides some useful insights into the level of measles immunity in Ontario, it does not consider the geographical distribution of measles immunity in Ontario. With low levels of protection among certain age groups, it is possible that levels of immunity to measles may be even lower in regions that are within measles clusters identified by the geospatial analysis presented earlier. Therefore, public-health officials should consider performing regular sero-epidemiology surveys to assess the geographical distribution of measles immunity in the province.

If conducting regular sero-epidemiology surveys across the province is beyond the capabilities of public-health authorities, then conducting thorough assessments of the geographic distribution of vaccination status in Ontario should be carried out. A recent study by Wilson et al. provided useful insights by assessing the geographic distribution of unvaccinated children in Ontario during the 2016–2017 school year using DHIR data (146). Interestingly, the identified regions with a higher prevalence of unvaccinated children that are outlined in the paper do not coincide with the same regions of the measles clusters that were identified in this thesis. The reasons for this discordance are plentiful. First, Wilson et al. conducted the study for the 2016–2017 year only, whereas this thesis included measles cases that were reported in Ontario from 2008 to 2019. Secondly and most importantly, Wilson et al. only assessed vaccination status among school-aged children (i.e., 7–17 year olds) since DHIR only houses vaccination data that is collected on school-aged pupils through Panorama. This is problematic, since this thesis showed that the majority of measles cases (i.e., 72%) were in children under five years old or adults older than 21, indicating that these age groups probably have lower vaccination rates or immunity and are more at risk for measles. Hence, expanding Ontario's vaccine-surveillance

practices using Panorama should be considered. Current practices involve compiling measles-vaccination rates within each public-health unit to compute the mean vaccination rate among school pupils in Ontario (88,89,142,143). These practices should be expanded to include individuals of all age groups to hopefully capture immigrants who did not attend elementary or secondary school in Ontario. If obtaining vaccination records from all individuals in Ontario is not feasible, then a representative sample of the Ontario population could be surveyed using weighted sampling techniques. Additionally, the geographical distribution of vaccination rates should be assessed during routine vaccine surveillance. As the results from the geospatial analysis showed that the risk of reported measles is not equally distributed across the province, it would be prudent to consider the geospatial distribution of vaccination rates among all age groups to identify more pockets of under-immunized communities.

Conducting thorough sero-epidemiological surveys or implementing robust vaccine-surveillance practices would open the door to possible targeted proactive interventions. Indeed, despite high overall vaccination rates among school pupils, the results of such practices would allow outlying clusters of under-immunized communities that may require additional public-health resources to be identified. Such interventions could involve targeted vaccination campaigns through pop-up clinics in immigrant communities or targeted vaccine education in potential vaccine-hesitant communities to prevent or limit the probability of future measles outbreaks. Given the low cost of MMR vaccines in Canada, targeted vaccination campaigns might actually be cost-effective. Indeed, the Public Health Agency of Canada estimates that for every dollar spent on immunization programs against measles, mumps and rubella, $16 are saved through the prevention of outbreaks (147,148). However, a robust cost-utility analysis would have to be performed in order to establish whether such campaigns would be truly cost-effective.

These applications could also be extended to diseases other than measles, such as COVID-19. This virus causes serious systemic manifestations that often leads to severe disease and necessary hospitalization in approximately 5% of cases (149). This has compelled governments and public-health authorities to introduce province-wide lockdowns, mass-vaccination campaigns and border closures to limit the spread of the virus and ease the burden on healthcare infrastructures (150). However, given that SARs-CoV-2 spreads in a similar way to MeV, it is possible that COVID-19 cases may be clustering in certain regions of the province rather than arising randomly across the province (151). Therefore, public-health officials may want to consider a targeted approach instead of implementing indiscriminate province-wide interventions such as province-wide lockdowns. The geographical distribution of COVID-19 cases should be assessed to determine the need for focalized lockdowns and targeted vaccine campaigns across geographical regions in the province.

Finally, the results of this thesis also bring forth some implications for health equity. Indeed, the results from the secondary objective showed that communities with certain characteristics are disproportionally affected by measles outbreaks. Of the utmost concern are the findings that FSAs within measles clusters tend to convey a higher proportion of migrants. Although robust conclusions cannot be made regarding the level of immunity among migrant communities using these findings, it is possible that these communities may be under-immunized against measles. Not only are adult migrants likely missed from vaccine-surveillance practices but their vaccination status is not thoroughly assessed when entering Canada (133). The government provides some guidelines for primary healthcare physicians to assess and update immunizations of new immigrants to Canada (133). However, this may not be enough for migrant communities, as it is well established that economic, cultural and linguistic barriers,

among others, hinder the accessibility to primary-care physicians and healthcare in general for immigrants (152–157). This issue is especially concerning for recent immigrants, who are less likely to have access to a regular healthcare provider than established immigrants (158). This is not the only body of work to find health inequities related to infectious diseases. Indeed, most recently a study that assessed vaccination coverage for the COVID-19 vaccine among people aged 12 years and older in all neighbourhoods in Ottawa, Canada, found that disadvantaged, racialized and migrant neighbourhoods of lower socio-economic status have significantly lower vaccination rates (159). For instances, the neigbourhood with the highest vaccination rate had a first-dose vaccination rate of 99.7% at the time of writing while the neighbourhood with the lowest vaccination rate had a first-dose vaccination rate of 62%. Residents and partner organizations of the study outlined various reasons for these trends, such as inequitable access, lack of time, lack of trust, inequitable access to clear and accurate information, unclear or mixed messages and misinformation. Therefore, when determining resource allocations in healthcare policies, special consideration should be given to immigrant and racialized communities of lower socio-economic status to ensure resources are distributed equitably and immunization schedules are up to date.

3.3 Strengths and Limitations

As is the case with most research projects, this thesis has some limitations that should be considered in conjunction with the results. First, the geospatial analysis used measles cases from 2008 to 2019 in Ontario to identify spatial clusters of measles, whereas the population at risk was determined from the 2016 Canadian census. This is problematic because real-world populations are dynamic in nature. Indeed, the population of Ontario experienced a population growth of approximately 1% year over year from 2008 to 2019 (160). This may bias the results in multiple

ways. For instance, given that the Ontario population in 2016 was larger than from 2008 through 2015, the population at risk used may be larger than the true population at risk when the cases arose in some instances. This could lead to an underestimation of the risk of measles in some regions and therefore some clusters going undetected. Another issue is that there might have been population migrations between FSAs, leading to significant changes in population size within FSAs that are not necessarily reflected in overall population changes in Ontario. The inability to capture these within-province migrations may bias the results towards or away from the null hypothesis. Indeed, as the true population at risk within FSAs may have been larger or smaller than the 2016 Census population when cases arose, the risk of measles might be over- or under-estimated in some regions.

Similarly, the exclusion of the 28 measles cases with missing FSAs from the geospatial analysis may bias the results in various ways. On the one hand, excluding these 28 measles cases lowered the overall rate of measles in Ontario. This means that when SaTScan compares the rate of measles in a potential cluster to the overall rate in the province, it is comparing it to an overall rate that is lower than the true rate, which could lead to misidentifying certain regions as a cluster. In other words, this may have biased the results away from the null hypothesis. On the other hand, if all of these 28 cases belonged to a specific region—which is the most likely scenario given the evidence of measles spread in Ontario following imported index cases—then this exclusion may have also led to undetected clusters of measles. In other words, this could have also biased the results towards the null hypothesis.

Furthermore, as the exposure variables did not respect the linearity assumption, they could not be introduced into the logistic-regression model as continuous variables. This left the options of either categorizing the continuous variables into two or more strata, or introducing

polynomial terms and fitting a spline that allows the relationship to change at given points and interpreting the results accordingly. Although the latter option was the optimal one in this situation, it was not pursued as this was not within the scope of the thesis. Therefore, the exposure variables were categorized into binary variables, which may have led to some lost information regarding the relationship between the exposures and the outcome.

Another important limitation that should be discussed is the small sample size of the case-control study. Given that the 2016 Census did not provide individual-level data but rather demographic data congregated at the FSA level, the unit of analysis had to be individual FSAs rather than each individual person within FSAs. This led to a fixed sample size of 511 FSAs, of which only 40 were in cluster regions.. The small number of cases hindered the capability of introducing covariates in the logistic regression model and test for confounding or variable interaction. Furthermore, one can note that the confidence intervals for each OR estimate are relatively wide, especially for immigrant status, population density and low-income prevalence. This is due to the small number of cases; more specifically, one can note that there are a very small number of cases in the unexposed category of each of these variables, which explains the wide confidence intervals.

There are also some concerns regarding FSA boundary changes over time that were considered in conjunction with the results of the geospatial analysis. Indeed, as the geospatial analysis made use of all measles cases that were reported from 2008 to 2019, there is a reasonable concern that the boundaries of FSAs in Ontario changed considerably during that time period. It may be argued that those changes may not be reflected in the 2016 FSA shapefile that was used for the study, which may influence the results of the study. However, upon further exploration of this potential bias, it was determined that it likely didn't have a significant impact

on the results for multiple reasons. First, if there were any FSA boundary changes over time, these changes are more likely to occur at the periphery of growing metropoles where the population counts are relatively low and the likelihood of observing measles cases are probably lower, while little to no changes would occur at the core of established metropoles. Second, when comparing the number of FSAs in the 2011 shapefile to the 2016 shapefile, there was only a difference of 3 FSAs, which shows that there were very little changes made to FSAs during that time period. Finally, all measles cases that had an FSA in iPHIS were successfully linked to their respective FSA on the 2016 census and the 2016 shapefile, which indicates that if any FSAs were added or removed from 2008 through 2016, such FSAs did not report any measles cases and wouldn't significantly influence the detection of measles clusters.

Finally, it is also important to take note of some strengths that arise from this study. To our knowledge, this study is the first to assess the geospatial distribution of measles cases in Ontario, which provides new insights of public-health relevance. Additionally, the geospatial analysis made use of the novel MCHS-P technique to select the scanning-window size. Traditionally, the scanning window has been selected by researchers using intuition and guidance from the unit of analysis used (e.g., FSA, census tract, public-health unit). The Gini coefficient is also useful for some researchers, but the MCHS-P has demonstrated improvements over the Gini coefficient, as it improves the sensitivity and specificity of the analysis, which strengthens the results. Furthermore, nearly 100,000 Monte Carlo simulations were performed for the geospatial analysis, which provides ample statistical power to identify measles clusters throughout the province. Finally, the geospatial analysis made use of laboratory-confirmed measles cases from the iPHIS database, which ensures that only true reported cases of measles

are included in the analysis. However, this could also be a limitation, as measles cases that go unreported to public-health units are not considered in the analysis.

3.4 Future studies

In light of these strengths and limitations, multiple possibilities arise to build off from the analyses presented above and improve future studies. First, future studies should consider using a geospatial model that takes into consideration the dynamic nature of human populations. This could be done by obtaining data of multiple census datasets from previous census years and computing the average population at risk through the years of the study across the province but also within each individual FSA.

Secondly, future studies may also want to consider including other provinces into the geospatial analysis in an attempt to increase the fixed sample size of FSAs. Another possibility would be to obtain individual-level census data through a formal data request to Statistics Canada (161). With such individual-level data, the study sample would become the bulk of the population living in Ontario, and the cases would be individuals living in measles cluster zones. No matter the option future studies undertake, increasing the sample size would allow for the inclusion of more exposure categories for each variable considered or even use continuous variables, which would allow for a better appreciation of the exposure-outcome relationships. A larger sample size would also open the door to test for confounding variables and interaction between various variables of interest. Advanced statistical methods should also be considered to allow for the use of non-linear variables when assessing exposure-outcome relationships.

Furthermore, more data may prove to be helpful to identify additional factors that may be associated with measles clusters. Indeed, additional studies may want to consider obtaining

measles-immunization data to assess how this factor influences measles clusters. With measles-immunization data, one could also identify clusters of under-immunized communities and assess how they may be associated with clusters of measles.

Additionally, appraising sources of measles importation would add a much-needed perspective to inform targeted public health policies. Future studies may focus on accessing the MeaNS database to link genotype data from measles cases in Ontario to genotypes reported around the world in an attempt to identify the main regions that are contributing to measles importation in Ontario. This would open the door for targeted interventions for migrant communities that are entering Canada from the countries that contribute the most to measles importation.

