The Epidemiology of Human *Salmonella* Enteritidis Infections in Ontario, Canada, 2007-2009

by

Csaba Varga

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Doctor of Philosophy
in
Population Medicine

Guelph, Ontario, Canada

© Csaba Varga, December, 2014
In this thesis, 1,932 *Salmonella* Enteritidis infections reported to Ontario’s surveillance system from 2007 to 2009 were evaluated. Annual age-and sex-adjusted incidence rates were calculated across Ontario’s public health units. Poisson regression was used to estimate incidence rate ratios of cases among years, seasons, age groups, and sexes. A scan statistic was used to identify spatial clusters of high infection rates within the Greater Toronto Area. Negative binomial regression was used to identify area-level associations between infection rates and socioeconomic status indicators. In Toronto, global and local clustering of infections were evaluated. Across Ontario’s health regions, cases with major phage types were classified by their incidence rates, clinical symptoms, and exposure settings. A scan statistic was applied to detect phage type-specific spatial, temporal, and space-time clusters. The annual age-and-sex-adjusted incidence rates per 100,000 person-years were 4.4 in 2007, and 5.2 in both 2008 and 2009. Higher incidence of infections was observed in the years 2008 and 2009, in the spring, and in children. Within three public health units of the Greater Toronto Area, a high rate spatial cluster was identified in downtown Toronto. Areas with high average number of children at home per family, areas with high and areas with low average median family income, and
areas with a medium proportion of visible minority population had the highest infection rates. In Toronto, the global cluster analysis showed significant maximum spatial clustering of high infection rates at 3.3 km. The local cluster analysis identified distinct areas with high infection rates, mainly in downtown Toronto. The major phage types in Ontario were 8, 13a, 13, 1, 5b, and 4. Diarrhea, abdominal pain, and fever were the most common symptoms. International travel and unknown exposure settings were the most frequently reported settings for PT 5b, 4, and 1 cases. Unknown, private home, food premise, and international travel were the most frequently reported exposure settings for PT 8, 13, and 13a cases. A number of PT-specific spatial, temporal and space-time clusters were identified. The study results will aid public health authorities to design effective disease prevention and control programs, and early detection systems.
ACKNOWLEDGEMENTS

I would like to express my deep appreciation and sincere gratitude to my advisor, Dr. Michele Guerin, for her friendship, mentorship, and guidance over the course of my graduate studies. I truly value her support that has assisted my development as an epidemiologist. I would also like to thank the members of my graduate committee, Dr. David Pearl, Dr. Scott McEwen, Dr. Jan Sargeant, and Dr. Frank Pollari, for their knowledge and advice throughout my program. I will always be thankful for your support and assistance. Thank you to Nina Arron, Director, Public Health Protection & Prevention Branch who approved the data sharing agreement that made this thesis possible. I also want to thank the staff of all public health units and laboratories that tested samples, followed up with cases, and entered information into the public health surveillance database that I was privileged to access and examine. I also want to acknowledge the Data Resource Centre at the University of Guelph library for their GIS support. The faculty and staff of the Department of Population Medicine at the Ontario Veterinary College provided a world-class graduate program that offered networking and superior learning opportunities. The graduate students from the Population Medicine Department generated an inspiring environment that helped me to grow professionally. Finally, I would like to extend my most genuine thanks to my wife Emese, sons Aron and Adam, my sister Csilla, my parents Ferenc and Edit for their never-ending love, encouragement, and support. This would not have been possible without them.
STATEMENT OF WORK

Dr. Varga wrote the proposal and acquired the passive surveillance data from the Ontario Ministry of Health and Long-Term Care. All data cleaning, merging, and assessment were performed by Dr. Csaba Varga under the supervision of Dr. Guerin. Dr. Varga developed the study design, analyzed the data, interpreted the results, wrote the first draft, and prepared the final manuscripts of each chapter. Dr. Guerin and Dr. Pearl were consulted on data analysis, study design, interpretation of results, and reviewed and commented on manuscript drafts. Drs. McEwen, Sargeant and Pollari provided advice on the data analysis, interpretation of results, and reviewed and commented on manuscript drafts. Editing assistance for the thesis in its entirety was received from Drs. Guerin, Pearl, McEwen, Sargeant, and Pollari. Staff at the Data Resource Centre, University of Guelph provided GIS and census support.
TABLE OF CONTENTS

ABSTRACT ......................................................................................................................... i

ACKNOWLEDGEMENTS ................................................................................................. iv

STATEMENT OF WORK ................................................................................................. v

TABLE OF CONTENTS ................................................................................................. vi

LIST OF TABLES ........................................................................................................ xxi

LIST OF FIGURES ......................................................................................................... xiii

CHAPTER ONE
Introduction, literature review, study rationale, and research objectives

INTRODUCTION .............................................................................................................. 1

LITERATURE REVIEW ................................................................................................. 4

1. Microbiology and classification of Salmonella species ................................................. 4

2. Characterization of S. Enteritidis isolates ..................................................................... 5

3. Worldwide occurrence of S. Enteritidis infections ....................................................... 6

4. Infection sources and exposure settings ...................................................................... 8

5. Risk factors for infection ............................................................................................ 10

  5.1. Individual-level risk factors ................................................................................ 10

  5.2. Population-level risk factors ................................................................................ 12

6. Seasonal characteristics of Salmonella infections ...................................................... 14

7. Clinical symptoms and severity of infections ............................................................ 15

8. Spatial analytical methods used in infectious disease research .................................. 17

  8.1. Exploratory spatial analysis ................................................................................ 18

  8.1.2. Standardization of disease rates ...................................................................... 19

  8.1.3. Smoothing of disease rates .......................................................................... 20

  8.2. Spatial statistics of area-level data ...................................................................... 20

  8.2.1. Evaluating global disease clustering ............................................................ 20

  8.2.2. Evaluating local disease clustering ............................................................... 21
CHAPTER TWO

Incidence, distribution, seasonality, and demographic risk factors of *Salmonella* Enteritidis human infections in Ontario, Canada, 2007-2009

Abstract ...................................................................................................................... 49

Background .............................................................................................................. 50

Methods ................................................................................................................... 52
  Data sources ........................................................................................................... 52
  Statistical methods ............................................................................................... 53

Results ....................................................................................................................... 56
  Descriptive statistics and direct standardized incidence rates ......................... 56
  Poisson regression ............................................................................................... 57

Discussion ............................................................................................................... 58

Conclusions .............................................................................................................. 61

References .............................................................................................................. 63

CHAPTER THREE


Abstract ...................................................................................................................... 78

Background .............................................................................................................. 80

Methods ................................................................................................................... 82
  Study population and study design .................................................................. 82
CHAPTER FOUR

Area-level global and local clustering of human *Salmonella* Enteritidis infection rates in the city of Toronto, Canada, 2007–2009

Abstract ....................................................................................................................... 125

Background .................................................................................................................. 127

Methods ......................................................................................................................... 128
  Study design and data sources .................................................................................. 128
  Statistical analysis ...................................................................................................... 129
  Exploratory spatial analysis ....................................................................................... 129
  Spatial statistics .......................................................................................................... 130

Results .......................................................................................................................... 133
  Descriptive statistics .................................................................................................. 133
  Exploratory spatial analysis ....................................................................................... 134
  Non-smoothed and smoothed standardized incidence rates .................................... 134
Spatial statistics .......................................................................................................................... 134
Global clustering (Getis-Ord General G) .................................................................................. 134
Local clustering .......................................................................................................................... 134

Discussion .................................................................................................................................. 135

Conclusions .................................................................................................................................. 140

References .................................................................................................................................... 142

CHAPTER FIVE

Spatial-temporal epidemiology of human *Salmonella* Enteritidis infections with major phage types (PTs 1, 4, 5b, 8, 13, and 13a) in Ontario, Canada, 2008-2009

Abstract ...................................................................................................................................... 159

Background ................................................................................................................................... 160

Methods ...................................................................................................................................... 163
  Study setting and data sources .................................................................................................. 163
  Statistical analysis .................................................................................................................... 164
  Data management ..................................................................................................................... 164
  Phage type-specific incidence rates .......................................................................................... 165
  Scan statistic ............................................................................................................................... 165

Results ........................................................................................................................................ 167
  Health region-level incidence rates ......................................................................................... 167
  Monthly raw and smoothed incidence rates ............................................................................ 168
  Clinical symptoms ................................................................................................................... 168
  Exposure settings ..................................................................................................................... 169
  Scan statistics ............................................................................................................................ 170
  Purely spatial clusters of *S*. Enteritidis cases .......................................................................... 170
  Purely temporal clusters of *S*. Enteritidis cases ..................................................................... 170
  Space-time clusters of *S*. Enteritidis cases ............................................................................ 170
  Space-time cluster cases’ exposure settings ............................................................................ 171

Discussion .................................................................................................................................... 172

Conclusions .................................................................................................................................. 178

References .................................................................................................................................... 181
CHAPTER SIX

Conclusions and summary discussion ................................................................. 201
Advantages and limitations of study design ...................................................... 208
Future research and recommendations ............................................................ 209
Final conclusions ............................................................................................... 212
References .......................................................................................................... 213

APPENDIX A

Legend 2.1. Example of data structure of the Poisson regression model used to evaluate associations between \( S. \) Enteritidis infections and demographic and seasonal factors... 220

APPENDIX B

Legend 2.2 for Figure 2.1. Ontario Public Health Units labels and names. .......... 221

APPENDIX C

Legend 4.1 for Table 4.1 and Figures 4.6 and 4.7. Toronto forward sortation area labels. ........................................................................................................... 222

APPENDIX D

Legend 5.1 for Figure 5.1. Ontario Public Health Unit labels, names, and population estimates ........................................................................................................... 223
LIST OF TABLES

Chapter 1

Table 1.1. Examples of the most frequent human *Salmonella* serotypes classified by surface antigens and relevant clinical syndromes (Based on data from Murray, et al., 1995; Bale, et al., 2007; and Grimont and Weill, 2007). .............................................. .... 28

Table 1.2. Worldwide distribution of *Salmonella* Enteritidis human infections. .............. 29

Chapter 2

Table 2.1. Direct standardized incidence rates of *Salmonella* Enteritidis infections in Ontario, 2007-2009 (n = 1,932 cases). ........................................................................................................ 73

Table 2.2. Risk factors for *Salmonella* Enteritidis infections in humans, Ontario, Canada, 2007-2009 (n = 1,932 cases). ........................................................................................................ 74

Chapter 3

Table 3.1. Spatial cluster of high *Salmonella* Enteritidis infection rates in the Greater Toronto Area, Ontario, Canada ............................................................ 112

Table 3.2. Results of univariable negative binomial regression models (n = 846 cases from 153 FSAs). ........................................................................................................ 113

Table 3.3. Results of the final multivariable negative binomial regression model (n = 846 cases from 153 FSAs). .................................................................................. 115

Chapter 4

Table 4.1. Forward sortation areas identified by different local cluster detection methods ........................................................................................................... 151

Chapter 5

Table 5.1. Frequency of *Salmonella* Enteritidis cases with different phage types in Ontario, Canada, 2008-2009 (n=1,336). ........................................................................ 189

Table 5.2. Clinical symptoms of *Salmonella* Enteritidis cases with major phage types in Ontario, Canada, 2008-2009 (N=1,123). .................................................................. 190

Table 5.3. Exposure settings of *Salmonella* Enteritidis cases with major phage types in Ontario, Canada, 2008-2009. ........................................................................ 191

Table 5.4. Clusters of *Salmonella* Enteritidis cases with the six most frequent phage types in Ontario, Canada, 2008-2009. ........................................................................ 192
Table 5.5. Exposure settings of the *Salmonella* Enteritidis cases included in the space-time clusters for the six most frequent phage types in Ontario, Canada, 2008-2009. .... 194
LIST OF FIGURES

Chapter 2

Figure 2.1. Mean age-and-sex-adjusted incidence rates of *Salmonella* Enteritidis infections, across Ontario public health units, 2007-2009 .................................................. 75

Figure 2.2. Seasonal distribution of *Salmonella* Enteritidis (S. Enteritidis) cases in Ontario, 2007-2009 (n = 1,932) .................................................................................. 76

Figure 2.3. Number of *Salmonella* Enteritidis (S. Enteritidis) cases by month in Ontario, 2007-2009 (n = 1,932) .................................................................................. 77

Chapter 3

Figure 3.1. Location of the study area within the Greater Toronto Area in Ontario, Canada ....................................................................................................................... 116

Figure 3.2. Unadjusted *Salmonella* Enteritidis incidence per 100,000 person-years, by forward sortation area (FSA) ................................................................. 117

Figure 3.3. Relative risk of *Salmonella* Enteritidis infections by forward sortation area. .................................................................................................................. 118

Figure 3.4. Spatial cluster of high *Salmonella* Enteritidis infection rates. .......... 119

Figure 3.5. Association between average number of children at home per census family and *Salmonella* Enteritidis infections. .......................................................... 120

Figure 3.6. Distribution of the average number of children at home per census family by forward sortation area (FSA). ................................................................. 121

Figure 3.7. Distribution of the average median family income by forward sortation area (FSA). ............................................................................................................. 122

Figure 3.8. Distribution of the visible minority population proportion by forward sortation area (FSA). .......................................................................................... 123

Figure 3.9. Distribution of forward sortation area (FSA)-level Anscombe residuals from the multivariable negative binomial regression model. .................................................. 124

Chapter 4

Figure 4.1. Map of Ontario, Canada highlighting the location of the study area. .... 152

Figure 4.2. Flow chart outlining the analytical steps used to evaluate area-level *Salmonella* Enteritidis infection rates. ................................................................. 153
Figure 4.3. Distribution of non-smoothed and smoothed *Salmonella* Enteritidis infection rates in Toronto, 2007-2009 (n = 473 cases; n = 95 forward sortation areas). .......................... 154

Figure 4.4. Global clusters of areas with high *Salmonella* Enteritidis infection rates in Toronto at different distances. ........................................................................ 155

Figure 4.5. Maximum spatial clustering of areas with high *Salmonella* Enteritidis infection rates in Toronto at 3.3 kilometers. ........................................................................ 156

Figure 4.6. Local clusters of *Salmonella* Enteritidis infection rates in Toronto identified by the Getis-Ord Gi* statistic. ........................................................................ 157

Figure 4.7. Local clusters of *Salmonella* Enteritidis infection rates in Toronto identified by the Moran’s I statistic................................................................. 158

**Chapter 5**

Figure 5.1. Health regions in Ontario, Canada. ................................................................. 194

Figure 5.2. Flow chart outlining the analytical steps used to evaluate *Salmonella* Enteritidis cases with major phage types................................................................. 195

Figure 5.3. Health region-level raw incidence rates of *Salmonella* Enteritidis cases with major phage types in Ontario, Canada, 2008-2009.............................................. 196

Figure 5.4. Monthly raw and smoothed incidence rates of *Salmonella* Enteritidis cases with major phage types in Ontario................................................................. 197

Figure 5.5. Spatial clusters of *Salmonella* Enteritidis cases with major phage types cases in Ontario, Canada, 2008-2009................................................................. 198

Figure 5.6. Space-time clusters of *Salmonella* Enteritidis cases with major phage types in Ontario, Canada, 2008-2009................................................................. 199
CHAPTER ONE

Introduction, literature review, study rationale, and research objectives

INTRODUCTION

In Ontario and Canada salmonellosis in humans is the second most common cause of bacterial gastroenteritis (Vrbova, et al., 2012; Thomas, et al., 2013), and the top foodborne bacterial infection causing hospitalization and death (Ruzante, et al., 2011). The majority of human infections are foodborne, however, person to person, animal to person, and waterborne infections have also been reported (Vrbova, et al., 2012; Thomas, et al., 2013). Children, seniors, and immuno-compromised populations are at highest risk of infection (Keegan, et al., 2009; Younus, et al., 2009). The most common symptom of salmonellosis is diarrhea, which is usually mild, self-limited, and rarely requires treatment (Heymann, 2009). In more severe cases, the illness can last for several months, can cause sequelae (e.g., chronic arthritis), and can result in hospitalization and death (Heymann, 2009).

During the last decade in Ontario, and Canada, surveillance systems have detected an increase in the number of Salmonella enterica serotype Enteritidis (S. Enteritidis) infections, such that S. Enteritidis became the highest reported serotype among non-typhoidal Salmonellae (NTS) (Nesbitt, et al., 2012). Despite the large number of case-control studies conducted in Canada and worldwide that analyzed individual-level risk factors of S. Enteritidis cases, few studies have evaluated area-level associations and clustering of cases. Population-based ecological studies are useful in identifying areas with high infection rates where future studies can be conducted, and future disease
prevention and control programs can be targeted (Arcury, et al., 2005; Green, et al., 2006). Moreover, population-based ecological studies are effective in assessing area-level associations between infection rates and various sociodemographic and socioeconomic status indicators (Younus, et al., 2007; Chang, et al., 2009).

**Salmonella** Enteritidis infections are common in Canada; therefore, differentiating between sporadically occurring infections and outbreak cases is difficult. Applying subtyping methods to **S. Enteritidis** infections is essential to make it possible to differentiate more effectively different strains (Kafatos, et al., 2009).

In this thesis, the incidence, seasonality, and demographic risk factors of **S. Enteritidis** human infections across Ontario’s public health units (PHUs) from 2007 to 2009 were described. The forward sortation area (FSA; first three digits of the postal code)-level clustering of **S. Enteritidis** infection rates in three PHUs (City of Toronto Health Unit, Peel Regional Health Unit, and York Regional Health Unit) within the greater Toronto area was evaluated, and underlying area-level associations between **S. Enteritidis** infection rates and socioeconomic status (SES) indicators were evaluated. In the city of Toronto, the global and local spatial clustering of **S. Enteritidis** infection rates were assessed, and the effectiveness of various local cluster detection methods in identifying high rate infection clusters were evaluated. Across Ontario’s health regions the health region-level incidence, monthly incidence, exposure settings, clinical symptoms, and spatial-temporal clustering of **S. Enteritidis** cases with the six most common phage types that predominate in Ontario were evaluated.
The results of this thesis will provide information on the incidence, demographic risk factors, area-level socioeconomic determinants, distribution, seasonality, phage type distribution, exposure settings, clinical symptoms, and clustering of human *S. Enteritidis* infections in Ontario that will support public health authorities and policy makers in designing effective disease detection, control and prevention programs. This literature review briefly summarizes published research pertaining to the epidemiology of *S. Enteritidis* infections. The main objectives of this review are to:

1) Describe the microbiology, and classification of *Salmonella* species.
2) Describe the subtypes of *S. Enteritidis* isolates (phage typing).
3) Briefly summarize published research pertaining to the worldwide occurrence, infection sources, and risk factors of *S. Enteritidis* cases.
4) Describe the seasonal characteristics of *Salmonella* cases.
5) Describe clinical symptoms and severity of illness.
6) Present the main spatial analytical methods used in infectious disease research.
LITERATURE REVIEW

1. Microbiology and classification of *Salmonella* species

The genus *Salmonella* belongs to the family of *Enterobacteriaceae*. *Salmonella* are separated into two species, *Salmonella enterica* and *Salmonella bongori*. *Salmonella enterica* is further classified into six subspecies: *S. enterica* subsp. *enterica* (I); *S. enterica* subsp. *salamae* (II); *S. enterica* subsp. *arizonae* (IIIa); *S. enterica* subsp. *diarizonae* (IIIb); *S. enterica* subsp. *houtenae* (IV); and *S. enterica* subsp. *indica* (VI) (Grimont and Weill, 2007). *Salmonellae* are motile, Gram negative, facultative anaerobic bacilli, which rarely ferment lactose (Nataro, et al., 2011). The genus *Salmonella*, according to the Kauffman and White classification scheme are classified into serotypes, based on their surface antigens (Bale, et al., 2007). First the "O" (somatic) antigen type is determined based on a polysaccharide associated with a lipopolysaccharide. Then the "H" (flagellar) antigen is determined based on flagellar proteins. *Salmonella* exhibit phase variation between motile and non-motile phenotypes (or specific and non-specific phases), and different "H" antigens may be expressed. Non-motile isolates can be induced to switch to the motile phase using a Cragie tube experiment (Murray, et al., 1995; Bale, et al., 2007). Agglutination reactions based on the “O” antigen are used to divide *Salmonella* into serogroups, which include A, B, C1, C2, D, E1, E2, E3, E4, F, G, H, I and others (Murray, et al., 1995; Grimont and Weill, 2007). Examples of the most frequent human *Salmonella* serotypes classified by surface antigens and relevant clinical syndromes are presented in Table 1.1.
Salmonella enterica subspecies enterica contains almost all the serotypes pathogenic for humans (Nataro, et al., 2011). Salmonella serotypes can be categorized as typhoidal or non-typhoidal Salmonella (NTS) based on the clinical syndromes of which they are mainly associated (Grimont and Weill, 2007). The typhoidal Salmonella include the S. enterica subspecies enterica serotypes Typhi and Paratyphi, which generally cause systemic illness (enteric fever) with little or no diarrhea, while NTS, a much larger serotype group, mainly cause acute, self-limiting gastroenteritis (Grimont and Weill, 2007; Haeusler and Curtis, 2013). In Ontario, diagnosis of Salmonella spp. are done through the isolation of Salmonella organisms from stool, rectal swabs, urine, blood or any other sterile site (MOHLTC, 2013). Non-typhoidal Salmonella, particularly S. enterica subspecies enterica serotype Enteritidis (S. Enteritidis) are the primary focus of this review.

2. Characterization of S. Enteritidis isolates

One of the most common subtyping methods of S. Enteritidis is the phage-typing scheme that provides enhanced strain discrimination, and is considered a valuable tool for epidemiological studies (Kafatos, et al., 2009). A phage-typing scheme for S. Enteritidis was first described by Ward and colleagues (Ward, et al., 1987). This method currently differentiates 27 phage types using 10 typing phages, and has a high discriminative power by assigning 97% of the strains to one of the 27 distinct phage type (Ward, et al., 1987). In Ontario, phage typing is done in reference laboratories, and it is a rapid and economical method of strain differentiation. Subtyping is particularly valuable for common Salmonella serotypes like Enteritidis because a large number of infections occur sporadically throughout the year (Kafatos, et al., 2009). Strain differentiation helps
investigators to discriminate between endemic and outbreak strains. Differences in reservoirs and exposure settings within *S.* Enteritidis infections might also occur, and further differentiation could help investigators to identify phage type-specific risk factors (Parmley, et al., 2013; Varga, et al., 2013). Moreover, there might be phage type-specific differences in spatial, temporal and spatio-temporal characteristics of different *S.* Enteritidis infections (Parmley, et al., 2013).

**3. Worldwide occurrence of *S.* Enteritidis infections**

Non-typhoidal salmonellosis has an important health impact worldwide, causing globally an estimated 93.8 million human infections and 155,000 deaths each year (Majowicz, et al., 2013). In the United States of America (US), NTS cause an estimated 1 million domestically acquired foodborne illnesses and 350 deaths annually (Scallan, et al., 2011). In Canada, an estimated 87,510 domestically acquired foodborne illnesses occur annually, equaling an estimated 269.26 cases per 100,000 persons (Thomas, et al., 2013). In Canada (PHAC, 2009), the United States of America (US) (CDC, 2011), and the European Union (EU) (EFSA, 2007; EFSA, 2010) NTS is the second most commonly reported foodborne bacterial illness. In Canada (Ruzante, et al., 2011) and the US (Barton Behravesh, et al., 2011), salmonellosis is the top foodborne bacterial infection that causes hospitalization and death. Data from the World Health Organization (WHO) Global Foodborne Infections Network indicated that *S.* Typhimurium and *S.* Enteritidis accounted for nearly 80% of all NTS human isolates reported globally (Vieira, et al., 2009). According to the WHO Global Salm-Surv program from 2000 to 2002, *S.* Enteritidis was the most commonly reported *Salmonella* serotype globally, accounting for 65% of all human isolates that included typhoidal and non-typhoidal *Salmonella*.
serotypes (Galanis, et al., 2006). Regional differences in the reported proportion of S. Enteritidis isolates have been observed, with the highest proportion identified in Europe (85% of cases) followed by Asia (38%), and then Latin America and the Caribbean (31%) (Galanis, et al., 2006). Recently conducted studies in Canada (Nesbitt, et al., 2012), US (Chai, et al., 2012), Israel (Bassal, et al., 2012), and EU (EFSA, 2010) have shown that S. Enteritidis was the most commonly reported serotype among human NTS isolates. The proportion of S. Enteritidis of all NTS serotypes ranged from 25.1 to 52.3%. In New Zealand, S. Enteritidis was the 2\textsuperscript{nd} most common NTS serotype, consisting of 10% of all NTS serotypes (Lal, et al., 2012). The distribution of S. Enteritidis infections in industrialized countries is shown in Table 1.2.

Differences in the most commonly reported human S. Enteritidis PTs have been described worldwide. In The Netherlands from 1997 to 2001, S. Enteritidis infections with PT 4 were most commonly reported, followed by PTs 21, 28, 1, and 6 (van Duijkeren, et al., 2002). In Spain from 1997 to 2001, S. Enteritidis infections with PTs 1 (18%), 4 (15%), and 6a (5%) were the most commonly reported (Echeita, et al., 2005). In the EU, an international multi-center study conducted from 2001 to 2004 identified S. Enteritidis infections with PT 4 as the most commonly reported, followed by PTs 1, 8, 21, and 14b (Peters, et al., 2007). In Australia in 2009, the majority of S. Enteritidis infections were travel related, and the most common PT cases were 6a, 1b, 21, and 1 (OzFoodNet, 2010). In Canada in 2010, S. Enteritidis infections with PTs 8, 13a, 13, 1, 4 and 5b were the most frequently reported (PHAC, 2010).
4. Infection sources and exposure settings

*Salmonella* bacteria are widespread in the environment, and can colonize the intestinal tracts of humans and various animals including poultry, livestock, frogs, turtles, wildlife, and domestic pets (Heymann, 2008). Humans generally are infected with *Salmonella* through the fecal-oral route from several sources, including consumption of contaminated food or water, contact with infected animals, and person to person transmission (Heymann, 2008). Consumption of contaminated food is the major infection source of salmonellosis, with an estimated proportion of 73.1-85% of cases reported to be foodborne (Majowicz, et al., 2010; Vrbova, et al., 2012). A recently published Ontario study described that among sporadic domestically acquired reportable salmonellosis cases the most common sources of infections were food (73.1%), contact with animals (15.2%), contact with persons ill with similar symptoms (9.9%), and water (1.3%) (Vrbova, et al., 2012).

*Salmonella* Enteritidis has an increased ability relative to other *Salmonella* serotypes to colonize and persist in the intestinal and extra-intestinal sites of chicken without causing any clinical sign or illness (Gantois, et al., 2008; Han, et al., 2013). The preferential extra-intestinal site of *S.* Enteritidis colonization is the reproductive organs of hens. This enables the pathogen to infect the internal contents of eggs, which facilitates vertical transmission of the bacteria (Gast, et al., 2007; Gantois, et al., 2008). Horizontal transmission of *S.* Enteritidis might also occur if the egg is contaminated by external factors such as feces, vectors (e.g., insects, rodents, humans), transportation equipment, food preparation utensils, or environment (Wray, et al., 1999; Bailey, et al., 2001). Compared to other *Salmonella* serotypes, *S.* Enteritidis survives in egg white more
efficiently (De Vylder, et al., 2013), and has an enhanced survival rate at 42 °C (the body temperature of the hen) (Gantois, et al., 2008), which allows this serotype to effectively contaminate the egg. In the EU, contaminated eggs are considered to be the most important source of human S. Enteritidis infections (Pires, et al., 2011). In the US, a major egg-associated outbreak occurred in 2010, where 1,939 known S. Enteritidis cases were identified, and over 500 million shell eggs were recalled (Kuehn, 2010). In Brazil, a study evaluated Salmonella prevalence in eggs obtained from commercial egg laying farms and from surplus broiler hatching eggs sold at farmers’ market and small shops. (Kottwitz, et al., 2013). Salmonella were not detected in samples obtained from commercial egg laying farms; however, 52.0% of samples obtained from surplus broiler hatching eggs were positive for Salmonella spp., of which the most common serovar was Enteritidis (84.6% of samples) (Kottwitz, et al., 2013). In the US, among 144 S. Enteritidis outbreaks, consumption of eggs (93 outbreaks) and chicken (19 outbreaks) were the most commonly reported food items that were associated with illness (Jackson, et al., 2013).

International travel was demonstrated to be a major exposure setting of S. Enteritidis infections in Ontario. Among all cases, 51.9% reported international travel during the illness incubation period; among them 88.9% reported travel to Caribbean and Mexico regions, and 90.1% reported staying on a resort (Tighe, et al., 2012). The major phage types reported from these cases were 1, 4, 5b, 7a (Tighe, et al., 2012).

Restaurant and private homes were demonstrated to be the most common exposure settings for endemic S. Enteritidis infections. Previous studies identified that infected but
asymptomatic food handlers contaminated food served at restaurants connected to outbreaks (Beatty, et al., 2009; Hedican, et al., 2009; Medus, et al., 2009). Exposures in private homes were demonstrated to be an important exposure setting for sporadic, home-based infections (Scott, et al., 2003; Wagner et al., 2014). Person-to-person (Ethelberg, et al., 2004), and animal-to-person (Aiken, et al., 2010; Leonard, et al., 2011) transmissions were identified as the main sources for infections that occur in private homes.

