THE RELATIONSHIP BETWEEN WATER AND HELICOBACTER PYLORI AND THE BURDEN OF RELATED ILLNESSES IN THE TOWNSHIP OF LANGLEY, BRITISH COLUMBIA

by

Neil Bellack

B.Sc., McGill University, 2001

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

FACULTY OF GRADUATE STUDIES

(Health Care and Epidemiology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2008

© Neil Bellack, 2008
ABSTRACT

The objectives of this dissertation are to (1) propose a conceptual framework describing the role of water in *H. pylori* transmission; (2) construct a database of water system and environmental characteristics using a geographic information system (GIS); (3) investigate the prevalence and incidence of *H. pylori*-related illnesses; and (4) examine the association between water system and environmental variables and *H. pylori* infection. The setting for this work was the Township of Langley, British Columbia.

Based on findings from epidemiological and microbiological studies, a conceptual model of water’s role in *H. pylori* transmission was developed (Chapter 2). A population-based approach was employed in the construction of the GIS database to assign the risk factors outlined in the model to each Township resident (Chapter 3).

Using administrative health services records for *H. pylori*-related illness, the average annual prevalence of gastric cancer, peptic ulcer disease and gastritis was 20, 450 and 1,777 cases per 100,000 population respectively, and average annual incidence was 13, 268 and 899 cases per 100,000 respectively. There was a significant decrease in the prevalence and incidence of peptic ulcer disease and the incidence of gastritis in the study population over the follow-up period, however, the costs to the health care system remained high (Chapter 4).

In a nested case-control study, variables from the GIS database were linked to a database of serological results used to ascertain positive *H. pylori* infection. *H. pylori* infection was associated with mixed water source (adjusted OR = 0.63, versus groundwater only) and non-residential land use (adjusted OR=1.58, versus residential) among younger cases (Chapter 5). Odds ratios diminished in the older age groups, suggesting the presence of a cohort effect and that the exposures investigated were most relevant for younger individuals.

The findings highlight possible transmission routes, relevant within the Canadian (and North American) context and evidence for public health interventions with regards to water supply and land use. Further, the methodology linked a variety of administrative data to cover all residents
of the study area and assigned environmental variables over time, and can serve as a model for other environmental epidemiologic studies.
# TABLE OF CONTENTS

Abstract ........................................................................................................................................... ii

Table of contents ............................................................................................................................ iv

List of tables ................................................................................................................................... viii

List of figures .................................................................................................................................. ix

Acknowledgements ......................................................................................................................... x

Dedication ...................................................................................................................................... xi

Co-authorship statement ............................................................................................................... xii

1 Chapter 1: Introduction and literature review ........................................................................ 1
  1.1 Background ..................................................................................................................... 1
  1.2 Rationale for the current work ........................................................................................ 2
    1.2.1 Thesis structure ....................................................................................................... 4
    1.2.2 Chapters and objectives ........................................................................................... 5
      1.2.2.1 Chapter 2: A conceptual model of water’s role as a reservoir in Helicobacter pylori transmission .......................................................... 5
      1.2.2.2 Chapter 3: Construction and characteristics of an environmental database for population health research ........................................ 5
      1.2.2.3 Chapter 4: The prevalence and incidence of Helicobacter pylori-related illnesses in the Township of Langley, British Columbia ................ 5
      1.2.2.4 Chapter 5: The influence of residential environment on Helicobacter pylori infection ............................................................... 6
  1.3 Literature review ............................................................................................................. 8
    1.3.1 Helicobacter pylori .................................................................................................. 8
      1.3.1.1 Characteristics ................................................................................................. 8
      1.3.1.2 Diagnosis ........................................................................................................ 8
      1.3.1.3 Epidemiology of H. pylori infection ................................................................. 9
      1.3.1.4 Risk factors for H. pylori infection ................................................................. 11
      1.3.1.5 Associated health outcomes ............................................................................. 14
        1.3.1.5.1 Gastric cancer .......................................................................................... 14
        1.3.1.5.2 Peptic ulcer disease ............................................................................... 15
        1.3.1.5.3 Gastritis ................................................................................................. 17
      1.3.1.6 Environmental microbiology ........................................................................... 17
1.3.2 Use of administrative data for epidemiological research ........................................ 17
  1.3.2.1 Validity and reliability of administrative data .................................................. 18
  1.3.2.2 Validity of H. pylori-related diagnoses .......................................................... 18
  1.3.2.3 The BC Linked Health Database ...................................................................... 20
1.3.3 Use of geographic information systems for epidemiological studies ................. 20

1.4 Introduction summary ................................................................................................... 21

1.5 References ..................................................................................................................... 22

2 Chapter 2: A conceptual model of water’s role as a reservoir in Helicobacter pylori transmission ................................................................. 35
  2.1 Background .................................................................................................................. 35
  2.2 Detection of H. pylori in water and the environment .............................................. 36
    2.2.1 Methods of detection .......................................................................................... 37
    2.2.2 Experimental studies ......................................................................................... 38
    2.2.3 Detection in the natural environment ................................................................. 38
    2.2.4 Viability of H. pylori in water ............................................................................. 39
  2.3 The epidemiology of transmission of H. pylori in water ........................................ 40
    2.3.1 Water transmission factors .............................................................................. 41
    2.3.2 Evidence against waterborne transmission ...................................................... 44
  2.4 Difficulties in investigating waterborne transmission of H. pylori ......................... 44
  2.5 Waterborne H. pylori transmission model ............................................................. 45
  2.6 Conclusion .................................................................................................................... 48
  2.7 References ..................................................................................................................... 49

3 Chapter 3: Construction and characteristics of an environmental database for population health research .................................................................................................................. 56
  3.1 Background .................................................................................................................. 56
  3.2 Methods ........................................................................................................................ 57
    3.2.1 Overview ............................................................................................................ 57
    3.2.2 Environmental database .................................................................................... 58
      3.2.2.1 Variable descriptions .................................................................................. 62
      3.2.2.2 Geographic information systems ................................................................. 66
        3.2.2.2.1 Variables linked by spatial join ................................................................. 68
        3.2.2.2.2 Variables linked by property identifier .................................................. 71
3.2.2.3 Variables assigned according to water system ........................................... 71
3.3 Distribution of variables .................................................................................... 72
3.4 Discussion .......................................................................................................... 76
3.5 References ........................................................................................................... 82

4 Chapter 4: The prevalence and incidence of Helicobacter pylori-related illnesses in the Township of Langley, British Columbia ................................................................. 85
4.1 Background ......................................................................................................... 85
4.2 Methods .............................................................................................................. 86
4.3 Results ................................................................................................................ 89
4.3.1 Prevalence of H. pylori illnesses .................................................................. 89
4.3.2 Incidence ....................................................................................................... 92
4.4 Discussion .......................................................................................................... 97
4.5 References ........................................................................................................ 102

5 Chapter 5: The influence of drinking water and residential environment on Helicobacter pylori infection .............................................................................................................. 106
5.1 Introduction ....................................................................................................... 106
5.2 Methods ............................................................................................................ 107
5.2.1 Sampling frame for cases and controls ....................................................... 107
5.2.2 Case identification ..................................................................................... 108
5.2.3 Case-control matching .............................................................................. 108
5.2.4 Water system and related environmental variables ..................................... 108
5.2.5 Analysis ....................................................................................................... 110
5.3 Results ................................................................................................................ 111
5.4 Discussion ........................................................................................................ 116
5.5 References ........................................................................................................ 124