Finally, a thorough qualitative research project may prove to be beneficial to understand some of the factors that lead to subpar immunization in certain communities. Indeed, by conducting thorough interviews with members living in underimmunized communities, it would be possible to conduct a thematic analysis of interview extracts to compile common themes and gain valuable insights into the unique inequalities that may be afflicting such communities. It is by understanding such factors that one can address them effectively.

3.5 Conclusion

In the era of chronic diseases, infectious diseases continue to burden human health. As accelerated human activity drives the spread of new and established infectious diseases, contributing factors should be explored carefully to address them accordingly and ease the hampering effects of such diseases. This thesis focused on studying such factors for the measles virus, a deadly respiratory pathogen that has shared a relationship with human civilizations for

millennia. Although safe and effective vaccines for this virus have existed for decades and humans are the only known reservoirs for the disease, the route to eradication has been a long and challenging one, as vaccine uptake remains suboptimal among certain communities. Building on theories in social psychology, it was hypothesized that individuals with certain characteristics that heighten their risk for measles were grouping together within certain geographical regions in Ontario, which may be driving the majority of observed measles cases. The geospatial analysis corroborated this hypothesis, as it showed that the risk of reported measles cases was not randomly distributed across the province, and six statistically significant clusters of reported measles cases were identified that contributed to 40% of all measles cases from 2008 through 2019 in Ontario. The case-control study provided further insights into this tendency, as it showed that measles clusters usually bore a higher proportion of immigrants, a higher prevalence of low-income earners and a higher population density. These insights lay the groundwork for future research projects and open the door for targeted and equitable interventions as we move forward in the path to measles eradication. Such interventions could also be used for other respiratory viruses that share similar characteristics with MeV. Just as John Snow successfully halted the Soho Cholera epidemic by disabling the Broad Street water pump and interrupting the chain of transmission of the disease, the figurative pump that contributes to global measles transmission can be disabled through collective efforts.

3.6 Acknowledgements and Contributions

A master's thesis represents a comprehensive piece of work that cannot be completed without the time and devotion of various stakeholders. First, it is important to recognize the important contributions of Dr. Ann Jolly and Dr. Cindy Feng in the development of this project and establishing a strong conceptual framework from which this thesis was built. Endless

gratitude is also in place for Dr. Sarah Wilson from Public Heath Ontario who acted as a consultant in the development of the thesis hypothesis. Special consideration should also be given to Wei Wang from the Western China School of Public Health at Sichuan University, who provided the R package necessary to compute the MCHS-P and also gave his time to assist with bugs in the code. Also, a very special thank you to all faculty members from the School of Epidemiology and Public Health at the University of Ottawa who put endless efforts to provide high-quality training for aspiring epidemiologists and were available to answer any questions throughout the completion of this thesis. Next, it is important to recognize the role that Public Health Ontario played in providing access to measles case data for Ontario. Finally, Dr. Stacey Smith? and Dr. Alice Zwerling made significant contributions to this work by providing ongoing support through the entirety of this project.

Marcel Miron-Celis wrote and contributed to this thesis in its entirety, completed all the coding for the analysis and worked to obtain access to various data sets used. Dr. Stacey Smith? was the primary supervisor and contributed to the development of the conceptual framework while also providing ongoing support throughout the entirety of project through weekly laboratory meetings. Dr. Stacey Smith? also contributed to this work by reading and revising the final version of the thesis. Dr. Alice Zwerling also provided support as the secondary supervisor of this project and read a draft of this thesis to provide feedback and revisions where necessary.

3.7 Funding and Conflicts of Interest

A merit scholarship was provided to Marcel Miron-Celis by the University of Ottawa as part of the MSc Epidemiology program through which this thesis was developed. No other

funding was received to complete this project. The authors have no conflicts of interest to declare.

References

1. Geddes AM. The history of smallpox. Clin Dermatol. 2006 May 1;24(3):152–7.

2. Halioua B. Histoire de la médecine. 3rd ed. Paris: Elsevier Masson SAS; 2009. 278 p.

3. Koch A, Brierley C, Maslin MM, Lewis SL. Earth system impacts of the European arrival and Great Dying in the Americas after 1492. Quat Sci Rev. 2019 Mar 1;207:13–36.

4. Coronavirus Update (Live): 92,493,946 Cases and 1,980,825 Deaths from COVID-19 Virus Pandemic - Worldometer. Available from: https://www.worldometers.info/coronavirus/.[Accessed 2021 Jan 13].

5. Baron S. Medical Microbiology. 4th ed. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Available from: http://www.ncbi.nlm.nih.gov/books/NBK7627/. [Accessed 2021 Jan 13].

6. Haydon D. Identifying Reservoirs of Infection: A Conceptual and Practical Challenge. Emerg Infect Dis. 2002 Dec;8(12):1468–73.

7. Snow J. On the mode of communication of cholera. London : John Churchill; 1855. 216 p. Available from: http://archive.org/details/b28985266. [Accessed 2021 Jan 14].

8. Legros D, Partners of the Global Task Force on Cholera Control. Global Cholera Epidemiology: Opportunities to Reduce the Burden of Cholera by 2030. J Infect Dis. 2018 Oct 15;218(suppl_3):S137–40.

9. Riedel S. Edward Jenner and the history of smallpox and vaccination. Proc Bayl Univ Med Cent. 2005 Jan;18(1):21–5.

10. Organization WH. The global eradication of smallpox. Final Report of the Global Commission for the Certification of Smallpox Eradication. Hist Int Public Health. 1980; 4. Available from: https://www.cabdirect.org/cabdirect/abstract/19812902815. [Accessed 2020 Jun 10].

11. Fleming A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. Br J Exp Pathol. 1929 Jun;10(3):226–36.

12. Gaynes R. The Discovery of Penicillin—New Insights After More Than 75 Years of Clinical Use. Emerg Infect Dis. 2017 May;23(5):849–53.

13. Tan SY, Tatsumura Y. Alexander Fleming (1881–1955): Discoverer of penicillin. Singapore Med J. 2015 Jul;56(7):366–7.

14. Pandey N, Cascella M. Beta Lactam Antibiotics. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2020. Available from: http://www.ncbi.nlm.nih.gov/books/NBK545311/. [Accessed 2021 Jan 18].

15. Sanders JW, Fuhrer GS, Johnson MD, Riddle MS. The Epidemiological Transition: The Current Status of Infectious Diseases in the Developed World versus the Developing World. Sci Prog. 2008 Mar 1;91(1):1–37.

16. CDC. Measles | History of Measles | CDC. 2018. Available from: https://www.cdc.gov/measles/about/history.html. [Accessed 2020 May 9].

17. WHO. More than 140,000 die from measles as cases surge worldwide. 2019. Available from: https://www.who.int/news-room/detail/05-12-2019-more-than-140-000-die-from-measles-as-cases-surge-worldwide. [Accessed 2020 May 9].

18. Public Health Agency of Canada PHA of C. Measles in Canada. aem. 2019. Available from: https://www.canada.ca/en/public-health/services/diseases/measles/measles-in-canada.html. [Accessed 2020 May 9].

19. De Serres G, Markowski F, Toth E, Landry M, Auger D, Mercier M, et al. Largest measles epidemic in North America in a decade--Quebec, Canada, 2011: contribution of susceptibility, serendipity, and superspreading events. J Infect Dis. 2013 Mar 15;207(6):990–8.

20. Public Health Ontario. Measles in Ontario. Public Health Ontario. 2018. Available from: https://www.publichealthontario.ca/Diseases and Conditions/Infectious Diseases/Vaccine Preventable Diseases/Measles. [Accessed 2020 May 19].

21. Pitt D, Sevane N, Nicolazzi EL, MacHugh DE, Park SDE, Colli L, et al. Domestication of cattle: Two or three events? Evol Appl. 2019 Jan;12(1):123–36.

22. Morand S, McIntyre KM, Baylis M. Domesticated animals and human infectious diseases of zoonotic origins: Domestication time matters. Infect Genet Evol. 2014 Jun 1;24:76–81.

23. Roeder P, Mariner J, Kock R. Rinderpest: the veterinary perspective on eradication. Philos Trans R Soc Lond B Biol Sci. 2013 Aug 5;368(1623):20120139.

24. Kumar N, Maherchandani S, Kashyap SK, Singh SV, Sharma S, Chaubey KK, et al. Peste Des Petits Ruminants Virus Infection of Small Ruminants: A Comprehensive Review. Viruses. 2014 Jun 6;6(6):2287–327.

25. Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE. Host Phylogeny Constrains Cross-Species Emergence and Establishment of Rabies Virus in Bats. Science. 2010 Aug 6;329(5992):676–9.

26. Couacy-Hymann E, Koffi MY, Kouadio VK, Mossoum A, Kouadio L, Kouassi A, et al. Experimental infection of cattle with wild type peste-des-petits-ruminants virus – Their role in its maintenance and spread. Res Vet Sci. 2019 Jun 1;124:118–22.

27. Lembo T, Oura C, Parida S, Hoare R, Frost L, Fyumagwa R, et al. Peste des Petits Ruminants Infection among Cattle and Wildlife in Northern Tanzania. Emerg Infect Dis. 2013 Dec;19(12).

28. Wertheim JO, Kosakovsky Pond SL. Purifying Selection Can Obscure the Ancient Age of Viral Lineages. Mol Biol Evol. 2011 Dec;28(12):3355–65.

29. Furuse Y, Suzuki A, Oshitani H. Origin of measles virus: divergence from rinderpest virus between the 11 th and 12 th centuries. BioMed Central; 2010 Dec p. 1–4. Report No.: 1. Available from: http://virologyj.biomedcentral.com/articles/10.1186/1743-422X-7-52

30. Modelski G. World Cities: –3000 to 2000. United States: Faros 2000; 2003. 245 p.

31. Düx A, Lequime S, Patrono LV, Vrancken B, Boral S, Gogarten JF, et al. Measles virus and rinderpest virus divergence dated to the sixth century BCE. Science. 2020 Jun 19;368(6497):1367–70.

32. Black FL. Measles endemicity in insular populations: Critical community size and its evolutionary implication. J Theor Biol. 1966 Jul 1;11(2):207–11.

33. Cohen SG. Measles and immunomodulation. J Allergy Clin Immunol. 2008 Feb 1;121(2):543–4.

34. Longrigg J. The great plague of Athens. Hist Sci Annu Rev Lit Res Teach. 1980 Sep;18(41 pt 3):209–25.

35. Shrewsbury JFD. The Plague of Athens. Bull Hist Med. 1950 Jan 1;24:1–25.

36. Nardin T. Thucydides, On Justice, Power, and Human Nature. Ethics Int Aff. 1994 Mar;8:216–216.

37. Cunha BA. The cause of the plague of Athens: plague, typhoid, typhus, smallpox, or measles? Infect Dis Clin North Am. 2004 Mar;18(1):29–43.

38. Duncan-Jones RP. The impact of the Antonine plague. J Roman Archaeol. 1996 ed;9:108–36.

39. Harper K. Pandemics and passages to late antiquity: rethinking the plague of c.249–270 described by Cyprian. J Roman Archaeol. 2015 ed;28:223–60.

40. McVean A. Measles: the plague that ruined Rome. Office for Science and Society. 2018. Available from: https://www.mcgill.ca/oss/article/did-you-know-history/measles-plague-ruined-rome. [Accessed 2021 Feb 1].

41. Byrne J. Encyclopedia of Pestilence, Pandemics and Plagues. Vols. 1, A–M. Westport, Connecticut: Greenwood Press; 2008. 917 p. Available from: https://www.yumpu.com/en/document/read/36727655/encyclopedia-of-pestilence-pandemics-and-plagues. [Accessed 2021 Feb 3].

42. Shanks GD, Lee S-E, Howard A, Brundage JF. Extreme Mortality After First Introduction of Measles Virus to the Polynesian Island of Rotuma, 1911. Am J Epidemiol. 2011 May 15;173(10):1211–22.

43. WHO. Measles — Global situation. WHO. World Health Organization; 2019. Available from: http://www.who.int/csr/don/26-november-2019-measles-global_situation/en/. [Accessed 2021 Apr 14].