5. Risk factors for infection

5.1. Individual-level risk factors

A recent systematic review of case-control studies evaluated the source attribution of sporadic S. Enteritidis infections and found that travel, consumption of undercooked eggs, consumption of eggs, and consumption of chicken were the major risk factors for human illnesses (Domingues, et al., 2012). A case-control study conducted from January to August 2011 in Ontario identified consumption of any poultry meat, consumption of processed chicken, and not washing hands following the handling of raw eggs as being statistically significant risk factors for sporadic domestically acquired S. Enteritidis infections (Middleton, et al., 2012). Consumption of any poultry meat had a population attributable fraction of 46% of all S. Enteritidis infections (Middleton, et al., 2012). In a population-based case-control study, conducted during a 12-month period from 2002 to 2003 in five FoodNet Sites in the US, travel outside the US was the greatest risk factor for S. Enteritidis infections, based on the size of the odds ratio (Marcus, et al., 2007). Eating undercooked eggs inside the home and eating chicken outside the home in the 5 days before illness onset were also significantly associated with domestically acquired S. Enteritidis infections (Marcus, et al., 2007). In Minnesota, US, a case-control study
identified associations between S. Enteritidis infections and eating stuffed chicken products during 7 days prior to disease onset (Smith, et al., 2008). In a case-case study conducted in England between 2004 and 2007, recent reptile exposure was significantly associated with Salmonella infections (Aiken, et al, 2010). Children under five years of age were at increased risk of infection. Reptile exposure had a population attributable fraction of 0.9 %.

Eating undercooked eggs was demonstrated to be a major risk factor for human S. Enteritidis infections. In British Columbia, Canada an investigation conducted between 2007 and 2010 identified eggs as the most likely source of human illnesses (Taylor, et al., 2012). The majority of cases consumed ungraded eggs at home or in foodservice establishments (Taylor, et al., 2012). In Alberta, Canada an outbreak was detected between October 2010 and February 2011, involving 91 cases who purchased food from mobile lunch trucks (CDC, 2013). Food was most likely contaminated directly or indirectly by illegally obtained S. Enteritidis contaminated eggs (CDC, 2013). In Connecticut, US, a multiagency outbreak investigation team identified cases of S. Enteritidis in a long-term care facility with an indistinguishable pulsed-field gel electrophoresis pattern. The most likely cause of infections was consumption of undercooked shell eggs (Styles, et al., 2012).

A number of individual-level demographic and socioeconomic risk factors of salmonellosis have been identified. In Michigan, US, researchers analyzed health surveillance data between 1995 and 2001 using Poisson regression and found that children <1 year of age, and children 1-4 years of age were at highest risk for S.
Enteritidis infections compared to adults aged 15-39 years (Younus, et al., 2006). In the Waterloo region, Canada, researchers assessed reportable disease data for 13 enteric diseases between 1990 and 2004 using Poisson regression and found that children aged 0-4 years were at the highest risk for salmonellosis (Keegan, et al., 2009). A Danish study followed the entire population of Denmark from 1993 to 2004 and assessed associations between laboratory-confirmed S. Enteritidis infections and various socioeconomic risk factors (Simonsen, et al., 2008). High-income groups compared to low income groups had greater risks of S. Enteritidis infections. Married persons were at higher risk of contracting S. Enteritidis infections than unmarried persons without a partner. Immigrant individuals whose parents were born outside Denmark had lower risk of infections compared to persons born in Denmark (Simonsen, et al., 2008). Researchers concluded that differences in infection risks might be explained by cases’ variations in diet, travel activity, and health seeking behavior (Simonsen, et al., 2008). A Canadian study evaluated community-level risk factors in the Northwest Territories between 1991 and 2008 using negative binomial regression, and detected high Salmonella infection rates in communities with higher proportions of 'households in core need' (unsuitable, inadequate, and/or unaffordable housing) (Pardhan-Ali, et al., 2013).

5.2. Population-level risk factors

A limited number of population-based ecological studies assessed area-level associations between foodborne bacterial infections and socioeconomic and sociodemographic risk factors. A Spanish study examined associations among municipality-level incidence rates of foodborne disease outbreaks and demographic and socioeconomic characteristics in each municipality (Broner, et al., 2010). Regions with medium and high proportions of

12
residents aged ≥65 years, regions with a high proportion of women and unemployed had lower incident rates of outbreaks compared to regions where these risk factors were classified as low. Regions with a high proportion of people with higher education level had lower incidence of outbreaks compared to areas with a low proportion of highly educated residents (Broner, et al., 2010).

An US ecological study evaluated associations between area-level socioeconomic and demographic factors and salmonellosis rates (Younus, et al., 2007). These researchers demonstrated a higher salmonellosis rate in areas with a higher proportion of black, Hispanic or Latino populations, in areas with a high number of physician rate per 100,000 persons, in areas with a high proportion of community hospital beds per 100,000 persons, and in areas with high proportion of the population aged ≥65 years and <5 years (Younus, et al., 2007). On the other hand, a decreased salmonellosis rate was observed in areas with high percentage of population aged 45–64 years, in areas with high unemployment rate, in areas with a high proportion of population living in cities, and in areas with a high proportion of adults with less than a ninth-grade education (Younus, et al., 2007). High salmonellosis rates in areas with high proportion of young children and older adults were explained by higher severity of salmonellosis in these age groups; consequently increased likelihood of visiting a physician, and getting tested and reported. The low salmonellosis rate in areas with low education attainment and high unemployment rate were explained by these socioeconomic groups having decreased access to healthcare, lower health seeking behavior; consequently, their underrepresentation in the surveillance system (Younus, et al., 2007). Another US population-level ecological study demonstrated that areas with a high proportion of
residents with high education levels were more commonly represented among salmonellosis cases than areas with a high proportion of residents with lower education levels (< high school degree vs. > or = college degree) (Chang, et al., 2009). It was hypothesized that residents with higher education levels had higher incomes, which enabled them to eat outside of their homes more frequently, which predisposed them of contacting *Salmonella* infections (Chang, et al., 2009).

6. Seasonal characteristics of *Salmonella* infections

A recent systematic review of published literature between 1960 and 2010 examined seasonal patterns of five enteric zoonotic diseases and found that *Salmonella* infections had a distinct summer peak (Lal, et al., 2012). In Alberta, Canada, Guerin et al. (2005) analyzed *Salmonella* infections that were reported to Alberta’s passive surveillance system from January 1990 to January 2002. The researchers detected two short duration temporal *S.* Enteritidis infection clusters, indicating point source infection outbreaks; a primary cluster occurred in February and a secondary cluster occurred in April (Guerin, et al., 2005). In Gwangju area, Korea, Kim, et al. (2012) evaluated the seasonal patterns of different *Salmonella* serotypes that were reported to Korea’s surveillance system from 2000 to 2009. The largest numbers of *S.* Enteritidis infections were reported during the summer and fall months (Kim, et al., 2012). In North Dakota, US, Oloya, et al. (2007) assessed the occurrence of *Salmonella* infections in humans and domestic animals from 2000 to 2005 and found that humans had the largest number of *Salmonella* infections during the summer months. In Ontario, Canada, Tighe, et al. (2012) evaluated *S.* Enteritidis infections reported to Ontario’s surveillance system in 2010 and 2011 and established that a large number of cases were reported during the late winter and early
spring months. The majority of these cases during their disease incubation periods visited all-inclusive resorts in the Caribbean and Mexico regions (Tighe, et al., 2012).

7. Clinical symptoms and severity of infections

The incubation period of non-typhoidal salmonellosis ranges from half to three days, the illness usually lasts for 4 to 7 days, and most patients recover without treatment (Heymann, 2008; Jones, et al., 2008). Depending on the NTS serotype and the age and health status of the host, the infective dose could be as low as one cell (Jones, et al., 2008). Salmonellosis among healthy people with competent immune systems generally results in acute, self-limiting gastroenteritis, with diarrhea being the most common clinical symptom. Fever, abdominal cramping, nausea and vomiting may also occur (Heymann, 2008; Jones, et al., 2008). The most common complications of acute salmonellosis are fluid and electrolyte disturbances (Heymann, 2008; Jones, et al., 2008), but asymptomatic gastrointestinal infections can also occur (Jones, et al., 2008). In Ontario, the most commonly reported symptoms among individuals with S. Enteritidis infections were: diarrhea (98.7 %), abdominal cramps (87.6 %), fever (52 %), vomiting (39.5 %), and nausea (59.4 %), with 10% of cases requiring hospitalization (Tighe, et al., 2012).

In industrialized countries, an estimated 9% of acute NTS gastroenteritis cases develop bacteremia, an extra-intestinal infection defined as the presence of Salmonella bacteria in the blood stream (Haeusler, et al., 2013). Fever and/or septicemia usually accompany Salmonella bacteremia, which may result in focal infections at any site, including, meningitis, and bone and joint infections (Schutze, et al., 1997; Sauteur, et al., 2013). The
frequency of invasive NTS infections varies largely by serotype, antimicrobial susceptibility profile of the bacteria, region, age, exposure source, co-infection, and immune status of the host (Schutze, et al., 1997; Gordon, et al., 2008; Jones, et al., 2008; Sauteur, et al., 2013; Phoba, et al., 2014). Differences in the severity of infections among common NTS serotypes have been previously demonstrated (Jones, et al., 2008). The case fatality rate for S. Dublin (3.0%) and S. Typhimurium (0.6%) was significantly higher than S. Newport (0.3%), whereas the hospitalization rate for S. Enteritidis (7%), S. Heidelberg (13%), S. Choleraesuis (57%), and S. Dublin (64%) were higher than for S. Typhimurium (6%). In economically-constrained countries a high rate of invasive NTS infections with multidrug resistant organisms were observed, with a mortality rate of 11 - 24% (Blomberg, et al., 2005; Gordon, et al., 2008; Phoba, et al., 2014). In Central-Africa, an epidemic increase of invasive Salmonella bloodstream infections has been observed lately, with S. Enteritidis accounting for most of the increase (Phoba, et al., 2014). The majority of cases were children < 5 years of age; the case fatality rate was 11.1 %, while 69.7 % of cases had Plasmodium falciparum co-infection (Phoba, et al., 2014). The majority of Salmonella organisms isolated from the blood stream were multidrug resistant (Phoba, et al., 2014). A recent study evaluated the published research since 1965 on reptile-related salmonellosis, and found that 15 % of cases presented invasive infections, which were associated with cases’ young age (mostly <6 months of age (Sauteur, et al., 2013). In Canada, 13% of reported Salmonella spp. cases were hospitalized, and the percentage of cases hospitalized was greatest in those aged <5 and >59 years (Ruzante, et al., 2011).
Long-term sequelae to various organ systems after acute NTS infections are documented in the literature (Batz, et al., 2013). Reactive arthritis was the most commonly reported clinical syndrome. In Finland, among 496 *Salmonella*-positive cases 4.4% had reactive arthritis. Among NTS serotypes *S. Enteritidis* was the most common causative agent (Tuompo, et al., 2013). In the US, among 624 *Salmonella*-positive cases, 0.5% had reactive arthritis (Porter, et al., 2013).

*Salmonella* carriage, which is defined as asymptomatic excretion of the organism, might occur following acute NTS infection (Heymann, 2008). *Salmonella* carriage can be divided into convalescent and chronic carriage (Buchwald and Blaser, 1984). Convalescent carriage follows symptomatic or asymptomatic NTS infections. During convalescent carriage, the duration of *Salmonella* excretion is longer in young children, in cases with symptomatic infections, and in cases treated with antimicrobials (Buchwald and Blaser, 1984). Chronic carriage is described as the excretion of *Salmonella* for more than one year. Overall, chronic carriage occurs in less than 1% of cases; however it might occur in up to 2.6 % of children under 5 years of age (Buchwald and Blaser, 1984).

8. Spatial analytical methods used in infectious disease research

A number of spatial statistical methods combined with geographic information system (GIS) applications have been used effectively in infectious disease research studies. A literature review conducted in 2007 identified disease surveillance (n=227), risk analysis (n=189), health access and planning (n=138), and community health profiling (n=115) as the main uses of GIS applications in health promotion and public health (Nykiforuk and
Flaman, 2011). These studies highlighted how GIS-related techniques can be effectively used in public health policy, promotion, and practice research.

8.1. Exploratory spatial analysis

8.1.1. Crude disease counts and the application of disease mapping

Tract count analysis evaluates disease counts within a study period in a well-defined study area. These areas usually are arbitrary regions because they are demarcated for administrative purposes (e.g., census tracts, counties, municipalities, postal code areas, and public health unit regions) (Lawson, 2006). The researchers’ goal is to choose a geographical unit as small as possible to be able to illustrate and evaluate small area-level disease heterogeneity (Elliott and Savitz, 2008).

Choropleth maps are commonly used to illustrate area-level count data, and they can show how population counts vary across geographic areas (Slocum, et al., 2009). These maps are effective to illustrate regional changes in disease rates over time, and they can highlight disease trends that tabular representation of data might miss (Berke, 2001). Choropleth maps illustrate the spatial variation of disease counts and rates over a well demarcated area grouped into classes, such as the number of *S. Enteritidis* cases per postal code area. Each class represents a range of values and is illustrated by a sequence of colors (Slocum, et al., 2009). Classification of values could be done by quantiles, equal intervals, defined intervals, and Jenk’s natural breaks (Slocum, et al., 2009). It is advisable to use Jenk's optimization classification method to produce high quality choropleth maps. Jenk's optimization arranges data into classes based on their distribution by using an algorithm that reduces variance within groups and maximizes variance between groups (Jenks, 1967). Public health studies use choropleth maps frequently to
illustrate the spatial distribution of disease rates, to detect areas with increased disease rates where future studies can be conducted to identify factors causing the increase, and to generate hypotheses about what factors these areas have that are not found in other areas.

8.1.2. Standardization of disease rates

Crude incidence rate maps do not account for area-level differences in age and sex distribution that might impact area-level disease rates due to younger and older resident’s higher salmonellosis rates (Chui, et al., 2009; Younus, et al., 2010). To overcome this problem calculating age- and sex-adjusted disease rates have been proposed (Ahmad, et al., 2009). There are two main approaches to standardization: direct and indirect. The direct method uses the observed age-and-sex-specific rates for each group in the study population (e.g. females 0-4 years of age) and relates it to the standard population distribution with respect to age and sex (Kahn and Semos, 1989). The indirect method uses the age-and-sex-specific rates from the standard population and relates it to the observed age-and-sex-specific distribution of the study population. The latter method generates an “expected” rate, assuming the standard set of age-and-sex-specific infection rates were in effect in that population (AIHW, 2011). Area-level infection rates are extensively used in research studies, however, they have limitations. The disease rate is a ratio, and can be influenced by small changes in the observed number of cases. A large ratio will be obtained when several cases are detected in an area with small background population numbers (Lawson, 2006; Elliott and Savitz, 2008).
8.1.3. Smoothing of disease rates

Disease rate smoothing has been proposed to account for unstable disease rates (Clayton and Kaldor, 1987). One of the most frequently used smoothing methods in spatial analytical studies is the Spatial Empirical Bayes smoothing, which accounts for unstable disease rates by reducing the unbalanced rates toward the local mean if regional clustering was detected, and toward the global mean if no regional clustering was detected (Clayton and Bernardinelli, 1997; Waller and Gotway, 2004). The Empirical Bayes smoothing is another common method; however, it does not account for local clustering and shrinks each local disease rate toward the global population mean (Clayton and Kaldor, 1987). Smoothing techniques highlight the overall disease trends, and increase researchers’ ability to identify disease clusters.

8.2. Spatial statistics of area-level data

8.2.1. Evaluating global disease clustering

Maps are used to visualize and evaluate spatial patterns of disease distributions; however, maps cannot indicate the statistical significance of disease patterns. To overcome this problem various spatial statistical pattern analyses methods have been proposed to quantify, evaluate, and compare different non-random disease distributions (Lawson, 2006; Anselin, et al., 2007; Pfeiffer, et al., 2008). These spatial statistical methods rely on inferential statistics, which assumes that all area-level disease patterns are exhibiting spatial random distribution. When disease patterns are not random the null hypothesis can be rejected, and a statistically significant cluster is identified (Getis and Ord, 1992; Pfeiffer, et al., 2008). The two major spatial statistical methods that detect global disease clustering are the Getis-Ord General G and the Global Moran's I (Getis and Ord, 1992;
Mitchell, 2005). The Global Moran's I statistic evaluates whether disease rates are clustered, dispersed, or random. A statistically significant positive Moran's I index indicates disease clustering, whereas a significant negative Moran's I index indicates disease dispersion; however, this method does not specify whether low or high disease rates were clustered or dispersed in the study area (Getis and Ord, 1992; Mitchell, 2005). To overcome this problem the Getis-Ord General G method was proposed, which quantifies broad spatial patterns and trends by measuring the extent of clustering for either high (positive z-score) or low (negative z-score) disease rates across a study area (Getis and Ord, 1992). The intensity of disease clustering is indicated by the scale of the observed z-score. The Getis-Ord General G method has been used effectively to assess the global clustering for a variety of infectious and non-infectious diseases including: schistosome infection rates in Jiangxi Province, China (Tang, et al., 2012), pulmonary tuberculosis in Zhejinag Province, China (Ying and Chen, 2012), acute respiratory infections among indigenous peoples in Chile (Rojas, 2007), and worldwide distribution rates of multiple sclerosis (Wade, 2014).

8.2.2. Evaluating local disease clustering

Local cluster detection methods show “where the clusters or outliers are located” and “what type of spatial correlation is most important” (Anselin, et al., 2007). These methods are useful in identifying local areas with high, low or dissimilar infection rates where public health authorities can conduct future studies to identify individual-level risk factors, or to target prevention and control programs.
8.2.2.1. Local Moran’s I

Anselin Local Moran’s I (Anselin, et al., 2007) method is commonly used to detect local infection clusters. This technique detects areas with high infection rates [hot spots (high value areas surrounded by high value areas, HH)], low infection rates [“cold spots” (low value areas surrounded by low value areas, LL)], and outlier areas [(high value areas amongst low value areas, HL or vice versa, LH)]. Local Moran’s I statistic has been used successfully to identify areas with high rate Campylobacter infections in Denmark (Jepsen, et al., 2009), county-based high human West Nile virus incidence in the US (Sugumaran, et al., 2009), areas with high influenza-associated mortality among US elderly (Greene, et al., 2006), geographical distribution of human giardiasis in Ontario, Canada (Odoi, et al., 2003), and areas with high overweight and obesity rates in Canada (Pouliou and Elliott, 2009).

8.2.2.2. Getis-Ord Gi*

Researchers have used the Getis-Ord Gi* statistic extensively when there was a strong prior assumption that infections were clustered within a study area (Anselin et al., 2007). This method identifies local high (hot spot) or low (cold spot) areas by comparing the local mean rates to the global mean rates, and identifies a hot spot if the target area has a high infection rate and it is surrounded by high rate areas. A cold spot is identified when a low infection rate area is surrounded by low infection rate areas (Anselin and Getis, 1992). Getis-Ord Gi* statistics have been used previously to detect county-level hotspots of prostate cancer incidence in Georgia, US (Wagner, et al., 2013), to identify hot spots of colorectal cancer cases in Kuala Lumpur (Shah, et al., 2014), to detect district-level malaria hotspots in Bangladesh (Haque et al., 2014), to identify hot and cold spot areas of
kala-azar disease in Vaishali district, India (Bhunia et al., 2013), and to identify hot spots of human immunodeficiency virus infections among injection drug users in Tijuana, Mexico (Brouwer, et al., 2012).

8.2.2.3. Scan statistic

Scan statistical methods have been successfully used by various researchers worldwide to test whether a disease is randomly distributed over space, over time or over space and time (Kulldorff, 1997). Using scan statistics to identify statistically significant non-random distribution of infections has several advantages. This method allows scanning for disease clusters across the whole study area without pre-determining their size or location, which reduces pre-selection bias (Kulldorff, et al., 1998). Scan statistic also allows adjusting for confounding variables (e.g., age, sex) that might bias disease clustering. For the disease cluster identified by this method, a relative risk and a p-value obtained through Monte Carlo hypothesis testing are given, which allows for infection cluster comparison and characterization (Kulldorff, 1997; Kulldorff, et al., 1998).

Researchers have effectively applied the scan statistic to identify spatial, temporal or space-time \textit{S}. Enteritidis infection clusters. A recent study used space-time permutation models to detect disease clusters among cases reported to Thailand’s passive surveillance system between 2002 and 2007 (Domingues, et al., 2014). Investigators identified several \textit{S}. Enteritidis outbreaks. In Belgium, researchers used a temporal multivariate scan to simultaneously search for and evaluate \textit{S}. Enteritidis infection clusters of humans and layer birds (Welby, et al., 2011). Overlapping infection clusters of PT 24 of humans and layers birds were detected. The concomitant monthly occurrence of infection clusters in
humans and layer birds suggested that humans were most likely infected by consuming infected eggs. In Alberta, Canada, purely temporal scans using a Bernoulli model identified a number of time periods were identified with higher than expected S. Enteritidis infection rates (Guerin, et al., 2005).

Scan statistical models have been also applied effectively to evaluate a cholera outbreak in Papua New Guinea (Horwood, et al., 2014), to identify human Cryptosporidium disease outbreaks in England (Briggs, et al., 2014), to detect county-level spatial and temporal clusters of tuberculosis in China (Zhao, et al., 2013), to detect localized community outbreaks of Shigella spp. in Argentina (Viñas, et al., 2013), and to identify the location of high and low rate areas of campylobacteriosis in Manitoba, Canada (Green, et al., 2006).

**STUDY RATIONALE**

No previous study in Ontario or Canada has analyzed passive disease surveillance data to evaluate the epidemiology of S. Enteritidis infections. Despite the widespread use of spatial statistical methods in detecting infectious disease clusters few studies analyzed the spatial epidemiology of serotype- and subtype-specific Salmonella spp. infections. Public health authorities, to reduce the burden of S. Enteritidis infections in Ontario and Canada, need information on the spatial and temporal clustering, seasonality, demographic and socioeconomic determinants, subtype characteristics, clinical symptoms, and exposure settings of S. Enteritidis cases. This thesis will provide evidence-based information on the epidemiology of S. Enteritidis infections in Ontario that will assist public health
authorities in developing effective prevention and control programs, and early disease
detection systems.

The following sections describe the epidemiology of human *S. Enteritidis* infections in
Ontario; however the findings of this thesis might share similarities with other foodborne
bacterial infections in Canada and other industrialized countries. The study approach and
statistical analysis may be appropriate to examine other data from passive foodborne
disease surveillance systems.
RESEARCH OBJECTIVES

The principal objectives of this thesis are to:

1. Describe annual age-and-sex-adjusted incidence rates (IRs), and annual and mean age-adjusted sex-specific IRs of S. Enteritidis cases in Ontario (Chapter 2).
2. Describe the mean age-and-sex-adjusted IR for each public health unit (PHU) in Ontario (Chapter 2).
3. Identify associations between S. Enteritidis IRs and demographic and seasonal factors (Chapter 2).
4. Assess forward sortation area (FSA; first three digits of the postal code)-level clustering of S. Enteritidis infections within three PHUs within the Greater Toronto Area (Chapter 3).
5. Identify underlying area-level associations between standardized incidence rates (SIRs) of S. Enteritidis infections and socioeconomic status indicators in these three PHUs that might explain the clustering of S. Enteritidis infections (Chapter 3).
6. Evaluate FSA-level geographical variations of raw and smoothed SIRs of S. Enteritidis infections in the city of Toronto (Chapter 4).
7. Identify FSA-level global and local spatial clusters of high SIRs of S. Enteritidis infections in the city of Toronto (Chapter 4).
8. Estimate phage type-specific health region-level IRs (Chapter 5).
9. Estimate phage type-specific monthly IRs (Chapter 5).

10. Describe phage type-specific exposure settings and clinical symptoms (Chapter 5).

11. Detect phage-type-specific spatial, temporal, and space-time clusters of cases (Chapter 5).

12. Examine phage-type-specific space-time cluster cases’ exposure settings to identify common exposures (Chapter 5).
Table 1.1. Examples of the most frequent human *Salmonella* serotypes classified by surface antigens and relevant clinical syndromes (Based on data from Murray, et al., 1995; Bale, et al., 2007; and Grimont and Weil, 2007).a.

<table>
<thead>
<tr>
<th>&quot;O&quot;-group</th>
<th>Serotypes</th>
<th>&quot;O&quot; antigens</th>
<th>Phase 1 (motile) &quot;H&quot; antigens</th>
<th>Phase 2 (non-motile) &quot;H&quot; antigens</th>
<th>Clinical symptoms b</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td><em>S. Paratyphi</em> A</td>
<td>1,2,12</td>
<td>a</td>
<td>no phase 2 antigen</td>
<td>Enteric fever</td>
</tr>
<tr>
<td>B</td>
<td><em>S. Paratyphi</em> B</td>
<td>1,4,5,12</td>
<td>b</td>
<td>1,2</td>
<td>Enteric fever</td>
</tr>
<tr>
<td></td>
<td><em>S. Typhimurium</em></td>
<td>1,4,5,12</td>
<td>i</td>
<td>1,2</td>
<td>NTS</td>
</tr>
<tr>
<td></td>
<td><em>S. Heidelberg</em></td>
<td>1,4,5,12</td>
<td>r</td>
<td>1,2</td>
<td>NTS</td>
</tr>
<tr>
<td></td>
<td><em>S. Saint-paul</em></td>
<td>1,4,5,12</td>
<td>e,h</td>
<td>1,2</td>
<td>NTS</td>
</tr>
<tr>
<td>C$_1$</td>
<td><em>S. Paratyphi</em> C</td>
<td>6,7</td>
<td>c</td>
<td>1,5</td>
<td>Enteric fever</td>
</tr>
<tr>
<td></td>
<td><em>S. Cholerae-suis</em></td>
<td>6,7</td>
<td>c</td>
<td>1,5</td>
<td>NTS</td>
</tr>
<tr>
<td></td>
<td><em>S. Bareilly</em></td>
<td>6,7</td>
<td>y</td>
<td>1,5</td>
<td>NTS</td>
</tr>
<tr>
<td></td>
<td><em>S. Infantis</em></td>
<td>6,7</td>
<td>r</td>
<td>1,5</td>
<td>NTS</td>
</tr>
<tr>
<td></td>
<td><em>S. Montevideo</em></td>
<td>6,7</td>
<td>g,m,s</td>
<td>no phase 2 antigen</td>
<td>NTS</td>
</tr>
<tr>
<td></td>
<td><em>S. Oranienburg</em></td>
<td>6,7</td>
<td>m,t</td>
<td>no phase 2 antigen</td>
<td>NTS</td>
</tr>
<tr>
<td></td>
<td><em>S. Thompson</em></td>
<td>6,7</td>
<td>k</td>
<td>1,5</td>
<td>NTS</td>
</tr>
<tr>
<td>C$_2$</td>
<td><em>S. Newport</em></td>
<td>6,8</td>
<td>e,h</td>
<td>1,2</td>
<td>NTS</td>
</tr>
<tr>
<td>D</td>
<td><em>S. Typhi</em></td>
<td>9,12,Vi</td>
<td>d</td>
<td>no phase 2 antigen</td>
<td>Enteric fever</td>
</tr>
<tr>
<td></td>
<td><em>S. Dublin</em></td>
<td>1,9,12</td>
<td>g,p</td>
<td>no phase 2 antigen</td>
<td>NTS</td>
</tr>
<tr>
<td></td>
<td><em>S. Enteritidis</em></td>
<td>1,9,12</td>
<td>g,m</td>
<td>no phase 2 antigen</td>
<td>NTS</td>
</tr>
</tbody>
</table>

a All *Salmonella* serotypes listed are members of *Salmonella enterica* subspecies *enterica*.

b NTS = non-typhoidal Salmonellae - Clinical symptoms: acute, self-limiting gastroenteritis, diarrhea being the most common symptom. Fever, abdominal cramping, nausea and vomiting may also occur.
Table 1.2. Worldwide distribution of *Salmonella* Enteritidis human infections.

<table>
<thead>
<tr>
<th>Country</th>
<th>Total # isolates</th>
<th>Mean Annual Incidence (100,000)</th>
<th>% of NTS</th>
<th>Ranked among NTS</th>
<th>Study period</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>NA</td>
<td>5.79</td>
<td>32.1</td>
<td>1st</td>
<td>2009</td>
<td>Nesbitt <em>et al.</em>, 2012</td>
</tr>
<tr>
<td>Italy</td>
<td>57,499</td>
<td>2.99</td>
<td>25.1</td>
<td>1st</td>
<td>1980 to 2011</td>
<td>Graziani <em>et al.</em>, 2013</td>
</tr>
<tr>
<td>US</td>
<td>6,777</td>
<td>2.5</td>
<td>NA</td>
<td>1st</td>
<td>2004 to 2009</td>
<td>Chai <em>et al.</em>, 2012</td>
</tr>
<tr>
<td>EU</td>
<td>53,382</td>
<td>NA</td>
<td>52.3</td>
<td>1st</td>
<td>2009</td>
<td>EFSA, 2010</td>
</tr>
</tbody>
</table>

*NTS= non-typhoidal Salmonellae*
References


Bailey JS, Stern NJ, Fedorka-Cray P, Craven SE, Cox NA, Cosby DE, Ladely S, Musgrove MT: Sources and movement of Salmonella through integrated poultry...