6 Chapter 6: Concluding chapter ............................................................................. 129
6.1 Thesis summary ................................................................................................ 129
6.1.2 Summary of results ..................................................................................... 129
6.1.2.1 Chapter 2: A conceptual model of water’s role as a reservoir in Helicobacter pylori transmission ........................................................................................................ 129
6.1.2.2 Chapter 3: Construction and characteristics of an environmental database for population health research ................................................................. 130
6.1.2.3 Chapter 4: The prevalence and incidence of Helicobacter pylori-related illnesses in the Township of Langley, British Columbia........................................ 131

6.1.2.4 Chapter 5: The influence of drinking water and residential environment on Helicobacter pylori infection................................................................. 131

6.2 Contributions .............................................................................................................. 132

6.2.1 Helicobacter pylori literature .................................................................................. 132

6.2.2 Environmental epidemiology research ................................................................... 133

6.2.3 Stakeholder involvement ....................................................................................... 136

6.3 Limitations and future work ....................................................................................... 136

6.3.1 Data access ............................................................................................................. 136

6.3.2 Data quality .......................................................................................................... 137

6.3.3 Methodology and study design ............................................................................. 138

6.3.4 Interpretation of results ....................................................................................... 139

6.4 Impact on policy and future research ....................................................................... 140

6.5 References ................................................................................................................ 143

Appendices .......................................................................................................................... 144

Appendix A – Behavioural Research Ethics Board Certificate of Approval.................. 144
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>ICD9 codes for <em>H. pylori</em>-related illnesses</td>
<td>19</td>
</tr>
<tr>
<td>Table 2.1</td>
<td>Important factors in hypothesized waterborne <em>H. pylori</em> transmission cycle</td>
<td>47</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Variable descriptions and origin</td>
<td>60</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Variable descriptions and origin (continued)</td>
<td>61</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Distribution of key primary water variables at beginning and end of study period</td>
<td>74</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Prevalence, health care utilization and costs of gastric cancer, peptic ulcer disease and gastritis</td>
<td>90</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Incidence of gastric cancer, peptic ulcer disease and gastritis</td>
<td>94</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Demographic characteristics of incident cases of <em>H. pylori</em>-related illnesses</td>
<td>96</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Proportional distribution of demographic and environmental variables for cases and controls</td>
<td>112</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Correlations between water system and water source variables</td>
<td>113</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Percentage of cases and controls with one exposure for duration of study period</td>
<td>113</td>
</tr>
<tr>
<td>Table 5.4</td>
<td>Unadjusted odds ratios and 95% confidence intervals for all age groups</td>
<td>115</td>
</tr>
<tr>
<td>Table 5.5</td>
<td>Adjusted odds ratios and 95% confidence intervals in parsimonious logistic regression models</td>
<td>116</td>
</tr>
<tr>
<td>Table 6.1</td>
<td>Cohort members and person-days linked per year of study</td>
<td>135</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1.2  The prevalence of *H. pylori* by age in the developed and developing world ... 10
Figure 2.1  Methods of *H. pylori* detection from carriers and environmental samples...... 37
Figure 2.2  Conceptual model of waterborne *H. pylori* transmission pathways ............ 43
Figure 3.1  GIS shapefile description ........................................................................ 67
Figure 3.2A Spatial join of data from surface elevation layer to parcels layer ............. 69
Figure 3.2B Spatial join of data from soil type layer to parcels layer ......................... 70
Figure 3.2C Spatial join of data from water wells layer to parcels layer ...................... 70
Figure 3.3  Parcel history during study period ................................................................ 73
Figure 3.4  Land use distribution at start and end of study period ............................... 76
Figure 4.1  Annual active prevalence of peptic ulcer disease by age .......................... 91
Figure 4.2  Annual active prevalence of gastritis by age ............................................. 92
Figure 4.3  Annual incidence of peptic ulcer disease by age ......................................... 93
Figure 4.4  Annual incidence of gastritis by age ........................................................... 93
Figure 4.5  Incident gastritis cases by month for children under 16 years of age .......... 95
ACKNOWLEDGEMENTS

I would like to acknowledge many people for their support in pursuit of this degree. I would especially like to thank my supervisor, Mieke Koehoorn, for generously sharing her time and guiding me through this process. Throughout my doctoral work, she encouraged me to develop independent thinking and research skills, stimulated my analytical thinking and assisted me with scientific writing. I am also very grateful for the support of Ying MacNab, Hans Schreier and Muhammad Morshed for their help and encouragement on my thesis committee.

I wish to thank the students and staff in the Department of Health Care and Epidemiology. This experience would not have been nearly as rich and fulfilling without a group of such intelligent, diverse and motivated individuals to learn from. The Department staff has made my life easier on too many occasions to count and for that I am grateful.

I would also like to express my gratitude to the Bridge Strategic Training Program and to the Michael Smith Foundation for Health Research for their support. Specifically I would like to recognize the hard work of Kay Teschke as she made the Bridge Program a great success and a wonderful learning environment.

Finally, I would like to acknowledge my mother (Jill), father (Allan) and my sisters (Sydney and Licorice) for everything they’ve done for me throughout this degree and before. Of course, I would like to thank Anna-Maria, for putting up with my many late nights in front of the computer.
DEDICATION

To Anna-Maria
CO-AUTHORSHIP STATEMENT

This dissertation is nested within the larger Langley Water and Health Study funded by the Canadian Institutes of Health Research (CIHR). The investigators of the Langley Study are Judy Isaac-Renton (principal investigator), Kay Teschke, Hans Schreier, Ying Macnab and Jim Atwater. Hans Schreier and Ying Macnab were part of my thesis committee in addition to my supervisor, Mieke Koehoorn, and Muhammad Morshed.

With the guidance of my committee, I developed the plan for this thesis independently of the Langley Study, however, with the idea of using much of the data outlined in the original proposal to CIHR.

I conducted all data analysis and prepared the manuscripts that make up the chapters of this work. This was done in consultation with some of the aforementioned individuals. For example, statistical advice was provided throughout by Mieke Koehoorn and Ying Macnab. Jim Atwater and Hans Schreier were consulted for their expertise and knowledge of water systems and the Township of Langley.

A version of Chapter 2 has been published in the journal titled Epidemiology and Infection (Cambridge University Press) with co-authors Mieke Koehoorn, Ying Macnab and Muhammad Morshed. The article was published in 2006, issue 134, pages 439-449 under the title “A conceptual model of water’s role as a reservoir in Helicobacter pylori transmission: a review of the evidence”.

1  CHAPTER 1:  INTRODUCTION AND LITERATURE REVIEW

1.1  Background

A discovery made by two Australian physicians in the early 1980s altered the way we look at diseases of the upper gastrointestinal tract. Barry Marshall and Robin Warren reported the isolation of Campylobacter-like organisms in the gastric mucosa of individuals with chronic gastritis and the species became known as *Campylobacter pylori* [1]. Subsequent microscopic investigation revealed differences between this organism and other Campylobacter species. As a result, the genus Helicobacter was termed and the organism called *Helicobacter pylori* (*H. pylori*) [2]. In the years that followed, epidemiological methods would be used to demonstrate an association of *H. pylori* infection and gastritis, peptic ulcer disease and gastric cancer [3-8]. The discovery earned Marshall and Warren the 2005 Nobel Prize in Medicine.