44. Chen RT, Goldbaum GM, Wassilak SGF, Markowitz LE, Orenstein WA. An explosive point-source measles outbreak in a highly vaccinated population. Modes of transmission and risk factors for disease. Am J Epidemiol. 1989 Jan 1;129(1):173–82.

45. WHO. Manual for the Laboratory-based Surveillance of Measles, Rubella, and Congenital Rubella Syndrome. WHO. World Health Organization; 2018. Available from: http://www.who.int/immunization/monitoring_surveillance/burden/laboratory/manual_section1.1/en/. [Accessed 2021 May 1].

46. Malvoisin E, Wild TF 1993. Measles virus glycoproteins: studies on the structure and interaction of the haemagglutinin and fusion proteins. J Gen Virol. 74(11):2365–72.

47. Yanagi Y, Takeda M, Ohno S 2006. Measles virus: cellular receptors, tropism and pathogenesis. J Gen Virol. 87(10):2767–79.

48. Watanabe A, Yoneda M, Ikeda F, Terao-Muto Y, Sato H, Kai C. CD147/EMMPRIN Acts as a Functional Entry Receptor for Measles Virus on Epithelial Cells. J Virol. 2010 May 1;84(9):4183–93.

49. Iwasaki M, Takeda M, Shirogane Y, Nakatsu Y, Nakamura T, Yanagi Y. The Matrix Protein of Measles Virus Regulates Viral RNA Synthesis and Assembly by Interacting with the Nucleocapsid Protein. J Virol. 2009 Oct 15;83(20):10374–83.

50. Tober C, Seufert M, Schneider H, Billeter MA, Johnston IC, Niewiesk S, et al. Expression of measles virus V protein is associated with pathogenicity and control of viral RNA synthesis. J Virol. 1998 Oct;72(10):8124–32.

51. Liston P, DiFlumeri C, Briedis DJ. Protein interactions entered into by the measles virus P, V, and C proteins. Virus Res. 1995 Oct;38(2–3):241–59.

52. Devaux P, von Messling V, Songsungthong W, Springfeld C, Cattaneo R. Tyrosine 110 in the measles virus phosphoprotein is required to block STAT1 phosphorylation. Virology. 2007 Mar 30;360(1):72–83.

53. Lessler J, Reich NG, Brookmeyer R, Perl TM, Nelson KE, Cummings DA. Incubation periods of acute respiratory viral infections: a systematic review. Lancet Infect Dis. 2009 May 1;9(5):291–300.

54. Griffin DE. Measles virus-induced suppression of immune responses. Immunol Rev. 2010;236(1):176–89.

55. de Swart RL, Ludlow M, Witte L de, Yanagi Y, Amerongen G van, McQuaid S, et al. Predominant Infection of CD150+ Lymphocytes and Dendritic Cells during Measles Virus Infection of Macaques. PLOS Pathog. 2007 Nov 16;3(11):e178.

56. Moss WJ, Griffin DE. Measles. The Lancet. 2012 Jan 14;379(9811):153–64.

57. Wolfson LJ, Grais RF, Luquero FJ, Birmingham ME, Strebel PM. Estimates of measles case fatality ratios: a comprehensive review of community-based studies. Int J Epidemiol. 2009 Feb 1;38(1):192–205.

58. Semba RD, Bloem MW. Measles blindness. Surv Ophthalmol. 2004 Mar 1;49(2):243–55.

59. NIH. Subacute Sclerosing Panencephalitis Information Page. National Institute of Neurological Disorders and Stroke. 2019. Available from: https://www.ninds.nih.gov/Disorders/All-Disorders/Subacute-Sclerosing-Panencephalitis-Information-Page. [Accessed 2020 Jun 24].

60. Johnson RT, Griffin DE, Hirsch RL, Wounsky JS, Roedenbeck S, de Soriano IL, et al. Measles Encephalomyelitis — Clinical and Immunologic Studie. N Engl J Med. 1984 Jan 19;310(3):137–41.

61. Riddell MA, Moss WJ, Hauer D, Monze M, Griffin DE. Slow clearance of measles virus RNA after acute infection. J Clin Virol. 2007 Aug 1;39(4):312–7.

62. Ota MO, Ndhlovu Z, Oh S, Piyasirisilp S, Berzofsky JA, Moss WJ, et al. Hemagglutinin Protein Is a Primary Target of the Measles Virus–Specific HLA-A2–Restricted CD8+ T Cell Response during Measles and after Vaccination. J Infect Dis. 2007 Jun 15;195(12):1799–807.

63. Cáceres VM, Strebel PM, Sutter RW. Factors Determining Prevalence of Maternal Antibody to Measles Virus throughout Infancy: A Review. Clin Infect Dis. 2000 Jul 1;31(1):110–9.

64. Griffin DE, Oldstone MBA. Measles: History and Basic Biology. Springer Science & Business Media; 2008. 201 p.

65. Gay NJ. The Theory of Measles Elimination: Implications for the Design of Elimination Strategies. J Infect Dis. 2004 May 1;189(Supplement_1):S27–35.

66. WHO. Measles. 2019. Available from: https://www.who.int/news-room/fact-sheets/detail/measles. [Accessed 2021 May 15].

67. Shulman ST, Shulman DL, Sims RH. The tragic 1824 journey of the Hawaiian king and queen to London: history of measles in Hawaii. Pediatr Infect Dis J. 2009 Aug;28(8):728–33.

68. WHO. The Measles Initiative vaccinates one billion children in first decade. WHO. World Health Organization; 2011. Available from: https://www.who.int/immunization/newsroom/press/measles_initiative_vaccinates_billion_4august2011/en/. [Accessed 2021 May 19].

69. WHO. Worldwide measles deaths climb 50% from 2016 to 2019 claiming over 207 500 lives in 2019. 2020. Available from: https://www.who.int/news/item/12-11-2020-worldwide-measles-deaths-climb-50-from-2016-to-2019-claiming-over-207-500-lives-in-2019. [Accessed 2021 May 20].

70. Guerra FM, Bolotin S, Lim G, Heffernan J, Deeks SL, Li Y, et al. The basic reproduction number (R0) of measles: a systematic review. Lancet Infect Dis. 2017 Dec 1;17(12):e420–8.

71. van den Hof S, Meffre CM, Conyn-van Spaendonck MA, Woonink F, de Melker HE, van Binnendijk RS. Measles outbreak in a community with very low vaccine coverage, the Netherlands. Emerg Infect Dis. 2001;7(3 Suppl):593–7.

72. Keeling MJ, Grenfell BT. Disease Extinction and Community Size: Modeling the Persistence of Measles. Science. 1997 Jan 3;275(5296):65–7.

73. Fine PEM, Clarkson JA. Measles in England and Wales—I: An Analysis of Factors Underlying Seasonal Patterns. Int J Epidemiol. 1982 Mar 1;11(1):5–14.

74. Varughese P. Measles in Canada—1986 update. CMAJ Can Med Assoc J. 1986 Sep 1;135(5):499–501.

75. Varughese P, Ellis E, Carter A. Measles in Canada. Can Dis Wkly Rep Rapp Hebd Mal Au Can. 1989 Jan 1;15(1):1–7.

76. Boulianne N, De Serres G, Duval B, Joly JR, Meyer F, Déry P, et al. Major measles epidemic in the region of Quebec despite a 99% vaccine coverage. Can J Public Health Rev Can Sante Publique. 1991 Jun;82(3):189–90.

77. Sutcliffe PA, Rea E. Outbreak of measles in a highly vaccinated secondary school population. CMAJ Can Med Assoc J. 1996 Nov 15;155(10):1407–13.

78. King K, Varughese P, De Serres G, Tipples G, Waters J. Measles Elimination in Canada. J Infect Dis. 2004 May 1;189(Supplement_1):S236–42.

79. Coulby C, Reyes Domingo F, Hiebert J, Squires SG. Measles surveillance in Canada, 2019. Can Commun Dis Rep. 2021 Mar 31;47(03):149–59.

80. Coulby C, Reyes Domingo F, Hiebert J, MacDonald D. Measles surveillance in Canada: 2018. Can Commun Dis Rep. 2020 Apr 2;46(04):77–83.

81. Public Health Agency of Canada. Measles surveillance in Canada: 2017. Government of Canada. 2019. Available from: http://www.canada.ca/en/public-health/services/publications/diseases-conditions/measles-surveillance-canada-2017.html. [Accessed 2021 May 25].

82. Public Health Agency of Canada. Measles surveillance in Canada: 2016. Government of Canada. 2017. Available from: http://www.canada.ca/en/public-

health/services/publications/diseases-conditions/measles-surveillance-canada-2016.html. [Accessed 2021 May 25].

83. Sherrard L, Hiebert J, Cunliffe J, Mendoza L, Cutler J. Measles surveillance in Canada: 2015. Can Commun Dis Rep Releve Mal Transm Au Can. 2016 Jul 7;42(7):139–45.

84. Sherrard L, Hiebert J, Squires S. Measles surveillance in Canada: Trends for 2014. Can Commun Dis Rep. 2015 Jul 2;41(7):157–68.

85. Shane A, Hiebert J, Sherrard L, Deehan H. Measles surveillance in Canada: Trends for 2013. Can Commun Dis Rep. 2014 Jun 12;40(12):219–32.

86. Lim GH, Deeks SL, Fediurek J, Gubbay J, Crowcroft NS. Documenting the elimination of measles, rubella and congenital rubella syndrome in Ontario: 2009–12. Can Commun Dis Rep Releve Mal Transm Au Can. 2014 Apr 17;40(8):143–51.

87. Statistics Canada. Population estimates, quarterly. 2021. Available from: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000901. [Accessed 2021 May 30].

88. Hobbs L, Seo CY, Paul L, Johnson C, Harris T, Wilson S. Technical Annex: Immunization Coverage Report for School Pupils in Ontario: 2018–19 School Year. Ontario, Canada: Public Health Ontario; 2020 Aug. Available from: https://www.publichealthontario.ca/-/media/documents/i/2020/immunization-coverage-technical-2019-20.pdf?la=en

89. Hobbs L, Seo CY, Paul L, Johnson C, Harris T, Wilson S. Immunization Coverage Report for School Pupils in Ontario, 2018-19 School Year. Ontario,Canada: Public Health Ontario; 2020 Aug. Available from: https://www.publichealthontario.ca/-/media/documents/i/2020/immunization-coverage-2018-19.pdf?la=en

90. Ministry of Finance. Ontario Demographic Quarterly: Highlights of Second Quarter 2019. Government of Ontario. Ministry of Finance: Government of Ontario; 2019. Available from: https://www.fin.gov.on.ca/en/economy/demographics/quarterly/dhiq2.html. [Accessed 2021 Jun 7].

91. Kreps GL, Sparks L. Meeting the health literacy needs of immigrant populations. Patient Educ Couns. 2008 Jun 1;71(3):328–32.

92. Zanchetta MS, Poureslami IM. Health Literacy Within the Reality of Immigrants' Culture and Language. Can J Public Health Rev Can Santee Publique. 2006;97:S26–30.

93. Opel DJ, Taylor JA, Mangione-Smith R, Solomon C, Zhao C, Catz S, et al. Validity and reliability of a survey to identify vaccine-hesitant parents. Vaccine. 2011 Sep 2;29(38):6598–605.

94. Gust DA, Woodruff R, Kennedy A, Brown C, Sheedy K, Hibbs B. Parental perceptions surrounding risks and benefits of immunization. Semin Pediatr Infect Dis. 2003 Jul;14(3):207–12.

95. Kennedy AM, Brown CJ, Gust DA. Vaccine beliefs of parents who oppose compulsory vaccination. Public Health Rep Wash DC 1974. 2005 Jun;120(3):252–8.

96. Shui IM, Weintraub ES, Gust DA. Parents concerned about vaccine safety: Differences in race/ethnicity and attitudes. Am J Prev Med. 2006 Sep;31(3):244–51.

97. Prislin R, Dyer JA, Blakely CH, Johnson CD. Immunization status and sociodemographic characteristics: the mediating role of beliefs, attitudes, and perceived control. Am J Public Health. 1998 Dec;88(12):1821–6.