Bhunia GS, Kesari S, Chatterjee N, Kumar V, Das P: **Spatial and temporal variation and hotspot detection of kala-azar disease in Vaishali district (Bihar), India.** BMC Infect Dis 2013, **13**:64.


Briggs AD, Boxall NS, Van Santen D, Chalmers RM, McCarthy ND: **Approaches to the detection of very small, common, and easily missed outbreaks that together contribute substantially to human Cryptosporidium infection.** Epidemiol Infect 2014, **142**(9):1869-76.


Kuehn BM: *Salmonella* cases traced to egg producers: findings trigger recall of more than 500 million eggs. JAMA 2010, 304(12):1316.


Younus M, Hartwick E, Siddiqi AA, Wilkins M, Davies HD, Rahbar M, Funk J, Saeed M: The role of neighborhood level socioeconomic characteristics in Salmonella


CHAPTER TWO

Incidence, distribution, seasonality, and demographic risk factors of 

(A Varga C, Pearl DL, McEwen SA, Sargeant JM, Pollari F, Guerin MT: Incidence, distribution, 
seasonality, and demographic risk factors of Salmonella Enteritidis human infections in Ontario, 

Abstract

Background

In Canada, surveillance systems have highlighted the increasing trend of Salmonella 
enterica serovar Enteritidis (S. Enteritidis) human infections. Our study objectives were 
to evaluate the epidemiology of S. Enteritidis infections in Ontario using surveillance data 
from January 1, 2007 through December 31, 2009.

Methods

Annual age-and-sex-adjusted incidence rates (IRs), annual and mean age-adjusted sex-
specific IRs, and mean age-and-sex-adjusted IRs by public health unit (PHU), were 
calculated for laboratory-confirmed S. Enteritidis cases across Ontario using direct 
standardization. Multivariable Poisson regression with PHU as a random effect was used 
to estimate incidence rate ratios (IROs) of S. Enteritidis infections among years, seasons, 
age groups, and sexes.

Results

The annual age-and-sex-adjusted IR per 100,000 person-years was 4.4 [95% CI 4.0-4.7] 
in 2007, and 5.2 [95% CI 4.8-5.6] in both 2008 and 2009. The annual age-adjusted sex-
specific IRs per 100,000 person-years ranged from 4.5 to 5.5 for females and 4.2 to 5.2
for males. The mean age-adjusted sex-specific IR was 5.1 [95% CI 4.8-5.4] for females and 4.8 [95% CI 4.5-5.1] for males. High mean age-and-sex-adjusted IRs (6.001-8.10) were identified in three western PHUs, one northern PHU, and in the City of Toronto. Regression results showed a higher IR of *S. Enteritidis* infections in 2009 [IRR = 1.18, 95% CI 1.06-1.32; P = 0.003] and 2008 [IRR = 1.17, 95% CI 1.05-1.31; P = 0.005] compared to 2007. Compared to the fall season, a higher IR of *S. Enteritidis* infections was observed in the spring [IRR = 1.14, 95% CI 1.01-1.29; P = 0.040]. Children 0-4 years of age (reference category), followed by children 5-9 years of age [IRR = 0.64, 95% CI 0.52-0.78; P < 0.001] had the highest IRR. Adults ≥ 60 years of age and 40-49 years of age [IRR = 0.31, 95% CI 0.26-0.37; P < 0.001] had the lowest IRR.

**Conclusions**

The study findings suggest that there was an increase in the incidence of *S. Enteritidis* infections in Ontario from 2007 to 2008-2009, and indicate seasonal, demographic, and regional differences, which warrant further public health attention.

**Keywords:** *Salmonella* Enteritidis, incidence, Poisson regression, mixed model, direct standardization, demographic risk factors, Ontario, Canada

**Background**

Salmonellosis remains an important public health issue worldwide [1-3], causing considerable health costs [4-7] and financial losses to all members of the food supply chain [2]. Globally, non-typhoidal salmonellae (NTS) cause an estimated 93.8 million human infections and 155,000 deaths annually [3]. Non-typhoidal salmonellae are the second most frequently reported enteric bacterial pathogens in Canada [8-9], the United
States of America (US) [10], and Europe [11]; and they are the top foodborne bacteria causing hospitalization and death in Canada [12] and the US [13-14].

*Salmonella enterica* serovar Enteritidis (*S. Enteritidis*) recently became the most common serotype among the NTS in the US [10], with a significantly increased incidence in 2009 compared with the periods 2006-2008 and 1996-1998 [6, 14]. Moreover, in Canada, surveillance systems have highlighted the increasing trend of *S. Enteritidis* human infections, such that *S. Enteritidis* has become the most prevalent NTS serotype [8-9, 15]. Considering the under-reporting rate of salmonellosis in Canada (an estimated 13 to 37 cases are unreported per reported case), the burden of infections is even higher [16].

The epidemiology of human *S. Enteritidis* infections is complex due to the multitude of risk factors that could be associated with illness. Previous epidemiological studies have revealed the following individual-level risk factors for *S. Enteritidis* infections: eating chicken outside of the home [17-18]; eating breaded, stuffed chicken products [19] and raw or undercooked eggs [18, 20, 21]; another infected person in the home [22]; eating food prepared by an infected food handler [23-26]; contact with birds and reptiles [26]; international travel [18, 26-28]; young age [29-30]; and exposures during June and July [31]; although other risk factors might also be important.

In Canada, health regions are administrative zones demarcated by provincial ministries of health according to provincial legislations [32]. In Ontario, Canada there are 36 public health units (PHUs) that oversee health promotion and disease prevention programs. In Ontario, salmonellosis is a reportable disease under provincial legislation and all laboratory-confirmed cases are reported to the local PHU; personnel at each PHU are
required to perform case investigations and enter their findings into the Ontario Ministry of Health and Long-Term Care’s (MOHLTC) integrated Public Health Information System (iPHIS) surveillance database. In addition, all clinical *Salmonella* isolates are sent to Public Health Ontario’s Toronto Public Health Laboratories for confirmation and serotyping using conventional methods [33].

Although passive surveillance systems represent an underestimation of disease burden, they provide invaluable data on enteric disease incidence and trends [34-35]. There is a need to better understand the demographic, geographic, and seasonal factors associated with the increase in human *S. Enteritidis* infections in Ontario and to provide evidence-based information for policy makers to prioritize future efforts in addressing the increasing number of infections. Thus, the objectives of this study were to 1) describe annual age-and-sex-adjusted incidence rates (IRs), and annual and mean age-adjusted sex-specific IRs of *S. Enteritidis* cases in Ontario; 2) describe the mean age-and-sex-adjusted IR for each PHU; and 3) identify associations between *S. Enteritidis* IRs and demographic and seasonal factors.

**Methods**

**Data sources**

In Ontario, a confirmed case of salmonellosis is defined as the isolation of *Salmonella* (excluding *Salmonella* Typhi or Paratyphi) from an appropriate clinical sample (e.g., stool, urine, blood) with or without clinically compatible signs and symptoms [36]. Data pertaining to the *S. Enteritidis* cases’ age, sex, reporting PHU, and date of illness onset were acquired from the iPHIS database. The University of Guelph Ethics Review Board was consulted because our research involved human participants; however, ethics
approval was not required because our data did not contain any personal or health information that could be linked back to the original identifiers. The data represent all cases of *S.* Enteritidis that were captured within the database between January 1, 2007 and December 31, 2009. Travel-related (i.e., those who had traveled outside of Canada within 3 days before the onset of illness) and outbreaks (two or more epidemiologically-linked cases) were included in the analysis because the study objectives were to describe the overall epidemiology of *S.* Enteritidis infections in Ontario.

The Census of Canada is administered every five years by Statistics Canada, to collect demographic and socioeconomic information on Canadian residents [37]. Estimates based on the 2006 Census of Population for each year, age category, sex, and PHU were obtained from Statistics Canada, Demography Division [38].

**Statistical methods**

The distribution of values was examined, missing data and improbable values were identified, and the data were corrected wherever possible or eliminated from the analysis. Descriptive and statistical analyses were performed using Microsoft Excel 2000 (Microsoft Corporation, Redmond, WA, USA) and STATA Intercooled statistical software, version 10.1 (Stata Corporation, College Station, TX, USA).

Direct standardization [39-41] was used with the PHU-, year-, age-, and sex-based population as the reference population to calculate annual age-and-sex-adjusted IRs, and annual and mean age-adjusted sex-specific IRs for *S.* Enteritidis cases in Ontario, and the mean age-and-sex-adjusted IR for each PHU.
To identify associations between *S. Enteritidis* IRs in Ontario and demographic and seasonal factors, a multivariable Poisson regression analysis was conducted. The dependent variable was the number of *S. Enteritidis* cases by year, season, age group, sex, and PHU (see Appendix A - Legend 2.1 for an example of the data structure). The categorical independent variables were year, season, age group, sex, and PHU. The variable PHU represented the 36 PHUs in Ontario. The PHU was included as a fixed effect because of the observed variability of the IRs across PHUs (Fig. 2.1). The District of Algoma Health Unit, because it had the lowest IR, was used as the reference category to which the other PHUs were compared. The date of onset of illness reported by each *S. Enteritidis* case was used to assign the case to a particular year and season. When the date of onset was missing, the date when the sample was received by the laboratory or when the case was reported into the iPHIS database was used. Season was categorized as winter (December, January, and February), spring (March, April, and May), summer (June, July, and August), and fall (September, October, and November). The variable year was defined as a consecutive 12-month period from January 1st to December 31st; thus, there were three categories for year (2007, 2008, and 2009). The variable age included ten-year age categories, with the exception of children < 4 years of age and those 5-9 years of age, which were retained because of their biological importance [43-44], and adults 60 years of age and older, which were pooled into one category because of the small number of cases in this age group. Pair-wise correlation coefficients using the Spearman’s rank test among all variables were examined. If the independent variables were highly correlated (Spearman’s rho > 0.70), variables with the smallest p-value were considered for the model building process. To address the differences in year-, age group-, sex-, and PHU-
based population size estimates, we used the natural log-transformed population estimates as the offset, which accounted for the denominator when calculating incidence rate ratios (IRRs). An IRR was the IR in the category of interest compared to the IR in the reference category. Variables with a p-value equal to or less than 0.05 were considered significant and were kept in the model. Incidence rate ratios and their corresponding 95% confidence intervals were estimated. Interaction terms were created between each independent variable and tested for significance. If the interaction term was significant (p ≤ 0.05) it was retained in the final model. The model was evaluated by identifying influential observations (i.e. large values of Cook’s distance) and outliers (i.e. large values of Pearson, deviance, or Anscombe residuals) using residual plots. The overall fit of the model was assessed using Deviance and Pearson $\chi^2$ goodness-of-fit tests [42].

To account for lack of fit, a multi-level mixed-effects Poisson regression model was then constructed using the xtmepoisson command in STATA [45], which uses adaptive Gaussian quadrature to approximate the log likelihood. The model included the same dependent and independent (year, season, age group, and sex) variables as the first model with the exception that PHU was included as a random intercept instead of a fixed effect. The structure of the multi-level model included an offset representing the natural log-transformed year-, age group-, sex-, and PHU-based population size estimates. As part of assessing model fit, we examined the normality of the best linear unbiased predictors (BLUPs) [46]. Outlier and influential observations were assessed using residual plots. Bayesian information criterion (BIC) was used to compare the fit of the two models.
ArcGIS 10 (Environmental Systems Research Institute, Inc., Redlands, CA, USA) was used to create a choropleth map for mean age-and-sex-adjusted IRs across Ontario’s PHUs; Jenk's optimization classification method [47] was employed for defining the critical intervals. This method arranges data into classes based on their distribution by using an algorithm that reduces variance within groups and maximizes variance between groups.

**Results**

**Descriptive statistics and direct standardized incidence rates**

Between January 1, 2007 and December 31, 2009, 1,935 S. Enteritidis cases were reported into iPHIS in Ontario. Three cases were excluded because they lacked age or sex information. The date of onset of illness was reported for 1,670 (86.4%) cases; 230 (11.9%) and 32 (1.7%) cases only had information on the date when the sample was received by the laboratory and the date when the case was reported into the iPHIS database, respectively. The iPHIS collects all reportable diseases throughout Ontario’s PHUs, and no major changes in salmonellosis reporting requirements or testing protocols were noted during the study period, which makes our data robust and reliable.

Information on specimen type was not available; however, based on our working experience at the MOHLTC, and the literature, the majority of specimens were stool samples. No major outbreaks were declared during the study period.

The age of cases ranged from < 1 year to > 90 years. Children < 4 years of age and adults ≥ 60 years of age represented 13.1% and 13.6% of cases, respectively, while adults 20-29
years of age represented 16.8% of cases. Overall, 51.6% and 48.4% of cases were females and males, respectively.

The annual age-and-sex-adjusted IR per 100,000 person-years was 4.4 [95% CI 4.0-4.7] in 2007, and 5.2 [95% CI 4.8-5.6] in both 2008 and 2009 (Table 2.1). Over the study period, the annual age-adjusted sex-specific IR per 100,000 person-years ranged between 4.5 and 5.5 for females and between 4.2 and 5.2 for males (Table 2.1). The mean age-adjusted sex-specific IR per 100,000 person-years was 5.1 [95% CI 4.8-5.4] for females and 4.8 [95% CI 4.5-5.1] for males (Table 2.1).

Seasonal counts ranged from 135 to 187 cases in winter (mean over 3-year period = 160 cases), 155 to 189 in spring (mean = 173), 156 to 166 in summer (mean = 160), and 121 to 177 in fall (mean = 151) (Fig. 2.2). The highest monthly count was 75 cases in March 2007 and the lowest was 28 cases in November 2007 (Fig. 2.3).

The mean age-and-sex-adjusted IRs for the entire study period across Ontario’s PHUs ranged from 1.9 to 8.1 (Fig. 2.1; Appendix B - Legend 2.2.). Visually exploring the map, the highest IRs (> 6.0 per 100,000 person-years) were observed in three south-western PHUs (Halton Regional Health Unit, Huron County Health Unit, and Waterloo Health Unit), one northern PHU (Thunder Bay District Health Unit), and in the City of Toronto Health Unit.

**Poisson regression**

The Deviance and Pearson $\chi^2$ goodness-of-fit test statistics for the Poisson model with PHUs included as fixed effect were 6,212.6 (P = 1.00) and 17,727.6 (P = 0.004),
respectively. Several outlier and influential observations were identified; however, re-running the model without these observations did not change any of the coefficients. The BIC for the model was 9,906.4.

Because one of the two goodness-of-fit tests for the Poisson model indicated lack of fit, we used a multi-level model. No outlier or influential observations were identified for the upper level residuals of the multi-level model. The BLUPs for the PHU random intercept were normally distributed. The BIC for the multi-level model was 9,655.2, indicating a better fit.

The results of the multi-level model are shown in Table 2.2. Significantly higher IRs of \textit{S. Enteritidis} infections were reported in 2009 [IRR = 1.18, 95% CI 1.06-1.32] and 2008 [IRR = 1.17, 95% CI 1.05-1.31] compared to 2007. Compared to the fall season, a significantly higher IR of \textit{S. Enteritidis} infections was reported in the spring [IRR = 1.14, 95% CI 1.01-1.29]. Children 0-4 years of age (reference category), followed by children 5-9 years of age [IRR = 0.64, 95% CI 0.52-0.78] had the highest IR of infection. Adults $\geq$ 60 years of age and 40-49 years of age [IRR = 0.31, 95% CI 0.26-0.37] had the lowest rates of infection. No statistically significant difference in \textit{S. Enteritidis} infection rates were detected between sexes.

**Discussion**

Our study is the most current and geographically diverse investigation from Ontario, and fills information gaps related to current knowledge of the incidence, demographic determinants, distribution, and seasonality of human \textit{S. Enteritidis} infections.
From 2007 to 2008-2009, an increase in annual age-and-sex-adjusted IRs of \textit{S. Enteritidis} infections was identified. Moreover, the model revealed a significantly higher rate of \textit{S. Enteritidis} infections in 2008 and 2009 compared to 2007. This finding is in agreement with the results of current Canadian [15] and US [6] surveillance that have shown an increase in \textit{S. Enteritidis} infections. \textit{Salmonella} Enteritidis continues to be a key cause of human enteric illness and poses a substantial health burden to the North American population [3]. Reducing the incidence of \textit{S. Enteritidis} infections is challenging due to the variety of transmission routes and contaminated food sources [26], the possible increase in environmental reservoir(s), and changes in food processing and safety practices [48]; however, increased efforts should be directed toward mitigation strategies for this pathogen.

Our study demonstrated that young children 0-4 years of age had the highest \textit{S. Enteritidis} infection IRs, which is in agreement with results of other studies from developed countries [29-30, 43, 49-50]. Previous studies identified several risk factors for \textit{S. Enteritidis} infections for this age group, including international travel [51-52], riding in shopping carts and exposure to raw meat and poultry products [53], and contact with reptiles [44, 52, 54] and cats [44]. In our study, adults 60 years of age and older had the lowest IR among all age groups, which is in contrast with other studies [29, 47]. This finding was unexpected because typically the two age group extremes have the highest rates of enteric infections. Prospective research studies are needed in Ontario to assess differences in \textit{S. Enteritidis} infection rates between age groups that are attributed to various exposures.
Examination of seasonal differences in *S.* Enteritidis rates in our study revealed a higher IR of infections during the spring (March through May). The higher incidence in spring might be associated with international travel. Travel has been identified as an important risk factor for *S.* Enteritidis, and it was shown in recent Ontario studies that a large proportion of *S.* Enteritidis cases, especially in the winter and spring, were travel-related [28,56].

We did not find a statistically significant difference in the IR of *S.* Enteritidis infections between females and males, which is consistent with a previous US study [29].

When analysing the differences in the incidence rates of *S.* Enteritidis infections among Ontario’s PHUs, we calculated mean age-and-sex-adjusted IRs using direct standardization. This method is useful when the prevalence of exposures might differ among age groups, sexes, and PHUs. We used geographic information system software to create choropleth maps for IRs of *S.* Enteritidis infections across PHUs in Ontario. This is a useful technique to visualize the findings of conventional statistical analysis, and by using Jenk's optimization classification for defining the critical intervals for mapping the IRs, it allowed us to identify high risk PHUs. Future research studies should be conducted to identify and assess novel transmission routes, spatio-temporal trends, and socioeconomic status indicators that might have an impact on the emergence of *S.* Enteritidis infections in these regions.

Before extrapolating our results to the whole Ontario population, a few limitations need to be noted. It is essential to mention that laboratory surveillance systems generally underestimate the true burden of enteric diseases in a population for several reasons.
There might be differences in underreporting across age groups, because children and older adults are more likely to visit a physician, and physicians are more likely to request stool samples from them for testing. Moreover, there might be geographic variation in underreporting of *S*. Enteritidis infections due to differences in health care providers’ accessibility, and in the sensitivity of laboratory methods used at different laboratories [34-35, 57]. Finally, misclassification of cases might have occurred when cases were categorized into year and season. However this bias was likely minor because the majority of cases had date of illness onset (or date of sample reception) information. The difference between date of illness onset and the date when the samples were received by the laboratory could be estimated to be a maximum of one week, considering the time of delivery of samples within Ontario, and the incubation period of *S*. Enteritidis that ranges from half to three days [58].

**Conclusions**

Our results showed higher IRs of *S*. Enteritidis infections in 2008 and 2009 compared to 2007, and indicate seasonal and regional differences, with a higher IR of *S*. Enteritidis infections in the spring. In Ontario, we found that children 0-4 years of age were at the highest risk for *S*. Enteritidis infections. These results provide evidence-based information that will assist policy makers to prioritize future efforts in addressing the increase in the number of *S*. Enteritidis infections in the human population in Ontario. We recommend that children, and PHUs with high *S*. Enteritidis rates, be targeted for prevention and control programs designed to decrease the incidence of *S*. Enteritidis. Further case-control and ecological studies are needed to identify novel risk factors (food
sources, socioeconomic determinants, and transmission routes) and spatio-temporal trends for *S*. Enteritidis infections in Ontario.

**List of Abbreviations Used**

BIC - Bayesian information criterion

BLUP - best linear unbiased predictor

iPHIS - integrated Public Health Information System

IR - incidence rate

IRR – incidence rate ratio

MOHLTC - Ontario Ministry of Health and Long-Term Care

NTS - non-typhoidal salmonellae

PHU - Public Health Unit

*S*. Enteritidis - *Salmonella enterica* serovar Enteritidis

US – United States of America

**Acknowledgements**

The authors acknowledge the MOHLTC for providing the data. We thank the staff of all PHUs and public health laboratories that tested samples, followed up with cases, and entered information into the public health surveillance database. The views expressed in
this study are the views of the authors and do not necessarily reflect those of the MOHLTC.

References


6. Centers for Disease Control and Prevention (CDC): Vital signs: incidence and trends of infection with pathogens transmitted commonly through food--
foodborne diseases active surveillance network, 10 U.S. sites, 1996-2010.


17. Kimura AC, Reddy V, Marcus R, Cieslak PR, Mohle-Boetani JC, Kassenborg HD, Segler SD, Hardnett FP, Barrett T, Swerdlow DL: Chicken consumption is


Canada: who is counted in provincial communicable disease statistics?


Table 2.1. Direct standardized incidence rates of *Salmonella* Enteritidis infections in Ontario, 2007-2009 (n = 1,932 cases).

<table>
<thead>
<tr>
<th>Year</th>
<th>N</th>
<th>IR</th>
<th>95 % CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>12,792,937</td>
<td>4.4</td>
<td>4.0-4.7</td>
</tr>
<tr>
<td>2008</td>
<td>13,070,584</td>
<td>5.2</td>
<td>4.8-5.6</td>
</tr>
<tr>
<td>2009</td>
<td>13,064,900</td>
<td>5.2</td>
<td>4.9-5.6</td>
</tr>
<tr>
<td>2007</td>
<td>Female 6,480,556</td>
<td>4.5</td>
<td>4.0-5.1</td>
</tr>
<tr>
<td></td>
<td>Male 6,312,381</td>
<td>4.2</td>
<td>3.7-4.7</td>
</tr>
<tr>
<td>2008</td>
<td>Female 6,565,166</td>
<td>5.2</td>
<td>4.7-5.8</td>
</tr>
<tr>
<td></td>
<td>Male 6,505,418</td>
<td>5.2</td>
<td>4.7-5.8</td>
</tr>
<tr>
<td>2009</td>
<td>Female 6,625,568</td>
<td>5.5</td>
<td>5.0-6.1</td>
</tr>
<tr>
<td></td>
<td>Male 6,439,332</td>
<td>5.0</td>
<td>4.5-5.6</td>
</tr>
<tr>
<td>2007-2009</td>
<td>Female 19,671,290</td>
<td>5.1</td>
<td>4.8-5.4</td>
</tr>
<tr>
<td></td>
<td>Male 19,257,131</td>
<td>4.8</td>
<td>4.5-5.1</td>
</tr>
</tbody>
</table>

a) Year: A consecutive 12-month period from January 1st to December 31st
b) N: Reference population estimates obtained from the 2006 Census of Canada
c) IR: Direct standardized incidence rate per 100,000 person-years. Denominator included year-, age-, sex-, and public health unit-based population estimates.
d) CI: Confidence interval of the adjusted IR
e) Annual age-and-sex-adjusted IR
f) Annual age-adjusted sex-specific IR
g) Mean age-adjusted sex-specific IR
Table 2.2. Risk factors for *Salmonella* Enteritidis infections in humans, Ontario, Canada, 2007-2009 (n = 1,932 cases).

<table>
<thead>
<tr>
<th>Variable</th>
<th>IRR b)</th>
<th>95% CI c)</th>
<th>P-value d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>Reference</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2008</td>
<td>1.17</td>
<td>1.05-1.31</td>
<td>0.005</td>
</tr>
<tr>
<td>2009</td>
<td>1.18</td>
<td>1.06-1.32</td>
<td>0.003</td>
</tr>
<tr>
<td>Season</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>Reference</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Spring</td>
<td>1.14</td>
<td>1.01-1.29</td>
<td>0.040</td>
</tr>
<tr>
<td>Summer</td>
<td>1.06</td>
<td>0.93-1.20</td>
<td>0.377</td>
</tr>
<tr>
<td>Winter</td>
<td>1.06</td>
<td>0.93-1.20</td>
<td>0.413</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-4</td>
<td>Reference</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5-9</td>
<td>0.64</td>
<td>0.52-0.78</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>10-19</td>
<td>0.44</td>
<td>0.34-0.51</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>20-29</td>
<td>0.51</td>
<td>0.43-0.60</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>30-39</td>
<td>0.34</td>
<td>0.28-0.40</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>40-49</td>
<td>0.31</td>
<td>0.26-0.37</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>50-59</td>
<td>0.33</td>
<td>0.28-0.40</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>≥ 60</td>
<td>0.31</td>
<td>0.26-0.37</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>Reference</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Male</td>
<td>0.95</td>
<td>0.87-1.04</td>
<td>0.273</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.00002</td>
<td>0.00002-0.00003</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

a) Multi-level mixed-effects Poisson regression model using adaptive Gaussian quadrature. Public health unit (PHU) was included as a random intercept. Variance of PHU random effects = 0.074 [95% CI = 0.036-0.153]. Dependent variable: number of *Salmonella* Enteritidis cases by year, season, age group, sex, and PHU. Offset: natural log-transformed year-, age group-, sex-, and PHU-based population size estimates. b) IRR: Incidence rate ratio for categorical independent variables, in which the incidence rate (IR) in the category of interest was compared to the IR in the reference category. c) CI: Confidence interval of the IRR. d) Statistically significant at P ≤ 0.05
Figure 2.1. Mean age-and-sex-adjusted incidence rates of *Salmonella* Enteritidis infections, across Ontario public health units, 2007-2009.

IR- Incidence rate. Public health unit labels and names are presented in Appendix B – Legend 2.2
Figure 2.2. Seasonal distribution of *Salmonella* Enteritidis (S. Enteritidis) cases in Ontario, 2007-2009 (n = 1,932) a).

a) Winter (December, January, and February), spring (March, April, and May), summer (June, July, and August), and fall (September, October, and November). Three-year-average: number of S. Enteritidis cases for each season divided by the number of years.
Figure 2.3. Number of *Salmonella* Enteritidis (S. Enteritidis) cases by month in Ontario, 2007-2009 (n = 1,932) \(^a\).

\(^a\) Three-year-average: number of *Salmonella* Enteritidis cases for each month divided by the number of years
CHAPTER THREE


Abstract

Background

There have been only a few region-level ecological studies conducted in Canada investigating enteric infections in humans. Our objectives were to assess spatial clustering of *Salmonella enterica* serotype Enteritidis (*S.* Enteritidis) infections in three public health units (PHUs) within the Greater Toronto Area (GTA), Ontario, Canada, and identify underlying area-level associations between standardized incidence rates (SIRs) of *S.* Enteritidis infections and socioeconomic status (SES) indicators that explain the clustering of infections.

Methods

Retrospective data on *S.* Enteritidis infections from 2007 to 2009 were obtained from the Ontario Ministry of Health and Long-Term Care’s surveillance database and grouped at the forward sortation area (FSA) - level. A spatial scan statistic was used to identify FSA-level spatial clusters of high infection rates. A multivariable negative binomial regression model was used to identify FSA-level associations between SIRs of *S.* Enteritidis infections and SES risk factor variables obtained from the 2006 Census of Canada. The
final model was evaluated for spatial clustering after removing the effect of SES variables by assessing global Moran’s I statistic of the Anscombe residuals.

**Results**

A spatial cluster of high rates that included nine neighbouring FSAs was identified in downtown Toronto. A significant positive curvilinear relationship was observed between the SIR of *S. Enteritidis* infections and FSA-level average number of children at home per census family. Areas with high and areas with low average median family income had higher SIRs of *S. Enteritidis* infections than FSAs with medium average median family income. Areas with a high proportion of visible minority population had lower SIRs of *S. Enteritidis* infections than FSAs with a medium proportion of visible minority population. The Moran's I statistic was not significant indicating that no residual spatial autocorrelation was present after accounting for these SES variables.

**Conclusions**

Our study demonstrated that FSAs with high and low average median family income, medium proportion of visible minority population, and high average number of children at home per census family had the highest SIRs of *S. Enteritidis* infections. These areas should be targeted when designing disease control and prevention programs. Future studies are needed in areas with high SIRs of *S. Enteritidis* infections to identify novel individual-level risk factors.

**Keywords:**

*S. Enteritidis*, socioeconomic status, spatial scan statistic, GIS, choropleth map, negative binomial regression, Moran’s I, ecological study
Background

Non-typhoidal salmonellosis (NTS) is a major foodborne zoonotic infection that poses a significant public health risk [1-5]; on a global basis, it affects an estimated 93.8 million people and causes 155,000 deaths annually [1]. In Canada, NTS is the second most frequently reported enteric infection [6], and results in the largest number of hospitalizations and deaths among foodborne diseases [7]. Surveillance systems in Canada have identified an increasing trend of Salmonella enterica serovar Enteritidis (S. Enteritidis) infections in humans, with a threefold increase between 2003 and 2009 [8]. As a result of this increase, S. Enteritidis became the most common Salmonella serotype in Ontario and Canada.