It is estimated that over 50% of the global population are carriers of *H. pylori*, making it one of the most common bacterial infections in the world. Prevalence of infection is much greater in the developing world, approaching 100% in some areas [9-11]. In the developed world, prevalence in lower socio-economic subgroups of the population may be high, nearing that of developing countries [12]. In Canada, estimates of infection come from a few published studies. In a group of adults in Nova Scotia, prevalence ranged from 21% for study participants aged 20-29 years to 50% for those aged 70-79 years. Other studies conducted in native and First Nation communities have measured prevalence similar to developing countries [12-14].

Although only a reported 15-20% of carriers develop symptoms, the burden on the population and health care system is considerable [15]. Worldwide, there are an estimated six million incident cases of duodenal ulcer each year and 900,000 cases of gastric carcinoma [16]. In Canada, estimates of infection come from a few published studies. In Canada, peptic ulcers affect 3-5% of the population and there are an estimated 2,800 cases of gastric cancer diagnosed each year [17, 18].

Although the role of *H. pylori* in the development of certain health outcomes is well documented, the route of transmission into susceptible individuals is unknown making primary
prevention efforts difficult. Though the bacteria can be treated with antibiotics, disease can progress significantly before detection. Antibiotic resistant \emph{H. pylori} strains are increasingly common which reduces the ability to combat infection. Acid-reducing drugs that are used for symptom management may temporarily alleviate symptoms but do not eliminate the bacteria and can be harmful after chronic use.

Numerous studies have demonstrated that water may play a role in \emph{H. pylori} transmission. Water may be an intermediate in faecal-oral transmission by acting as a reservoir in which the bacteria can remain for periods of time before it is ingested [19]. If water is indeed a reservoir in \emph{H. pylori} transmission, it may represent an opportunity for a public health intervention to reduce transmission and therefore the burden of disease in the population.

\section*{1.2 Rationale for the current work}

Waterborne disease has long been a serious public health concern for British Columbians and Canadians [20]. Many harmful agents, such as Cryptosporidium, Escherichia coli, and Campylobacter are known waterborne pathogens. These pathogens have been implicated in numerous outbreaks all over Canada and continue to be the source of both endemic and epidemic cases of acute gastrointestinal illness [21].

What is less clear is the role that water and other environmental exposures play in the transmission of \emph{Helicobacter pylori} (\emph{H. pylori}). Experimental studies and analysis of environmental samples have demonstrated that \emph{H. pylori} has been detected in water and other environmental media [22-33]. Numerous cross-sectional epidemiological studies have indicated that drinking water is associated with \emph{H. pylori} infection and may be a vehicle for transmission [26, 27, 34-41]. In Canada, studies have been set in native communities that are often located in rural environments and have distinct social organization [12, 13, 42]. In general, epidemiological studies have lacked spatial characterization of an individuals’ environment as a means of describing relevant risk factors. Further, no previous studies employed a population-based approach to examine the risk of \emph{H. pylori} infection across a number of factors related to water supply and environment.
To address this gap, this thesis will study the relationship between *H. pylori* infection and environmental variables. It will take place in the Township of Langley in British Columbia, Canada, and will make extensive use of geographic information system-based data on water distribution systems, sewage disposal systems, and land use as exposure variables as well as other factors identified as important in the *H. pylori* transmission pathway between the environment and the individual. Outcome data will be derived from a database of serology results and the BC Linked Health Database that combines health data from a variety of sources and is able to link data for individuals across these sources. Based on results from previous studies, a higher proportion of *H. pylori* infection is expected among individuals relying on a private well-water system (versus municipal), whose water originated from a combination of surface and groundwater sources (versus groundwater only), and whose water was untreated (versus treatment, in the form of chlorination in the study area).

The Township of Langley is an ideal place to examine the relationship between *H. pylori* infection and water system characteristics such as water supply and water treatment. First, there is a mixture of urban and rural settings that produces a variety of exposures. For example, approximately two thirds of addresses are connected to the municipal water supply and half are connected to the municipal sewer system. The range of land uses in the Township also presents the opportunity for a variety of exposures. Second, the Township has a Water Resources Management Strategy that was developed in order to determine the best approach for the management and protection of its water supply. This initiative has produced an abundance of useful data describing characteristics related to water supply.

This work will focus on increasing our understanding of *H. pylori* transmission in order to identify potential targets for prevention measures that may decrease the burden of *H. pylori* infection on the population and health care system. A clear description of this phenomenon is especially important as population growth puts increased stress on our resources. As residential development expands into agricultural lands, it raises the potential for exposures that may increase the risk of *H. pylori* infection. Prevention is especially important as *H. pylori* has started to demonstrate resistance to antibiotics that have been used to eradicate it from the gastrointestinal systems of infection individuals. From a treatment perspective, the identification
of susceptible populations on the basis of specific risk factors would facilitate detection and treatment efforts.

1.2.1 Thesis structure

This work, consisting of six chapters, focuses broadly on the field of environmental epidemiology research and specifically on the relationship between water system characteristics and the risk of *H. pylori* infection, and the burden of *H. pylori* related-illness in a Canadian context.

The introductory chapter (Chapter 1) reviews the relevant literature on the epidemiology of *H. pylori*. Chapter 2 proposes a theoretical framework of *H. pylori* transmission through a review of the studies investigating water’s role as a reservoir in the transmission cycle. Chapter 3 describes the data sources and methods used to assemble the environmental database consisting of water system characteristics that may play a role in the transmission of waterborne illnesses. In addition to being used in this dissertation, the environmental database is of use for other studies looking at health outcomes related to water systems. Chapter 4 is a cohort study investigating the burden of *H. pylori*-related illnesses in the study area. Chapter 5 is an analysis of the relationship between environmental risk factors, focusing on water supply and water system characteristics, and *H. pylori* infection. The theoretical model proposed (Chapter 2) provides a rationale for many of the variables included in the environmental database (Chapter 3) and the study design used in the multivariable analysis (Chapter 5). The concluding chapter (Chapter 6) summarizes the findings of these articles and discusses the contributions, limitations and public health relevance of findings.

The following section in this chapter (Section 1.3) provides an overview of the material relevant to the issues discussed in these chapters including a description of the clinical characteristics, associated health outcomes, and risk factors for infection associated with *H. pylori*. Other important topics are also described including the use of administrative data in epidemiological research and the role of geographic information systems (GIS) in health research.
1.2.2 Chapters and Objectives

1.2.2.1 Chapter 2: A conceptual model of water’s role as a reservoir in Helicobacter pylori transmission

The primary objective of Chapter 2 is to review the literature on the role that water plays in the transmission of \textit{H. pylori}. Epidemiological and microbiological studies are reviewed and factors for \textit{H. pylori} infection related to water supply and other environmental exposures are highlighted. A conceptual model is introduced that summarizes the risk factors and discusses how the water supplies and water system characteristics could produce potential mechanisms for \textit{H. pylori} transmission.

1.2.2.2 Chapter 3: Construction and characteristics of an environmental database for population health research

Chapter 3 is a description of the methods used for the construction of the environmental database that is necessary for the epidemiological analysis of \textit{H. pylori} infection (Chapter 5). The variables on water supply and water system characteristics included in the database are surrogates for the presence of \textit{H. pylori} in the environment, but establishing a relationship between these variables and \textit{H. pylori} infection will provide evidence of the role of the water environment in transmission. Characteristics of the database, such as the distribution of variables across space and time, are discussed. Advantages and disadvantages of using administrative data in population health research are reviewed and challenges encountered during the construction of the database are highlighted.