98. Sartorius B, Cohen C, Chirwa T, Ntshoe G, Puren A, Hofman K. Identifying high-risk areas for sporadic measles outbreaks: lessons from South Africa. Bull World Health Organ. 2013 Mar 1;91(3):174–83.

99. Public Health Ontario. iPHIS Resources. Public Health Ontario. Available from: https://www.publichealthontario.ca/en/Diseases and Conditions/Infectious Diseases/CCM/iPHIS. [Accessed 2020 May 31].

100. Public Health Ontario. Data Requests. Public Health Ontario. 2017. Available from: https://www.publichealthontario.ca/en/Data and Analysis/Using Data/Data Requests. [Accessed 2021 Jun 11].

101. Government of Ontario. Infectious Diseases Protocol Appendix B: Provincial Case Definitions for Diseases of Public Health Significance Disease: Measles. Ontario, Canada: Ministry of Health and Long-Term Care; 2019 Feb. Available from: https://www.health.gov.on.ca/en/pro/programs/publichealth/oph_standards/docs/measles_cd.pdf

102. Statistics Canada. Census Profile, 2016 Census. Government of Canada. 2017. Available from: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/download-telecharger/comp/page_dl-tc.cfm?Lang=E. [Accessed 2021 Jun 11].

103. Statistics Canada. Census Data Collection. Government of Canada. 2018. Available from: https://www12.statcan.gc.ca/census-recensement/2016/ref/98-306/ch1-eng.cfm. [Accessed 2021 Jun 11].

104. Statistics Canada. Census Short-Form (2A) Questionnaire. Government of Canada. 2016. Available from: https://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=assembleInstr&a=1&&lang=en&Item_Id=295241. [Accessed 2021 Jun 10].

105. Statistics Canada. Estimation from the census long form sample - Sampling and Weighting Technical Report, Census of Population, 2016. Government of Canada. 2018.

Available from: https://www12.statcan.gc.ca/census-recensement/2016/ref/98-306/ch4-eng.cfm. [Accessed 2021 Jun 11].

106. Statistics Canada. Census Long-Form (2A-L) Questionnaire. Government of Canada. 2016. Available from: https://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=assembleInstr&a=1&&lang=en&Item_Id=295122. [Accessed 2021 Jun 10].

107. Statistics Canada. Forward Sortation Area Boundary File. Government of Canada. 2017. Available from: https://www150.statcan.gc.ca/n1/en/catalogue/92-179-X. [Accessed 2021 Jun 10].

108. Natural Resources Canada. Geolocation Service. Government of Canada. Natural Resources Canada; 2015. Available from: https://www.nrcan.gc.ca/science-and-data/science-and-research/earth-sciences/geography/topographic-information/web-services/geolocation-service/17304. [Accessed 2021 Jun 11].

109. R Core Team. R: A Language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org/

110. Wickham H. Tidyverse. Comprehensive R Archive Network (CRAN); 2021. Available from: https://CRAN.R-project.org/package=tidyverse. [Accessed 2021 Jun 13]

111. Wickham H, Bryan J, Kalicinski M, Valery K, Leitienne C, Colbert B, et al. Readxl. Comprehensive R Archive Network (CRAN); 2019. Available from: https://CRAN.R-project.org/package=readxl. [Accessed 2021 Jun 13].

112. Wickham H. Rvest. Comprehensive R Archive Network (CRAN); 2021. Available from: https://CRAN.R-project.org/package=rvest. [Accessed 2021 Jun 13].

113. Bivand R. Spdep. Comprehensive R Archive Network (CRAN); 2021. Available from: https://CRAN.R-project.org/package=spdep. [Accessed 2021 Jun 13].

114. Hijmans RJ. Raster. Comprehensive R Archive Network (CRAN); 2021. Available from: https://CRAN.R-project.org/package=raster. [Accessed 2021 Jun 13].

115. Government of Canada. Forward Sortation Area—Definition. Innovation, Science and Economic Development Canada; 2015. Available from: https://www.ic.gc.ca/eic/site/bsf-osb.nsf/eng/br03396.html. [Accessed 2021 Jun 15].

116. Kleinman K. rsatscan. 2015. Available from: https://CRAN.R-project.org/package=rsatscan. [Accessed 2021 Jun 17].

117. Warnes GR, Bolker B, Lumley T, SAIC-Frederick and RCJC from RCJ are C (2005), Program IF by the IR, NIH of the, et al. gmodels. 2018. Available from: https://CRAN.R-project.org/package=gmodels. [Accessed 2021 Jun 17].

118. Wilke CO. cowplot. 2020. Available from: https://CRAN.R-project.org/package=cowplot. [Accessed 2021 Jun 17].

119. Kulldorff M. SaTScan. 2005. Available from: https://www.satscan.org/download.html. [Accessed 2021 Jun 17].

120. Kulldorff M. A spatial scan statistic. Commun Stat—Theory Methods. 1997 Jan 1;26(6):1481–96.

121. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. Stat Med. 1995 Apr 30;14(8):799–810.

122. Kulldorff M. SaTScan User Guide. 2021 Jan. Available from: https://www.satscan.org/cgi-bin/satscan/register.pl/SaTScan_Users_Guide.pdf?todo=process_userguide_download

123. Lihou K, Rose Vineer H, Wall R. Distribution and prevalence of ticks and tick-borne disease on sheep and cattle farms in Great Britain. Parasit Vectors. 2020 Aug 10;13(1):406.

124. Read S, Bath P, Willett P, Maheswaran R. Measuring the spatial accuracy of the spatial scan statistic. Spat Spatio-Temporal Epidemiol. 2011 Jun 1;2(2):69–78.

125. Han J, Zhu L, Kulldorff M, Hostovich S, Stinchcomb DG, Tatalovich Z, et al. Using Gini coefficient to determining optimal cluster reporting sizes for spatial scan statistics. Int J Health Geogr. 2016 Aug 3;15(1):27.

126. Wang W, Zhang T, Yin F, Xiao X, Chen S, Zhang X, et al. Using the maximum clustering heterogeneous set-proportion to select the maximum window size for the spatial scan statistic. Sci Rep. 2020 Mar 17;10(1):4900.

127. Statistics Canada. Immigration and ethnocultural diversity: Key results from the 2016 Census. Government of Canada. 2017. Available from: https://www150.statcan.gc.ca/n1/daily-quotidien/171025/dq171025b-eng.htm?indid=14428-1&indgeo=0. [Accessed 2021 Jun 21].

128. Statistics Canada. Education in Canada: Key results from the 2016 Census. Government of Canada. 2017. Available from: https://www150.statcan.gc.ca/n1/daily-quotidien/171129/dq171129a-eng.html. [Accessed 2021 Jun 21].

129. Durrheim DN, Crowcroft NS, Strebel PM. Measles - The epidemiology of elimination. Vaccine. 2014 Dec 5;32(51):6880–3.

130. Rota PA, Brown K, Mankertz A, Santibanez S, Shulga S, Muller CP, et al. Global Distribution of Measles Genotypes and Measles Molecular Epidemiology. J Infect Dis. 2011 Jul;204(suppl_1):S514–23.

131. Bahns AJ, Crandall CS, Gillath O, Preacher KJ. Similarity in relationships as niche construction: Choice, stability, and influence within dyads in a free choice environment. J Pers Soc Psychol. 2017;112(2):329–55.

132. Thibaut JW. The Social Psychology of Groups. New York: Routledge; 2017. 342 p.

133. Public Health Agency of Canada. Immunization of persons new to Canada: Canadian Immunization Guide. Government of Canada. Available from: http://www.canada.ca/en/public-health/services/publications/healthy-living/canadian-immunization-guide-part-3-vaccination-specific-populations/page-10-immunization-persons-new-canada.html. [Accessed 2021 Jul 5].

134. Smith? RJ, Li J, Gordon R, Heffernan JM. Can we spend our way out of the AIDS epidemic? A world halting AIDS model. BMC Public Health. 2009 Nov 18;9(1):S15.

135. Furuse Y, Oshitani H. Global Transmission Dynamics of Measles in the Measles Elimination Era. Viruses. 2017 Apr;9(4):82.

136. Picot G, Hou F. The rise in low-income rates among immigrants in Canada. 2003 Jun;58.

137. Picot G, Hou F. Immigration, Low Income and Income Inequality in Canada: What's New in the 2000s?. 2014 Dec. Available from: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.668.7071&rep=rep1&type=pdf

138. Statistics Canada. Census in Brief: Does education pay? A comparison of earnings by level of education in Canada and its provinces and territories. Government of Canada. 2017. Available from: https://www12.statcan.gc.ca/census-recensement/2016/as-sa/98-200-x/2016024/98-200-x2016024-eng.cfm. [Accessed 2021 Jul 6].

139. Ramsay LC, Crowcroft NS, Thomas S, Aruffo E, Teslya A, Heffernan JM, et al. Cost-effectiveness of measles control during elimination in Ontario, Canada, 2015. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2019 Mar;24(11).

140. Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies: Canada. Ottawa: CADTH; 2017 Jul. Available from: https://www.cadth.ca/sites/default/files/pdf/economic_guidelines_worked_example.pdf

141. Bolotin S, Severini A, Hatchette T, McLachlan E, Savage R, Hughes SL, et al. Assessment of population immunity to measles in Ontario, Canada: a Canadian Immunization Research Network (CIRN) study. Hum Vaccines Immunother. 2019;15(12):2856–64.

142. Bunko A, Harris T, Seo CY, Wilson S. Immunization Coverage Report for School Pupils in Ontario: 2016-17 School Year. Ontario, Canada: Public Health Ontario; 2018 Aug p. 65. Available from: https://www.publichealthontario.ca/-/media/documents/i/2018/immunization-coverage-2016-17.pdf?la=en

143. Wilson S, Seo CY, Lim G. 2012–13 Immunization Coverage Report. 2014 Sep. Available from: https://www.publichealthontario.ca/-/media/documents/I/2014/immunization-coverage-2012-13.pdf

144. Whittle H, Aaby P, Samb B, Cissé B, Kanteh F, Soumaré M, et al. Poor serologic responses five to seven years after immunization with high and standard titer measles vaccines. Pediatr Infect Dis J. 1999 Jan;18(1):53–7.

145. World Health Organization. The immunological basis for immunization series: module 7: measles: update 2020. World Health Organization; 2020. x, 36. Available from: https://apps.who.int/iris/handle/10665/331533. [Accessed 2021 Aug 26].

146. Wilson SE, Bunko A, Johnson S, Murray J, Wang Y, Deeks SL, et al. The geographic distribution of un-immunized children in Ontario, Canada: Hotspot detection using Bayesian spatial analysis. Vaccine. 2021 Feb 22;39(8):1349–57.

147. Public Health Agency of Canada. Benefits of immunization: Canadian Immunization Guide. 2016. Available from: http://www.canada.ca/en/public-health/services/publications/healthy-living/canadian-immunization-guide-part-1-key-immunization-information/page-3-benefits-immunization.html. [Accessed 2021 Jul 8].

148. Pelletier L, Chung P, Duclos P, Manga P, Scott J. A benefit-cost analysis of two-dose measles immunization in Canada. Vaccine. 1998 May 1;16(9–10):989–96.

149. Public Health Ontario. Daily Epidemiological Summary–COVID-19 in Ontario: January 15, 2020 to July 7, 2021. Ontario: Public Health Ontario; 2021 Jul. Available from: https://www.publichealthontario.ca/-/media/documents/ncov/epi/covid-19-daily-epi-summary-report.pdf?la=en

150. Canadian Institute for Health Information. Covid-19 Intervention Timeline in Canada. Canadian Institute for Health Information; 2021 Jun. Available from: https://www.cihi.ca/en/covid-19-intervention-timeline-in-canada

151. WHO. Coronavirus disease (COVID-19): How is it transmitted?. 2020. Available from: https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted. [Accessed 2021 Jul 8].

152. Turin TC, Rashid R, Ferdous M, Naeem I, Rumana N, Rahman A, et al. Perceived barriers and primary care access experiences among immigrant Bangladeshi men in Canada. Fam Med Community Health. 2020 Sep 1;8(4):e000453.