There are an abundance of case-control studies worldwide that have evaluated individual-level associations between S. Enteritidis infections and potential risk factors including: consumption of chicken [9-10], consumption of raw or undercooked eggs [9,11-12], person-to-person transmission through infected food handlers [13-15], animal-to-person transmission [16-17], and international travel [9, 18-20]. In contrast, there are a limited number of population-based ecological studies that have evaluated area-level associations between enteric infections and socioeconomic status (SES) indicators. Ecological and individual-level studies from Canada and Europe [21-24] have demonstrated associations between enteric infections and SES determinants (e.g. household income, education level, unemployment rate, number of children per household, cultural group, population density). Two ecological studies from the United States of America (US) analyzed associations between salmonellosis and area-level SES and sociodemographic factors. Younus et al. [25] found that areas with higher education attainment had a greater
incidence of salmonellosis compared to areas with a lower education attainment, which was partly explained by the overrepresentation of highly educated persons in the surveillance system due to their better access to health care and their willingness to visit a doctor even for mild symptoms. It has also been hypothesized that groups with higher education tend to eat outside of the home more frequently and are more likely to own pets [25], which are considered to be sources of *Salmonella* for individuals [26-27]. In another US ecological study, Chang et al. [28] showed that areas with a higher black, Hispanic, or Latino population had a higher likelihood of *Salmonella* infections than areas with a predominantly white population. However, *S*. Enteritidis infections might have different SES risk factors than other *Salmonella* serotypes.

We previously examined the incidence, seasonality, and demographic risk factors of *S*. Enteritidis human infections across Ontario’s health regions from 2007 to 2009 [29] and identified three public health units (PHUs) within the Greater Toronto Area (GTA) with moderate to high incidence rates that were of interest for further assessment (Figure 3.1). Therefore, our objectives were to assess forward sortation area (FSA; first three digits of the postal code)-level clustering of *S*. Enteritidis infections within these three PHUs, and identify underlying area-level associations between standardized incidence rates (SIRs) of *S*. Enteritidis infections and SES risk factors. The findings of this study are expected to assist public health authorities in designing effective area-based prevention and control programs for *S*. Enteritidis.
Methods

Study population and study design

In Ontario, there are 36 PHUs mandated by the provincial ministry of health to manage health promotion and disease prevention programs. Our study area included three PHUs within the GTA, namely, the City of Toronto Health Unit, Peel Regional Health Unit, and York Regional Health Unit. The analysis was conducted at the FSA-level, a well-defined zone within a larger geographic region, represented by the first three characters of the postal code. We excluded FSAs with less than 500 residents (n = 12) to obtain stable incidence rates, and FSAs with centroids outside of the three PHU boundaries (n = 7) to attain a well-delimited and uniform geographical study area. We included 153 FSAs containing a combined population of 4,570,151 that accounted for 40% of Ontario’s population in 2006 (Figure 3.1). The annual population sizes of the 153 FSAs ranged from 2,172 to 84,180 persons, with a mean of 29,870 persons.

Data sources

i) Ontario Ministry of Health and Long-Term Care, integrated Public Health Information System

In Ontario, a confirmed case of salmonellosis is defined as the isolation of *Salmonella* spp. (excluding *Salmonella Typhi* or Paratyphi) from an appropriate clinical sample (e.g., stool, urine, blood) with or without clinically-compatible signs and symptoms [30]. Salmonellosis is a reportable disease under provincial legislation [31]; all infections confirmed by hospital, private, and public health laboratories must be reported to the local PHU, whose personnel are required to investigate each case and enter the case’s demographic information (age, sex, area of residence), clinical features, and exposure
history to the MOHLTC through iPHIS. In addition, all clinical *Salmonella* isolates are sent to Toronto Public Health Laboratories for confirmation and serotyping using conventional methods [32]. Information pertaining to *S*. Enteritidis cases’ PHU, FSA, and date of onset of illness were extracted from iPHIS; our study data represent all *S.* Enteritidis infections from the three PHUs captured within the database between January 1, 2007 and December 31, 2009. Reports obtained from iPHIS are useful for evaluating food source attribution and demographic risk factors, yet they lack SES indicators, such as family income, family size, or level of education, due to privacy legislation. Our data did not have any personal or health information that could be linked back to the original identifiers; therefore, the University of Guelph Ethics Review Board did not require ethics approval for this study.

**ii) Census of Canada**

The Census of Canada is administered every five years by Statistics Canada and collects self-reported demographic and socioeconomic characteristics on Canadian residents [33]. Census data are available for various geographic areas, including country, provinces and territories, health regions, and FSAs. Forward sortation area was selected as the unit of interest for our study because FSAs in the GTA have well-defined geographical boundaries; FSA-level SES indicators for the three PHUs of interest within the GTA were obtained from the 2006 Census.

**Statistical analysis**

**Descriptive statistics**

The unadjusted *S*. Enteritidis incidence per 100,000 person-years per FSA over the study period was calculated by dividing the sum of the annual number of *S*. Enteritidis cases
per FSA by the sum of the annual FSA population estimates obtained from the 2006 Census. A choropleth map of the unadjusted incidence was created using ArcGIS 10 (ESRI Inc., Redlands, CA, US); Jenk’s optimization classification method was used to define the critical intervals for mapping [34]. This method arranges data into classes based on their distribution by using an algorithm that reduces variance within groups and maximizes variance between groups.

In addition, the relative risk per FSA was calculated by dividing the observed number of cases per FSA by the expected number of cases per FSA (calculation described in the ‘Dependent variable’ subsection), and then mapped using ArcGIS 10 to illustrate FSAs with excess risk after adjusting for the age and sex distribution of the population.

**Spatial analysis**

Date of illness onset was used to assign each case to a particular month and year. Each case was categorized into one of five age groups: 0-9, 10-24, 25-34, 35-49, and 50+ years. Cartesian coordinates of latitude and longitude for each FSA centroid were calculated in ArcGIS 10. A database was created in a Microsoft Excel 2010 (Microsoft Corporation, Redmond, WA, US) spreadsheet containing each case identifier and its corresponding information (FSA, FSA population estimate, Cartesian coordinates of the FSA centroid, illness onset month and year, age, and sex), which was subsequently imported into the SaTScan™ software [35] for analysis.

To identify clusters with high infection rates we employed the retrospective, purely spatial scan statistic that used the discrete Poisson model [36]. Age and sex were added to the model to account for possible confounding effects. The model assumes that the
number of cases in each FSA are Poisson-distributed, based on a known underlying population at risk. The scan statistic uses a circular scanning window of variable radii. The null hypothesis assumes equal infection rates inside the scanning window compared to outside the scanning window, whereas the alternative hypothesis expects a higher infection rate inside the scanning window compared to outside [37]. The scan statistic uses a maximum likelihood function that identifies high rate clusters that are least likely to have happened by chance alone. A p-value was obtained through Monte Carlo hypothesis testing [38] using 999 replications and a relative risk is allocated to this cluster. For our analysis, the maximum spatial circular window size was set to include up to 50% of the population at risk [37]. Secondary clusters were also reported if they did not overlap with the primary cluster. Spatial clusters significant at $\alpha = 0.05$ were illustrated in a map using ArcGIS 10.

**Regression analysis**

A regression analysis (described below) was employed to identify area-level associations between SIRs of S. Enteritidis infections and SES variables in order to better understand the underlying socioeconomic factors that might contribute to spatial clustering of S. Enteritidis infections within the three PHUs.

**Variable selection and creation**

i) SES variables: According to the Census of Canada, a *census family* is a married couple (with or without children of either or both spouse), common-law couple (with or without children of either or both partner), or lone-parent family regardless of sex of parent. A couple may be of
opposite or same sex. All members of a specific census family must live in the same household.

The following FSA-level SES indicators were obtained from the 2006 Census of Canada [33] and calculated as follows:

1) Average number of children at home per census family, which was calculated by summing the number of children per family then dividing by the total number of families.

2) Average number of persons per census family, which was calculated by summing the number of persons per family then dividing by the total number of families.

3) Average number of rooms per home, which was calculated by summing the number of rooms in a home then dividing by the total number of homes.

4) Immigrant population proportion, which was calculated by dividing the immigrant population by the total population and multiplying by 100. The term immigrant refers to a person who is, or has ever been, a landed immigrant in Canada. A landed immigrant is a person born outside of Canada who has been granted the right to live in Canada permanently by immigration authorities.

5) Average median family income, which was calculated by dividing the sum of the median income of families in a household by the total number of households. The Census dictionary defines the median income of families in a household as “that amount which divides family member's income size distribution, ranked by size of income, into two halves. Median incomes of families in a household are normally calculated for all members of the family, whether or not they reported
income”. The total income of the family members was defined as the sum of the income obtained from wages, salaries, bonuses, interest, and any provincial - federal benefits and welfare payments.

6) Proportion of university graduates of persons between 25 and 64 years of age, which was calculated by dividing the number of persons between 25 and 64 years of age with a university degree by the total population between 25 and 64 years of age and multiplying by 100.

7) Unemployment rate of persons 15 years of age or older, which was calculated by dividing the number of unemployed by the number of people in the labour force and multiplying by 100.

8) Visible minority population proportion, which was calculated by dividing the visible minority population by the total population and multiplying by 100.

Visible minority, according to the Employment Equity Act, refers to a person, other than an Aboriginal person, who is non-Caucasian in race or non-white in colour.

ii) Dependent Variable:

The dependent variable (observed number of cases) was the number of laboratory-confirmed S. Enteritidis human infections reported in iPHIS from January 1, 2007 through December 31, 2009 from the three PHUs within the GTA, aggregated to the FSA-level. Using age-and-sex-based population estimates from the 2006 Census, the expected number of cases for each FSA was calculated using indirect standardisation [59].
**Modeling approach**

A database containing the SES variables, dependent variable, and expected number of cases for each FSA was constructed in Microsoft Office Excel 2010, and subsequently imported into STATA Intercooled 10.1 statistical software (Stata Corporation, College Station, TX, US) for analysis. We initially analyzed the data using a Poisson regression model. To estimate SIRs of *S*. Enteritidis infections, the natural log-transformed expected number of cases was used as the offset to account for differences in FSA-level age-and-sex-based population that might affect FSA-level infection rates [39].

We assessed the relationship between the natural log of the SIR of *S*. Enteritidis infections and each of the continuous SES variables by a locally weighted regression method using lowess curves. If non-linearity was observed, we introduced a quadratic term. If the quadratic term was statistically significant ($p \leq 0.05$), collinearity between the linear and quadratic terms was reduced by subtracting the mean value from each individual value of the linear term (centering), and then squaring each centered value [40]. If the non-linear relationship was not adequately modelled with a quadratic term, the variable was categorized into three equal groups - high, medium, and low - based on its distribution over all FSAs. The medium category was chosen as the reference category.

Univariable screening ($\alpha = 0.10$) consisted of individually regressing each SES variable on the dependent variable. Pair-wise correlation coefficients using the Spearman’s rank test among all significant independent variables ($p < 0.10$) were examined to prevent the inclusion of collinear variables in the multivariable model. If two variables were highly
correlated (rho > 0.70), the variable with the smallest p-value on univariable analysis was considered for the multivariable model.

All unconditionally significant variables identified on univariable screening, and a categorical variable signifying the three PHUs, were offered to a multivariable model, and a manual backward elimination process was performed. Variables with p > 0.05 were removed unless there was evidence of confounding (when removal of a non-intervening variable changed the coefficients of remaining variables by more than 25%) [40]. Once a main effects model was created, two-way interaction terms between significant SES variables were tested and retained in the final model if they were significant (p ≤ 0.05).

The fit of the Poisson model was assessed using the Pearson $\chi^2$ goodness-of-fit test [41]. The test was significant ($\chi^2 = 178.84, P = 0.03$) indicating that the model did not fit the data; therefore, the analysis was repeated using a negative binomial regression model. The final negative binomial model was evaluated to assess model fit and to measure spatial autocorrelation among residuals. First, we assessed Anscombe residuals for each FSA to identify outliers, and then a normal quantile plot (quantiles of the residuals against the quantiles of the normal distribution) was used to assess model fit [42]. In addition, we visually assessed Anscombe residuals in a choropleth map using ArcGIS 10 to identify FSAs with over- or under-predicted values (> 3 or < - 3). Secondly, spatial autocorrelation of FSA-level residuals was evaluated by applying the global Moran's I statistic [43]. A significant global Moran's I (p ≤ 0.05) would indicate that FSA-level residuals are not randomly distributed across the study area and accounting for spatial clustering among FSAs would be required.
For each SES variable from the final model with a significant quadratic term, we plotted the variable (linear and quadratic terms) against the predicted log SIR of *S. Enteritidis* infections in order to describe the curvilinear relationship while controlling for the other variables in the model. For each SES variable from the final model, choropleth maps were created in ArcGIS 10 to visualize the FSA-level distribution of the data.

**Results**

**Descriptive statistics**

The total number of *S. Enteritidis* infections for the study period was 846; the count per FSA ranged from 0 to 18 (mean = 5.53), and the unadjusted incidence per 100,000 person-years per FSA over the study period ranged from 0 to 23 (mean = 6.45) (Figure 3.2).

Figure 3.3 illustrates the FSA-level distribution of the relative risk (ratio of the observed number of cases to the age-and-sex-adjusted expected number of cases). Values < 1 indicate lower incidence than expected, whereas values > 1 indicate higher incidence than expected. There were 11 FSAs with a relative risk greater than 2.

**Spatial clustering**

We identified a single cluster of higher than expected infection rates located in the south-central area (downtown) of the City of Toronto Health Unit, which included nine neighbouring FSAs [Relative Risk = 1.90, *P* = 0.050] (Table 3.1; Figure 3.4).
Negative binomial regression

Variables significantly associated with the SIR of S. Enteritidis infections at the FSA-level on univariable screening included average number of children at home per census family, average number of persons per census family, average number of rooms per home, immigrant population proportion, average median family income, and visible minority population proportion (Table 3.2). The average number of children at home per census family and the average number of persons per census family were highly correlated (rho = 0.94, P < 0.01); average number of children at home was kept for the multivariable model because of its smaller p-value on univariable analysis and because children were identified as having the highest incidence rate of S. Enteritidis infections in Ontario during the study period [29].

Variables significantly associated with the SIR of S. Enteritidis infections at the FSA-level in the final multivariable model included average number of children at home per census family, average median family income, and visible minority population proportion (Table 3.3). No significant interaction terms were identified. Public health unit was not significant on the likelihood ratio (LR) test (LR $\chi^2 = 2.77$, P = 0.250); therefore, we did not keep this variable in our final model. Figure 3.5 shows a plot of the average number of children at home per census family against the predicted log SIR of S. Enteritidis infections, holding all other variables from the final model at their referent level. The positive curvilinear convex shape of the line indicates an exponential increase in the SIR of S. Enteritidis infections as the FSA-level average number of children at home per census family increases. Forward sortation areas with low average median family income [Incidence Rate Ratio (IRR): 1.34, P = 0.002] and high average median family income...
[IRR: 1.24, P = 0.020] had higher SIRs of S. Enteritidis infections than FSAs with medium average median family income. Forward sortation areas with high visible minority population proportion [IRR: 0.76, P = 0.002] had a lower SIR of S. Enteritidis infections than FSAs with medium visible minority population proportion. Figures 3.6, 3.7, and 3.8 display the FSA-level distribution of the average number of children at home per census family, average median family income, and visible minority population proportion, respectively. Visual assessment of the maps showed that there were clusters of high values for the FSA-level average number of children at home per census family in the north-western and eastern parts of the City of Toronto Health Unit, central and southern parts of the York Regional Health Unit, and most of the Peel Regional Health Unit except the southern part. Low average median family income FSAs clustered in the north-western, south-central, and eastern parts of the City of Toronto Health Unit, and in the mid-eastern part of the Peel Regional Health Unit. High average median family income FSAs clustered in the central part of the City of Toronto Health Unit, and in most of York Regional Health Unit except the southern part. Forward sortation areas with medium visible minority population proportion clustered in central parts of the City of Toronto Health Unit, and western parts of Peel Regional Health Unit.

Visually inspecting a normal quantile residual plot, we concluded that the FSA-level Anscombe residuals from the final regression model were normally distributed, indicating the model fit the data. Figure 3.9 illustrates the spatial distribution of the residuals; visually, no clustering of over- or under-predicted values was evident. The global Moran's I statistic of the Anscombe residuals was not significant [I = -0.007, P = 0.98] indicating that no significant spatial autocorrelation of the residuals was present.
after accounting for the SES variables in the model. One outlier was identified; however, we kept it in the model because re-running the model without that observation did not modify any of the variable coefficients.

**Discussion**

Our study used a spatial scan statistic to identify high rate clusters of *S*. Enteritidis infections in three PHUs within the GTA in Ontario. Multivariable negative binomial regression was employed to identify associations between *S*. Enteritidis infections and SES variables to better understand the underlying SES risk factors that might contribute to the spatial clustering of *S*. Enteritidis infections. Our unit of analysis was FSAs from these PHUs, where the *S*. Enteritidis case data acquired from the Ontario Ministry of Health and Long-Term Care (MOHLTC) integrated Public Health Information System (iPHIS) database, and the SES information obtained from the 2006 Census of Canada, were aggregated, merged, and analyzed. We combined traditional statistical methods with GIS technology (choropleth maps) for the statistically significant high rate cluster and the distribution of statistically significant SES indicators, which was a useful approach that assisted with visualizing FSA-level associations across the GTA.

We found that FSAs with high and FSAs with low average median family income had a higher incidence of *S*. Enteritidis infections compared to FSAs with medium average median family income (the reference category). Increased salmonellosis rates in high income neighbourhoods might be explained by residents’ more frequent international travel, which has been shown to be an important risk factor for salmonellosis [9, 18-20]. Further, an ecological study conducted in the US demonstrated higher salmonellosis rates
in areas with high family income population compared to areas with low family income population [25]. The researchers hypothesized that residents from higher income areas had greater access to healthcare, health-seeking behaviour, and pet ownership, and had eating behaviours that increased resident’s salmonellosis risk. In contrast to our study, the US study had the lowest salmonellosis rates in low median family income areas. This discrepancy might be partly explained by the fact that all Canadian citizens, permanent residents, and landed immigrants have access to provincially-funded healthcare, which likely eliminates or reduces bias related to healthcare accessibility, compared to the US, where a large proportion of the population does not have access to healthcare [44]. Moreover, the US study combined all Salmonella serotypes. The pathogenicity and source attribution of Salmonella serotypes can differ [20, 45-46], which might have led to non serotype-specific associations in the US study that are not representative of S. Enteritidis infections. Additionally, the higher rates of infection from low income areas in our study might be explained by poorer microbial quality of foods consumed [47], or by more frequent retail food safety violations [48-49] that could increase residents’ risk of foodborne diseases, including S. Enteritidis infections.

We found a positive curvilinear association between the FSA-level average number of children at home per census family and the SIR of S. Enteritidis infections. This finding could be explained by childrens’ higher susceptibility to salmonellosis, which has been demonstrated by individual-level studies [29, 50-51]. Risk factors associated with S. Enteritidis infections in children include international travel [52-53], riding in shopping carts and exposure to raw meat and poultry products [54], person-to-person transmission in daycare centres and in private homes [55], and contact with reptiles [53,56-57] and
cats [56]. In addition, an ecological study conducted in the US identified a positive association between salmonellosis and counties with a high proportion of young children, which appears to be a consistent risk factor for all Salmonella serotypes [28]. This finding might be explained by the higher severity of salmonellosis in this age group, and consequently an increased likelihood of visiting a physician and getting tested and reported [28]. However, we acknowledge that the average number of children at home per census family and the average number of persons per census family could be substitutes for each other due to their high correlation, which might suggest that home density and eating habits of larger families might have a role in the positive association with the SIR of S. Enteritidis infections. Future individual-level research is needed in Ontario to elucidate our finding.

Forward sortation areas with high visible minority population proportion had a lower SIR of S. Enteritidis infections compared to FSAs with medium visible minority population proportion. This finding is in contrast with a previous US ecological study, in which researchers demonstrated increased salmonellosis rates in areas with higher black, Hispanic, or Latino populations compared to areas with a predominantly Caucasian population [28]. However, comparing non-Caucasian populations between Canada and the US should be done with caution, because in the GTA, a high proportion of the non-Caucasian population is South Asian or Chinese [58]. Of note, in our study no significant difference was found between the low and high visible minority population proportion FSAs, which warrants further individual-level studies to better understand the effect of ethnicity on the incidence of S. Enteritidis infections.
A single spatial cluster of higher than expected incidence rates of *S*. Enteritidis infections was identified in the south-central area (downtown) of the City of Toronto Health Unit that included nine neighbouring FSAs. The majority of these FSAs had SES characteristics that were positively associated with the SIR of *S*. Enteritidis infections in the regression model (low average median family income, medium visible minority population proportion), which is consistent with these SES indicators playing a significant role in the clustering of *S*. Enteritidis infections. However, many of the FSAs in this cluster had one SES characteristic that was inconsistent with the regression model (low average number of children at home per census family), which warrants further spatial epidemiological assessment. Applying GIS technology to visualize the FSA-level spatial cluster of high incidence rates of *S*. Enteritidis infections and the FSA-level distribution of significant SES variables was an useful technique to highlight areas where prevention and control programs should be targeted, and where future studies should be conducted to understand the underlying individual-level risk factors.

First- and second-order spatial effects must be considered when analyzing spatial epidemiological data [59]. First-order effects in our analysis could be defined as variation in the mean value of the *S*. Enteritidis infection rate in the study area (i.e. a global or large-scale trend). We addressed these effects by using a multivariable negative binomial regression model. Second-order effects in our study could be defined as spatial dependence in the *S*. Enteritidis infection rates (i.e. local or small-scale effects). It could be expected that neighbouring FSAs have more common infection rates and SES indicator features than distant FSAs. We assessed the presence of second-order effects by evaluating the spatial autocorrelation of FSA-level Anscombe residuals using the global
Morans’s I statistic. The statistic was not significant indicating that the SES variables explained the variation in S. Enteritidis infection rates across the study area well and no second-order effects remained once these fixed effects were accounted for in our model.

Before generalizing our findings, a few limitations need to be considered. Passive laboratory-based surveillance programs underestimate the true burden of enteric infections. In Canada, it was estimated that for every reported salmonellosis case there were 26.1 cases in the general population that remained unreported [60]. There might also be geographic variations in under-reporting of S. Enteritidis infections due to differences in utilization of health care providers or the sensitivity of testing methods used at different laboratories [61-62]. However, in our study, these factors might not be as influential because our study area enclosed three neighbouring PHUs with similar healthcare providers.

Another possible limitation is that ecological studies only consider variation between groups and not within groups [63], making them unsuitable to relate population level risk factors to the individual level. However, ecological studies can be considered as a cost-effective and useful alternative to individual-level studies, where regional differences in SES risk factors can be evaluated and identified, which can be used by public health authorities to further assess these risk factors and target prevention and control programs [25, 28]. It is also recognized that analyzing associations among S. Enteritidis infections and SES variables at higher geographic levels (e.g. health unit, province) might have given different results [64]. However, it is more accurate to analyze the data at a lower
Our main objective was to identify area-level associations between *S. Enteritidis* infection rates and SES indicators. Several variables that we investigated were based on census families (average number of children per census family, average number of persons per census family, average median family income); consequently, for these three SES indicators, single person homes were not investigated in our study, which could limit the external validity of our estimates. However, within the study area, the proportion of single person households was considerably lower than census family households [33].

Finally, we analyzed *S. Enteritidis* case data from 2007 to 2009; however, the only available socioeconomic data were from the 2006 Census of Canada. Population changes over the study period could have introduced bias into our study. However, it is likely that SES indicators and population sizes did not change drastically over this relatively short time period. We excluded FSAs with a population of less than 500 residents; however, these areas accounted for less than one percent of the total population, and therefore it is unlikely that these exclusions affected our results.

**Conclusions**

Our study demonstrated the usefulness of combining GIS technology and conventional and spatial statistical methods to identify high rate clusters and to analyze area-level associations between *S. Enteritidis* infection rates and SES variables in three PHUs within the GTA by using an ecological study approach. We found a higher incidence rate of *S. Enteritidis* infections in FSAs with low and high average median family income.
compared to FSAs with medium average median family income that might be explained by the poorer quality of food consumed in the low income areas, in differences in food preparation and consumption practices among various areas, and in high income area residents’ more frequent international travel that increases their risk of salmonellosis. A positive curvilinear relationship was observed between the FSA-level SIR of \textit{S. Enteritidis} infections and average number of children at home per census family that might be attributable to children’s greater susceptibility to enteric infections. We did not detect any significant spatial dependency of residuals indicating that our fixed effects model explained the spatial dependency well.

The high rate cluster of \textit{S. Enteritidis} infections detected by the scan statistic contained mainly FSAs in which the majority of the SES variables identified in the regression model were positively associated with \textit{S. Enteritidis} infections, suggesting that these SES indicators (low average median family income, medium visible minority population proportion) significantly contributed to the spatial clustering of \textit{S. Enteritidis} infections.

These findings will aid public health policy makers and practitioners to further evaluate individual-level SES indicators. Areas with low and high average median family income, medium proportion of visible minority population, and high average number of children at home per census family should be targeted when designing disease control and prevention programs within these PHUs. Further studies are needed to identify novel individual-level risk factors and spatio-temporal trends of \textit{S. Enteritidis} infections across the GTA.
List of Abbreviations

CI – confidence interval

FSA - forward sortation area

GIS - geographic information system

GTA - Greater Toronto Area

iPHIS - integrated Public Health Information System

IRR – incidence rate ratio

LR – likelihood ratio

MOHLTC - Ontario Ministry of Health and Long-Term Care

PHU – Public Health Unit

S. Enteritidis - Salmonella enterica subspecies enterica serotype Enteritidis

SES – socioeconomic status

SIR - standardized incidence rate

US – United States of America

Acknowledgements

The authors acknowledge the MOHLTC for providing the data. We thank the staff of all PHUs and public health laboratories that tested samples, followed up with cases, and
entered information into the public health surveillance database. We also acknowledge
the Data Resource Centre at the University of Guelph library for their GIS and census
support. The views expressed in this study are the views of the authors and do not
necessarily reflect those of the MOHLTC.
References


11. Kuehn BM: Salmonella cases traced to egg producers: findings trigger recall of more than 500 million eggs. JAMA 2010, 304(12):1316.

12. Doorduyn Y, Van Den Brandhof WE, Van Duynhoven YT, Wannet WJ, Van Pelt W: Risk factors for Salmonella Enteritidis and Typhimurium (DT104 and


[http://www.health.gov.on.ca/english/providers/program/pubhealth/oph_standards/ophs/infdispro.html]
31. Health Protection and Promotion Act, R.S.O. 1990, c. H.7
   [http://www.elaws.gov.on.ca/html/statutes/english/elaws_statutes_90h07_e.htm]
   World Health Organization Collaborating Centre for Reference and Research on
   Salmonella. Pasteur Institute, Paris, France. 1997
34. Jenks, G F. "The Data Model Concept in Statistical Mapping", International
35. Kulldorff M and Information Management Services, Inc. SaTScanTM v9.1.1:
   Software for the spatial and space-time scan statistics. http://www.satscan.org
38. Dwass M: Modified randomization tests for nonparametric hypotheses. Ann
   Math Statist 1957, 28:181-187
   Wiley and Sons Ltd; 2006.
48. Signs RJ, Darcey VL, Carney TA, Evans AA, Quinlan JJ: Retail food safety risks for populations of different races, ethnicities, and income levels. J Food Prot 2011, 74(10):1717-23
49. Darcey VL, Quinlan JJ: Use of geographic information systems technology to track critical health code violations in retail facilities available to populations


Canada: who is counted in provincial communicable disease statistics?


Table 3.1. Spatial cluster of high *Salmonella* Enteritidis infection rates in the Greater Toronto Area, Ontario, Canada *a)*.