1.2.2.3 Chapter 4: The prevalence and incidence of Helicobacter pylori-related illnesses in the Township of Langley, British Columbia

In Chapter 4, administrative data documenting health care utilization for all residents of the Township of Langley for the duration of the study period are analyzed. The objective of this chapter is to summarize the burden of gastric cancer, peptic ulcer disease and gastritis for the cohort. The accuracy and reliability of administrative data are reviewed and an algorithm is proposed that aims to increase the sensitivity and specificity of the outcomes investigated.
1.2.2.4 Chapter 5: The influence of residential environment on Helicobacter pylori infection

In this chapter, the relationship between water supply and water system characteristics described in Chapters 2 and 3 and \( H. \text{ pylori} \) infection are investigated in a nested case-control study among residents of the Township of Langley using a multiple logistic regression analysis. The methodological strengths and limitations of the study design are discussed as well as the public health implication of the findings.
Figure 1.1  Thesis framework of research questions and data sources
1.3 Literature review

The first sections below will provide background information on different aspects of *H. pylori* infection including means of diagnosis, patterns of infection in the population, associated health outcomes and the detection of the bacteria in the environment. The following section discusses the use of administrative data in epidemiological research as this study relies on the use of secondary data collection. Research investigating the validity and reliability of *H. pylori*-related illnesses is also summarized. Finally, the use of geographic information systems (GIS) for health research is reviewed. Some topics are briefly introduced below as they are discussed in greater detail later in other sections of this dissertation.

1.3.1 *Helicobacter pylori*

1.3.1.1 Characteristics

*H. pylori* are spiral-shaped, motile, gram-negative and microaerophilic (requiring very little oxygen) bacteria. They measure approximately 2.4 to 4.0 μm in length and 0.5 to 1.0 μm in width and have between two to six flagella that provide them with mobility [44]. They are able to survive the hostile, acidic environment of the stomach by embedding themselves in the mucous layer covering the stomach and duodenum [45]. Using the urease enzyme, they break down urea into ammonia (NH4+) that neutralizes the area around the bacteria. Ammonia serves as source of nitrogen (i.e. nourishment) for the bacteria and also causes damage to epithelial cells in the stomach [46]. While strains of *H. pylori* are phenotypically similar (i.e. they appear similar to one another) there is substantial genotypic diversity [47]. Nearly every carrier harbours a unique bacterial population and multiple strain infections have been observed repeatedly [48-51]. The high variety of strains may be explained by abnormally high rates of mutation and recombination [52].

1.3.1.2 Diagnosis

Diagnosis of *H. pylori* can be made using invasive or non-invasive methods. The gold standard involves the invasive method of gastroesophagoduodenoscopy and biopsy followed by one or more of culture, staining methods, or the detection of urease activity [45]. Recently, Graham et
al developed a means to obtain a culture using a brush that can be swallowed and extended into the stomach in order to minimize the invasiveness of this procedure [53].

Non-invasive methods include serologic techniques, the urea breath test, and a stool antigen test. Serologic tests detect *H. pylori* antibodies and indicate current or past infection and are therefore not useful to assess the success of antibiotic treatment for eradication. A meta-analysis of commercially available serological tests demonstrated a sensitivity of 85% and a specificity of 79% [54]. Test performance is dependent upon factors such as age, ethnicity and the population providing the reference sera [55]. Restandardizing the serologic test in the population in which it will be used is a good practice to achieve maximum sensitivity and specificity [56]. The accuracy of serology tests in children is not as good as in adults as sensitivity appears to increase with age, however specificity is similar for all age groups [57].

For the urea breath test, patients drink a urea solution labeled with carbon-13 or carbon-14. *H. pylori* produces carbon dioxide that is captured in the breath of the patient and measured. Samples are taken at different intervals after ingestion and compared to baseline samples [58]. The test has a sensitivity and a specificity of approximately 95 and 96%, respectively [54]. The stool antigen test can be used to detect current infection [59]. A monoclonal test has been shown to be more sensitive and specific that a polyclonal test (96% versus 91% and 97% versus 93%, respectively) [60].

He BC Centre for Disease Control (Vancouver, BC) conducts analysis of serological samples taken in the province. Blood samples are analyzed using an Immulite 2000 IgG (Siemens Medical Solutions Diagnostics, Los Angeles, California) that has a sensitivity and specificity of 91% and 100%, respectively [61].

### 1.3.1.3 Epidemiology of *H. pylori* infection

*H. pylori* remains one of the most common bacterial infections in the world. It is estimated that approximately half of the world’s population are carriers; however, there are distinct patterns of infection in developing and developed countries (Figure 1.2). In developing countries, most people are infected as children and remain carriers for the duration of their lifetime. In developed countries, the profile indicates that risk of infection is continuous throughout life. It is
also possible that a cohort effect exists, and rate of acquisition in childhood has decreased over time [62].

**Figure 1.2** The prevalence of *H. pylori* by age in the developed and developing world *

![Graph showing the prevalence of *H. pylori* by age in different countries.]

* From cross-sectional survey data [9, 63-69]

There are little data available to describe the prevalence and incidence of *H. pylori* infection in Canada. A study conducted in Eastern Canada indicated that adults seroconvert at a rate of 1% per year and prevalence of infection ranged from 21% for those aged 20-29 years up to 50% for those aged 70-79 years [70]. Other studies in Canada have focused on native communities in rural settings and prevalence estimates ranged from 51% to 95% [15, 42]. Among native children in Manitoba, 67% were positive by their second year [71] and rate of acquisition among children aged 0 to 12 years was estimated at 16% per year [12]. Estimates of prevalence vary from place to place and by study population, however a higher prevalence among older populations is a nearly universal finding regardless of the location or study sample. Patterns of prevalence in Canada most likely resemble those in other industrialized parts of the world such as the United States, Europe, Australia and New Zealand. In general, less than 10% of children under ten years of age are carriers [34, 72-74]. A population-based survey in the U.S. found an overall prevalence of 25% among children aged six to nineteen years old [75]. In older
populations in the U.S., Australia and Western Europe, prevalence ranges from 20% to over 50% [63, 76-80].

Countries in Asia such as Japan and Korea have unique \textit{H. pylori} prevalence patterns as they have transitioned from developing to developed countries. Younger populations have proportions of carriers similar to developed countries while the profile for older populations resembles developing countries [81, 82].

Measuring incidence of \textit{H. pylori} infection is more difficult since there is no known clinical sign of recent \textit{H. pylori} acquisition. In Finland, incidence of infection was estimated at 0.3% per year for children aged three and 12 years by examining serology samples taken at various cross-sections of time [83]. In the U.S., the highest rate of \textit{H. pylori} acquisition was found in children aged four to five years (2.1% annually) and decreased slightly with age until early adulthood, when 25% of individuals were carriers [84]. Incident rates between 1-3% have been measured in adults, indicating ongoing susceptibility throughout life [70, 85]. Spontaneous (i.e. untreated) clearance of infection appears to be relatively common and has been observed in people of all ages [70, 74, 78, 83, 86-88] but also may be explained by the use of a test with poor sensitivity or specificity or unreported use of antibiotics.

Some individuals, even those in high prevalence areas, never become carriers. The reason for this is mostly unknown, however, colonization is not favoured in individuals with atrophic gastritis and achlorhydria, the deficiency of hydrochloric acid in gastric secretions [89].