153. Wang L, Guruge S, Montana G. Older Immigrants' Access to Primary Health Care in Canada: A Scoping Review. Can J Aging Rev Can Vieil. 2019 Jun;38(2):193–209.

154. Wang L, Roisman D. Modeling Spatial Accessibility of Immigrants to Culturally Diverse Family Physicians. Prof Geogr. 2011 Feb 1;63(1):73–91.

155. Wang L, Rosenberg M, Lo L. Ethnicity and utilization of family physicians: a case study of Mainland Chinese immigrants in Toronto, Canada. Soc Sci Med 1982. 2008 Nov;67(9):1410–22.

156. Salami B, Mason A, Salma J, Yohani S, Amin M, Okeke-Ihejirika P, et al. Access to Healthcare for Immigrant Children in Canada. Int J Environ Res Public Health. 2020 May 10;17(9):E3320.

157. Ridde V, Aho J, Ndao EM, Benoit M, Hanley J, Lagrange S, et al. Unmet healthcare needs among migrants without medical insurance in Montreal, Canada. Glob Public Health. 2020 Nov;15(11):1603–16.

158. Ravichandiran N. Utilization of Healthcare by Immigrants in Canada: A Cross-Sectional Analysis of the Canadian Community Health Survey [dissertation]. London, Ontario: The University of Western Ontario; 2020 Jun 22. Available from: https://ir.lib.uwo.ca/etd/7155

159. COVID-19 Vaccination Coverage in Ottawa Neighbourhoods | Ottawa Neighbourhood Study. 2021. Available from: https://www.neighbourhoodstudy.ca/covid-19-vaccination-coverage-in-ottawa-neighbourhoods/. [Accessed 2021 Aug 12].

160. Governement of Ontario. Ontario's Long-Term Report on the Economy: Chapter 1: Demographic Trends and Projections | Ontario.ca. 2020. Available from: https://www.ontario.ca/document/ontarios-long-term-report-economy/chapter-1-demographic-trends-and-projections. [Accessed 2021 Jul 9].

161. Statistics Canada. Access to microdata. Government of Canada. 2015. Available from: https://www.statcan.gc.ca/eng/microdata. [Accessed 2021 Jul 10].