<table>
<thead>
<tr>
<th>Location Ids (Forward Sortation Areas)</th>
<th>Observed Cases</th>
<th>Expected Cases</th>
<th>Observed / Expected</th>
<th>Relative Risk</th>
<th>Log Likelihood</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>M5T, M5G, M5S, M5B, M6J, M5V, M5C, M4Y, M5E</td>
<td>40</td>
<td>21.56</td>
<td>1.85</td>
<td>1.90</td>
<td>6.49</td>
<td>0.050</td>
</tr>
</tbody>
</table>

*a) Results of a spatial discrete Poisson model using the SatScan™ software. A circular scanning window containing up to 50% of the population at risk was used. Age groups (0-9, 10-24, 25-34, 35-49, and 50+ years) and sex were used as covariates to account for confounding. Significant at p ≤ 0.05
Table 3.2. Results of univariable negative binomial regression models (n = 846 cases from 153 FSAs).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>IRR (95% CI)</th>
<th>P - value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of children at home per census family</td>
<td>Linear (X)</td>
<td>1.03 (0.71, 1.50)</td>
<td>0.881</td>
</tr>
<tr>
<td></td>
<td>Quadratic (X^2)</td>
<td>4.45 (1.81, 10.92)</td>
<td>0.001</td>
</tr>
<tr>
<td>Average number of persons per census family</td>
<td>Linear (X)</td>
<td>0.89 (0.64, 1.25)</td>
<td>0.515</td>
</tr>
<tr>
<td></td>
<td>Quadratic (X^2)</td>
<td>4.47 (1.82, 10.98)</td>
<td>0.001</td>
</tr>
<tr>
<td>Average number of rooms per home</td>
<td>Linear (X)</td>
<td>0.97 (0.91, 1.04)</td>
<td>0.398</td>
</tr>
<tr>
<td></td>
<td>Quadratic (X^2)</td>
<td>1.08 (1.03, 1.13)</td>
<td>0.002</td>
</tr>
<tr>
<td>Immigrant population proportion</td>
<td>Linear (X)</td>
<td>0.99 (0.99, 1.00)</td>
<td>0.284</td>
</tr>
<tr>
<td></td>
<td>Quadratic (X^2)</td>
<td>0.99 (0.99, 0.99)</td>
<td>0.014</td>
</tr>
<tr>
<td>Average median family income</td>
<td>Categorical</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low (&lt; CAD d) 65,000</td>
<td>1.30 (1.08, 1.57)</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>Medium (CAD 65,000 - 85,000)</td>
<td>Reference</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>High (&gt; CAD 85,000)</td>
<td>1.31 (1.08, 1.58)</td>
<td>0.006</td>
</tr>
</tbody>
</table>
Table 3.2. Continued.

<table>
<thead>
<tr>
<th>Variable a)</th>
<th>Type</th>
<th>IRR c) (95% CI)</th>
<th>P - value</th>
<th>0.446</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of university graduates of persons between 25 to 64 years of age</td>
<td>Categorical</td>
<td>Low (10.4 - 28.6)</td>
<td>0.93 (0.77, 1.12)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium (28.7 – 40.0)</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>High (40.4 – 72.5)</td>
<td>0.95 (0.78, 1.16)</td>
<td>0.640</td>
</tr>
<tr>
<td>Unemployment rate of persons 15 years of age or older</td>
<td>Categorical</td>
<td>Low (0 - 6.0)</td>
<td>1.06 (0.87, 1.29)</td>
<td>0.544</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium (6.1 - 7.5)</td>
<td>Reference</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High (7.6 - 11.0)</td>
<td>1.08 (0.89, 1.32)</td>
<td>0.423</td>
</tr>
<tr>
<td>Visible minority population proportion</td>
<td>Categorical</td>
<td>Low (2.1 - 29.4)</td>
<td>0.78 (0.64, 0.94)</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium (29.5 - 51.8)</td>
<td>Reference</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High (51.9 - 93.4)</td>
<td>0.77 (0.64, 0.92)</td>
<td>0.004</td>
</tr>
</tbody>
</table>

a) Dependent variable: Number of *Salmonella* Enteritidis infections by forward sortation area (FSA). Offset: natural log-transformed FSA-based expected number of cases. b) Each linear (continuous) variable was centered by substracting the mean value from each value. A quadratic term for the centered linear variable was introduced and kept in the model if it was significant. If the linear and quadratic terms were not statistically significant, the variable was categorized into three equal groups (low, medium, and high). c) IRR: incidence rate ratio; CI: confidence interval. d) CAD: Canadian Dollar. Significant at p ≤ 0.05.
Table 3.3. Results of the final multivariable negative binomial regression model (n = 846 cases from 153 FSAs).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Estimate (95% CI)</th>
<th>IRR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of children at home per census family</td>
<td>Linear (X) b)</td>
<td>0.22 (-0.20, 0.64)</td>
<td>1.24 (0.82, 1.89)</td>
<td>0.313</td>
</tr>
<tr>
<td></td>
<td>Quadratic (X^2)</td>
<td>1.69 (0.84, 2.53)</td>
<td>5.40 (2.32, 12.54)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Average median family income</td>
<td>Categorical Low (&lt; CAD d) 65,000)</td>
<td>0.29 (0.11, 0.47)</td>
<td>1.34 (1.12, 1.61)</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Medium (CAD 65,000 - 85,000)</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High (&gt; CAD 85,000)</td>
<td>0.22 (0.03, 0.40)</td>
<td>1.24 (1.03, 1.49)</td>
<td>0.020</td>
</tr>
<tr>
<td>Visible minority population proportion</td>
<td>Categorical Low (2.1 - 29.4)</td>
<td>-0.16 (-0.35, 0.036)</td>
<td>0.85 (0.70, 1.04)</td>
<td>0.110</td>
</tr>
<tr>
<td></td>
<td>Medium (29.5 - 51.8)</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High (51.9 - 93.4)</td>
<td>-0.27 (-0.44, -0.10)</td>
<td>0.76 (0.64, 0.91)</td>
<td>0.002</td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td>-0.12 (-0.29, 0.05)</td>
<td></td>
<td>0.178</td>
</tr>
</tbody>
</table>

a) Dependent variable: Number of *Salmonella* Enteritidis infections by forward sortation area (FSA). Offset: natural log-transformed FSA-based expected number of cases.
b) Each linear (continuous) variable was centered by subtracting the mean value from each value. A quadratic term for the centered linear variable was introduced and kept in the model if it was significant. If the linear and quadratic terms were not statistically significant, the variable was categorized into three equal groups (low, medium, and high).
c) IRR: incidence rate ratio; CI: confidence interval.
d) CAD: Canadian Dollar. Significant at P \(\leq 0.05\)
Figure 3.1. Location of the study area within the Greater Toronto Area in Ontario, Canada.
Figure 3.2. Unadjusted *Salmonella* Enteritidis incidence per 100,000 person-years, by forward sortation area (FSA)\(^a\):

\(^a\) 846 cases from 153 FSAs from three public health units (City of Toronto Health Unit, Peel Regional Health Unit, and York Regional Health Unit) within the Greater Toronto Area, Ontario, Canada were reported in the integrated Public Health Information System from January 1, 2007 through December 31, 2009 and included in the study. HR-Health Region.
Figure 3.3. Relative risk\(^a\) of *Salmonella* Enteritidis infections by forward sortation area.

\(^a\) The ratio of the observed number of cases to the age-and-sex-adjusted expected number of cases. Values < 1 indicate lower incidence than expected, whereas values > 1 indicate higher incidence than expected.
Figure 3.4. Spatial cluster \(^a\) of high *Salmonella* Enteritidis infection rates.

\(^a\) Results of a spatial discrete Poisson model using the SatScanTM software. Study period: January 1, 2007 to December 31, 2009. A circular scanning window containing up to 50\% of the population at risk was used. Age groups (0-9, 10-24, 25-34, 35-49, and 50+ years) and sex were included as covariates to account for confounding. Significant at \(p \leq 0.05\)
Figure 3.5. Association between average number of children at home per census family and *Salmonella* Enteritidis infections

The graph displays the effect of a 0.1 unit change in the average number of children at home per census family on the predicted log of the standardized incidence rate (SIR) of *Salmonella* Enteritidis (*S*. Enteritidis) infections, keeping all other variables in the final negative binomial regression model at their referent level. The final model included the linear and quadratic terms of average number of children at home per census family, average median family income, and visible minority population proportion. Dependent variable: Number of *S*. Enteritidis infections by forward sortation area (FSA). Offset: natural log-transformed FSA-based expected number of cases.
Figure 3.6. Distribution of the average number of children at home per census family by forward sortation area (FSA).
Figure 3.7. Distribution of the average median family income by forward sortation area (FSA).
Figure 3.8. Distribution of the visible minority population proportion by forward sortation area (FSA).
Figure 3.9. Distribution of forward sortation area (FSA)-level Anscombe residuals from the multivariable negative binomial regression model a).

a) Positive values signify under-prediction whereas negative values signify over-prediction. Dependent variable: Number of *Salmonella* Enteritidis infections aggregated to the Forward Sortation Area (FSA)-level. Offset: natural log-transformed FSA-based expected number of cases. Independent categorical variables: FSA-level average median family income and visible minority population proportion. Independent continuous variable: FSA-level average number of children at home per census family linear and quadratic terms.
CHAPTER FOUR

Area-level global and local clustering of human *Salmonella Enteritidis* infection rates in the city of Toronto, Canada, 2007–2009


Abstract

Background

*Salmonella enterica* serotype Enteritidis (*S. Enteritidis*) remains a major foodborne pathogen in North America yet studies examining the spatial epidemiology of salmonellosis in urban environments are lacking. Our ecological study combined a number of spatial statistical methods with a geographic information system to assess area-level heterogeneity of *S. Enteritidis* infection rates in the city of Toronto.

Methods

Data on *S. Enteritidis* infections between January 1, 2007 and December 31, 2009 were obtained from Ontario’s surveillance system, and were grouped and analyzed at the forward sortation area (FSA)-level. Incidence rates were directly standardized using the FSA-level age- and sex-based standard population. A spatial empirical Bayes method was used to smooth the standardized incidence rates (SIRs). Global clustering of FSAs with high or low non-smoothed SIRs was evaluated using the Getis-Ord G method. Local clustering of FSAs with high, low, or dissimilar non-smoothed SIRs was assessed using the Getis-Ord Gi* and the Local Moran’s I methods.
Results

Spatial heterogeneity of *S.* Enteritidis infection rates was detected across the city of Toronto. The non-smoothed FSA-level SIRs ranged from 0 to 16.9 infections per 100,000 person-years (mean = 6.6), whereas the smoothed SIRs ranged from 2.9 to 11.1 (mean = 6.3). The global Getis-Ord G method showed significant ($p \leq 0.05$) maximum spatial clustering of FSAs with high SIRs at 3.3 km. The local Getis-Ord Gi* method identified eight FSAs with significantly high SIRs and one FSA with a significantly low SIR. The Local Moran’s I method detected five FSAs with significantly high-high SIRs, one FSA with a significantly low-low SIR, and four significant outlier FSAs (one high-low, and three low-high).

Conclusions

*Salmonella* Enteritidis infection rates clustered globally at a small distance band, suggesting clustering of high SIRs in small distinct areas. This finding was supported by the local cluster analyses, where distinct FSAs with high SIRs, mainly in downtown Toronto, were detected. These areas should be targeted with prevention and control programs. We demonstrated the usefulness of combining several spatial statistical techniques with a geographic information system to evaluate spatial processes that influenced *S.* Enteritidis infection rates. Our study methodology could be applied to other foodborne disease surveillance data.

Keywords:

*Salmonella* Enteritidis, spatial, direct standardization, Toronto, Canada, geographic information system, retrospective study, population-based study, foodborne illness.
Background

Salmonellosis continuously poses a significant health burden to human populations globally, affecting annually an estimated 93.8 million persons worldwide [1]. In Canada, an estimated 109,384 non-typhoidal *Salmonella* infections are acquired domestically, of which 80% are considered to be foodborne [2]. Within the last decade, an increase in the number of *Salmonella enterica* serotype Enteritidis (*S.* Enteritidis) infections has been reported in Canada [3], the United States of America [4], and the European Union [5], such that *S.* Enteritidis has become the top serotype among the non-typhoidal salmonellae. *Salmonella* Enteritidis infections in humans have typically been associated with consumption of contaminated chicken products [6-7] and eggs [8-9]. However, salmonellosis has recently been linked to other factors, including international travel [10-11], demographic [12-13] and socioeconomic [14-15] characteristics, and animal contact [7, 16].

Country- or region-level studies have used various spatial epidemiological methods to identify clustering of health conditions, including notifiable gastrointestinal illness [17], giardiasis [18], campylobacteriosis [19-20], influenza B [21], *Escherichia coli* O157 [22-23], traumatic brain injury [24], stroke [25], and myocardial infarction [25]. Moreover, city-level studies have evaluated spatial differences in neighbourhood-level infection rates of rotavirus in Berlin, Germany [26], pandemic influenza A in Hong Kong [27], and tuberculosis in Linyi City, China [28].

Our study area involved the city of Toronto—the capital of Ontario, Canada located on the shore of Lake Ontario in the southern part of the province [Figure 4.1]. In 2009, an
estimated 2.7 million people lived in the city, accounting for 21% of Ontario’s total population [29]. Toronto’s forward sortation areas (FSAs; see Methods section 1) have diverse age- and sex-based populations that can affect area-level infection rates, due to younger and older residents’ higher salmonellosis rates [13, 16, 30]. Standardization of area-level infection rates based on the age and sex distribution of the population has been recommended to overcome this problem [31]. Moreover, infection rates in small population areas can become unstable and unreliable. The spatial empirical Bayes (SEB) smoothing method has been proposed to reduce the random variation of infection rates linked with these areas [31-32].

Despite the abundance of research studies that have assessed large scale (country- or region-level) risk factors for various foodborne infections, few studies have assessed small scale (city- or FSA-level) spatial clustering of salmonellosis rates. Small area studies in urban environments are valuable for understanding the geographical distribution of infections, which can assist in the design of local prevention and control programs [21, 26]. Our retrospective, population-based, ecological study used a systematic approach that combined spatial exploratory and statistical methods with a geographic information system (GIS) [Figure 4.2], to evaluate the spatial heterogeneity of S. Enteritidis infection rates across the city of Toronto.

Methods

Study design and data sources

Forward sortation areas are well-delimited areas signified by the first three characters of the postal code; they are established by the Canada Post Corporation based on the mail
distribution zones of postal facilities. Forward sortation area-level population estimates and FSA cartographical boundary files were acquired from the 2006 Census of Canada [29, 33].

In Ontario, salmonellosis is a reportable disease under provincial legislation [34]. A diagnosis of salmonellosis is made after isolation of *Salmonella* spp. (excluding *Salmonella* Typhi or Paratyphi) from an appropriate clinical sample (the majority are stool samples) by public health, hospital, or private laboratory staff [35]. All isolates are sent to the Public Health Ontario Laboratories for confirmation and serotyping using the Kauffmann-White scheme [36]. Salmonellosis cases must be followed up by local public health unit staff, and investigation findings must be reported to the Ontario Ministry of Health and Long-Term Care (MOHLTC) through the integrated Public Health Information System (iPHIS). We obtained case information from all reported *S.* Enteritidis infections from the city of Toronto between January 1, 2007 and December 31, 2009 that were captured within iPHIS.

**Statistical analysis**

Spatial heterogeneity of *S.* Enteritidis infection rates was assessed by following several analytical steps, which are outlined in Figure 4.2, and described in detail below.

**Exploratory spatial analysis**

In order to obtain stable infection rate estimates, we excluded FSAs with less than 500 residents. Annual standardized incidence rates (SIRs) were calculated for each FSA using direct standardization [37-38] in STATA Intercooled 10.1 statistical software (Stata Corporation, College Station, TX, USA). The annual SIR was estimated by calculating
the observed rate for each age-sex category within each FSA, and multiplying it by the age-sex population numbers, which were obtained from the 2006 Census of Canada [29]. To account for unstable SIRs of areas with small populations [39], we smoothed the rates using the SEB method [40] with 2nd order queen contiguity weights [41] in GeoDa version 095i software (Spatial Analysis Lab, University of Illinois Urbana-Champaign, IL, USA). The non-smoothed and smoothed annual SIRs were presented as the number of S. Enteritidis infections per 100,000 person-years per FSA, and were visualized using choropleth maps with ArcGIS 10.1 (ESRI Inc., Redlands, CA, USA) using Jenk’s categorization [42] to define the critical intervals for mapping. Jenk’s optimization classification method arranges areas into categories based on the distribution of the SIRs, using a procedure that decreases variance within, and maximizes variance between SIRs.

**Spatial statistics**

Each FSA was represented by a polygon, its centroid, and its distinct non-smoothed SIR. The *Spatial Statistics Tool* in ArcGIS 10.1 was used to identify global and local spatial clusters. Euclidean distance bands were used to measure distances from each FSA’s centroid to neighbouring FSAs’ centroids (see Methods subsection 2.2.1). To avoid the omission of local factors by imposing sharp neighbourhood boundaries, the “zone of indifference” conceptualization parameter was chosen for our global and local cluster analyses. Using this parameter, the target FSA and all neighbouring FSAs within a specified distance band are given a maximum weight; once this critical distance is exceeded, neighbouring FSAs are assigned smaller and smaller weights as the distance from the target FSA increases [43-44]. The null hypothesis for both global and local cluster analyses is that there is complete spatial randomness (i.e. FSAs with high or low
SIRs are randomly distributed across the study area). The null hypothesis is rejected when FSAs with high or low SIRs are more spatially clustered than would be expected if the underlying spatial processes were truly random. When the null hypothesis is rejected, a Z-score and a p-value are given for the identified cluster [43-44].

**Global clustering (Getis-Ord General G)**

Global spatial clustering of FSAs with high or low SIRs across Toronto was evaluated using the Getis-Ord General G statistic [44]. Several Euclidean distances (3.3 to 5.9 km, with 100 m increments) were included in the model to identify the distance bands with the highest and lowest statistically significant Z-scores. Distance bands that required each FSA to have at least one neighbour were selected. A large, positive Z-score (values $\geq 1.96$) and a significant p-value (p $\leq 0.05$) signified that FSAs with high SIRs were clustered in the study area, whereas a large, negative Z-score (values $\leq -1.96$) and a significant p-value signified that FSAs with low SIRs were clustered in the study area [44].

**Local clustering**

For the local cluster analyses, we used the distance band identified at the global clustering step that showed maximum spatial clustering of FSAs with high SIRs (see Results subsection 3.1).

**Hot Spot Analysis (Getis-Ord Gi*)**

Local spatial clusters of FSAs with high or low SIRs were examined using the Getis-Ord Gi* statistic [44-45]. The statistic compares the local sum of SIRs (the sum of the SIR of
the targeted FSA and its neighbouring FSAs) to the sum of SIRs of all FSAs within the study area. A statistically significant large, positive Z-score signifies a local high-rate cluster \((\text{hot spot})\). Hot spots are detected when FSAs with high rates are surrounded by FSAs with high rates; the observed local sum of SIRs is higher than the expected local sum and the difference is too large to be the result of chance alone. Similarly, a statistically significant large, negative Z-score signifies a local low-rate cluster \((\text{cold spot})\), where FSAs with low rates are surrounded by FSAs with low rates [42-45]. Statistically significant hot and cold spots were visualized using a map with FSA boundaries.

**Cluster and Outlier Analysis (Anselin Local Moran's I)**

We also used the Local Moran’s I statistic to identify local spatial clusters of FSA-level \(S.\) Enteritidis SIRs during the study period [46]. The statistic identifies hot spots \((\text{high-high})\), cold spots \((\text{low-low})\), and spatial outliers \((\text{high-low} \text{ and low-high})\). A positive Local Moran’s I value indicates that the target FSA is surrounded by FSAs with similar rates \((\text{high-high}: \text{FSA with a high rate surrounded by FSAs with high rates}; \text{low-low}: \text{FSA with a low rate surrounded by FSAs with low rates})\). A negative Local Moran’s I value indicates that the target FSA is surrounded by FSAs with dissimilar rates \((\text{high-low}: \text{FSA with a high rate surrounded by FSAs with low rates}; \text{low-high}: \text{FSA with a low rate surrounded by FSAs with high rates})\) [46]. The designation of FSAs to these four classes depends on the results of a statistical test. This test performs random comparisons among the target FSA’s and its neighbours Moran’s I values to all FSAs’ Moran’s I values within the study area, and compares the observed Moran’s I value to the value corresponding to the random permutations (expected Moran’s I value) [46]. If the test is
significant \( (p \leq 0.05) \), the observed Moran’s I value is significantly larger (or smaller in the case of a negative relationship) than the expected Moran’s I value. If the test is not significant, the FSA remains in a neutral class (no spatial dependence) [46]. Statistically significant high-high, low-low, and outlier local clusters were visualized using a map with FSA boundaries.

**Ethics Review**

The University of Guelph Ethics Review Board was consulted since we used surveillance data for a reportable disease of humans; however, ethics approval was not required because our data did not contain any personal or health information that could be connected back to the original identifiers.

**Results**

**Descriptive statistics**

Based on the 2006 Census, there were a total of 102 FSAs in the city of Toronto; the FSA-level population size ranged from 5 to 65,125 persons. Ninety-five FSAs met the inclusion criteria, for which the population size ranged from 2,165 to 65,125 persons (mean = 26,345). A total of 495 laboratory confirmed \( S. \) Enteritidis infections were identified in the MOHLTC’s iPHIS database during the study period (165 cases in 2007, 168 in 2008, and 162 in 2009). In total, 22 cases (4.4%) were excluded because of missing FSA data (14 cases in 2007, 4 in 2008, and 4 in 2009). Thus, there were 473 cases (151 in 2007, 164 in 2008, and 158 in 2009) available for analysis. No outbreaks were declared by the MOHLTC during the study period.
Exploratory spatial analysis

Non-smoothed and smoothed standardized incidence rates

Figure 4.3 illustrates the non-smoothed and smoothed annual SIRs of S. Enteritidis infections per FSA in Toronto. The non-smoothed FSA-level SIRs ranged from 0 to 16.9 infections per 100,000 person-years (mean = 6.6). The smoothed SIRs ranged from 2.9 to 11.1 (mean = 6.3).

Spatial statistics

Global clustering (Getis-Ord General G)

The Getis-Ord General G statistic results are shown in Figures 4.4 and 4.5. Statistically significant positive Z-scores (1.99 - 2.34) were observed between 3.3 and 4.7 km. The highest statistically significant positive Z-score was observed at 3.3 km (Z = 2.34, p = 0.019), signifying maximum spatial clustering of FSAs with high SIRs at this distance band [Figure 4.5]. There were no statistically significant negative Z-scores.

Local clustering

Hot Spot Analysis (Getis-Ord Gi*)

Eight FSAs with high SIRs (hot spots) (M5C, M5E, M5G, M5M, M5R, M5S, M5T, M9R) and one FSA with a low SIR (cold spot) (M3H) were detected using the Getis-Ord Gi* method (Table 4.1, Figure 4.6, Appendix C - Legend 4.1). The majority of hot spots (6 of 8) were located in south-central (i.e. downtown) Toronto.
Cluster and Outlier Analysis (Anselin Local Moran's I)

Five FSAs with high-high SIRs (M4Y, M5E, M5G, M5M, M9R), one FSA with a low-low SIR (M3H), and four outlier FSAs (one high-low (M3M) and three low-high (M4G, M5C, M5R)) were identified using the Local Moran’s I method (Table 4.1, Figure 4.7, Appendix C - Legend 4.1). Three FSAs with high-high SIRs were detected in downtown Toronto.

Discussion

*Salmonella* Enteritidis infection rates clustered globally and locally in the city of Toronto. The small distance band at which high *S.* Enteritidis infection rates clustered globally suggests that infection rates were localized to small distinct areas. This finding was subsequently supported by the local cluster analyses, where distinct FSAs, mainly in downtown Toronto, were identified as areas with significantly high SIRs. The two local cluster detection methods (Getis-Ord Gi* and Local Moran’s I) identified a number of the same clusters, suggesting consistency between these methods, and indicating the robustness of our study results.

We assessed the area-level spatial heterogeneity of *S.* Enteritidis infection rates across the city of Toronto by combining spatial exploratory and spatial statistical methods with GIS. A systematic approach was used, in which analytical steps succeeded each other, starting from more general to more specific stages that increased our study’s specificity. Each step provided additional information to enhance our understanding of the spatial epidemiology of *S.* Enteritidis infection rates in Toronto. However, these steps were
sometimes connected and difficult to delineate; consequently, a holistic approach that considers the results of all steps should be followed when interpreting our findings.

The variability of small scale infection rate estimates was accounted for by using the SEB smoothing method. This method reduces the variation of infection rate estimates of areas with unbalanced rates, by shrinking the less stable estimates toward the local mean if local clustering of high-rate areas are detected, and toward the global mean if no local clustering is present [39]. The major advantage of smoothing is that it focuses attention on the overall spatial disease trends, which increases the ability to identify areas with high or low rates. However, as noted with our data, areas can be misclassified by the smoothing method. For example, one high-rate area (M9R) that was evident on the non-smoothed SIR map and subsequently detected by both local cluster detection methods, was hidden by the smoothing process. Fewer FSAs with high rates were identified using the smoothed SIRs compared to the non-smoothed SIRs; nonetheless, in the central and south-central parts of the city, both methods identified many of the same high-rate FSAs. The SEB smoothing method reduced the highest non-smoothed SIR by 5.8 units, indicating that there were FSAs with unstable SIR estimates.

When analyzing small scale area-level data, the spatial estimates can become unbalanced at the study area limits where FSAs do not have neighbours. Moreover, because FSA boundaries are arbitrary delimitations based on the mail distribution zones of postal facilities, they might not always delineate areas based on their spatial characteristics. To account for potential “edge” and “zoning” effects, we used the “zone of indifference” conceptualization parameter, which does not force sharp boundaries on neighbouring
FSA’s spatial characteristics nor limit the number of neighbours [43]. This conceptualization parameter considers every FSA to be a neighbour of every other FSA, yet it assigns a maximum weight to areas within a pre-determined distance band, and reduces the intensity of spatial relationships once this distance is passed.

The Getis-Ord G method was valuable for identifying the extent of global clustering. Although Toronto is a large city (area of approximately 630 km²), maximum spatial clustering of FSAs with high SIRs was detected at 3.3 km, which suggests that clustering of S. Enteritidis infections was localized to relatively small areas within the city. This result might suggest that local clusters were driven by small outbreaks (e.g. exposures in homes, local daycares, or restaurants) and not by widespread contamination of food or water supplies.

The Getis-Ord Gi* and Local Moran’s I methods identified several of the same clusters. Specifically, four hot spots (M5E, M5G, M5M, M9R) and one cold spot (M3H) were identified by both methods, highlighting the robustness of our study findings. Moreover, our study results are generally in agreement with our previous study [15], in which we evaluated area-level spatial clustering of S. Enteritidis infection rates within three public health units (the city of Toronto, Peel Region, and York Region) in the Greater Toronto Area using a spatial discrete Poisson model within a spatial scan statistic. In that study, a single cluster of significantly higher than expected infection rates located in the south-central part (downtown) of the city of Toronto Health Unit was identified, which included nine neighbouring FSAs (M4Y, M5B, M5C, M5E, M5G, M5S, M5T, M5V, M6J). By comparison, in the current study, the Getis-Ord Gi* method detected five hot spots (M5C,
M5E, M5G, M5S, M5T), and the Local Moran’s I method detected three high-high clusters (M4Y, M5E, M5G) and one low-high cluster (M5C) in downtown Toronto. Taken together, these findings show that these spatial methods could be used in real-time for foodborne disease surveillance data analysis or retrospectively for prevention and control program planning.

However, it is important to understand the specifics of each method to avoid making misleading conclusions. The Getis-Ord Gi* method is ideal when there is an assumption that infection rates cluster within the study area, when investigators are only interested in detecting local high- or low-rate clusters, and when there are a limited number of neighbouring areas with dissimilar rates [44-45]. Because the Getis-Ord Gi* statistic includes the target FSA’s rate when calculating the local sum of rates, it is not as useful in study areas in which there are several small areas with dissimilar rates. For example, if the target FSA has a sufficiently high rate, it can be designated as a hot spot even though it is surrounded by FSAs with low rates. Likewise, some of its neighbouring low-rate FSAs will also be identified as hot spots; or high- or low-rate FSAs will not be identified at all. These issues explain why two of the hot spots (M5C, M5R) identified by the Getis-Ord Gi* method were identified as low-high clusters by the Local Moran’s I method, and why an FSA with a high SIR (M3M) and an FSA with a low SIR (M4G) were undetected by the Getis-Ord Gi* method yet were identified as a high-low and a low-high cluster, respectively, by the Local Moran’s I method. The latter method identifies local areas with dissimilar rates and excludes these from the local high- or low-rate clusters, thus preventing misclassification of FSAs in study areas with relatively high numbers of dissimilar neighbouring areas.
This study was a hypothesis-generating study and did not aim to identify individual-level risk factors that might influence the spatial heterogeneity of S. Enteritidis infection rates. However, both demographic and socioeconomic characteristics have been identified as important risk factors for salmonellosis, some of which include eating behaviours (e.g. frequency of eating outside the home) [6], international travel patterns of local residents [10-11], ethnicity (e.g. proportion of the population that is non-Caucasian) [12], and the proportion of the population with a high income [12-13, 15]. Moreover, local clusters of high S. Enteritidis infection rates could be explained by differences in environmental contamination of food products in local retail facilities and restaurants [47], variations in microbial quality of food consumed [48], or food safety practices followed by local residents [49]. Future hypothesis-testing studies should be conducted in high-rate FSAs to identify area- and individual-level environmental, behavioural, and socioeconomic risk factors that impact S. Enteritidis infection rates. Areas identified as spatial outliers should be investigated using case-control studies (e.g. high-rate areas designated as cases and low-rate areas designated as controls) to identify risk factors that contribute to infection rate increases.