1.3.1.4 Risk factors for \textit{H. pylori} infection

The variability in \textit{H. pylori} prevalence and incidence across time and space can be explained by the presence of risk factors that have been identified from epidemiological studies. In general, differences in \textit{H. pylori} status have been identified for groups of individuals relying on different water supplies (i.e. municipal versus private supply), water sources (groundwater versus surface water), water usage behaviours and water storage patterns [26, 27, 34-41]. These findings have been replicated in numerous populations in a variety of settings and form the rationale for this dissertation. Water and environmental characteristics related to \textit{H. pylori} transmission and infection are discussed in detail in Chapter 2.
Age is a strong predictor of *H. pylori* infection with a much higher prevalence in older populations. A continuous risk of infection exists for adults, however, there appears to be periods of susceptibility during childhood [74, 84, 85]. Gender is not considered a significant risk factor for infection [44].

There is an inverse association between socioeconomic status (SES) and *H. pylori* infection. This applies both at the individual level, with people of lower SES at a greater risk of infection, and at the group level, where poorer communities and developing countries have higher prevalence and incidence rates. This relationship has been examined by using a variety of characteristics to describe SES such as occupation. Children in Italy whose parents were farmers, shepherds or semi-skilled workers were at a much greater risk of infection than children whose parents had graduated from university (OR = 8.7, 95% CI: 1.1 – 71.0) [90]. In a large study made up of individuals from 11 countries in Europe as well as Japan and the United States, only 34% of people with postsecondary education were seropositive versus 47% with secondary education and 62% with only primary school education [91]. This pattern of *H. pylori* infection with respect to education has been observed repeatedly [41, 75, 92, 93].

A greater availability of household appliances as a proxy of SES status has been identified as a protective factor after adjustment for other important variables (OR = 0.79, 95% CI: 0.67 – 0.92) [94]. Rural versus urban residence has also been investigated with children in rural areas being at greater risk of infection (OR = 3.8, 95% CI: 3.2 – 4.7) [90]. As with many measures of SES, a variety of other factors may have contributed to the difference observed by SES status.

Ethnicity has also been found to have an effect on carrier status. In some cases, this can be explained by differences in SES between ethnic groups [75]. For example, in children aged seven to nine years from the same socio-economic stratum, a substantially greater proportion of black children than white children were seropositive (40% versus 11%, p = 0.0001) [95]. Another study found that children living in the developed world whose fathers came from a developing country were at an increased risk of infection (OR = 6.2; 95% CI: 2.0 – 19.5), which probably is indicative of differences in SES [93].
Lower SES is not a proximal cause of *H. pylori* infection. The association between lower SES and hygiene is hypothesized to account for an increased exposure to the bacterium. Other factors related to SES and living conditions have also been identified as risk factors for infection. For example, a crowded household, measured by the number of people per room, increases risk of infection (OR = 2.58, 95% CI: 1.4-4.6) [96]. The presence of more than three children in a household was also a significant risk factor in multiple logistic regression (OR = 4.2; 95% CI: 1.2 – 14.6) [34]. Birth order has shown to be a significant predictor of *H. pylori* status with later children being at an increased risk of infection (OR = 1.19, 95% CI: 1.00 – 1.41 for 4th or 5th in birth order; OR = 1.47, 95% CI: 1.17 – 1.85 for 6th or more), as these children grow up in a more crowded environment [97]. Crowding in environments outside the home also increases risk of infection. Children attending day care or nursery had greater risk of infection compared to those who do not (OR = 1.5; 95% CI: 1.1 – 2.0) [90].

The presence of multiple carriers in a single family is a common observation. A greater proportion of children are colonized by *H. pylori* when a parent or other sibling is identified as a carrier [98-100]. It is hard to determine whether this is a result of direct person-to-person transmission or exposure to a common source.

Hygiene and *H. pylori* infection are also influenced by the availability of a water supply and sewage disposal in the home, with a higher risk for those in homes lacking these services [94, 101]. The presence of flush toilets in the home during childhood has also been shown to decrease the likelihood of infection (OR = 0.74, 95% CI: 0.55 – 0.96) [97]. The gradual drop in *H. pylori* prevalence observed in developed countries over the past half-century, a period coinciding with improvements in sanitation, supports this association from an ecological perspective [84].

Exposure to animals is another potential risk factor, however, it has been shown to both increase and decrease risk of infection. For example, an increased risk has been measured for people with dogs (OR = 1.8; 95% CI = 1.3 – 2.6) [90] and hamsters (OR = 2.4, 95% CI: 1.2 – 4.7) [34] and a decreased risk in other studies [102, 103]. In the former circumstance, the presence of pets in the home may exacerbate lower levels of hygiene while being an indicator of elevated SES in the latter.
Other variables have also been identified as risk factors for infection. Patterns in food consumption, such as the consumption of uncooked vegetables, have been associated with \textit{H. pylori} infection [40, 41]. The consumption of uncooked shellfish has been identified as an independent risk factor [41]. Altitude has been identified as a significant predictor of infection with those living at higher altitudes (above 1200 metres) being less likely to be infected (OR = 0.81; 95% CI: 0.71-0.92) [94]. Foreign travel, which may increase exposure to the bacteria depending on destination, has been shown to increase odds of infection [34].

This work will examine the influence of water supply and system factors associated with an individuals’ residence and the relations to the risk of \textit{H. pylori} infection. These factors such as such as water supply, (e.g. well water), water source (e.g. groundwater) and water treatment (e.g. chlorination) have an impact on the ability of water to act as a reservoir for \textit{H. pylori} (i.e to be present and to survive) and ultimately to be transmitted to humans. Multiple logistic regression allows for the measurement of the risk of infection associated with water supply and system characteristics, after adjusting for other important factors such as SES.

\textbf{1.3.1.5 Associated health outcomes}

While the prevalence of \textit{H. pylori} infection is high, not all carriers present clinically for the illnesses associated with it. This may be explained by characteristics of the bacterium and host and may include age of acquisition, other environmental exposures of host, differences in genetic makeup of hosts and differences in virulence among strains of \textit{H. pylori} [47]. \textit{H. pylori} has been identified as playing a role in the development of gastric cancer, peptic ulcer disease and gastritis.

\textbf{1.3.1.5.1 Gastric cancer}

The International Agency for Research on Cancer (IARC) has listed \textit{H. pylori} as a Class I carcinogen [104]. In Canada, an estimated 2,800 individuals were diagnosed with stomach cancer in 2007 and 1,850 died as a result of it [17]. There are an estimated 900,000 incident cases of gastric carcinoma per year worldwide [14].
The evidence for the relationship with cancer comes from epidemiological studies conducted in the early 1990s that were able to demonstrate the presence of \textit{H. pylori} infection prior to the development of disease. Analysis of a cohort of Japanese men enrolled in a cardiac health study from 1967 to 1970 revealed an odds ratio of 6.0 (95% CI: 2.1 – 17.3) for gastric carcinoma among those seropositive for \textit{H. pylori} antibodies and an additional risk identified among individuals with higher levels of antibodies [105]. In another cohort of people enrolled at a health maintenance organization in the U.S., patients with a positive serology test for \textit{H. pylori} demonstrated a higher odds for the development of gastric cancer (OR = 3.6, 95% CI: 1.8 – 7.3) with a much higher risk for blacks (OR = 9.0, 95% CI: 1.1 – 71.0) and females (OR = 18.0, 95% CI: 2.4 – 134.8) [106]. These findings have been replicated in other settings and have produced risk estimates in the same range or even higher [107, 108]. An ecological study of data taken from 13 countries found that the prevalence of \textit{H. pylori} infection is significantly related to gastric cancer incidence and mortality [109]. The proportion of gastric cancer cases associated with \textit{H. pylori} has been estimated to be between 50% and 90% [14, 110, 111].