Appendix A – R Code for Data Curation & Analysis

```
#These are the packages we will use to load some excel files, perform data wr
angling techniques and conduct data analysis

library(tidyverse)

library(readxl)

library(rvest)

library(spdep)

library(raster)

library(ParSatscan)

library(rsatscan)

library(gmodels)

library(cowplot)

library(sp)

library(ggforce)


sessionInfo()
```

```
#These are the datasets that will be used

#IPHIS measles cases data

Thesis_Data_R <- read_excel("~/MSc Epidemiology/Thesis/Data/Measles Data/Thes
is Data_R.xlsx")

head(Thesis_Data_R,100)

str(Thesis_Data_R)


#2016 census dataset

Census2016<-read.csv("~/MSc Epidemiology/Thesis/Data/2016 Census/98-401-X2016
046_English_CSV_data.csv")

head(Census2016,100)

str(Census2016)


#Shape data file of Canada's FSAs

FSAshp<-st_read( "C:\\Users\\marce\\Documents\\MSc Epidemiology\\Thesis\\Data
\\FSA shape file")

summary(FSAshp)
```

```
plot(FSAshp)
```

```
#This selects only data relevant for SaTScan analysis from the measles datase
t and groups all measles cases by FSA

SaTScanDat<-select(Thesis_Data_R,caseno,FSA)%>%

  mutate(count(Thesis_Data_R,caseno)) %>%

  group_by(FSA) %>%

  summarize(sum(n))


#Now we can view the new dataset and its structure to ensure everything is in
order

view(SaTScanDat)

str(SaTScanDat)

summary(SaTScanDat)


#This saves the dataset as a csv file in the desired location

write.csv(SaTScanDat, "/Users/marce/Documents/MSc Epidemiology/Thesis/Data/Sa
TScanDat.csv")
```

```
#Data wrangling the census dataset to select only FSAs from Ontario and the t
otal population in each FSA

Census2016 %>%

  filter(grepl("^[KLMNP]",GEO_CODE..POR.))%>%

  dplyr::select(GEO_CODE..POR.,DIM..Profile.of.Forward.Sortation.Areas..2247.
,Dim..Sex..3...Member.ID...1...Total...Sex,Dim..Sex..3...Member.ID...2...Male
,Dim..Sex..3...Member.ID...3...Female) -> Census2016ON


Census2016ON%>% filter(grepl("Population\\, 2016",DIM..Profile.of.Forward.Sor
tation.Areas..2247.))->Census2016ON_Totpop


colnames(Census2016ON_Totpop)[c(1,3)]<-c("FSA","PopulationTotal2016")


Census2016ON_Totpop %>% dplyr::select(FSA,PopulationTotal2016)->Census2016ON_
Totpop


as.numeric(Census2016ON_Totpop$PopulationTotal2016)->Census2016ON_Totpop$Popu
lationTotal2016
```

```
#Now we can view the new dataset and its structure to ensure everything is in
order

length(unique(Census2016ON_Totpop$FSA))

sum(Census2016ON_Totpop$PopulationTotal2016)

str(Census2016ON_Totpop)

view(Census2016ON_Totpop)
```

```
#Merging the measles cases dataset created earlier and the census dataset to
have the total cases of measles and the total population by FSA

merge(Census2016ON_Totpop,SaTScanDat,all=TRUE) %>% unique() %>% replace_na(li
st(`sum(n)`=0))->merged

colnames(merged)[c(3)]<-c("MeaslesCases")


#Now we can view the new dataset and its structure to ensure everything is in
order

view(merged)

summary(merged)

sum(merged$MeaslesCases)
```

```
#Now we need to perform web scrapping to obtain the centroid coordinates of e
ach FSA

#This is the link we will be extracting coordinates from

GeoLocURL<-"http://geogratis.gc.ca/services/geolocation/en/locate?q="


#This is the string pattern of the coordinates we will be scrapping

copat<-"[- ]*([0-9]{2})[.]([0-9]{3,7})[,][- ]*([0-9]{2})[.]([0-9]{3,7})"


#We meed to remove measles cases for which there are missing FSAs

merged2<-na.omit(merged)

view(merged2)

summary(merged2)

sum(merged2$MeaslesCases)


#This portion conducts the webscrapping
```

```
GeoLocURL %>%
  map2_chr(merged2$FSA,paste0) %>%
  map(. %>%
    read_html() %>%
    html_nodes("body") %>%
    html_text()) %>% unlist() -> coordinates


head(coordinates)
```

```
#This portion cleans the extracted web data using the string pattern describe
d above and transforms it into a format we can work with
str_extract(coordinates,copat) %>%
  as.data.frame() %>%
  separate("." , into=c("Longitude","Latitude"), sep = ",") %>%
  mutate(Longitude = as.numeric(Longitude)) %>%
  mutate(Latitude = as.numeric(Latitude))->COclean
head(COclean)
```

```
#this portion merges the extracted coordinates with the existing measles case
s and total population dataset by FSA
SatScanDat2<-add_column(merged2,COclean)


#Now we can view the new dataset and its structure to ensure everything is in
order
view(SatScanDat2)
summary(SatScanDat2)
SatScanDat2[rowSums(is.na(SatScanDat2))>0,]


#We can see that we are missing Longitude and Latitude values for the FSA L9E
. After a quick google search It appears that L9E and L9T are the same area s
o we will remove the L9E FSA.
```

```
#This removes the L9E FSA from the data set and we can see that there are no
longer any missing coordinates.
na.omit(SatScanDat2)->FinalSatScanDat
```

```
FinalSatScanDat[rowSums(is.na(FinalSatScanDat))>0,]

view(FinalSatScanDat)

str(FinalSatScanDat)
```

```
#Using the FSA shapefile loaded earlier, we select only FSAs within ontario t
o create a shapefile of Ontario FSAs

ONFSAshp<-FSAshp[FSAshp$PRUID==35,]

summary(ONFSAshp)

plot(ONFSAshp)



#Now we need to create a list indicating contiguous relationships between FSA
s for the MCHS-P technique

ContiONFSA<-poly2nb(ONFSAshp$geometry, row.names = ONFSAshp$CFSAUID)

summary(ContiONFSA)
```

```
#Now we need to make sure that the FSAs in the Ontario shapefile and the SatS
can data file are the same

#From the function below we can see that the FSAs K1A,L5S,L5T,M5K,M5L,M5W,M5X
,M7A,M7Y are in the FinalSatScan Data but not in the shapefile, we need to re
move them or else there will error outputs when computing the MCHS-P. The mis
sing FSAs indicate the location of institutial or governmental property and p
resent low population counts and zero measles cases.

filter(FinalSatScanDat,!(FinalSatScanDat$FSA %in% ONFSAshp$CFSAUID))


#To remove the FSAs that are in FinalSatScanDat but not in the shapefile, we
run the following code and we can see that those FSAs have been removed.

FinalSatScanDat%>% filter(FinalSatScanDat$FSA %in% ONFSAshp$CFSAUID)->FinalSa
tScanData

filter(FinalSatScanData,!(FinalSatScanData$FSA %in% ONFSAshp$CFSAUID))

summary(FinalSatScanData)

sum(FinalSatScanData$PopulationTotal2016)


#This saves the final SatScan dataset as a csv file in the desired location

write.csv(FinalSatScanData, "/Users/marce/Documents/MSc Epidemiology/Thesis/D
ata/SatScanData/FinalSatScanData.csv")
```

```
#Now we need to compute the MCHS-P


#This shows the SatScan parameters

ss.options()


#This resets the SatScan parameters

invisible(ss.options(reset = TRUE))


#And now we write new parameters and SatScan data files using the FinalSatSca
nData file

setwd(tempdir())

write.cas(FinalSatScanData[,c("FSA","MeaslesCases")],filename = "ONM", locati
on = getwd())

write.geo(FinalSatScanData[,c("FSA","Latitude","Longitude")],filename = "ONM"
, location = getwd())

write.pop(data.frame(FinalSatScanData[[1]],"unspecified",FinalSatScanData[[2]
]), filename = "ONM", location = getwd())


#This is a list of the parameters we are going to use

ss.options(list(CaseFile=paste0(getwd(),"/ONM.cas"), PrecisionCaseTimes=0,Coo
rdinatesFile=paste0(getwd(),"/ONM.geo"),  PopulationFile=paste0(getwd(),"/ONM
.pop"), CoordinatesType=1, AnalysisType=1,  ModelType=0, MonteCarloReps=999,
ReportGiniClusters="n"))




#The location where SatScan is saved on the OS

sslocation0<-"C:/Program Files (x86)/SaTScan"


#To compute the MCHS-P

ONMCHSP <-selpar_mchsp(pop = FinalSatScanData$PopulationTotal2016, case =Fina
lSatScanData$MeaslesCases,sizes=seq(0.1,50,by=0.1), id = FinalSatScanData$FSA
, prmsave = F, sslocation=sslocation0, gallist=ContiONFSA, verbose=T)


#The MCHS-P is 4.1

ONMCHSP$Osize

ONMCHSP$Oresult
```

```
#Now we can run a geospatial analysis with SatScan

#This shows a list of the potential input parameters and clears all parameter
s previously used to make way for new ones.

ss.options()

invisible(ss.options(reset = TRUE))


#And now we write new parameters and SatScan data files using the FinalSatSca
nData file

#These are the data files that are read in by SatScan

setwd(tempdir())

write.cas(FinalSatScanData[,c("FSA","MeaslesCases")],filename = "ONM", locati
on = getwd())

write.geo(FinalSatScanData[,c("FSA","Latitude","Longitude")],filename = "ONM"
, location = getwd())

write.pop(data.frame(FinalSatScanData[[1]],"unspecified",FinalSatScanData[[2]
]), filename = "ONM", location = getwd())


#This is a list of the parameters we are going to use

ss.options(list(CaseFile=paste0(getwd(),"/ONM.cas"), PrecisionCaseTimes=0,Coo
rdinatesFile=paste0(getwd(),"/ONM.geo"), PopulationFile=paste0(getwd(),"/ONM.
pop"), CoordinatesType=1, AnalysisType=1,  ModelType=0, MonteCarloReps=99999,
MaxSpatialSizeInPopulationAtRisk=4.1, ReportGiniClusters="n", StartDate="2008
/01/01", EndDate="2019/01/01"))

write.ss.prm(getwd(),"ONM")


#The location where SatScan is saved on the OS

sslocation0<-"C:/Program Files (x86)/SaTScan"


#The line that runs SatScan

ONM_GSA=satscan(getwd(),"ONM",sslocation = sslocation0, verbose = TRUE)


#SatScan results can be accessed with the following code but are masked to pr
event identification of case FSAs

summary(ONM_GSA)

ONM_GSA
```

```
#Now we need to get frequencies of various variables of interest from the Can
adian Census and Identify case FSAs from the SatScan Analysis
```

```
#We need counts for immigrant status

filter(Census2016ON,grepl("^Immigrants",DIM..Profile.of.Forward.Sortation.Areas..2247.)) %>% view()


Census2016ON %>% filter(grepl("^Immigrants",DIM..Profile.of.Forward.Sortation.Areas..2247.))->Census2016ON_Imigrnt

  Census2016ON_Imigrnt[match(unique(Census2016ON_Imigrnt$GEO_CODE..POR.),Census2016ON_Imigrnt$GEO_CODE..POR.),]->Census2016ON_Imigrnt

  colnames(Census2016ON_Imigrnt)[c(1,3)]<-c("FSA","Imigrnt")

  Census2016ON_Imigrnt$Imigrnt<-as.numeric(Census2016ON_Imigrnt$Imigrnt)

  dplyr::select(Census2016ON_Imigrnt,c(1,3))->Census2016ON_Imigrnt

  str(Census2016ON_Imigrnt)

  summary(Census2016ON_Imigrnt)


Census2016ON %>% filter(grepl("^Non-immigrants",DIM..Profile.of.Forward.Sortation.Areas..2247.))->Census2016ON_NImigrnt

  Census2016ON_NImigrnt[match(unique(Census2016ON_NImigrnt$GEO_CODE..POR.),Census2016ON_NImigrnt$GEO_CODE..POR.),]->Census2016ON_NImigrnt

  colnames(Census2016ON_NImigrnt)[c(1,3)]<-c("FSA","NImigrnt")

  Census2016ON_NImigrnt$NImigrnt<-as.numeric(Census2016ON_NImigrnt$NImigrnt)

  dplyr::select(Census2016ON_NImigrnt,c(1,3))->Census2016ON_NImigrnt

  str(Census2016ON_NImigrnt)

  summary(Census2016ON_NImigrnt)


Census2016ON %>% filter(grepl("^Non-permanent residents",DIM..Profile.of.Forward.Sortation.Areas..2247.))->Census2016ON_Npermres

  Census2016ON_Npermres[match(unique(Census2016ON_Npermres$GEO_CODE..POR.),Census2016ON_Npermres$GEO_CODE..POR.),]->Census2016ON_Npermres

  colnames(Census2016ON_Npermres)[c(1,3)]<-c("FSA","Npermres")

  Census2016ON_Npermres$Npermres<-as.numeric(Census2016ON_Npermres$Npermres)

  dplyr::select(Census2016ON_Npermres,c(1,3))->Census2016ON_Npermres

  str(Census2016ON_Npermres)

  summary(Census2016ON_Npermres)
```

```
#Now we need to extract the variable for income

filter(Census2016ON,grepl("^Prevalence of low income based on the Low-income
cut-offs",DIM..Profile.of.Forward.Sortation.Areas..2247.)) %>% view()


Census2016ON %>% filter(grepl("^Prevalence of low income based on the Low-inc
ome cut-offs",DIM..Profile.of.Forward.Sortation.Areas..2247.))->Census2016ON_
LowIncome

  colnames(Census2016ON_LowIncome)[c(1,3)]<-c("FSA","Low-Income Prevalence (%
)")

  Census2016ON_LowIncome$`Low-Income Prevalence (%)`<-as.numeric(Census2016ON
_LowIncome$`Low-Income Prevalence (%)`)

  dplyr::select(Census2016ON_LowIncome,c(1,3))->Census2016ON_LowIncome

  str(Census2016ON_LowIncome)

  summary(Census2016ON_LowIncome)

  view(Census2016ON_LowIncome)
```

```
#Now for the education variable

filter(Census2016ON,grepl("^No certificate, diploma or degree",DIM..Profile.o
f.Forward.Sortation.Areas..2247.)) %>% view()


Census2016ON %>% filter(grepl("^No certificate, diploma or degree",DIM..Profi
le.of.Forward.Sortation.Areas..2247.))->Census2016ON_NoDiploma

colnames(Census2016ON_NoDiploma)[c(1,3)]<-c("FSA","NoDiploma")

subset(Census2016ON_NoDiploma,duplicated(Census2016ON_NoDiploma$FSA))->Census
2016ON_NoDiploma

dplyr::select(Census2016ON_NoDiploma,c(1,3))->Census2016ON_NoDiploma

as.numeric(Census2016ON_NoDiploma$NoDiploma)->Census2016ON_NoDiploma$NoDiplom
a

str(Census2016ON_NoDiploma)

summary(Census2016ON_NoDiploma)

view(Census2016ON_NoDiploma)



filter(Census2016ON,grepl("^Secondary",DIM..Profile.of.Forward.Sortation.Area
s..2247.)) %>% view()
```

```
Census2016ON %>% filter(grepl("Secondary",DIM..Profile.of.Forward.Sortation.Areas..2247.))->Census2016ON_HighSchoolDiploma

colnames(Census2016ON_HighSchoolDiploma)[c(1,3)]<-c("FSA","HighSchoolDiploma")

subset(Census2016ON_HighSchoolDiploma,duplicated(Census2016ON_HighSchoolDiploma$FSA))->Census2016ON_HighSchoolDiploma

subset(Census2016ON_HighSchoolDiploma,duplicated(Census2016ON_HighSchoolDiploma$FSA))->Census2016ON_HighSchoolDiploma

dplyr::select(Census2016ON_HighSchoolDiploma,c(1,3))->Census2016ON_HighSchoolDiploma

as.numeric(Census2016ON_HighSchoolDiploma$HighSchoolDiploma)->Census2016ON_HighSchoolDiploma$HighSchoolDiploma

str(Census2016ON_HighSchoolDiploma)

summary(Census2016ON_HighSchoolDiploma)

view(Census2016ON_HighSchoolDiploma)



filter(Census2016ON,grepl("^Postsecondary",DIM..Profile.of.Forward.Sortation.Areas..2247.)) %>% view()


Census2016ON %>% filter(grepl("Postsecondary",DIM..Profile.of.Forward.Sortation.Areas..2247.))->Census2016ON_PostsecondaryDiploma

colnames(Census2016ON_PostsecondaryDiploma)[c(1,3)]<-c("FSA","PostsecondaryDiploma")

subset(Census2016ON_PostsecondaryDiploma,duplicated(Census2016ON_PostsecondaryDiploma$FSA))->Census2016ON_PostsecondaryDiploma

Census2016ON_PostsecondaryDiploma[match(unique(Census2016ON_PostsecondaryDiploma$FSA),Census2016ON_PostsecondaryDiploma$FSA),]->Census2016ON_PostsecondaryDiploma

dplyr::select(Census2016ON_PostsecondaryDiploma,c(1,3))->Census2016ON_PostsecondaryDiploma

as.numeric(Census2016ON_PostsecondaryDiploma$PostsecondaryDiploma)->Census2016ON_PostsecondaryDiploma$PostsecondaryDiploma

str(Census2016ON_PostsecondaryDiploma)

summary(Census2016ON_PostsecondaryDiploma)

view(Census2016ON_PostsecondaryDiploma)
```

```
#Now we need to merge all our extracted variables into one single analysis set
```

```
merge(Census2016ON_Totpop,Census2016ON_Imigrnt,by="FSA") %>%
  merge(Census2016ON_NImigrnt,by="FSA") %>%
    merge(Census2016ON_Npermres,by="FSA") %>%
     merge(Census2016ON_LowIncome,by="FSA") %>%
      merge(Census2016ON_NoDiploma,by="FSA") %>%
       merge(Census2016ON_HighSchoolDiploma,by="FSA") %>%
        merge(Census2016ON_PostsecondaryDiploma,by="FSA")-> CaseCon


str(CaseCon)
```

```
#We remove FSA that were excluded from the geospatial analysis.
CaseCon%>% filter(CaseCon$FSA %in% ONFSAshp$CFSAUID)->CaseConData
```

```
#Check which FSAs are missing counts for different variables of interest
CaseConData[rowSums(is.na(CaseConData))>0,]


#L4V and P0Y are missing counts for varibales of interest, likely do to their
low population counts (5 and 30). This chunk removes them and checks wether t
hey were removed.
na.omit(CaseConData)->CaseConData
CaseConData[rowSums(is.na(CaseConData))>0,]
```

```
#We also need to identify case & control FSAs using the results from the SatS
can Analysis. Case FSAs are not shown below.
CaseConData$Cases=ifelse(CaseConData$FSA %in% c("Inlclude case FSAs here"),1,
0)


sum(CaseConData$Cases)
```

```
#We also want to create a second variable of low income prevalence divided by
100 to facilitate ploting later
CaseConData$LowIncome=CaseConData$`Low-Income Prevalence (%)`/100
```

```
#We also want a variable with the Immigrant proportion and a Binary immigrant variable

CaseConData$ImmiProp<-rowSums(CaseConData[,c("Imigrnt","Npermres")],na.rm=TRUE)/(rowSums(CaseConData[,c("Imigrnt","Npermres","NImigrnt")],na.rm=TRUE))

CaseConData$ImmiBin=ifelse(CaseConData$ImmiProp>0.219,1,0)
```

```
#We also want a variable with the post-secondary education proportion and a Binary education exposure variable

CaseConData$EduProp<-CaseConData$PostsecondaryDiploma/(rowSums(CaseConData[,c("PostsecondaryDiploma","NoDiploma","HighSchoolDiploma")],na.rm=TRUE))

CaseConData$EduBin=ifelse(CaseConData$EduProp<0.54,1,0)
```

```
#We also want a binary low-income variable

CaseConData$IncomeBin=ifelse(CaseConData$`Low-Income Prevalence (%)`>8.8,1,0)
```

```
#To get the population density we need to merge the shape file with the casecondata and make a few changes to the geometry properties.


#First We need to change region ID name to FSA in order to merge the two files

names(ONFSAshp)[names(ONFSAshp)=="CFSAUID"]<-"FSA"


#This merges the shapefile with the casecondata

CaseConData<-merge(ONFSAshp,CaseConData,by="FSA")

str(CaseConData)


#Now we see that the originial shapefile is not using NAD83 for the CRS, this needs to be changed on the file to the WGS84 system to make it easier to work with