As with every population-based ecological study, our research has limitations, which should be considered when interpreting our results. We recognize that analysis at a different scale might offer different results (the “modifiable areal unit problem”) [50-51]. However, previous studies have demonstrated that examining infection rates at small scales reduces ecological bias, and gives optimal estimates for area-level risk factors for foodborne diseases [51-53]. The “zoning effect” [54] might also have occurred if neighbourhood boundaries did not follow the area’s spatial characteristics. However,
Toronto’s FSAs are of a sufficiently small scale to highlight and delimit neighbourhoods with distinct spatial characteristics, and we also accounted for this issue by using the “zone of indifference” conceptualization parameter. Another limitation of our study is that passive surveillance systems underdiagnose and underreport the true level of infection [2, 55-56]. Population changes might also have occurred during the study period due to movement of residents into and out of the study area. However, this issue should be minor because populations generally do not change considerably in a relatively short time frame. Lastly, exclusion of cases due to missing information might have affected our results. However, 96% of available cases were included in our analysis; therefore, our estimates should be reliable.

**Conclusions**

To the best of our knowledge, this is the first study worldwide that investigated the spatial epidemiology of *S.* Enteritidis infections in an urban setting. *Salmonella* Enteritidis infection rates clustered globally at a small distance band of 3.3 km, suggesting clustering of high rates in small distinct areas. This finding was supported by the local cluster analyses, where distinct FSAs with high rates, mainly in downtown Toronto, were detected. The robustness of our research findings were demonstrated by linking a number of spatial data explorations and statistical methods with GIS. Our study findings will aid public health professionals to design targeted prevention and control programs for *S.* Enteritidis.

**List of Abbreviations**

FSA-forward sortation area
GIS-geographic information system

iPHIS-integrated Public Health Information System

MOHLTC-Ontario Ministry of Health and Long-Term Care

SEB-spatial empirical Bayes

S. Enteritidis-Salmonella enterica serovar Enteritidis

SIR-standardized incidence rate

Competing interests

The authors declare that they have no competing interests.
References


PFGE pattern 8 (National Designation ECXAI.0001) in Alberta as an example.  
Zoonoses Public Health 2013, 60(5):341-8

24. Colantonio A, Moldofsky B, Escobar M, Vernich L, Chipman M, McLellan B:


29. Statistics Canada: Population and dwelling counts, for Canada, provinces and territories, and forward sortation areas as reported by the respondents, 2006 Census. Available at: http://www12.statcan.ca/english/census06/data/popdwell/Table.cfm?T=1201&SR=1&S=0&O=A&RPP=9999&PR=0&CMA=0

30. Chui KK, Webb P, Russell RM, Naumova EN: Geographic variations and temporal trends of Salmonella-associated hospitalization in the U.S. elderly,


47. Signs RJ, Darcey VL, Carney TA, Evans AA, Quinlan JJ: **Retail food safety risks for populations of different races, ethnicities, and income levels.** *J Food Prot* 2011, 74(10):1717-23.


52. Weisent J, Rohrbach B, Dunn JR, Odoi A: **Detection of high risk campylobacteriosis clusters at three geographic levels.** *Geospat Health* 2011, 6(1):65-76.


**Authors' contribution**

CV developed the study design, analyzed the data, created the maps, interpreted the results, wrote the first draft of the manuscript, responded to editorial comments, and prepared the final manuscript for submission.

MTG and DLP were consulted on study design, data analysis, interpretation of results, and reviewed and commented on manuscript drafts.

SAM, FP, and JMS provided advice on the data analysis, interpretation of results, and reviewed and commented on manuscript drafts.

All authors read and approved the final manuscript.

**Acknowledgements**

The authors acknowledge the MOHLTC for providing the data. We thank the staff of all public health units and public health laboratories that tested samples, followed up with cases, and entered information into the public health surveillance database. We also acknowledge the Data Resource Centre at the University of Guelph library for their GIS and census support. The views expressed in this study are the views of the authors and do not necessarily reflect those of the MOHLTC.
Table 4.1. Forward sortation areas identified by different local cluster detection methods

<table>
<thead>
<tr>
<th>Type of clustera</th>
<th>Method</th>
<th>Forward Sortation Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (cold spot)</td>
<td>Getis-Ord Gi*</td>
<td>M3H</td>
</tr>
<tr>
<td>High-high</td>
<td>Local Moran’s I</td>
<td>M4Y, M5E, M5G, M5M, M9R</td>
</tr>
<tr>
<td>Low-low</td>
<td>Local Moran’s I</td>
<td>M3H</td>
</tr>
<tr>
<td>High-low</td>
<td>Local Moran’s I</td>
<td>M3M</td>
</tr>
<tr>
<td>Low-high</td>
<td>Local Moran’s I</td>
<td>M4G, M5C, M5R</td>
</tr>
</tbody>
</table>

a The Getis-Ord Gi* and Local Moran’s I methods identified four identical hot spots (M5E, M5G, M5M, M9R) and one identical cold spot (M3H)
Figures

Figure 4.1. Map of Ontario, Canada highlighting the location of the study area.
Figure 4.2. Flow chart outlining the analytical steps used to evaluate area-level Salmonella Enteritidis infection rates.
Figure 4.3. Distribution of non-smoothed and smoothed *Salmonella Enteritidis* infection rates in Toronto, 2007-2009 (n = 473 cases; n = 95 forward sortation areas)*a,b,c.*

*a* Direct standardization was used to calculate forward sortation area (FSA)-level annual standardized incidence rates (SIRs) of *Salmonella Enteritidis* infections.

*b* Spatial empirical Bayes smoothing method with 2nd order queen contiguity weights in GeoDa software (Spatial Analysis Lab, University of Illinois Urbana-Champaign, IL, USA) was used to smooth the SIRs.

Figure 4.4. Global clusters of areas with high *Salmonella* Enteritidis infection rates in Toronto at different distances \(^a\).

\(^a\)Results of the Getis-Ord G statistic. Large, positive Z-scores (e.g. values \(\geq 1.96\)) indicate global clustering of forward sortation areas with high standardized incidence rates. The zone of indifference conceptualization parameter was used for the analysis. Statistically significant at \(p \leq 0.05\).
Figure 4.5. Maximum spatial clustering of areas with high *Salmonella* Enteritidis infection rates in Toronto at 3.3 kilometers $^a$.

$^a$Result of the Getis-Ord G statistic. A large, positive Z-score (values $\geq 1.96$) indicates global clustering of forward sortation areas with high standardized incidence rates. The zone of indifference conceptualization parameter was used for the analysis. Statistically significant at $p \leq 0.05$. 
Figure 4.6. Local clusters of *Salmonella* Enteritidis infection rates in Toronto identified by the Getis-Ord Gi* statistic\(^a\).

\(^a\)Significant clusters of forward sortation areas (FSAs) with high standardized incidence rates (SIRs) (Z-score ≥ 1.96; \(p \leq 0.05\)). Significant clusters of FSAs with low SIRs (Z-score ≥ - 1.96; \(p \leq 0.05\)). A Euclidean distance band of 3.3 km, and the zone of indifference conceptualization parameter, were used for the analysis.
Figure 4.7. Local clusters of *Salmonella* Enteritidis infection rates in Toronto identified by the Moran’s I statistic\(^a\).

\(^a\) An Euclidean distance band of 3.3 km, and the zone of indifference conceptualization parameter, were used for the analysis. Statistically significant at \(p \leq 0.05\).
CHAPTER FIVE

Spatial-temporal epidemiology of human *Salmonella* Enteritidis infections with major phage types (PTs 1, 4, 5b, 8, 13, and 13a) in Ontario, Canada, 2008-2009.

Abstract

Background

In Ontario and Canada, the incidence of human *Salmonella enterica* serotype Enteritidis (*S.* Enteritidis) infections have increased steadily during the last decade. Our study evaluated the spatial and temporal epidemiology of the major phage types (PTs) of *S.* Enteritidis infections to aid public health practitioners design effective prevention and control programs.

Methods

Data on *S.* Enteritidis infections between January 1, 2008 and December 31, 2009 were obtained from Ontario’s disease surveillance system. *Salmonella* Enteritidis infections with major phage types were classified by their annual health region-level incidence rates (IRs), monthly IRs, clinical symptoms, and exposure settings. A scan statistic was employed to detect retrospective phage type-specific spatial, temporal, and space-time clusters of *S.* Enteritidis infections. Space-time cluster cases’ exposure settings were evaluated to identify common exposures.

Results

1,336 cases were available for analysis. The six most frequently reported *S.* Enteritidis PTs were 8 (n=398), 13a (n=218), 13 (n=198), 1 (n=132), 5b (n=83), and 4 (n=76). Reported rates of *S.* Enteritidis infections with major phage types varied by health region.
and month. International travel and unknown exposure settings were the most frequently reported settings for PT 5b, 4, and 1 cases, whereas unknown exposure setting, private home, food premise, and international travel were the most frequently reported settings for PT 8, 13, and 13a cases.

Diarrhea, abdominal pain, and fever were the most commonly reported clinical symptoms. A number of phage type-specific spatial, temporal, and space-time clusters were identified. Space-time clusters of PTs 1, 4, and 5b occurred mainly during the winter and spring months in the North West, North East, Eastern, Central East, and Central West regions. Space-time clusters of PTs 13 and 13a occurred at different times of the year in the Toronto region. Space-time clusters of PT 8 occurred at different times of the year in the North West and South West regions.

Conclusions

Phage type-specific differences in exposure settings, and spatial-temporal clustering of S. Enteritidis infections were demonstrated that might guide public health surveillance of disease outbreaks. Our study methodology could be applied to other foodborne disease surveillance data to detect retrospective high disease rate clusters, which could aid public health authorities in developing effective prevention and control programs.

Keywords: phage type, surveillance, retrospective, travel, restaurant, scan statistic, Salmonella Enteritidis, Canada

Background

Salmonellosis is a major foodborne bacterial infection that continuously poses a significant human health burden worldwide [1]. In Canada, salmonellosis is the main
cause of hospitalization and death among domestically acquired foodborne infections [2], causing an estimated 87,510 illnesses annually [3]. In the last decade, *Salmonella enterica* serotype Enteritidis (*S. Enteritidis*) became the top serovar among the non-typhoidal salmonellae in Canada [4], the United States of America (US) [5-6], and the European Union [7].

Currently in Canada, the predominant *S. Enteritidis* phage types (PTs) among human cases are PT 8, 13a, 13, 1, 4, and 5b [4]. Between 2006 and 2010, Canadian integrated surveillance systems identified the emergence of PT 13a and an increase in the number of cases of PT 8 [4].

Several research studies conducted in North America have evaluated phage type-specific risk factors for *S. Enteritidis* infections in humans. In Ontario, Canada, researchers demonstrated that cases with PT 8 were more likely to have had contact with dogs compared to cases with other phage types [8]. In British Columbia, Canada, a concurrent increase in the incidence of *S. Enteritidis* infections with PT 8 in humans and the prevalence of PT 8 in poultry was observed between 2007 and 2010 [9]. The researchers demonstrated increased odds of infection with PT 8 in human cases who consumed illegally-sourced ungraded eggs compared to controls [9]. In Alberta, Canada, an outbreak of PT 8, 13, and atypical PTs was linked to the consumption of food products purchased from mobile lunch trucks that were contaminated by illegally-obtained eggs and/or by infected food handlers [10]. In the US, PT 8 cases were more likely to have consumed chicken or be the owner of a lizard than controls, whereas PT 13 cases were more likely to have eaten undercooked eggs in their home than controls [11].
In Ontario and Canada, an increase in the reported number of human *S. Enteritidis* cases was observed during the last decade [12-13]. Current Ontario studies revealed that the majority of *S. Enteritidis* cases with PT 1, 4, or 6a were international travel-related, whereas cases with PT 8, 13, or 13a were mainly acquired domestically [14-15]. These studies provided valuable information on the seasonality and exposure locations of *S. Enteritidis* cases, although they lacked information on cases’ geographical distribution and spatial-temporal clustering. Identifying areas with high rates of reported *S. Enteritidis* cases can be useful for targeting prevention and control programs [12, 16].

There have been a limited number of studies that evaluated foodborne disease surveillance data by incorporating geographical information system (GIS) data, spatial-temporal scan statistic results, exposure setting information, and clinical syndrome history. Scan statistics have been effectively used to evaluate clustering and transmission dynamics of pandemic influenza A (H1N1) in Hong Kong, China [17], to detect *Escherichia coli* O157:H7 outbreaks involving common molecular subtypes in Alberta, Canada [18], to identify the location of high and low rate areas of campylobacteriosis incidence in Manitoba, Canada [19], to identify high incidence clusters of tuberculosis in Linyi City, China [20], and to find childhood cancer clusters in Alberta, Canada [21].

This study assesses the spatial and temporal epidemiology of the phage types of *S. Enteritidis* that predominate in Ontario health regions by: 1) estimating phage type-specific health region-level incidence rates (IRs); 2) estimating phage type-specific monthly IRs; 3) describing phage type-specific exposure settings and clinical symptoms;
4) detecting phage-type-specific spatial, temporal, and space-time clusters of cases; and
5) examining the exposure settings of cases identified within space-time clusters.

**Methods**

**Study setting and data sources**

Our study was conducted in Ontario, Canada. In 2009, an estimated 13 million people lived in Ontario, accounting for 39% of Canada’s total population [22]. There are 36 public health units (PHUs) in Ontario that are mandated by the provincial ministry of health to administer health promotion and disease prevention programs [23]. These PHUs are grouped into seven planning regions, which were used for the purposes of our study (Figure 5.1; Appendix D. Legend 5.1).

Salmonellosis is a reportable disease under provincial legislation [23], and is diagnosed by public health, hospital, and private laboratories after isolation of *Salmonella* spp. (excluding *Salmonella* Typhi or Paratyphi) from stool (the majority of samples), rectal swabs, urine, blood, or any other sterile site [24]. All *Salmonella* isolates are sent to the Public Health Ontario Laboratories-Toronto for confirmation and serotyping using serological confirmation of compatible somatic and flagellar antigens (Kauffmann-White classification) [25]. All isolates serotyped as *S. Enteritidis* are sent to the National Microbiology Laboratory in Winnipeg, Manitoba for phage typing using techniques defined by Ward and colleagues [26].

Staff at each PHU in Ontario must follow up with every *S. Enteritidis* case to identify exposure settings during the illness incubation period and the clinical symptoms during illness. Case investigation records must be reported to the Ontario Ministry of Health and
Long-Term Care (MOHLTC) through the integrated Public Health Information System (iPHIS). Each PHU has its own case follow-up protocol, and currently there is no standardized follow-up form or set timeline for initial case contact. The exposure setting information is based on what the case reported and was considered significant by the investigator. Exposure settings in the surveillance database were categorized as: international travel (i.e., travelled outside of Canada), private home, food premises (e.g., restaurant, grocery store, bakery, deli, caterer, mobile food premise), other (e.g., institution, hospital, farm, petting zoo, child care centre), or unknown (if the only exposure reported was “unknown”). Cases without exposure setting details were excluded from the exposure setting analysis. When more than one exposure setting was reported, the primary exposure was included in our analysis. Secondary exposure was only considered when the primary exposure was reported as “unknown”.

**Statistical analysis**

**Data management**

Data pertaining to the S. Enteritidis cases’ phage type, age, sex, reporting PHU, date of illness onset, exposure setting, and clinical symptoms were acquired from the iPHIS passive surveillance database. Data were entered into a spreadsheet program (Microsoft Excel 2010, Microsoft Corporation, Redmond, WA, US), reviewed for missing values, and subsequently imported into STATA Intercooled statistical software, version 10.1 (Stata Corporation, College Station, TX, US) for descriptive statistical analysis. Data were available from January 1, 2007 to December 31, 2009; however, due to the large amount of missing phage type information in 2007, all cases from 2007 were excluded from the analyses. Therefore, we evaluated all S. Enteritidis cases that were captured
within the iPHIS database between January 1, 2008 and December 31, 2009. The frequency of *Salmonella* Enteritidis phage types was calculated, the most commonly reported phage types were identified (> 5% of the total number of *S*. Enteritidis cases that were phage typed during the 2-year study period), and the spatial and temporal epidemiology of these phage types were assessed by following several analytical steps, which are outlined in Figure 5.2 and described in detail below. Cartographical boundary files and population estimates for each health region were acquired from Statistics Canada [27].

**Phage type-specific incidence rates**

Health region-level IRs for the six most commonly reported *S*. Enteritidis phage type cases were calculated by dividing the number of cases in a health region with the phage type during the 2-year study period by the population estimate for the health region for the 2-year study period. Health region-level phage-type specific IRs were illustrated in choropleth maps using ArcGIS 10 software (ESRI Inc., Redlands, CA, US).

For the entire province, monthly IRs for the six most frequent phage types were calculated by dividing the number of cases in a month with the phage type by the monthly population estimate. Smoothed IRs based on a simple 3-month moving average were calculated in Microsoft Excel 2010 and plotted together with the monthly raw IRs.

**Scan statistic**

Individual models were built for the 2-year study period for the six most frequent phage types in Ontario. Scan statistics using discrete Poisson models [28] in SaTScan software version 9.0 [29] were conducted to identify purely spatial, purely temporal, and space-
time clusters of *S. Enteritidis* cases. The assumption of the Poisson model is that the number of cases in each health region are Poisson-distributed, based on a known underlying population at risk [28, 30]. Cartesian coordinates of latitude and longitude for each health region centroid were calculated in ArcGIS 10. The smallest spatial and temporal unit was the centroid of a health region and the month of disease onset, respectively. Only high rate clusters were investigated. Secondary clusters were reported if they did not overlap in space with the primary cluster. The scan statistic uses a circular scanning window in space, an interval in time, and a cylinder with a circular spatial base and height corresponding to time in space-time [28, 30]. The scanning window of variable radii gradually moves through time and/or space comparing the rate of cases inside the scanning window to outside the window. When the rate inside the scanning window compared to outside is higher than expected by random chance alone, a high rate cluster is identified. A relative risk and a p-value obtained through Monte Carlo hypothesis testing using 999 replications were estimated for each cluster [31]. A p-value \( \leq 0.05 \) was considered to be significant. The maximum scanning window size was set to include up to 50\% of the population at risk and up to 50\% of study period [28, 30]. Analyses were adjusted for age (0-9, 10-24, 25-34, 35-49, \( \geq 50 \) years) and sex covariates [28]. Statistically significant spatial and space-time clusters were illustrated using a map with health region boundaries in ArcGIS 10. The exposure settings of cases that were part of statistically significant space-time clusters were obtained from iPHIS and examined to assess if a common exposure explained the clustering of cases in space-time.
Results

A total of 1,364 S. Enteritidis cases were recorded in the iPHIS database during the study period; of these, 28 cases were missing phage type information, leaving 1,336 cases (97.9%) available for analysis. The most commonly reported phage types were PT 8 (n=398), PT 13a (n=218), PT 13 (n=198), PT 1 (n=132), PT 5b (n=83), and PT 4 (n=76), which together accounted for 82.7% of all S. Enteritidis cases in Ontario with known phage types during the study period (Table 5.1). One PT 8 case was excluded from the scan statistics because of missing sex information. No outbreaks (e.g., two or more cases linked epidemiologically) were declared by the MOHLTC during the study period.

Health region-level incidence rates

Figure 5.3 illustrates the annual health region-level IRs of S. Enteritidis infections per 100,000 person-years for the six most frequent phage types in Ontario, and described below. For PT 1, the IR ranged from 0.25 to 0.62 units (mean=0.48), with the highest IRs observed in the Central West and Central East regions. For PT 4, the IR ranged from 0.09 to 0.44 units (mean=0.25), with the highest IRs observed in the Central West and Central East regions. For PT 5b, the IR ranged from 0.19 to 0.58 units (mean=0.31), with the highest IRs observed in the Central West and North East regions. For PT 8, the IR ranged from 0.79 to 4.57 units (mean=1.84), with the highest IRs observed in the North West and Toronto regions. For PT 13, the IR ranged from 0.35 to 1.39 units (mean=0.75), with the highest IRs observed in the Toronto and North West regions. For PT 13a, the IR ranged from 0 to 1.18 units (mean=0.74), with the highest IRs observed in the Toronto and Eastern regions.
Monthly raw and smoothed incidence rates

Time-series of raw and smoothed IRs of *S.* Enteritidis infections per 100,000 person-months for the six most frequent phage types in Ontario are illustrated in Figure 5.4, and are described below. The monthly IR ranged from 0 to 0.14 units (mean=0.04) for PT 1, 0 to 0.09 units (mean=0.02) for PT 4, 0 to 0.08 units (mean=0.03) for PT 5b, 0.05 to 0.20 units (mean=0.13) for PT 8, 0 to 0.12 units (mean=0.06) for PT 13, and 0 to 0.16 units (mean=0.07) for PT 13a.

Visually assessing the smoothed trend lines, a number of patterns were observed (Figure 5.4). For PT 1, there were steep up slopes and gradual down slopes, with peaks occurring in January 2008, May 2009, and December 2009. For PT 4, there were three small peaks, which occurred in February 2008, November 2008, and March 2009. For PT 5b, there was one high peak in January 2009. For PT 8, there were monthly variations with five peaks, which occurred in February 2008, November 2008, February 2009, May 2009, and October 2009. For PT 13, there were two peaks, which occurred in May 2008 and May 2009. For PT 13a, there was a high plateau between August 2008 and October 2008, a small plateau between August 2009 and October 2009, and a small peak in December 2009.

Clinical symptoms

Of the 1,336 *S.* Enteritidis cases with known phage types, 1,123 cases (84.1%) had clinical symptom information available. The most commonly reported symptoms were diarrhea (89-97% of cases depending on the phage type), abdominal pain (49-64%), fever (43-55%), vomiting (23-33%), and nausea (28-36%) (Table 5.2).
Exposure settings

Of the 1,336 S. Enteritidis cases with known phage types, 372 (27.8%) cases were missing exposure setting information, leaving 964 cases (72.2%) available for exposure setting analysis (Table 5.3). International travel (19.7% of 1,336 cases), private home (7.0%), and food premise (6.4%) were the most commonly reported known exposure settings. Unknown exposure setting was reported for 37.9% of cases.

Of the six most frequent S. Enteritidis phage types (n=1,105), 306 (27.7%) cases were missing exposure setting information, leaving 799 (72.3%) cases available for exposure setting analysis. Unknown exposure setting was reported for 434 (39.3%) of cases.

Known exposure setting information was reported for 365 (33.03%) of cases: 67 PT 1 cases (50.8% of all PT 1 cases), 37 PT 4 cases (48.7% of all PT 4 cases), 45 PT 5b cases (54.2% of all PT 5b cases), 104 PT 8 cases (26.2% of all PT 8 cases), 58 PT 13 cases (29.3% of all PT 13 cases), and 54 PT 13a cases (24.8% of all PT 13a cases).

Of the cases that had exposure setting information available, international travel and unknown exposure settings were the most frequently reported settings for PT 5b cases (76 and 22% of PT 5b cases, respectively), PT 4 cases (69 and 23% of PT 4 cases, respectively), and PT 1 cases (65 and 28% of PT 1 cases, respectively). Unknown, private home, food premise, and international travel were the most frequently reported exposure settings for PT 8 cases (64, 13, 12, and 10% of PT 8 cases, respectively), PT 13 cases (64, 13, 15, and 6% of PT 13 cases, respectively), and PT 13a cases (66, 14, 9, and 9% of PT 13a cases, respectively) (Table 5.3).
Scan statistics

Purely spatial clusters of S. Enteritidis cases

Four significant high rate spatial clusters were detected (Table 5.4 and Figure 5.5). A cluster of 29 PT 5b cases was identified in the Central West region (RR=2.26, p=0.003). A cluster of 22 PT 8 cases was identified in the North West region (RR=3.10, p≤0.001). A cluster of 74 PT 13 cases was identified in the Toronto region (RR=2.31, p≤0.001). A cluster of 63 PT 13a cases was identified in the Toronto region (RR=1.58, p=0.018).

Purely temporal clusters of S. Enteritidis cases

Five significant high rate temporal clusters were detected (Table 5.4). A cluster of 40 PT 1 cases occurred from January to March 2008 (RR=3.07, p=0.001). A cluster of 51 PT 4 cases occurred from January to November 2008 (RR=2.43, p=0.009). A cluster of 29 PT 5b cases occurred from December 2008 to March 2009 (RR=2.70, p=0.002). A cluster slightly above the rejection threshold of 29 PT 13 cases occurred from April to May 2008 (RR=1.89, p=0.051). A cluster of 63 PT 13a cases occurred from July to October 2008 (RR=2.02, p=0.001).

Space-time clusters of S. Enteritidis cases

Eight significant high rate space-time clusters were detected, including two secondary clusters (Table 5.4 and Figure 5.6). Two clusters of PT1 cases were identified: a primary cluster of 20 cases occurred from January to February 2008 in the North West, North East, Eastern, and Central East regions (RR=4.38, p≤0.001); and a secondary cluster of 16 cases occurred from January to May 2008 in the Central West region (RR=3.33, p=0.046). A cluster of 15 PT 4 cases occurred from February to April 2008 in the Eastern
and Central East regions (RR=4.55, p=0.010). A cluster of 17 PT 5b cases occurred from
September 2008 to April 2009 in the Central West region (RR=3.80, p=0.016). Two
clusters of PT 8 cases were identified: a primary cluster of 15 cases occurred from
February to May 2009 in the North West region (RR=12.91, p≤0.001); and a secondary
cluster of 21 cases occurred from September to December 2008 in the South West region
(RR=2.68, p=0.046). A cluster of 40 PT 13 cases occurred from April to October 2008 in
the Toronto region (RR=3.97, p≤0.001). A cluster of 18 PT 13a cases occurred from
October to December 2009 in the Toronto region (RR=3.39, p=0.018).

**Space-time cluster cases’ exposure settings**

Exposure setting information was unknown or missing for many of the cases that were
part of the space-time clusters (Table 5.5). For the primary PT 1 cluster, exposure setting
information was known for 9 of the 20 cases; seven cases reported international travel
and two cases reported food premises as their exposure setting. For the secondary PT 1
cluster, exposure setting information was known for 4 of the 16 cases; all four cases
reported international travel as their exposure setting. For the PT 4 cluster, exposure
setting information was known for 9 of the 15 cases; all nine cases reported international
travel as their exposure setting. For the PT 5b cluster, exposure setting information was
known for 9 of the 17 cases; all nine cases reported international travel as their exposure
setting. For the primary PT 8 cluster, no exposure setting information was known for the
15 cases. For the secondary PT 8 cluster, exposure setting information was known for 7
of the 21 cases; three cases reported food premises, two cases reported private homes,
one case reported other setting, and one case reported international travel as their
exposure setting. For the PT 13 cluster, exposure setting information was known for 14 of
the 40 cases; 10 cases reported food premises, two cases reported private homes, one case reported other setting, and one case reported international travel as their exposure setting. For the PT 13a cluster, exposure setting information was known for 11 of the 18 cases; seven cases reported food premises, three cases reported international travel, and one case reported private home as their exposure setting.

**Discussion**

Our study enhanced the current knowledge on the spatial and temporal epidemiology of the phage types of *S. Enteritidis* that predominate in Ontario health regions. We used a step-wise approach, starting with a general exploratory analysis followed by a more specific statistical analysis. A number of phage type-specific high rate areas and time periods were identified during the exploratory analysis that were confirmed by the statistical analysis as significant spatial, temporal, or space-time clusters of cases.

Foodborne disease clusters are generally defined as the occurrence of a higher than expected number of cases for a given location and/or time period. These clusters may or may not meet the definition of an outbreak [32-33]. Subtype-based surveillance systems frequently use the term “cluster” to describe a group of cases infected with identical microbial strains [32]. Subtyping is useful for differentiating between endemic and outbreak cases, especially for common *Salmonella* serotypes, such as Enteritidis, that occur sporadically throughout the year [34]. Differences in reservoirs and exposure settings might exist for different *S. Enteritidis* phage types, and molecular differentiation can help to understand potential sources of the different phage types [8, 34]. We defined a cluster as a health region, time period, or a health region during a particular time period...
with a statistically significant higher than expected phage type-specific *S. Enteritidis* infection rate. Thus, we demonstrated the effectiveness of using cluster detection tests, in conjunction with subtyping methods to understand the epidemiology of a foodborne pathogen.

A number of patterns were observed when assessing the geographical heterogeneity of health region-level IRs of *S. Enteritidis* infections for the most frequent phage types. The Central West region had the highest IRs for PTs 1, 4, and 5b, whereas the Toronto region had the highest IRs for PTs 13 and 13a. Several of these regions were later confirmed by the spatial scan statistic as regions with significant high rate clusters (e.g., cases of PT 5b significantly clustered in the Central West region and cases of PTs 13 and 13a significantly clustered in the Toronto region).

We used a smoothing method for our time-series graph to reduce the month-to-month random variation of infection rates and make the overall trends clearer. The observed trends were relatively consistent with the results of the purely temporal scan statistic, albeit not as definitive. With the exception of PT 5b, all temporal clusters occurred during 2008. Further, most clusters occurred during a distinct season. Cases of PTs 1 and 5b clustered during the winter months, cases of PT 13 clustered during the spring months, and cases of PT 13a clustered during the summer and fall months. Differences in the duration of the temporal clusters were also observed. The majority of clusters (PTs 1, 5b, 13, and 13a) were of relatively short duration (2-4 months), whereas the PT 4 cluster was of long duration (11 months). Of note, the most commonly reported phage type (PT 8) did not cluster temporally, suggesting a fairly even distribution of PT 8 cases over time.
throughout Ontario. A study conducted in Alberta, Canada, examining *Salmonella* serotypes rather than phage types, detected several serotype-specific temporal clusters during the 11-year study period (January 1990 to January 2002) [35]; for *S. Enteritidis*, the clusters were of short duration and occurred during the winter and spring months.