\textit{1.3.1.5.2 Peptic ulcer disease}

The role of \textit{H. pylori} in the development of peptic ulcer disease (PUD) is well known and represents a considerable burden to the population and health care system. In a Canadian survey, approximately 3-5% of individuals reported the presence of a peptic ulcer that was diagnosed by a physician [18]. Estimates from Europe, the United States and Australia indicate that ulcers affect approximately 2-6% of the population [76, 112-115].

Evidence for the causal role of \textit{H. pylori} in the development of PUD comes from epidemiological studies comparing \textit{H. pylori} status in PUD patients and controls as well as studies that have documented the effect of \textit{H. pylori} eradication on PUD. As in studies of gastric cancer, nested case-control analyses in prospective cohorts were used to demonstrate the presence of \textit{H. pylori} prior to onset of peptic ulcer disease in order to satisfy the temporality criteria of causality.

In a nested case-control study, the odds ratios for the development of gastric ulcer and duodenal ulcer associated with \textit{H. pylori} infection were found to be 3.2 (95% CI: 1.6 – 6.5) and 4.0 (1.1 – 14.2), respectively [8]. In another matched case-control study, a positive serology test resulted in
an OR = 3.83 (95% CI, 1.4 – 10.2) that was only slightly attenuated (OR = 3.6) after adjustment for other important variables [5].

In a group of pediatric duodenal ulcer patients who were initially H. pylori-positive, all children who had successfully eradicated the bacteria had their ulcers healed at follow-up [116]. In a group of eradication studies, recurrence rates of PUD ranged between 0-27% at one year for patients with eradicated H. pylori compared to recurrence rates ranging from 67-95% when eradication was unsuccessful [117].

The proportion of cases of PUD associated with H. pylori is different for duodenal ulcers than for gastric ulcers. Many studies have estimated that between 85-100% of duodenal ulcers are associated with H. pylori infection in both adults and children, however one large study showed that only 73% of duodenal ulcer cases had the presence of H. pylori confirmed by two positive diagnostic tests and 81% confirmed by one positive test [99, 118-128]. If duodenal ulcer patients using antibiotics and non-steroidal anti-inflammatory drugs are removed, the prevalence of H. pylori is 99% [129] that indicates why nearly all duodenal ulcers in children are H. pylori related. The mechanism is related to the ability of pylori infection to produce chronic gastritis, a well-established risk factor for duodenal ulcer [130].

The proportion of gastric ulcers related to H. pylori infection is lower than for duodenal ulcers. Estimates as high as 75-85% have been reported and as low as 11-33%, although lower estimates were based on small study samples (n < 10) [126, 131-135]. The prevalence of idiopathic ulcers in adults seems to be increasing [76, 127].
1.3.1.5.3 Gastritis

The link between *H. pylori* infection and gastritis has been well documented and is universally accepted as it is almost ubiquitous in carriers. Four people who were exposed to the bacteria, two through voluntary ingestion and two through a contaminated endoscope, subsequently developed gastritis and had proven colonization in their upper gastrointestinal tract [136, 137]. In a group of children with nodular gastritis, 98.5% were *H. pylori* carriers [127]. In a group of adults, every carrier had gastritis while all non-carriers were gastritis-free [138]. Another study of volunteers found *H. pylori* in all patients with histologic gastritis and a significantly higher proportion of carriers with histologic duodenitis [139]. Other factors, such as the backflow of bile into the stomach, autoimmune disorders and exposure to non-steroidal anti-inflammatory drugs, alcohol, and cigarette smoke are also known causes for gastritis.

The prevalence of gastritis is difficult to determine because of the preponderance of asymptomatic cases and therefore the relationship between *H. pylori* and gastritis may have little clinical relevance. However, human challenge experiments have demonstrated a relationship between recent *H. pylori* acquisition and abdominal pain, indicating that gastritis-like symptoms may be a clue to recent exposure to *H. pylori* [117]. A sign of recent *H. pylori* acquisition is the key factor to the determination and understanding of *H. pylori* transmission.

1.3.1.6 Environmental microbiology

For *H. pylori* transmission to have an environmental component – whether it is via a short-term reservoir in the person-to-person transmission route or via longer-term survival on its own in an environmental reservoir – it must be able to exist in the environment for at least a small period of time. A number of studies have demonstrated the plausibility that this occurs using experimental methods or by analyzing environmental samples of water, soil and feces. The methods used for these studies as well as their results are discussed in detail in Chapter 2.

1.3.2 Use of administrative data for epidemiological research

Administrative data is the term used for information collected by non-research organizations for an administrative purpose and not primarily for use in research or surveillance [145]. This includes data on health care utilization such as hospital discharge summaries and claims for physician billings and prescription drugs.
The use of administrative data for epidemiological research carries several advantages. First, it makes it feasible to conduct population-based studies, as nearly everyone living in an entire region can be captured, especially in the context of the single payer system that exists in Canada. Administrative data are longitudinal by nature and allows the researcher to observe health care utilization trends at the level of the group and individual. The biggest weakness of using these data is that they were not collected for the purpose of research and therefore the accuracy of the data is often questioned.

In many health-related administrative databases such as medical billing records, the purpose for visits to physicians or for hospitalizations are recorded using International Classification of Disease codes (ICD) [146]. These codes have undergone numerous revisions and are currently in the 10th revision. However, the majority of research in the late 20th century and early 21st century has been conducted using the codes from the ninth revision or ICD9.

Codes can contain as many as five digits, with the first three explaining the general grouping of a given health state, and the fourth and fifth providing more specific information. For example, the ICD9 code 535 generally describes gastritis and duodenitis whereas the code 535.1 describes atrophic gastritis.

1.3.2.1 Validity and reliability of administrative data

The validity and reliability of administrative data has been investigated to address concerns regarding coding practices. The validity of hospital separations data has been investigated by comparing ICD codes to data in medical charts [147, 148]. To assess reliability of physician billings data, algorithms have been developed that identify relevant procedure codes or frequencies of specific complementary ICD9 codes [149, 150]. Reliability of these data has also been assessed by measuring agreement with survey data [151, 152].