crs(ONFSAshp$geometry)


#We change the CRS of the geometry to the WGS84 system (lat/long system) to make it easier to work with

CaseConData$geometry<-st_transform(CaseConData$geometry,st_crs("WGS84"))

str(CaseConData)

crs(CaseConData)
```

```
#This approximates the land area in Km^2 and we can see that the sum of the 5
11 FSAs is approximately 1.05 million Km^2 which is close to the 1.08 million
km^2 landmass of Ontario reported online.
CaseConData$Area_KM2<-as.vector(st_area(CaseConData$geometry)/1000000)
sum(CaseConData$Area_KM2)


#This creates a population density variable
CaseConData$PopDens<-CaseConData$PopulationTotal2016/CaseConData$Area_KM2
summary(CaseConData$PopDens)


#Now lets create a binary variable with the population density, with >1000 in
d/Km^2 being 1 adn less being 0
CaseConData$PopDensBin=as.factor(ifelse(CaseConData$PopDens>1000,1,0))
str(CaseConData$PopDensBin)
summary(CaseConData$PopDensBin)


#We can look at our working dataset
view(CaseConData)


write.csv(CaseConData, "/Users/marce/Documents/MSc Epidemiology/Thesis/Data/C
aseConData.csv")
```

```
#Now we can analyse the data
#View which rows may have missing data
CaseConData[rowSums(is.na(CaseConData))>0,]
```

```
#Percentage of total population of case areas
sum(CaseConData[which(CaseConData[,12]==1),4])/(sum(CaseConData[,4]))
```

```
#Number of Case FSAs
sum(CaseConData$Cases)
```

```
#Counts and proportions of FSAs in each expsoure category
```

```
#For immigrant variable

CrossTable(CaseConData$ImmiBin,CaseConData$Cases,chisq = TRUE, fisher = TRUE)


#For population density

CrossTable(CaseConData$PopDensBin,CaseConData$Cases,chisq = TRUE, fisher = TR
UE)


#For low income prevalence

CrossTable(CaseConData$IncomeBin,CaseConData$Cases, chisq = TRUE, fisher = TR
UE)


#For education level

CrossTable(CaseConData$EduBin, CaseConData$Cases, chisq = TRUE, fisher = TRUE
)
```

```
#OR and 95% CIs for Immigrant proportion>21.9

ImmibinOR<-glm(Cases~ImmiBin,data = CaseConData,family = binomial)

zquantile<-qnorm(1-(1-0.95)/2)

lowerci1<-summary(ImmibinOR)$coefficients[,1]-zquantile*summary(ImmibinOR)$co
efficients[,2]

upperci1<-summary(ImmibinOR)$coefficients[,1]+zquantile*summary(ImmibinOR)$co
efficients[,2]

exp(summary(ImmibinOR)$coefficients[2,1])

exp(lowerci1[2])

exp(upperci1[2])

summary(ImmibinOR)

#FSAs with a proportion of immigrants greater than 21.9% have 9.92 times the
odds of being in a Measles hot spot than those with a proportion of immigrant
s equal to or lower than 21.9%
```

```
#OR and 95% CIs for population density>1000 people/km^2

PopdbinOR<-glm(Cases~PopDensBin, data = CaseConData, family = binomial)

lowerci2<-summary(PopdbinOR)$coefficients[,1]-zquantile*summary(PopdbinOR)$co
efficients[,2]

upperci2<-summary(PopdbinOR)$coefficients[,1]+zquantile*summary(PopdbinOR)$co
efficients[,2]

exp(summary(PopdbinOR)$coefficients[2,1])
```

```
exp(lowerci2[2])

exp(upperci2[2])

summary(PopdbinOR)
```

```
#OR and 95% CIs for Low Income prevalence >8.8%

IncomebinOR<-glm(Cases~IncomeBin, data = CaseConData, family = binomial)

lowerci3<-summary(IncomebinOR)$coefficients[,1]-zquantile*summary(IncomebinOR
)$coefficients[,2]

upperci3<-summary(IncomebinOR)$coefficients[,1]+zquantile*summary(IncomebinOR
)$coefficients[,2]

exp(summary(IncomebinOR)$coefficients[2,1])

exp(lowerci3[2])

exp(upperci3[2])

summary(IncomebinOR)
```

```
#OR and 95% CISs for Post-Secondary Education <54%

EdubinOR<-glm(Cases~EduBin,data = CaseConData,family = binomial)

lowerci4<-summary(EdubinOR)$coefficients[,1]-zquantile*summary(EdubinOR)$coef
ficients[,2]

upperci4<-summary(EdubinOR)$coefficients[,1]+zquantile*summary(EdubinOR)$coef
ficients[,2]

exp(summary(EdubinOR)$coefficients[2,1])

exp(lowerci4[2])

exp(upperci4[2])

summary(EdubinOR)
```

```
#Now to provide some informative histograms of cases within each significant
clusters we need to wrangle the iPHIs dataset.

#This creates a new iphis dataset we can manipulate

Thesis_Data_R->ClusterCharac


#This createa a new variable to identify the cluster that the case belongs to

ClusterCharac$Cluster=ifelse(is.na(Thesis_Data_R$FSA),NA,ifelse(Thesis_Data_R
$FSA %in% c("Insert FSAs from Cluster 1"),1,ifelse(Thesis_Data_R$FSA %in% c("
Insert FSAs from cluster 2"),2,ifelse(Thesis_Data_R$FSA %in% c("Insert FSAs f
rom cluster 3"),3,ifelse(Thesis_Data_R$FSA %in% c("Insert FSAs from Cluster 4
```

```
"),4,ifelse(Thesis_Data_R$FSA %in% c("Insert FSAs from Cluster 5 here"),5,ife
lse(Thesis_Data_R$FSA %in% c("Insert FSAs from Cluster 6 here"),6,0)))))))

as.factor(ClusterCharac$Cluster)

as.factor(ClusterCharac$year)


view(ClusterCharac)


ClusterCharac[ClusterCharac$Cluster==1,]
```

```
#Histogram for cluster 1

ggplot(filter(ClusterCharac,ClusterCharac$Cluster==1),aes(x=year,fill=genotyp
e))+geom_histogram(binwidth =1,colour="black") +theme_bw()+ scale_x_continuou
s(breaks = 2008:2019,name = "Year") + labs(fill = "Genotype") + theme(legend.
text = element_text(size = 7),legend.background = element_rect(color="black",
linetype = "solid"),panel.grid.major.x = element_blank(),panel.grid.minor.x =
element_blank() )+scale_fill_manual(labels =c("B3","D4","D8","NA"),values = c
("salmon","deepskyblue","darkgoldenrod1"),na.value="gray")+scale_y_continuous
(expand = c(0,0), limits=c(0,18),breaks = seq(0,18,by=2),name = "Measles Case
Count")
```

```
#Histogram for cluster 2

ggplot(filter(ClusterCharac,ClusterCharac$Cluster==2),aes(x=year,fill=genotyp
e))+geom_histogram(binwidth =1,colour="black") +theme_bw()+ scale_x_continuou
s(limits=c(2008,2019),breaks = 2008:2019,name = "Year") + labs(fill = "Genoty
pe") + theme(legend.text = element_text(size = 7),legend.background = element
_rect(color="black",linetype = "solid"),panel.grid.major.x = element_blank(),
panel.grid.minor.x = element_blank() )+scale_fill_manual(labels =c("D4","NA")
,values = c("deepskyblue"),na.value="gray")+scale_y_continuous(expand = c(0,0
), limits=c(0,18),breaks = seq(0,18,by=2),name = "Measles Case Count")
```

```
#Histogram for cluster 3

ggplot(filter(ClusterCharac,ClusterCharac$Cluster==3),aes(x=year,fill=genotyp
e))+geom_histogram(binwidth =1, colour="black") +theme_bw()+ scale_x_continuo
us(limits=c(2007,2019),breaks = 2008:2019,name = "Year") + labs(fill = "Genot
ype") + theme(legend.text = element_text(size = 7),legend.background = elemen
t_rect(color="black",linetype = "solid"),panel.grid.major.x = element_blank()
,panel.grid.minor.x = element_blank() )+scale_fill_manual(labels =c("D4","D8"
),values = c("deepskyblue","darkgoldenrod1"),na.value="gray")+scale_y_continu
ous(expand = c(0,0), limits=c(0,18),breaks = seq(0,18,by=2),name = "Measles C
ase Count")
```

```
#Histogram for cluster 4
```

```
ggplot(filter(ClusterCharac,ClusterCharac$Cluster==4),aes(x=year,fill=genotyp
e))+geom_histogram(binwidth =1, colour="black") +theme_bw()+ scale_x_continuo
us(limits=c(2007,2019),breaks = 2008:2019,name = "Year") + labs(fill = "Genot
ype") + theme(legend.text = element_text(size = 7),legend.background = elemen
t_rect(color="black",linetype = "solid"),panel.grid.major.x = element_blank()
,panel.grid.minor.x = element_blank() )+scale_fill_manual(labels =c("D8","NA"
),values = c("darkgoldenrod1"),na.value="gray")+scale_y_continuous(expand = c
(0,0), limits=c(0,18),breaks = seq(0,18,by=2),name = "Measles Case Count")
```

```
#Histogram for cluster 5

ggplot(filter(ClusterCharac,ClusterCharac$Cluster==5),aes(x=year,fill=genotyp
e))+geom_histogram(binwidth =1, colour="black") +theme_bw()+ scale_x_continuo
us(limits=c(2007,2020),breaks = 2008:2019,name = "Year") + labs(fill = "Genot
ype") + theme(legend.text = element_text(size = 7),legend.background = elemen
t_rect(color="black",linetype = "solid"),panel.grid.major.x = element_blank()
,panel.grid.minor.x = element_blank() )+scale_fill_manual(labels =c("B3","D8"
,"NA"),values = c("salmon","darkgoldenrod1"),na.value="gray")+scale_y_continu
ous(expand = c(0,0), limits=c(0,18),breaks = seq(0,18,by=2),name = "Measles C
ase Count")
```

```
#Histogram for cluster 6

ggplot(filter(ClusterCharac,ClusterCharac$Cluster==6),aes(x=year,fill=genotyp
e))+geom_histogram(binwidth =1, colour="black") +theme_bw()+ scale_x_continuo
us(limits=c(2007,2020),breaks = 2008:2019,name = "Year") + labs(fill = "Genot
ype") + theme(legend.text = element_text(size = 7),legend.background = elemen
t_rect(color="black",linetype = "solid"),panel.grid.major.x = element_blank()
,panel.grid.minor.x = element_blank() )+scale_fill_manual(labels =c("B3","D4"
,"D8","NA"),values = c("salmon","deepskyblue","darkgoldenrod1"),na.value="gra
y")+scale_y_continuous(expand = c(0,0), limits=c(0,18),breaks = seq(0,18,by=2
),name = "Measles Case Count")
```

```
#Now to create a map of the approximate cluster location in Ontario.

#First we need to create a data frame with the data required to create circle
s that encompass the general area of each cluster.

Clusters<-data.frame(Clusters = as.factor (c("Cluster 1","Cluster 2","Cluster
3","Cluster 4","Cluster 5","Cluster 6")),Longitude = as.numeric(c("Insert Lon
gitude Coordinates Here")),Latitude = as.numeric(c("Insert Latitude Coordinat
es Here")),r=as.numeric(c(0.25,0.25,0.25,0.25,0.25,0.25)))

view(Clusters)


#General cluster areas

ggplot()+geom_circle(aes(x0=Longitude,y0=Latitude,r=r,color=Clusters),data =
Clusters)
```

```
#This creates a main plot of Ontario with its FSAs and general cluster areas
added as a layer.

OntarioClusters<-ggplot(CaseConData)+geom_sf(color = alpha(c("black"),0.04),l
wd = 0)+coord_sf(expand = F)+theme(axis.text = element_blank(),axis.ticks = e
lement_blank(),rect = element_blank())+geom_circle(aes(x0=Longitude,y0=Latitu
de,r=r,color=Clusters),data = Clusters)+theme(axis.title.x = element_blank(),
axis.title.y = element_blank(),legend.position = c(-0.33,0.80),legend.text =
element_text(size = 7),legend.background = element_rect(color="black",linetyp
e = "solid"),legend.title = element_text(face = "bold"))



#Main plot of Ontario FSAs with general cluster areas and boxes in areas of i
nterest

OntarioClusters2<-OntarioClusters+geom_rect(xmin =-78.5, xmax =-82, ymin = 42
.5, ymax = 44.7,fill = NA, color = "grey", size = 0.6)+geom_rect(xmin =-75.2,
xmax =-76.5, ymin = 44.8, ymax = 45.8,fill = NA, color = "grey", size = 0.6)


#This adds a layer with other two areas of interest of the map zoomed in

OntarioClusters3<-ggdraw(OntarioClusters2)+draw_plot({OntarioClusters2 + coor
d_sf(xlim = c(-82,-78.5),ylim =c(42.5,44.7),expand = FALSE) +theme(legend.pos
ition = "none")},x =0.29,y=0.05,width = 0.2,height = 0.3)+draw_plot({OntarioC
lusters2 + coord_sf(xlim = c(-76.5,-75.2),ylim =c(44.8,45.8),expand = FALSE)
+theme(legend.position = "none")},x =0.66,y=0.33,width = 0.2,height = 0.3)


OntarioClusters3
```

```
#Now lets create a similar map but with the proprotion of immigrants in each
FSA

OnImmi<-ggplot(CaseConData)+geom_sf(aes(fill=ImmiProp), color = alpha(c("whit
e"),0.1),lwd = 0)+scale_fill_gradientn(colors = alpha(c("blue4","blue","purpl
e","orange","Orange1","orangered","red"),0.70),name = "% Immigrants",labels =
scales::percent)+coord_sf(expand = F)+theme(axis.text = element_blank(),axis.
ticks = element_blank(),rect = element_blank())+geom_circle(aes(x0=Longitude,
y0=Latitude,r=r,color=Clusters),data = Clusters)+theme(axis.title.x = element
_blank(),axis.title.y = element_blank(),legend.position = c(-0.29,0.59),legen
d.text = element_text(size = 7),legend.background = element_rect(color="black
",linetype = "solid"),legend.title = element_text(face = "bold"))


#Main plot of Ontario FSAs with general cluster areas, general cluster areas
and boxes in areas of interest

OnImmi2<-OnImmi+geom_rect(xmin =-78.5, xmax =-82, ymin = 42.5, ymax = 44.7,fi
ll = NA, color = "grey", size = 0.6)+geom_rect(xmin =-75.2, xmax =-76.5, ymin
= 44.8, ymax = 45.8,fill = NA, color = "grey", size = 0.6)
```

```
#This adds a layer with other two areas of interest of the map zoomed in

OnImmi3<-ggdraw(OnImmi2)+draw_plot({OnImmi2 + coord_sf(xlim = c(-82,-78.5),yl
im =c(42.5,44.7),expand = FALSE) +theme(legend.position = "none")},x =0.29,y=
0.05,width = 0.2,height = 0.3)+draw_plot({OnImmi2 + coord_sf(xlim = c(-76.5,-
75.2),ylim =c(44.8,45.8),expand = FALSE) +theme(legend.position = "none")},x
=0.66,y=0.33,width = 0.2,height = 0.3)


OnImmi3
```

```
#Now lets create a similar map but with the population density in each FSA

#Main plot of Ontario FSAs with

ONdens<-ggplot(CaseConData)+geom_sf(aes(fill=PopDens), color = alpha(c("white
"),0.1),lwd = 0)+scale_fill_gradientn(colors = alpha(c("blue4","blue","purple
","orange","Orange1","orangered","red"),0.70),name = expression(bold(paste("P
eople/Km"^"2"))),limits = c(250,3000),breaks=c(500,1000,1500,2000,2500,3000),
oob = scales::squish)+coord_sf(expand = F)+theme(axis.text = element_blank(),
axis.ticks = element_blank(),rect = element_blank())+geom_circle(aes(x0=Longi
tude,y0=Latitude,r=r,color=Clusters),data = Clusters)+theme(axis.title.x = el
ement_blank(),axis.title.y = element_blank(),legend.position = c(-0.30,0.58),
legend.text = element_text(size = 7),legend.background = element_rect(color="
black",linetype = "solid"),legend.title = element_text(face = "bold"))


#Main plot of Ontario FSAs with proportion of immigrants and boxes in areas o
f interest

ONdens2<-ONdens+geom_rect(xmin =-78.5, xmax =-82, ymin = 42.5, ymax = 44.7,fi
ll = NA, color = "grey", size = 0.6)+geom_rect(xmin =-75.2, xmax =-76.5, ymin
= 44.8, ymax = 45.8,fill = NA, color = "grey", size = 0.6)


#This adds a layer with other two areas of interest of the map zoomed in

ONdens3<-ggdraw(ONdens2)+draw_plot({ONdens2 + coord_sf(xlim = c(-82,-78.5),yl
im =c(42.5,44.7),expand = FALSE) +theme(legend.position = "none")},x =0.29,y=
0.05,width = 0.2,height = 0.3)+draw_plot({ONdens2 + coord_sf(xlim = c(-76.5,-
75.2),ylim =c(44.8,45.8),expand = FALSE) +theme(legend.position = "none")},x
=0.66,y=0.33,width = 0.2,height = 0.3)


ONdens3
```

```
#Now lets create a similar map but with the low income prevalence in each FSA

OnLowInc<-ggplot(CaseConData)+geom_sf(aes(fill=LowIncome), color = alpha(c("w
hite"),0.1),lwd = 0)+scale_fill_gradientn(colors = alpha(c("blue4","blue","pu
rple","orange","Orange1","orangered","red"),0.70),name = "Prevalence",labels
= scales::percent)+coord_sf(expand = F)+theme(axis.text = element_blank(),axi
s.ticks = element_blank(),rect = element_blank())+geom_circle(aes(x0=Longitud
```

```
e,y0=Latitude,r=r,color=Clusters),data = Clusters)+theme(axis.title.x = eleme
nt_blank(),axis.title.y = element_blank(),legend.position = c(-0.31,0.595),le
gend.text = element_text(size = 7),legend.background = element_rect(color="bl
ack",linetype = "solid"),legend.title = element_text(face = "bold"))


#Main plot of Ontario FSAs with general cluster areas, general cluster areas
and boxes in areas of interest

OnLowInc2<-OnLowInc+geom_rect(xmin =-78.5, xmax =-82, ymin = 42.5, ymax = 44.
7,fill = NA, color = "grey", size = 0.6)+geom_rect(xmin =-75.2, xmax =-76.5,
ymin = 44.8, ymax = 45.8,fill = NA, color = "grey", size = 0.6)


#This adds a layer with other two areas of interest of the map zoomed in

OnLowInc3<-ggdraw(OnLowInc2)+draw_plot({OnLowInc2 + coord_sf(xlim = c(-82,-78
.5),ylim =c(42.5,44.7),expand = FALSE) +theme(legend.position = "none")},x =0
.29,y=0.05,width = 0.2,height = 0.3)+draw_plot({OnLowInc2 + coord_sf(xlim = c
(-76.5,-75.2),ylim =c(44.8,45.8),expand = FALSE) +theme(legend.position = "no
ne")},x =0.66,y=0.33,width = 0.2,height = 0.3)


OnLowInc3
```

```
#Now lets create a similar map but with the prevalence of post-secondary educ
ation in each FSA.

ONEdu<-ggplot(CaseConData)+geom_sf(aes(fill=EduProp), color = alpha(c("white"
),0.1),lwd = 0)+scale_fill_gradientn(colors = alpha(c("blue4","blue","purple"
,"orange","Orange1","orangered","red"),0.70),name = "Prevalence",labels = sca
les::percent)+coord_sf(expand = F)+theme(axis.text = element_blank(),axis.tic
ks = element_blank(),rect = element_blank())+geom_circle(aes(x0=Longitude,y0=
Latitude,r=r,color=Clusters),data = Clusters)+theme(axis.title.x = element_bl
ank(),axis.title.y = element_blank(),legend.position = c(-0.31,0.595),legend.
text = element_text(size = 7),legend.background = element_rect(color="black",
linetype = "solid"),legend.title = element_text(face = "bold"))


#Main plot of Ontario FSAs with general cluster areas, general cluster areas
and boxes in areas of interest

ONEdu2<-ONEdu+geom_rect(xmin =-78.5, xmax =-82, ymin = 42.5, ymax = 44.7,fill
= NA, color = "grey", size = 0.6)+geom_rect(xmin =-75.2, xmax =-76.5, ymin =
44.8, ymax = 45.8,fill = NA, color = "grey", size = 0.6)


#This adds a layer with other two areas of interest of the map zoomed in

ONEdu3<-ggdraw(ONEdu2)+draw_plot({ONEdu2 + coord_sf(xlim = c(-82,-78.5),ylim
=c(42.5,44.7),expand = FALSE) +theme(legend.position = "none")},x =0.29,y=0.0
5,width = 0.2,height = 0.3)+draw_plot({ONEdu2 + coord_sf(xlim = c(-76.5,-75.2
),ylim =c(44.8,45.8),expand = FALSE) +theme(legend.position = "none")},x =0.6
6,y=0.33,width = 0.2,height = 0.3)
```

```
ONEdu3
```