The exposure setting information is rarely confirmed by data obtained through environmental health investigations or statistical associations obtained through case-control or cohort studies [36]; however, it is considered to be useful epidemiological data for foodborne illness source attribution [37]. Knowing when, where, and why clusters occurred can aid in the development of effective outbreak detection, prevention, and control programs. Our study identified differences between phage types with respect to the time and duration of the space-time clusters, even for clusters occurring in the same region. For example, the PT 13 and 13a clusters both occurred in the Toronto region, but during different time periods (the cluster of cases with PT 13 occurred in 2008, whereas the cluster of cases with PT 13a occurred in 2009). Moreover, the cluster of cases with PT 13 was of long duration (7 months), whereas the cluster of cases with PT 13a was of short duration (3 months). Short duration clusters might signify that cases were exposed to a single infection source (e.g., point source outbreak). Long duration clusters might signify that cases were exposed to a single source (e.g., contaminated food) over a longer time period (e.g., continuous common source outbreak) [32-33, 35], to multiple sources (e.g., continuous multiple source outbreak) [32-33], to the occurrence of secondary infections [35], to poor food preparation practices over a prolonged period, or that the typing method used was not of high enough resolution to differentiate between different strains.
Many of the cases with PT 13 or 13a that were part of a space-time cluster reported food premises (e.g., restaurant, grocery store, bakery, deli, caterer, mobile food premise) as their main exposure setting. In North America, restaurants have been shown to be an important exposure setting for *S. Enteritidis* infections [38-41]. A number of predisposing factors for food contamination with *S. Enteritidis* in restaurants were identified, including cross contamination from raw chicken meat to food server’s hands or cutting boards due to high food volumes and food handler’s improper food safety practices during food preparation [36, 38], inadequate heat treatment of foods [38], inappropriate food storage [38], and direct contamination of food served by infected food handlers [10, 39-41]. In Ontario, *S. Enteritidis* accounted for only 10.1% of the *Salmonella* isolates collected at pre-harvest from conventionally-raised broiler chicken flocks between July 2010 and April 2012; 65% of the isolates were PT 13a (Tara Roberts, 2014, personal communication).

A few of the cases that were part of a PT 8, 13, or 13a space-time cluster reported private homes as their exposure setting. Previous studies identified private homes as an important exposure setting for sporadic, home-based foodborne infections [42-44]. Several predisposing factors of home-based infections have been identified, including inappropriate food handling, storage, and food preparation [42-43]; consumption of contaminated raw and undercooked foods [42]; and person-to-person [44] and animal-to-person [45-46] transmission.

Space-time clusters of cases with PT 1 or 4 included several overlapping health regions, occurred during nearly identical winter and spring months, and were of short duration (2-
The majority of these cases reported international travel as their exposure setting. International travel was demonstrated by a number of studies as an important risk factor for \textit{S. Enteritidis} infections in North America [15, 47-48]. In the US, among all salmonellosis cases between 2004 and 2008, 11\% reported international travel as their exposure setting, and among those, the most commonly reported serotype was \textit{Enteritidis} (22\% of travel cases) [47]. In the region of Waterloo, Ontario, Canada, between June 2005 and May 2009, 48.7\% of \textit{S. Enteritidis} cases were international travel-related [48]. In Ontario, Canada, between July 2010 and June 2011, 51.9\% of \textit{S. Enteritidis} infections were international travel-related, and certain phage types (e.g., 1, 4, and 5b) were isolated from cases who visited all-inclusive resorts in the Caribbean or Mexico during the winter and spring months [15].

A number of limitations should be recognized before interpreting our study results. Surveillance programs underestimate the true burden of infections due to under-diagnosis and under-reporting of cases [3]. In Canada, it was estimated that for every reported salmonellosis case there were 26.1 unreported cases in the general population [3]. Under-reporting and under-diagnosis can be influenced by differences in populations’ medical care seeking behaviour and access to medical care [49], physicians’ specimen request and diagnosis practices [50], and laboratories testing protocols and reporting standards [52]. Regional differences in successful case follow-up should also be considered. Loss to follow-up of cases might be greater in low population density regions of the province due to difficulties encountered by public health staff in contacting cases. A large number of cases had missing or unknown exposure setting information, which might have biased our study results. The proportion and accuracy of known exposure setting information
reported by investigators can depend on several factors [36], including time passed from exposure to case interview and the related recall bias, difficulty and the effort made by the investigator to contact a case, follow-up protocol and questionnaire used by the investigator (e.g., face to face interview vs. phone interview vs. questionnaire sent through the mail), a case’s willingness to be interviewed, and possible survival bias. In our study, differences in unknown exposure setting among phage types were noted. The proportion of unknown exposure setting information was higher for cases with PT 8, 13, or 13a (64 - 66%) compared to cases with PT 1, 4, or 5b (22 - 28%), suggesting that international travel cases had more readily available exposure history; therefore in our study, the overall proportion of cases who reported international travel as their major exposure setting was likely slightly over-estimated. Lastly, misclassification of international travel-related cases might have also occurred, especially for cases for which the incubation period was short, and for cases with a longer disease incubation period who became infected before departure [48].

Obtaining exposure setting information is a first step toward developing effective prevention and control programs; however, the location and the primary source of contamination of food products that lead to infections are not always identical [52]. Therefore, future research studies are needed to identify the primary source of contamination, and the type of food products that cause infections.

This study demonstrated the utility of retrospective spatial and temporal analysis of subtype-based surveillance data using exploratory and statistical methods to detect clusters of cases. Phage type-specific spatial and spatial-temporal clusters should be
followed up by public health authorities to identify novel local individual-level risk factors. Prevention programs (e.g., travel advisories) that are targeted during the winter and spring months have the potential to decrease the incidence of S. Enteritidis, especially PTs 1, 4, and 5b. During the study period no outbreaks were reported in Ontario; thus, the evaluation of current outbreak detection methods used by public health staff at various PHUs is warranted. Future studies are needed to evaluate the frequency of false positive clusters, to assess the effectiveness of cluster detection using statistical methods, to compare the more traditional outbreak investigation procedures to scan statistic cluster detection techniques, and to measure the feasibility of statistical methods for identifying infection clusters. Purely spatial or purely temporal clusters might be the result of a space-time cluster, which should be considered when evaluating our study results. There is a need also for prospective research studies to identify clusters of S. Enteritidis infections in real-time (e.g., weeks, months), and to assess and evaluate individual-level risk factors for infections included in these clusters. Moreover, there is a need for high resolution molecular subtyping methods (e.g., multiple locus variable-number tandem repeat analysis or whole genome sequencing) to better understand relationships between cases in a cluster.

**Conclusions**

This is the first study that has evaluated the spatial and temporal epidemiology of the phage types of S. Enteritidis that predominate in Ontario health regions. This study demonstrated the value of using a number of spatial-temporal and subtyping methods to better understand the epidemiology of a foodborne illness, such as salmonellosis. Our study highlighted phage type-specific differences in spatial distributions, temporal trends,
clinical symptoms, exposure settings, and space-time clusters of \textit{S. Enteritidis} infections. Several health regions were identified with increased phage type-specific \textit{S. Enteritidis} infection rates where future studies should be conducted to identify novel individual-level risk factors, and where future prevention and control programs should be targeted to reduce the incidence of \textit{S. Enteritidis} infections. Our study methodology may be applicable to other foodborne disease surveillance data.

\textbf{List of Abbreviations}

\textbf{GIS} - geographic information system

\textbf{iPHIS} - integrated Public Health Information System

\textbf{IR} - incidence rate

\textbf{MOHLTC} - Ontario Ministry of Health and Long-Term Care

\textbf{PT} - phage type

\textbf{PHU} – public health unit

\textbf{S. Enteritidis} - \textit{Salmonella enterica} serovar Enteritidis

\textbf{US} – United States of America

\textbf{Competing interests}

The authors declare that they have no competing interests.
**Authors' contributions**

CV developed the study design, analysed the data, interpreted results, wrote the first draft of the manuscript, responded to editorial comments, and prepared the final manuscript for submission.

MTG was consulted on data analysis, study design, interpretation of results, and reviewed and commented on manuscript drafts.

DLP, SAM, FP, and JMS provided advice on the data analysis, interpretation of results, and reviewed and commented on manuscript drafts.

All authors read and approved the final manuscript.

**Acknowledgements**

The authors acknowledge the MOHLTC for providing the data. We thank the staff of all PHUs and public health laboratories that tested samples, followed up with cases, and entered information into the public health surveillance database. We also acknowledge the Data Resource Centre at the University of Guelph library for their GIS and census support. The views expressed in this study are the views of the authors and do not necessarily reflect those of the MOHLTC.
References


22. Statistics Canada: Table 051-0001. Estimates of population, by age group and sex for July 1, Canada, provinces and territories annual (persons unless otherwise noted) (Number). Available at: [http://www5.statcan.gc.ca/cansim/a26](http://www5.statcan.gc.ca/cansim/a26)


*Salmonella* Identification - Serotypes and Antigenic Formulae. London: Health 
Protection Agency; 2007.

26. Ward LR, de Sa JDH, Rowe B: A phage typing scheme for *Salmonella* 

27. Statistics Canada: Health region boundary files. Available at:  
http://www.statcan.gc.ca/pub/82-402-x/2009001/rg-eng.htm

Methods 1997, 26:1481-1496.

29. Kulldorff M. and Information Management Services, Inc. SaTScanTM v9.0:  
Software for the spatial and space-time scan statistics. Available at:  
http://www.satscan.org

space-time scan statistic and brain cancer in Los Alamos. American Journal of 

330.

32. World Health Organization: Foodborne disease outbreaks: Guidelines for 
investigation and control. 2008. Available at:  


47. Johnson LR, Gould LH, Dunn JR, Berkelman R, Mahon BE; Foodnet Travel Working Group: *Salmonella* infections associated with international travel: a


### Tables

**Table 5.1. Frequency of *Salmonella* Enteritidis cases with different phage types in Ontario, Canada, 2008-2009 (n=1,336).**

<table>
<thead>
<tr>
<th>Phage Type</th>
<th>n (n/N %)</th>
<th>Phage Type</th>
<th>n (n/N%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT 8</td>
<td>398 (29.8)</td>
<td>PT 1a</td>
<td>11 (0.8)</td>
</tr>
<tr>
<td>PT 13a</td>
<td>218 (16.3)</td>
<td>PT 21</td>
<td>11 (0.8)</td>
</tr>
<tr>
<td>PT 13</td>
<td>198 (14.8)</td>
<td>PT 1b</td>
<td>10 (0.7)</td>
</tr>
<tr>
<td>PT 1</td>
<td>132 (9.9)</td>
<td>PT 14b</td>
<td>10 (0.7)</td>
</tr>
<tr>
<td>PT 5b</td>
<td>83 (6.2)</td>
<td>PT 22</td>
<td>9 (0.7)</td>
</tr>
<tr>
<td>PT 4</td>
<td>76 (5.7)</td>
<td>PT 23</td>
<td>8 (0.6)</td>
</tr>
<tr>
<td>PT 6a</td>
<td>49 (3.7)</td>
<td>PT 19</td>
<td>7 (0.5)</td>
</tr>
<tr>
<td>atypical</td>
<td>33 (2.5)</td>
<td>PT 51</td>
<td>6 (0.4)</td>
</tr>
<tr>
<td>PT 6</td>
<td>20 (1.5)</td>
<td>other</td>
<td>57 (4.3)</td>
</tr>
</tbody>
</table>

n = number of *S. Enteritidis* cases with the phage type. N = total number of *S. Enteritidis* cases that were phage typed during the 2-year study period = 1,336
Table 5.2. Clinical symptoms of *Salmonella* Enteritidis cases with major phage types in Ontario, Canada, 2008-2009 (N=1,123).

<table>
<thead>
<tr>
<th>Phage Type</th>
<th>Diarrhea n (n/N %)</th>
<th>Bloody diarrhea n (n/N %)</th>
<th>Abdominal pain n (n/N %)</th>
<th>Fever n (n/N %)</th>
<th>Vomiting n (n/N %)</th>
<th>Nausea n (n/N %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT 1 (N=114)</td>
<td>106 (93)</td>
<td>3 (3)</td>
<td>67 (59)</td>
<td>49 (43)</td>
<td>34 (30)</td>
<td>36 (32)</td>
</tr>
<tr>
<td>PT 4 (N=63)</td>
<td>61(97)</td>
<td>3 (5)</td>
<td>38 (60)</td>
<td>33 (52)</td>
<td>21 (33)</td>
<td>19 (30)</td>
</tr>
<tr>
<td>PT 5b (N=75)</td>
<td>70 (93)</td>
<td>5 (7)</td>
<td>37 (49)</td>
<td>36 (48)</td>
<td>17 (23)</td>
<td>27 (36)</td>
</tr>
<tr>
<td>PT 8 (N=333)</td>
<td>315 (95)</td>
<td>33 (10)</td>
<td>213 (64)</td>
<td>178 (54)</td>
<td>96 (29)</td>
<td>95 (29)</td>
</tr>
<tr>
<td>PT 13 (N=169)</td>
<td>150 (89)</td>
<td>18 (11)</td>
<td>95 (56)</td>
<td>86 (51)</td>
<td>53 (31)</td>
<td>33 (20)</td>
</tr>
<tr>
<td>PT 13a (N=178)</td>
<td>171 (96)</td>
<td>19 (11)</td>
<td>111 (62)</td>
<td>96 (54)</td>
<td>48 (30)</td>
<td>49 (28)</td>
</tr>
<tr>
<td>PT Others (N=191)</td>
<td>180 (94)</td>
<td>12 (6)</td>
<td>117 (61)</td>
<td>105 (55)</td>
<td>59 (31)</td>
<td>62 (32)</td>
</tr>
<tr>
<td>All PTs (N=1,123)</td>
<td>1,053 (94)</td>
<td>93 (8)</td>
<td>678 (60)</td>
<td>583 (52)</td>
<td>328 (29)</td>
<td>321 (29)</td>
</tr>
</tbody>
</table>

Of the 1,336 *S. Enteritidis* cases with known phage types, 1,123 cases had clinical symptom information available. n = number of *S. Enteritidis* cases that had the symptom. N = number of *S. Enteritidis* cases with the phage type. Within a row, the percentages can add up to greater than 100% because a case could have more than one symptom.
Table 5.3. Exposure settings of *Salmonella* Enteritidis cases with major phage types in Ontario, Canada, 2008-2009.

<table>
<thead>
<tr>
<th>Phage Type</th>
<th>Private Home n (n/M %)</th>
<th>Food Premise n (n/M %)</th>
<th>International Travel n (n/M %)</th>
<th>Other Setting n (n/M %)</th>
<th>Unknown n (n/M %)</th>
<th>Missing (n/N %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT 1 (N=132) (M=93)</td>
<td>3 (3)</td>
<td>4 (4)</td>
<td>60 (65)</td>
<td>0</td>
<td>26 (28)</td>
<td>39 (30)</td>
</tr>
<tr>
<td>PT 4 (N=76) (M=48)</td>
<td>0</td>
<td>3 (6)</td>
<td>33 (69)</td>
<td>1 (2)</td>
<td>11 (23)</td>
<td>28 (37)</td>
</tr>
<tr>
<td>PT 5b (N=83) (M=58)</td>
<td>0</td>
<td>1 (2)</td>
<td>44 (76)</td>
<td>0</td>
<td>13 (22)</td>
<td>25 (30)</td>
</tr>
<tr>
<td>PT 8 (N=398) (M=285)</td>
<td>38 (13)</td>
<td>33 (12)</td>
<td>29 (10)</td>
<td>4 (1)</td>
<td>181 (64)</td>
<td>113 (28)</td>
</tr>
<tr>
<td>PT 13 (N=198) (M=158)</td>
<td>21 (13)</td>
<td>24 (15)</td>
<td>10 (6)</td>
<td>3 (2)</td>
<td>100 (64)</td>
<td>40 (20)</td>
</tr>
<tr>
<td>PT 13a (N=218) (M=157)</td>
<td>22 (14)</td>
<td>14 (9)</td>
<td>14 (9)</td>
<td>4 (2)</td>
<td>103 (66)</td>
<td>61 (28)</td>
</tr>
<tr>
<td>PT Others (N=231) (M=165)</td>
<td>9 (6)</td>
<td>7 (4)</td>
<td>73 (44)</td>
<td>4 (2)</td>
<td>72 (44)</td>
<td>66 (29)</td>
</tr>
<tr>
<td><strong>Total (N=1336) (M=964)</strong></td>
<td><strong>93 (10)</strong></td>
<td><strong>86 (9)</strong></td>
<td><strong>263 (27)</strong></td>
<td><strong>16 (2)</strong></td>
<td><strong>506 (52)</strong></td>
<td><strong>372 (28)</strong></td>
</tr>
</tbody>
</table>

Of the 1,336 *S. Enteritidis* cases with known phage types, 964 cases had exposure setting information available. n = number of *S. Enteritidis* cases with the exposure setting. M = number of *S. Enteritidis* cases that had exposure setting information available. N = number of *S. Enteritidis* cases with the phage type. Exposure settings in the surveillance database were categorized as: international travel (i.e., travelled outside of Canada), private home, food premise (e.g., restaurant, grocery store, bakery, deli, caterer, mobile food premise), other (e.g., institution, hospital, farm, petting zoo, child care centre), or unknown (if the only exposure reported was “unknown”).
Table 5.4. Clusters of *Salmonella* Enteritidis cases with the six most frequent phage types in Ontario, Canada, 2008-2009.

<table>
<thead>
<tr>
<th>Phage Type</th>
<th>Annual cases per 100,000</th>
<th>Cluster Type</th>
<th>Region</th>
<th>Time Frame (Year/Month)</th>
<th>Observed</th>
<th>Expected</th>
<th>O/E</th>
<th>RR</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT 1 (N=132)</td>
<td>0.05</td>
<td>Space-time</td>
<td>North West, North East, Eastern, Central East</td>
<td>2008/1 to 2008/2</td>
<td>20</td>
<td>5.17</td>
<td>3.87</td>
<td>4.38</td>
<td>≤0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Central West</td>
<td>2008/1 to 2008/5</td>
<td>16</td>
<td>5.24</td>
<td>3.05</td>
<td>3.33</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Temporal Central West</td>
<td>2008/1 to 2008/3</td>
<td>40</td>
<td>16.36</td>
<td>2.38</td>
<td>3.07</td>
<td>0.001</td>
</tr>
<tr>
<td>PT 4 (N=76)</td>
<td>0.03</td>
<td>Space-time</td>
<td>Eastern, Central East</td>
<td>2008/2 to 2008/4</td>
<td>15</td>
<td>3.89</td>
<td>3.85</td>
<td>4.55</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Temporal All</td>
<td>2008/1 to 2008/11</td>
<td>51</td>
<td>34.67</td>
<td>1.47</td>
<td>2.43</td>
<td>0.009</td>
</tr>
<tr>
<td>PT 5b (N=83)</td>
<td>0.03</td>
<td>Spatial</td>
<td>Central West</td>
<td>NA</td>
<td>29</td>
<td>15.93</td>
<td>1.82</td>
<td>2.26</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Space-time Central West</td>
<td>2008/9 to 2009/4</td>
<td>17</td>
<td>5.27</td>
<td>3.22</td>
<td>3.80</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Temporal All</td>
<td>2008/12 to 2009/3</td>
<td>29</td>
<td>13.77</td>
<td>2.11</td>
<td>2.70</td>
<td>0.002</td>
</tr>
<tr>
<td>PT 8 (N=397)</td>
<td>0.20</td>
<td>Spatial</td>
<td>North West</td>
<td>NA</td>
<td>22</td>
<td>7.37</td>
<td>2.98</td>
<td>3.10</td>
<td>≤0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Space-time North West</td>
<td>2009/2 to 2009/5</td>
<td>15</td>
<td>1.20</td>
<td>12.46</td>
<td>12.91</td>
<td>≤0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>South West</td>
<td>2008/9 to 2008/12</td>
<td>21</td>
<td>8.11</td>
<td>2.59</td>
<td>2.68</td>
<td>0.046</td>
</tr>
<tr>
<td>PT 13 (N=198)</td>
<td>0.08</td>
<td>Spatial</td>
<td>Toronto</td>
<td>NA</td>
<td>74</td>
<td>40.71</td>
<td>1.82</td>
<td>2.31</td>
<td>≤0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Space-time Toronto</td>
<td>2008/4 to 2008/10</td>
<td>40</td>
<td>11.87</td>
<td>3.37</td>
<td>3.97</td>
<td>≤0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Temporal All</td>
<td>2008/4 to 2008/5</td>
<td>29</td>
<td>16.46</td>
<td>1.76</td>
<td>1.89</td>
<td>0.051</td>
</tr>
</tbody>
</table>
Table. 5.4. cont.

<table>
<thead>
<tr>
<th>Phage Type</th>
<th>Annual cases per 100,000</th>
<th>Cluster Type</th>
<th>Region</th>
<th>Time Frame (Year/Month)</th>
<th>Observed</th>
<th>Expected</th>
<th>O/E</th>
<th>RR</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT 13a (N=218)</td>
<td>0.08</td>
<td>Spatial</td>
<td>Toronto</td>
<td>NA</td>
<td>63</td>
<td>44.70</td>
<td>1.41</td>
<td>1.58</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Space-time</td>
<td>Toronto</td>
<td>2009/10 to 2009/12</td>
<td>18</td>
<td>5.64</td>
<td>3.19</td>
<td>3.39</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Temporal</td>
<td>All</td>
<td>2008/7 to 2008/10</td>
<td>63</td>
<td>36.51</td>
<td>1.73</td>
<td>2.02</td>
<td>0.001</td>
</tr>
</tbody>
</table>

N = number of *S. Enteritidis* cases with the phage type. Results based on discrete Poisson models using the SaTScan™ software. Study period: January 1, 2008 to December 31, 2009. Time aggregation units: month. Time aggregation length: 1 month. Circular scanning window size: up to 50% of the population at risk and/or 50% of time the study period. Confounders controlled for: age (0-9, 10-24, 25-34, 35-49, ≥ 50 years) and sex. Criteria for reporting secondary clusters: no geographical overlap. Type of clusters investigated: high rate only. NA = not applicable. O/E = observed divided by expected. RR = relative risk. Significance level: p ≤ 0.05.
<table>
<thead>
<tr>
<th>Phage Type Cluster</th>
<th>Cases</th>
<th>Private Home</th>
<th>Food Premise</th>
<th>International Travel</th>
<th>Other setting</th>
<th>Unknown</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT 1 (N=132)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>North West, North East, Eastern, Central East</td>
<td>20</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Central West</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>PT 4 (N=76)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eastern, Central East</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PT 5b (N=83)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central West</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>PT 8 (N=397)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>North West</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>South West</td>
<td>21</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>PT 13 (N=198)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Toronto</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>PT 13a (N=218)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Toronto</td>
<td>18</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>7</td>
</tr>
</tbody>
</table>

N = number of S. Enteritidis cases with the phage type. n = number of S. Enteritidis cases with the exposure setting. Results based on discrete Poisson models using the SatScan™ software. Study period: January 1, 2008 to December 31, 2009. Time aggregation units: month. Time aggregation length: 1 month. Circular scanning window size: up to 50% of the population at risk and 50% of the study period. Confounders controlled for age (0-9, 10-24, 25-34, 35-49, ≥ 50 years) and sex. Criteria for reporting secondary clusters: no geographical overlap. Type of clusters investigated: high rate only. Exposure settings in the surveillance database were categorized as: international travel (i.e., travelled outside of Canada), private home, food premise (e.g., restaurant, grocery store, bakery, deli, caterer, mobile food premise), or other (e.g., institution, hospital, farm, petting zoo, child care centre).
The names and population estimates for the public health units (indicated by labels 0 through 35), are presented in Appendix D. Legend 5.1.
Figure 5.2. Flow chart outlining the analytical steps used to evaluate *Salmonella* Enteritidis cases with major phage types.
Figure 5.3. Health region-level raw incidence rates of *Salmonella* Enteritidis cases with major phage types in Ontario, Canada, 2008-2009.
Figure 5.4. Monthly raw and smoothed incidence rates of *Salmonella* Enteritidis cases with major phage types in Ontario

Smoothed IRs were based on a 3-month simple rolling average.
Figure 5.5. Spatial clusters of *Salmonella* Enteritidis cases with major phage types in Ontario, Canada, 2008-2009.

Results based on discrete Poisson models using the SaTScan™ software. Study period: January 1, 2008 to December 31, 2009. Circular scanning window size: up to 50% of the population at risk. Confounders controlled for: age (0-9, 10-24, 25-34, 35-49, ≥ 50 years) and sex. Criteria for reporting secondary clusters: no geographical overlap. Type of clusters investigated: high rate only. RR = relative risk. Significance level: p ≤ 0.05.
Figure 5.6. Space-time clusters of *Salmonella* Enteritidis cases with major phage types in Ontario, Canada, 2008-2009.

Results based on discrete Poisson models using the SaTScan™ software. Study period: January 1, 2008 to December 31, 2009. Circular scanning window size: up to 50% of the population at risk and 50% of the study period. Confounders controlled for: age (0-9, 10-24, 25-34, 35-49, ≥ 50 years) and sex. Time aggregation units: month. Time aggregation length: 1 month. Criteria for reporting secondary clusters: no geographical overlap. Type of clusters investigated: high rate only. RR = relative risk. Significance level: p ≤ 0.05.
CHAPTER SIX

Conclusions and summary discussion

Salmonellosis poses a significant human health burden globally, causing annually an estimated 93.8 million human infections and 155,000 deaths (Majowicz, et al., 2010). In Ontario and Canada salmonellosis is the second most commonly reported foodborne bacterial infection, causing an estimated 269.26 domestically acquired foodborne illnesses per 100,000 population each year (Thomas et al., 2013). In Ontario and Canada, the number of human S. Enteritidis infections has increased gradually during the last decade; consequently S. Enteritidis became the most common serovar among non-typhoidal Salmonellae (Nesbitt, et al., 2012; Middleton et al., 2013). There is a need to better understand various factors that might have an impact on the increase of S. Enteritidis infections in Ontario, and to identify areas with high S. Enteritidis infection rates where future prevention and control efforts should be directed.

The objectives of this thesis were to evaluate the epidemiology of S. Enteritidis infections in Ontario by analyzing passive surveillance data from January 1, 2007 through December 31, 2009. Direct standardization was used to calculate age-and-sex-adjusted incidence rates of S. Enteritidis infections in Ontario (Chapter 2). To identify associations between S. Enteritidis infection rates and demographic and seasonal factors across Ontario’s public health units a multivariable Poisson regression model was constructed (Chapter 2). Spatial clustering of S. Enteritidis infections across three public health units in the greater Toronto area was assessed, and underlying area-level associations between age- and sex-standardized incidence rates of S. Enteritidis infections and socioeconomic status indicators were evaluated (Chapter 3). In the city of Toronto forward sortation
area-level spatial heterogeneity of age and sex-adjusted \( S \). Enteritidis infection rates was assessed in choropleth maps (Chapter 4). Global and local clustering of age and sex-adjusted \( S \). Enteritidis infection rates in the city of Toronto were examined by using a number of spatial statistical methods (Chapter 4). Subtype-specific differences of \( S \). Enteritidis infections were evaluated by illustrating the incidence rates of the six main phage types that predominate in Ontario health regions in choropleth maps (Chapter 4). For these main phage types, phage type-specific exposure settings and clinical symptoms were described (Chapter 5). Scan statistic was used to detect phage-type-specific spatial, temporal, and space-time clusters of cases (Chapter 5). Exposure settings of cases identified within space-time clusters were examined to identify common exposures (Chapter 5).

A number of patterns can be observed throughout these thesis chapters when evaluating the epidemiology of \( S \). Enteritidis infections in Ontario. An increase in annual age-and-sex-adjusted IRs of \( S \). Enteritidis infections in Ontario from 2007 to 2008-2009, was identified (Chapter 2). High mean age-and-sex-adjusted IRs were detected in three southwestern PHUs (Halton Regional Health Unit, Huron County Health Unit, and Waterloo Health Unit), one northern PHU (Thunder Bay District Health Unit), and in the City of Toronto Health Unit (Chapter 2). Moreover, multivariable Poisson regression model with PHU as a random effect demonstrated a significantly higher incident rate ratio of \( S \). Enteritidis infections in 2008 and 2009 compared to 2007 (Chapter 2). This finding is in agreement with several Canadian and US studies that showed an increase in the incidence of human \( S \). Enteritidis infections during the same period (Chai, et al., 2012; Nesbitt, et al., 2012; Middleton et al., 2013; Parmley, et al., 2013).
The first study (Chapter 2) of this thesis identified seasonal differences in the reported rates of *S. Enteritidis* infections, with highest rates during the spring (March through May) months. The fourth study (Chapter 5) confirmed and explained this result, showing a number of phage type-specific space-time clusters across Ontario’s health regions that occurred mainly during the spring months. A secondary cluster of PT1 cases occurred from January to May 2008 in the Central West region, a primary cluster of PT 4 cases occurred from February to April 2008 in the Eastern and Central East regions, a primary cluster of PT 8 cases occurred from February to May 2009 in the North West region.