1.3.2.2 Validity of H. pylori-related diagnoses

Table 1.1 contains a list of the three-digit ICD9 codes for H. pylori-related diseases. The ICD9 code for H. pylori infection itself can only be specifically identified using a full five digit code (041.86).
Table 1.1  ICD9 codes for *H. pylori*-related illnesses

<table>
<thead>
<tr>
<th>ICD9</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>151</td>
<td>Malignant neoplasm of stomach</td>
</tr>
<tr>
<td>531</td>
<td>Gastric ulcer</td>
</tr>
<tr>
<td>532</td>
<td>Peptic ulcer</td>
</tr>
<tr>
<td>533</td>
<td>Peptic ulcer, site unspecified</td>
</tr>
<tr>
<td>534</td>
<td>Gastrojejunal ulcer</td>
</tr>
<tr>
<td>535</td>
<td>Gastritis and duodenitis</td>
</tr>
</tbody>
</table>

Three studies investigated the validity of peptic ulcer diagnoses in hospitalizations data. In an assessment of the validation of diagnosis for complicated peptic ulcer in the Saskatchewan Hospital Services Plan database, 88% of cases were confirmed by the presence of a positive result of diagnostic or treatment procedures showing a bleeding or perforated ulcer or a clinically-based diagnosis of melena or hematemesis [153]. For complicated gastric ulcer (531), duodenal ulcer (532) and peptic ulcer (533), positive predictive values (PPVs) were 95%, 94% and 90% respectively. Using a similar study design for an administrative database in Italy, PPVs were 97% for gastric and duodenal ulcers (531 and 532), 84% for gastrojejunal ulcers (534) and 80% for peptic ulcers (533). In an administrative database consisting of data from eight health maintenance organizations (HMOs) in the U.S., the PPVs of ICD-9 codes for peptic ulcers were evaluated using the presence of a description of surgery, endoscopy, X-ray or autopsy in the medical chart [154]. The PPVs for duodenal ulcer (532), and gastric and gastrojejunal ulcer (531, 534) were 66% and 61%. The wide ranges of PPVs from the different HMOs (20-100% for duodenal ulcers and 0-70% for gastric and gastrojejunal ulcers) suggests that varying record keeping practices exist even within small geographic areas.

The reliability of gastric and duodenal ulcers as well as gastritis was measured by examining the presence of fee codes describing upper gastrointestinal endoscopy in the Ontario Health Insurance Plan physician claims database [149]. Using presence of appropriate fee codes as a gold standard, a low proportion of false positives were present as indicated by high specificity (90.8% for gastritis, 98.7% for gastric ulcer, and 99.4% for duodenal ulcer), however each outcome had a low sensitivity (51.3%, 29.6% and 50.0% respectively).

The available evidence suggests that administrative data are a good means of providing estimates for the prevalence and incidence of *H. pylori*-related illnesses in a population.
1.3.2.3 The BC Linked Health Database

The BC Linked Health Database (BCLHD) is a collection of databases describing the utilization of health care services, vital statistics and other population health data for BC residents that are housed at UBC’s Centre for Health Service and Policy Research (CHSPR) [155]. In the BCLHD, data for individuals can be linked over time and across different databases using unique identifiers such as Personal Health Number (PHN), Social Insurance Number (SIN) or a combination of other descriptive fields.

For the purposes of this study, the BCLHD includes Medical Services Plan records and Hospital Separations on discharges and transfers for the majority of residents in British Columbia as the provincial health care plan is nearly universal. Numerous studies have been conducted using data from the BCLHD in a variety of fields including population health, workplace health and safety, primary health care and health technology assessment [156].

1.3.3 Use of geographic information systems for epidemiological studies

Geographic information systems (GIS) are “automated systems for the capture, storage, retrieval, analysis and display of spatial data” that have proven particularly useful as research tools in environmental epidemiological studies [157]. Data that are recorded in GIS databases are linked to latitude and longitude coordinates and can be portrayed spatially in map format. Therefore, a GIS has all the functional capabilities of a traditional database such as entering and editing data as well as being able to integrate data from different types of spatial data formats such as image maps or line maps [158]. This multi-level format of data storage allows the GIS user to search and select data based either on a trait of a specific variable or spatial characteristic. For example, in a GIS of traffic patterns, the user can select specific roadways by traffic volume or proximity to a chosen landmark. Information can therefore be added to the database by merging based on agreement of a single field of group of fields (i.e. deterministic linkage), as done with traditional databases, or by joining spatial data based on shared geographic coordinates.

The use of GIS in epidemiology is increasingly common because of the inherent advantages of being able to analyze spatial relationships between environment and disease. Some of the earliest applications of GIS in epidemiology and public health focused on the surveillance and
monitoring of infectious diseases such as malaria and Lyme disease, and the modeling of population exposures to lead and power lines [159-162].

GIS have also been used in the epidemiologic investigation of waterborne illness. For example, a GIS was used to analyze spatial variations in diarrhoeal illness with water supply characteristics and the study found an association between groundwater consumption and disease incidence [163]. It has also been used to construct risk maps to identify areas of poor water quality and determine causes, investigate the relationship between water supply and cryptosporidiosis and to describe spatial trends in the incidence of giardiasis [164-166].

1.4 Introduction Summary

From the numerous environmental risk factors identified for *H. pylori* infection, there is ample evidence indicating the possibility of a transmission route for the bacteria involving water. However, limited research has been conducted in developing countries where the relationship between individual and environment are likely different from in developed countries.

The use of administrative data for population health research is beneficial because it can be used to capture information on risk factors and outcomes for all individuals residing in a given area. The diversity of variables used for analysis can be enhanced by using data from multiple sources. Geographic information systems software is an increasingly important component of population-based studies because it allows data from different sources and in different formats to be combined into a single database.

The burden of *H. pylori* infection and related illnesses may be decreasing in many parts of the world; however, the situation in Canada is unknown as little attention has been to this issue. Investigating factors related to the transmission mechanisms for *H. pylori* infection and the burden of illness will provide policy makers with updated information on the severity of this problem in Canada and may identify areas where preventative measures can be adopted.
1.5 References


Hunt RH, Sumanac K and Huang JQ. Review article: should we kill or should we save *Helicobacter pylori*? Aliment Pharmacol Ther 2001;15 Suppl 1:51-9


15. Hunt RH, Sumanac K and Huang JQ. Review article: should we kill or should we save *Helicobacter pylori*? Aliment Pharmacol Ther 2001;15 Suppl 1:51-9


36. Karita M, Teramukai S and Matsumoto S. Risk of Helicobacter pylori transmission from drinking well water is higher than that from infected intrafamilial members in Japan. Dig Dis Sci 2003;48:1062-7


89. Mendall MA. Transmission of *Helicobacter pylori*. Semin Gastrointest Dis 1997;8:113-23


164. Zeilhofer P, Zeilhofer LV, Hardoim EL, Lima ZM and Oliveira CS. GIS applications for mapping and spatial modeling of urban-use water quality: a case study in District of Cuiaba, Mato Grosso, Brazil. Cad Saude Publica 2007;23:875-84


2 CHAPTER 2: A CONCEPTUAL MODEL OF WATER’S ROLE AS A RESERVOIR IN HELICOBACTER PYLORI TRANSMISSION

2.1 Background

The isolation of the bacterium *Helicobacter pylori* and subsequent findings of its role in the development of chronic gastritis, peptic ulcers and gastric cancer represent major medical breakthroughs. These have had a considerable impact on reducing the burden of gastrointestinal ailments. Current treatments aim to eradicate *H. pylori* infection using antibiotics, whereas previous treatments involved surgery or erroneously focused on stress reduction, dietary changes, and antacid therapies that often resulted in recurrence of symptoms [1, 2].

It is estimated that over half of the world’s population are carriers of *H. pylori*, making it one of the most common bacterial infections in the world [3]. However, many individuals harbouring the bacteria never develop clinical symptoms. Differences in outcome are thought to result from a variety of factors including age at acquisition, exposures to other environmental agents, and genetic characteristics of both host and *H. pylori* bacterium [4]. Though the incidence of *H. pylori* infection and peptic ulcers and gastric cancer has decreased over the past century, it still represents a major burden [5]. Worldwide, there are an estimated six million incident cases of duodenal ulcer each year and 900,000 cases of gastric carcinoma [6].