Appendix B – Letter of Ethics Approval

10/08/2020

Université d'Ottawa
Bureau d'éthique et d'intégrité de la recherche

University of Ottawa
Office of Research Ethics and Integrity

CERTIFICAT D'APPROBATION ÉTHIQUE | CERTIFICATE OF ETHICS APPROVAL

Numéro du dossier / Ethics File Number	H-08-20-6056
Titre du projet / Project Title	The Association Between Measles Cases and Migration/Settlement Patterns in Ontario
Type de projet / Project Type	Thèse de maîtrise / Master's thesis
Statut du projet / Project Status	Approuvé / Approved
Date d'approbation (jj/mm/aaaa) / Approval Date (dd/mm/yyyy)	10/08/2020
Date d'expiration (jj/mm/aaaa) / Expiry Date (dd/mm/yyyy)	09/08/2021

Équipe de recherche / Research Team

Chercheur / Researcher	Affiliation	Role
Marcel MIRON-CELIS	Département d'épidémiologie et santé publique / Department of Epidemiology and Public Health	Chercheur Principal / Principal Investigator
Robert SMITH	Département de mathématiques et statistiques / Department of Mathematics and Statistics	Superviseur / Supervisor

Conditions spéciales ou commentaires / Special conditions or comments

550, rue Cumberland, pièce 154 Ottawa (Ontario) K1N 6N5 Canada
550 Cumberland Street, Room 154 Ottawa, Ontario K1N 6N5 Canada

613-562-5387 • 613-562-5338 • ethique@uOttawa.ca / ethics@uOttawa.ca
www.recherche.uottawa.ca/deontologie | www.recherche.uottawa.ca/ethics

10/08/2020

Université d'Ottawa

Bureau d'éthique et d'intégrité de la recherche

University of Ottawa

Office of Research Ethics and Integrity

Le Comité d'éthique de la recherche (CÉR) de l'Université d'Ottawa, opérant conformément à l'*Énoncé de politique des Trois conseils* (2014) et toutes autres lois et tous règlements applicables, a examiné et approuvé la demande d'éthique du projet de recherche ci-nommé.

L'approbation est valide pour la durée indiquée plus haut et est sujette aux conditions énumérées dans la section intitulée "Conditions Spéciales ou Commentaires". Le formulaire « Renouvellement ou Fermeture de Projet » doit être complété quatre semaines avant la date d'échéance indiquée ci-haut afin de demander un renouvellement de cette approbation éthique ou afin de fermer le dossier.

Toutes modifications apportées au projet doivent être approuvées par le CÉR avant leur mise en place, sauf si le participant doit être retiré en raison d'un danger immédiat ou s'il s'agit d'un changement ayant trait à des éléments administratifs ou logistiques du projet. Les chercheurs doivent aviser le CÉR dans les plus brefs délais de tout changement pouvant augmenter le niveau de risque aux participants ou pouvant affecter considérablement le déroulement du projet, rapporter tout évènement imprévu ou indésirable et soumettre toute nouvelle information pouvant nuire à la conduite du projet ou à la sécurité des participants.

The University of Ottawa Research Ethics Board, which operates in accordance with the *Tri-Council Policy Statement* (2014) and other applicable laws and regulations, has examined and approved the ethics application for the above-named research project.

Ethics approval is valid for the period indicated above and is subject to the conditions listed in the section entitled "Special Conditions or Comments". The "Renewal/Project Closure" form must be completedfour weeks before the above-referenced expiry date to request a renewal of this ethics approval or closure of the file.

Any changes made to the project must be approved by the REB before being implemented, except when necessary to remove participants from immediate endangerment or when the modification(s) only pertain to administrative or logistical components of the project. Investigators must also promptly alert the REB of any changes that increase the risk to participant(s), any changes that considerably affect the conduct of the project, all unanticipated and harmful events that occur, and new information that may negatively affect the conduct of the project or the safety of the participant(s).

Kim THOMPSON
Responsable d'éthique en recherche / Protocol Officer
Pour/For **Daniel LAGAREC** Président(e) du/ Chair of the **Comité d'éthique de la recherche en sciences de la santé et sciences / Health Sciences and Sciences Research Ethics Board**

550, rue Cumberland, pièce 154
Ottawa (Ontario) K1N 6N5 Canada

550 Cumberland Street, Room 154
Ottawa, Ontario K1N 6N5 Canada

613-562-5387 • 613-562-5338 • ethique@uOttawa.ca / ethics@uOttawa.ca
www.recherche.uottawa.ca/deontologie | www.recherche.uottawa.ca/ethics