Many of the PT 1, 4 or 5b cases with known exposure settings that were identified in space-time clusters mainly reported international travel as their exposure settings, whereas space-time PT 8, 13, and 13a cases with known exposure settings frequently reported food premise or private homes as their exposure settings. In Ontario, international travel was shown to be a major risk factor for human *S. Enteritidis* infections (Ravel, et al., 2011; Tighe, et al., 2012). A recent study has revealed that 51.9% of *S. Enteritidis* infections between July 2010 and June 2011 that were reported to Ontario’s passive surveillance system were international travel-related (Tighe, et al., 2012). The majority of these *S. Enteritidis* cases were PTs 1, 4, and 5b. In the region of Waterloo, Ontario, between June 2005 and May 2009, 48.7% of *S. Enteritidis* infections reported international travel as their main exposure setting (Ravel, et al., 2011).

Restaurant-associated outbreaks with *S. Enteritidis* phage type 8 cases have been reported in Pennsylvania, US (Vugia, et al., 1993), in Portsmouth, United Kingdom (Severi, et al., 2012), in Montréal, Canada (Pilon and Laurin, 1997), and in British Columbia, Canada (Taylor, et al., 2012). In Alberta, Canada, consumption of food products purchased from

- 203 -
mobile lunch trucks was associated with *S. Enteritidis* infections with PTs 8 and 13 (CDC, 2013). A number of US studies described restaurant-related *S. Enteritidis* infection outbreaks where asymptomatic but *S. Enteritidis* positive food handlers contaminated food served at these restaurants (Hedican et al., 2009; Medus, et al., 2010).

Private home exposure was a frequently reported exposure setting for PT 8, 13, and 13a cases. Several risk factors of home-based *S. Enteritidis* infections have been revealed, including eating undercooked eggs (Marcus, et al., 2007), inappropriate food safety practices followed in the kitchens of cases (Scott, 2003), contact with reptiles (Sauteur, et al., 2013; Whitten, et al., 2014), and person-to-person transmission (Ethelberg, et al., 2004).

The first study (Chapter 2) identified three public health units (PHUs) within the Greater Toronto Area (GTA) with moderate (Peel Regional Health Unit and York Regional Health Unit) to high (City of Toronto Health Unit) infection rates that warranted further investigation. In the second study (Chapter 3), using spatial scan statistic we detected a high infection rate cluster in downtown Toronto that included nine neighbouring FSAs. Using a negative binomial regression model (Chapter 3), we identified associations between increased rates of *S. Enteritidis* infections and areas with high and low average median family income compared to areas with medium average median family income, medium proportion of visible minority population compared to areas with low proportion of visible minority population, and areas with high average number of children at home per census family. The high disease rate cluster in downtown Toronto identified in the spatial scan statistic contained primarily SES indicators (low average median family
income, medium visible minority population proportion) that were positively associated with increased rates of S. Enteritidis infections in the regression model, suggesting that these variables had an effect on the clustering of S. Enteritidis infections. A number of hypotheses exist for high income neighbourhood residents’ higher salmonellosis rates, including: better access to healthcare, greater health-seeking behaviour, and pet ownership (Younus, et al., 2007). Residents from low income neighbourhoods had high infection rates that might be explained by poorer microbial quality of foods available in these areas (Koro, et al., 2010), or by more frequent retail food safety violations at local restaurants and retail stores (Darcey and Quinlan, 2011) that predispose local residents to foodborne infections. The area-level distribution of SES indicators were illustrated in disease maps using GIS, which was a beneficial tool to visualize and highlight neighbourhoods where future prevention and control programs should be targeted, or future studies be conducted to evaluate the underlying individual-level risk factors.

Children 0-4 years of age (reference category), followed by children 5-9 years of age had the highest S. Enteritidis infection rates in the first study (Chapter 2). Children’s increased susceptibility to Salmonella infections has been demonstrated by other individual-level North American studies (Younus, et al., 2006; Arshad et al., 2007; Cellucci et al., 2010). Children’s high susceptibility to S. Enteritidis infection was supported in the second study (Chapter 3) where a significant positive curvilinear relationship between S. Enteritidis infection rates and area-level average number of children at home per census family was observed. Various risk factors have been shown to be associated with increased salmonellosis in children, including riding in shopping carts and exposure to raw meat and poultry products (Patrick, et al., 2010), person-to-
person transmission in daycare centers and in private homes (Ethelberg, et al., 2004), and contact with reptiles (Aiken, et al., 2010; Younus, et al., 2010; Sauteur, et al., 2013; Whitten, et al., 2014), cats (Younus, et al., 2010), and pet guinea pigs (Bartholomew, et al., 2014).

The second study (Chapter 3) identified a spatial cluster of high infection rates in downtown Toronto that included nine neighbouring FSAs (M4Y, M5B, M5C, M5E, M5G, M5S, M5T, M5V, and M6J). The third study’s local cluster detection methods (Chapter 4) identified several of the same local disease clusters [the Getis-Ord Gi* method detected five hot spots (M5C, M5E, M5G, M5S, M5T), and the Local Moran’s I method detected three high-high clusters (M4Y, M5E, M5G) and one low-high cluster (M5C)]. Moreover, the fourth study (Chapter 5) identified two high rate space-time clusters in the city of Toronto, a cluster of PT 13 cases from April to October 2008, and a cluster of PT 13a cases from October to December 2009. Several of these cases reported restaurant and private home as their exposure settings.

The city of Toronto was identified in the first and second studies (Chapter 2 and Chapter 3) as a public health unit with high infection rate that needed additional examination to understand different factors influencing the increased disease rates. The third study’s Global Getis-Ord G method assessed global indicators of spatial autocorrelation (Chapter 4), and identified a relatively small distance band (3.3 km) at which areas with high disease rates significantly clustered globally, suggesting that infections were clustering at small distinct areas. The two local cluster detection methods (Getis-Ord Gi* and the local
Moran’s I) confirmed this finding (Chapter 4) identifying several small distinct areas, largely in downtown Toronto with high infection rates.

The effectiveness of two local cluster detection methods (Getis-Ord Gi* and Local Moran’s I) in identifying local infection clusters were compared (Chapter 4). The Getis-Ord Gi* method was used first because there was a hypothesis that high or low infection rates clustered within the city of Toronto. However, this method included the target area rate when calculating the local sum of rates. A small number of areas were wrongly classified, because in downtown Toronto a number of areas were surrounded by areas with dissimilar rates. To prevent misclassification of areas with dissimilar rates the local Moran’s I method was applied, which identified these areas as outlier areas (areas with high rates surrounded by areas with low rates or areas with low rates surrounded by areas with high rates).

Across Ontario’s health regions phage type-specific variation in regional and monthly IRs were identified (Chapter 5). High IRs for PTs 1, 4, and 5b cases were observed in the Central West region, whereas high IRs for PTs 8, 13 and 13a were identified in Toronto region. Seasonal differences for high IRs of S. Enteritidis infections with major PTs were also observed; high IRs of PTs 1 and 5b cases were observed during the winter months, high IRs of PT 13 cases were detected during the spring months, and high IRs of PT 13a cases were identified during the summer and fall months

A retrospective phage type-specific scan statistical analysis found a number of spatial, temporal, and space-time clusters of S. Enteritidis cases (Chapter 5). Regions and time periods previously identified at the descriptive statistical analysis step as high S.
Enteritidis infection rate areas or months were confirmed by the spatial and temporal scan statistics as regions or time periods with statistically significant high rate *S*. Enteritidis infection clusters with specific phage types (e.g., cases of PT 5b significantly clustered in the Central West region; cases of PT 13 and 13a significantly clustered in the Toronto region; and cases of PT 1 had high incident rates from January 2008 to March, 2008).

In Ontario, *S*. Enteritidis cases with different phage types reported diarrhea, abdominal pain, and fever as their most common clinical symptoms, which is in agreement with other North American studies (Marcus et al., 2007; Tighe et al., 2012).

**Advantages and limitations of study design**

A main limitation of passive surveillance systems is that they underestimate the exact number of cases. In Canada it was estimated that for every salmonellosis case reported there were an estimated 26.1 cases that remained unreported (Thomas, et al., 2013). Differences in reporting could be influenced by variations in health seeking behaviour, differences in testing procedures used at various laboratories, and dissimilarities in health care providers’ accessibility (Scallan, et al., 2006; MacDougall, et al., 2008). Variations in underreporting across age groups might also occur, because children and older adults visit their physicians more frequently, physicians are more likely to request stool samples from these patients; consequently, they tend to be over represented in the surveillance database (Scallan, et al., 2006; Edge, et al., 2007; MacDougall, et al., 2008).

A number of biases must be taken into consideration when interpreting ecological studies (Chapters 3 & 4). Ecological studies only consider variation between groups and not within groups (ecological bias) that makes them unsuitable to relate population level risk
factors to the individual level (Morgenstern, 1995). However, ecological studies are cost-effective alternatives to individual-level studies, because they can identify population- or area-level risk factors that later can be followed up by individual-level studies. Another potential limitation of this study design is that data analysis at different levels (e.g. county or province) might provide different findings due to the “modifiable areal unit problem” (Wong, 2009). However, small area data analysis reduces the ecological bias, and it is an effective approach to identify area-level risk factors for foodborne infections (Morgenstern, 1995; Weisent, et al., 2011). Forward sortation area boundary was used to delimit neighbourhoods (Chapters 3 & 4), which is an arbitrary delimitation based on the mail distribution zones of postal facilities. Occasionally these borders do not follow neighbourhoods’ spatial features (“zoning effect”) (Waller, 2004). Moreover, at the study area boundaries forward sortation areas do not have neighbours, which can bias spatial estimates (“edge effect”) for certain spatial statistics (Waller, 2004). To account for the “zoning” and “edge” effects we used the zone of indifference conceptualization parameter, which accounts for neighbouring and border area spatial characteristics (Mitchell, 2005) by not forcing sharp boundaries on neighbouring areas spatial properties.

Case investigation records reported by staff at each public health unit to Ontario’s surveillance system can be biased (Chapter 5). The successful follow up of cases depends on several factors (Vrbova, et al., 2013), including the willingness of a case to be interviewed, the type of follow up protocol used (face to face vs phone call vs mail), time passed from illness to the case interview (recall bias), and location of the case (cases in low population density areas are more difficult to contact).
Future research and recommendations

This thesis identified high infection rates in children based on individual-and area-level analysis (Chapters 2 & 3). This finding should be followed up in Ontario by future prospective cohort or case-control studies to identify novel risk factors (e.g., environmental sources, transmission routes, exposure settings) that contribute to increased infection rates in this age group.

A number of area-level socioeconomic status indicators (high average number of children at home per census family, high and low average median family income, and medium visible minority population proportion) were positively associated with S. Enteritidis infection rates in three PHUs within the GTA (Chapter 3). Future individual-level studies should be conducted in these areas and in these socioeconomic status groups to identify potential infection sources and risk factors that might explain their high S. Enteritidis infection rates.

A hypothesis generating study identified several high S. Enteritidis infection rate areas across the city of Toronto (Chapter 4). Future hypothesis testing studies should be conducted in these areas to confirm our study finding, and to evaluate risk factors that had an effect on the clustering of S. Enteritidis infections.

Phage type-specific differences in the exposure settings of cases were identified in Ontario health regions (Chapter 5). Knowing the exposure setting of cases is valuable, however, the point of S. Enteritidis contamination of food products might be at different levels (e.g., farm, processing plant, transport truck, or food premise), and future studies should evaluate the most likely contamination point. Moreover, future studies should
identify the most probable type of food products (e.g., poultry meat, eggs, processed food, or raw vegetables) that caused infections.

A high proportion of PT 8, 13, and 13a cases had unknown exposure setting information that warrants further evaluation by public health authorities to identify likely exposure settings and risk factors of these phage types. Currently, case investigators at Ontario public health units are not using standardized case follow up protocols or questionnaires. The accuracy, successful follow-up and known exposure setting rates, and comparability between case investigations results could be improved by developing these resources.

This thesis applied several exploratory and spatial statistical tools and identified a number of *S. Enteritidis* infection clusters; however during the study period no outbreaks were identified in Ontario by public health authorities. Spatial statistical tools should be applied regularly by case investigators at each PHU to increase foodborne disease outbreak identification rates. The feasibility of using spatial statistical methods during outbreak investigations should be evaluated, and these methods should be compared to the more frequently used traditional outbreak investigation techniques.

Future studies are needed from other Canadian provinces to analyze provincial passive foodborne disease surveillance data. Evaluating the epidemiology of *S. Enteritidis* infections in other Canadian provinces will enable Canadian public health authorities to implement a national *S. Enteritidis* prevention and control strategy.
Final conclusions

This thesis is the first study in Ontario that evaluated the epidemiology of human *S. Enteritidis* infections that were reported to Ontario’s passive disease surveillance system from January 2007 to December 2009. A number of spatial, temporal, ecological, and subtyping methods were used to assess individual and area-level demographic and socioeconomic status indicators, exposure settings, clinical symptoms, and spatial-temporal clustering of *S. Enteritidis* infections in Ontario. A stepwise approach was used to analyze the surveillance data, starting with more general descriptive or spatial exploratory steps, followed by more specific statistical methods. Taken together, this thesis highlighted the effectiveness of combining GIS technology, conventional, subtyping, and spatial epidemiological methods to evaluate the epidemiology of human *S. Enteritidis* infections in Ontario. The findings of this thesis will aid public health policy makers and practitioners to prioritize and target prevention and control programs to reduce the burden of *S. Enteritidis* infections in Ontario. The study methodology can be applied by public health authorities prospectively in real-time for outbreak detection or retrospectively for prevention and control program planning. Although this thesis evaluated *S. Enteritidis* infections in Ontario, the study methodology could be applied to the analysis of other foodborne diseases in industrialized countries.
References


Darcey VL, Quinlan JJ: Use of geographic information systems technology to track critical health code violations in retail facilities available to populations of different socioeconomic status and demographics. J Food Prot 2011, 74(9):1524-30.


Chapter 2

Legend 2.1. Example of data structure of the Poisson regression model used to evaluate associations between *S. Enteritidis* infections and demographic and seasonal factors.

<table>
<thead>
<tr>
<th>Number of cases</th>
<th>Year</th>
<th>Season</th>
<th>Age-range</th>
<th>Gender</th>
<th>Public Health Unit</th>
<th>Population Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2007</td>
<td>Fall</td>
<td>30-34</td>
<td>Female</td>
<td>City of Toronto Health Unit</td>
<td>108866</td>
</tr>
<tr>
<td>0</td>
<td>2007</td>
<td>Spring</td>
<td>30-34</td>
<td>Female</td>
<td>City of Toronto Health Unit</td>
<td>108866</td>
</tr>
<tr>
<td>2</td>
<td>2007</td>
<td>Summer</td>
<td>30-34</td>
<td>Female</td>
<td>City of Toronto Health Unit</td>
<td>108866</td>
</tr>
<tr>
<td>1</td>
<td>2007</td>
<td>Winter</td>
<td>30-34</td>
<td>Female</td>
<td>City of Toronto Health Unit</td>
<td>108866</td>
</tr>
<tr>
<td>1</td>
<td>2007</td>
<td>Fall</td>
<td>30-34</td>
<td>Male</td>
<td>City of Toronto Health Unit</td>
<td>104869</td>
</tr>
<tr>
<td>1</td>
<td>2007</td>
<td>Spring</td>
<td>30-34</td>
<td>Male</td>
<td>City of Toronto Health Unit</td>
<td>104869</td>
</tr>
<tr>
<td>4</td>
<td>2007</td>
<td>Summer</td>
<td>30-34</td>
<td>Male</td>
<td>City of Toronto Health Unit</td>
<td>104869</td>
</tr>
<tr>
<td>0</td>
<td>2007</td>
<td>Winter</td>
<td>30-34</td>
<td>Male</td>
<td>City of Toronto Health Unit</td>
<td>104869</td>
</tr>
<tr>
<td>1</td>
<td>2008</td>
<td>Fall</td>
<td>30-34</td>
<td>Female</td>
<td>Waterloo Health Unit</td>
<td>18384</td>
</tr>
<tr>
<td>0</td>
<td>2008</td>
<td>Spring</td>
<td>30-34</td>
<td>Female</td>
<td>Waterloo Health Unit</td>
<td>18384</td>
</tr>
<tr>
<td>0</td>
<td>2008</td>
<td>Summer</td>
<td>30-34</td>
<td>Female</td>
<td>Waterloo Health Unit</td>
<td>18384</td>
</tr>
<tr>
<td>0</td>
<td>2008</td>
<td>Winter</td>
<td>30-34</td>
<td>Female</td>
<td>Waterloo Health Unit</td>
<td>18384</td>
</tr>
<tr>
<td>0</td>
<td>2008</td>
<td>Fall</td>
<td>30-34</td>
<td>Male</td>
<td>Waterloo Health Unit</td>
<td>19080</td>
</tr>
<tr>
<td>1</td>
<td>2008</td>
<td>Spring</td>
<td>30-34</td>
<td>Male</td>
<td>Waterloo Health Unit</td>
<td>19080</td>
</tr>
<tr>
<td>0</td>
<td>2008</td>
<td>Summer</td>
<td>30-34</td>
<td>Male</td>
<td>Waterloo Health Unit</td>
<td>19080</td>
</tr>
<tr>
<td>0</td>
<td>2008</td>
<td>Winter</td>
<td>30-34</td>
<td>Male</td>
<td>Waterloo Health Unit</td>
<td>19080</td>
</tr>
<tr>
<td>0</td>
<td>2008</td>
<td>Fall</td>
<td>30-34</td>
<td>Female</td>
<td>Waterloo Health Unit</td>
<td>18432</td>
</tr>
<tr>
<td>1</td>
<td>2008</td>
<td>Spring</td>
<td>30-34</td>
<td>Female</td>
<td>Waterloo Health Unit</td>
<td>18432</td>
</tr>
<tr>
<td>0</td>
<td>2008</td>
<td>Summer</td>
<td>30-34</td>
<td>Female</td>
<td>Waterloo Health Unit</td>
<td>18432</td>
</tr>
<tr>
<td>1</td>
<td>2008</td>
<td>Winter</td>
<td>30-34</td>
<td>Female</td>
<td>Waterloo Health Unit</td>
<td>18432</td>
</tr>
<tr>
<td>0</td>
<td>2009</td>
<td>Fall</td>
<td>30-34</td>
<td>Male</td>
<td>Waterloo Health Unit</td>
<td>18950</td>
</tr>
<tr>
<td>1</td>
<td>2009</td>
<td>Spring</td>
<td>30-34</td>
<td>Male</td>
<td>Waterloo Health Unit</td>
<td>18950</td>
</tr>
<tr>
<td>1</td>
<td>2009</td>
<td>Summer</td>
<td>30-34</td>
<td>Male</td>
<td>Waterloo Health Unit</td>
<td>18950</td>
</tr>
<tr>
<td>0</td>
<td>2009</td>
<td>Winter</td>
<td>30-34</td>
<td>Male</td>
<td>Waterloo Health Unit</td>
<td>18950</td>
</tr>
<tr>
<td>1</td>
<td>2008</td>
<td>Fall</td>
<td>55-59</td>
<td>Female</td>
<td>YorkRegional Health Unit</td>
<td>30605</td>
</tr>
<tr>
<td>1</td>
<td>2008</td>
<td>Spring</td>
<td>55-59</td>
<td>Female</td>
<td>YorkRegional Health Unit</td>
<td>30605</td>
</tr>
<tr>
<td>0</td>
<td>2008</td>
<td>Summer</td>
<td>55-59</td>
<td>Female</td>
<td>YorkRegional Health Unit</td>
<td>30605</td>
</tr>
<tr>
<td>2</td>
<td>2008</td>
<td>Winter</td>
<td>55-59</td>
<td>Female</td>
<td>YorkRegional Health Unit</td>
<td>30605</td>
</tr>
<tr>
<td>0</td>
<td>2008</td>
<td>Fall</td>
<td>55-59</td>
<td>Male</td>
<td>YorkRegional Health Unit</td>
<td>30007</td>
</tr>
<tr>
<td>0</td>
<td>2008</td>
<td>Spring</td>
<td>55-59</td>
<td>Male</td>
<td>YorkRegional Health Unit</td>
<td>30007</td>
</tr>
<tr>
<td>0</td>
<td>2008</td>
<td>Summer</td>
<td>55-59</td>
<td>Male</td>
<td>YorkRegional Health Unit</td>
<td>30007</td>
</tr>
<tr>
<td>0</td>
<td>2008</td>
<td>Winter</td>
<td>55-59</td>
<td>Male</td>
<td>YorkRegional Health Unit</td>
<td>30007</td>
</tr>
</tbody>
</table>
## APPENDIX B

### Chapter 2

**Legend 2.2 for Figure 2.1. Ontario Public Health Units labels and names.**

<table>
<thead>
<tr>
<th></th>
<th>The District of Algoma Health Unit</th>
<th>19</th>
<th>Northwestern Health Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Brant County Health Unit</td>
<td>20</td>
<td>City of Ottawa Health Unit</td>
</tr>
<tr>
<td>3</td>
<td>Durham Regional Health Unit</td>
<td>21</td>
<td>Oxford County Health Unit</td>
</tr>
<tr>
<td>4</td>
<td>Elgin-St. Thomas Health Unit</td>
<td>22</td>
<td>Peel Regional Health Unit</td>
</tr>
<tr>
<td>5</td>
<td>Grey Bruce Health Unit</td>
<td>23</td>
<td>Perth District Health Unit</td>
</tr>
<tr>
<td>6</td>
<td>Halimand-Norfolk Health Unit</td>
<td>24</td>
<td>Peterborough County-City Health Unit</td>
</tr>
<tr>
<td>7</td>
<td>Haliburton, Kawartha, Pine Ridge District Health Unit</td>
<td>25</td>
<td>Porcupine Health Unit</td>
</tr>
<tr>
<td>8</td>
<td>Halton Regional Health Unit</td>
<td>26</td>
<td>Renfrew County and District Health Unit</td>
</tr>
<tr>
<td>9</td>
<td>City of Hamilton Health Unit</td>
<td>27</td>
<td>The Eastern Ontario Health Unit</td>
</tr>
<tr>
<td>10</td>
<td>Hastings and Prince Edward Counties Health Unit</td>
<td>28</td>
<td>Simcoe Muskoka District Health Unit</td>
</tr>
<tr>
<td>11</td>
<td>Huron County Health Unit</td>
<td>29</td>
<td>Sudbury and District Health Unit</td>
</tr>
<tr>
<td>12</td>
<td>Chatham-Kent Health Unit</td>
<td>30</td>
<td>Thunder Bay District Health Unit</td>
</tr>
<tr>
<td>13</td>
<td>Kingston, Frontenac, and Lennox and Addington Health Unit</td>
<td>31</td>
<td>Timiskaming Health Unit</td>
</tr>
<tr>
<td>14</td>
<td>Lambton Health Unit</td>
<td>32</td>
<td>Waterloo Health Unit</td>
</tr>
<tr>
<td>15</td>
<td>Leeds, Grenville and Lanark District Health Unit</td>
<td>33</td>
<td>Wellington-Dufferin-Guelph Health Unit</td>
</tr>
<tr>
<td>16</td>
<td>Middlesex-London Health Unit</td>
<td>34</td>
<td>Windsor-Essex County Health Unit</td>
</tr>
<tr>
<td>17</td>
<td>Niagara Regional Area Health Unit</td>
<td>35</td>
<td>York Regional Health Unit</td>
</tr>
<tr>
<td>18</td>
<td>North Bay Parry Sound District Health Unit</td>
<td>36</td>
<td>City of Toronto Health Unit</td>
</tr>
</tbody>
</table>
APPENDIX C

Chapter 4

Legend 4.1 for Table 4.1 and Figures 4.6 and 4.7. Toronto forward sortation area labels.
### APPENDIX D

**Chapter 5**

Legend 5.1 for Figure 5.1. Ontario Public Health Unit labels, names, and population estimates.

<table>
<thead>
<tr>
<th>Label</th>
<th>Public Health Unit</th>
<th>Health Region</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>The District of Algoma Health Unit</td>
<td>North East</td>
<td>118,594</td>
</tr>
<tr>
<td>1</td>
<td>Brant County Health Unit</td>
<td>Central West</td>
<td>138,538</td>
</tr>
<tr>
<td>2</td>
<td>Durham Regional Health Unit</td>
<td>Central East</td>
<td>612,247</td>
</tr>
<tr>
<td>3</td>
<td>Elgin-St. Thomas Health Unit</td>
<td>South West</td>
<td>90,074</td>
</tr>
<tr>
<td>4</td>
<td>Grey Bruce Health Unit</td>
<td>South West</td>
<td>163,416</td>
</tr>
<tr>
<td>5</td>
<td>Haldimand-Norfolk Health Unit</td>
<td>Central West</td>
<td>111,823</td>
</tr>
<tr>
<td>6</td>
<td>Haliburton, Kawartha, Pine Ridge District Health Unit</td>
<td>Central East</td>
<td>177,736</td>
</tr>
<tr>
<td>7</td>
<td>Halton Regional Health Unit</td>
<td>Central West</td>
<td>493,450</td>
</tr>
<tr>
<td>8</td>
<td>City of Hamilton Health Unit</td>
<td>Central West</td>
<td>530,785</td>
</tr>
<tr>
<td>9</td>
<td>Hastings and Prince Edward Counties Health Unit</td>
<td>Eastern</td>
<td>162,777</td>
</tr>
<tr>
<td>10</td>
<td>Huron County Health Unit</td>
<td>South West</td>
<td>60,921</td>
</tr>
<tr>
<td>11</td>
<td>Chatham-Kent Health Unit</td>
<td>South West</td>
<td>110,458</td>
</tr>
<tr>
<td>12</td>
<td>Kingston, Frontenac, and Lennox and Addington Health Unit</td>
<td>Eastern</td>
<td>194,212</td>
</tr>
<tr>
<td>13</td>
<td>Lambton Health Unit</td>
<td>South West</td>
<td>132,173</td>
</tr>
<tr>
<td>14</td>
<td>Leeds, Grenville and Lanark District Health Unit</td>
<td>Eastern</td>
<td>168,219</td>
</tr>
<tr>
<td>15</td>
<td>Middlesex-London Health Unit</td>
<td>South West</td>
<td>451,884</td>
</tr>
<tr>
<td>16</td>
<td>Niagara Regional Area Health Unit</td>
<td>Central West</td>
<td>443,064</td>
</tr>
<tr>
<td>17</td>
<td>North Bay Parry Sound District Health Unit</td>
<td>North East</td>
<td>126,394</td>
</tr>
<tr>
<td>Label</td>
<td>Public Health Unit</td>
<td>Health Region</td>
<td>Population</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------------------------</td>
<td>---------------</td>
<td>------------</td>
</tr>
<tr>
<td>18</td>
<td>Northwestern Health Unit</td>
<td>North West</td>
<td>83,548</td>
</tr>
<tr>
<td>19</td>
<td>City of Ottawa Health Unit</td>
<td>Eastern</td>
<td>881,833</td>
</tr>
<tr>
<td>20</td>
<td>Oxford County Health Unit</td>
<td>South West</td>
<td>106,898</td>
</tr>
<tr>
<td>21</td>
<td>Peel Regional Health Unit</td>
<td>Central East</td>
<td>1,308,538</td>
</tr>
<tr>
<td>22</td>
<td>Perth District Health Unit</td>
<td>South West</td>
<td>77,228</td>
</tr>
<tr>
<td>23</td>
<td>Peterborough County-City Health Unit</td>
<td>Central East</td>
<td>138,516</td>
</tr>
<tr>
<td>24</td>
<td>Porcupine Health Unit</td>
<td>North East</td>
<td>86,717</td>
</tr>
<tr>
<td>25</td>
<td>Renfrew County and District Health Unit</td>
<td>Eastern</td>
<td>103,588</td>
</tr>
<tr>
<td>26</td>
<td>The Eastern Ontario Health Unit</td>
<td>Eastern</td>
<td>199,134</td>
</tr>
<tr>
<td>27</td>
<td>Simcoe Muskoka District Health Unit</td>
<td>Central East</td>
<td>513,367</td>
</tr>
<tr>
<td>28</td>
<td>Sudbury and District Health Unit</td>
<td>North East</td>
<td>200,041</td>
</tr>
<tr>
<td>29</td>
<td>Thunder Bay District Health Unit</td>
<td>North West</td>
<td>157,100</td>
</tr>
<tr>
<td>30</td>
<td>Timiskaming Health Unit</td>
<td>North East</td>
<td>34,593</td>
</tr>
<tr>
<td>31</td>
<td>Waterloo Health Unit</td>
<td>Central West</td>
<td>517,090</td>
</tr>
<tr>
<td>32</td>
<td>Wellington-Dufferin-Guelph Health Unit</td>
<td>Central West</td>
<td>272,091</td>
</tr>
<tr>
<td>33</td>
<td>Windsor-Essex County Health Unit</td>
<td>South West</td>
<td>402,663</td>
</tr>
<tr>
<td>34</td>
<td>York Regional Health Unit</td>
<td>Central East</td>
<td>1,013,932</td>
</tr>
<tr>
<td>35</td>
<td>City of Toronto Health Unit</td>
<td>Toronto</td>
<td>2,685,203</td>
</tr>
</tbody>
</table>