There are two distinct patterns of *H. pylori* acquisition. In developing countries, most people are infected as infants or pre-adolescents and remain infected for the duration of their life whereas in developed countries it appears that infection is acquired gradually with increasing age [7]. The higher overall prevalence of infection in developing countries suggest that hygiene and environmental factors may play a significant role in transmission, a theory supported by data showing lower socio-economic status as a factor associated with infection among defined populations [8, 9].

---

1 Published in *Epidemiology and Infection*: Bellack NR, Koehoorn MW, MacNab YC, Morshed MG. A conceptual model of water's role as a reservoir in *Helicobacter pylori* transmission: a review of the evidence. Epidemiol Infect. 2006; 134:439-49.
While the major health outcomes of *H. pylori* infection are well documented, routes of transmission remain unclear, making it difficult to implement public health measures to prevent infection. Primary prevention is particularly important as antibiotic-resistant strains exist and further antibiotic use may give rise to increased resistance among *H. pylori* and other bacteria [10]. Aside from iatrogenic modes of transmission, where patients have been infected via contaminated endoscope, person-to-person pathways have been proposed via faecal-oral or oral-oral routes [11, 12]. Water may be an intermediate in faecal-oral transmission, by acting as a reservoir in which the bacteria can remain for periods of time before it is ingested as drinking water, accidentally during bathing, or through other pathways involving food. The purpose of this paper is to review the microbiological and epidemiological evidence to determine what is currently known about the feasibility of water as a reservoir for transmission. A conceptual model of possible transmission pathways is proposed (Figure 2.2) which incorporates this evidence and provides direction for future research on the risk of transmission associated with environmental factors.

A literature search was conducted in the Medline, Web of Science and Environmental Sciences and Pollution Management databases for articles published from 1985 to the present. A combination of the following keywords were used: “*Helicobacter pylori*”, “water”, “faecal/fecal”, “transmission”, “epidemiology” and “microbiology”. Abstracts were read by the principal author and papers selected if they presented original research investigating (1) environmental factors associated with *H. pylori* infection or transmission; (2) other factors related to environmental transmission of *H. pylori* through water systems or sources; (3) methods used to detect *H. pylori* in the water supply, or; (4) environmental sampling of *H. pylori* in the natural environment. A total of 80 articles met the inclusion criteria and were reviewed for evidence to inform a hypothesized conceptual model of water as a reservoir in *H. pylori* transmission.

### 2.2 Detection of *H. pylori* in water and the environment

Over the past 15 years, substantial attention has been paid to the development of methods to detect *H. pylori* in water. These have been employed in a variety of settings and have contributed to the hypothesis of waterborne transmission. Figure 2.1 presents an overview of these methods and demonstrates that viable *H. pylori* has not been cultured from environmental
sources. However, sufficient evidence does exist to formulate a conceptual model of how water may be involved in the transmission pathway. Below is a summary of the evidence leading to these proposed pathways and the hypothesized conceptual model is presented in Figure 2.2.

**Figure 2.1  Methods of H. pylori detection from carriers and environmental samples**

*C*Possible morphological transformation into coccoid form;  **Unable to distinguish viable from non-viable forms

2.2.1 Methods of detection

Culturing *H. pylori* from areas outside the human stomach has been difficult because of a morphological change in the bacterium and overgrowth by competing micro-organisms [13, 14]. Attempts to culture the bacteria from inoculated water samples have been made using alternative growth media under variable conditions [15, 16]. The most common detection techniques are the polymerase chain reaction (PCR), autoradiography, or microscopic investigation of a stained sample. Sensitivity of these methods is estimated by testing *H. pylori*-inoculated samples while
specificity is determined by testing samples inoculated with other helicobacter organisms and structurally similar bacteria.

2.2.2 Experimental studies

Studies in controlled laboratory settings have revealed some information about the behaviour of *H. pylori* in the aqueous environment. Generally, these studies involve the inoculation of aqueous samples with *H. pylori* at different concentrations and under different conditions and follow-up tests are carried out to determine the effect on survival.

Survival times as long as twenty days have been reported in distilled water [17]. Increased survival times are associated with lower temperatures (< 10°C), pH levels of 5.8 to 7.0, and milk and saline solutions compared to water [17-22]. Survival may be dependent on even more specific conditions of the aqueous environment including the presence of other microorganisms such as zooplankton and free-living amoebae [23, 24]. Experimental work has also shown that *H. pylori* can incorporate into biofilms on surfaces in aqueous environments which is important since growth in water is probably negligible if it occurs at all, and biofilms may provide a substrate that would allow multiplication to occur [25-27]. The addition of chlorine to water supplies has proven effective at inactivating *H. pylori* [28].

Overall, these findings indicate that if introduced into an aqueous environment, *H. pylori* can survive for periods of time, and the length of that time is affected by a number of specific conditions of the aqueous environment.

2.2.3 Detection in the natural environment

Experimental studies have enabled the development of detection methods and demonstrated the possibility of water serving as a reservoir for *H. pylori*. PCR and microscopic techniques have been employed for the examination of environmental samples taken from all over the world.

*H. pylori* has been detected in wastewater systems, which is not surprising given that carriers are known to shed the bacteria in their faeces [29, 30]. However, positive results in both pre- and post-treated wastewater suggests that wastewater treatment may be ineffective in removing *H. pylori* from this aqueous environment [30]. *H. pylori* has been detected in surface water and
groundwater samples in various parts of the world. This includes places in both the developing and developed world, where human or animal faecal contamination was likely in water intended for human consumption [29-34]. *H. pylori* DNA detected in water pots and in a cast iron water pipe from a municipal water system suggests that biofilms may play a role in transmission in real world settings [35, 36]. Zero percent prevalence in chlorine-treated drinking water offers further evidence that it may be an effective method for disinfection [30].

The identification of *H. pylori* DNA on flies and in cow faeces, and in raw and pasteurized milk samples illustrates that animals may directly or indirectly play a role in zoonotic-related water transmission [32, 37]. The presence of *H. pylori* in sheep with the absence of an immune response indicates that they may be a natural host for the bacteria [38]. The isolation of *H. pylori* and other *Helicobacter* species from the stomachs of various animals, including dogs and cats, suggests that zoonotic transmission is a possibility. However, the role of non *pylori* species in this process is unknown [39-42].

Generally, the detection of *H. pylori* in the environment has shown that water and water sources linked to zoonotic reservoirs are a possibility that should be considered in the hypothesized link between water and *H. pylori* infection. However, these results alone do not verify that they are involved in transmission. The viability of *H. pylori* once it enters water and its capability to colonize the human stomach upon ingestion are key components to solving the question of waterborne transmission.

### 2.2.4 Viability of *H. pylori* in water

Although *H. pylori* becomes non-culturable by traditional methods after entering water, it may exist in a viable but non-culturable (VBNC) state. *H. pylori* undergoes a general morphological change from spiral to coccoid and U-shaped forms in water that probably contributes to the loss of culturability [17, 43, 44]. It has been suggested that transformation to the coccoid form is the manifestation of cell death and that it enables the bacterium to resist potentially adverse effects of entering aquatic environments, which has also been observed with other bacteria [17, 45-47].

Determining the viability of *H. pylori* in water is challenging because the methods used to detect the bacteria in water and the environment are unable to distinguish between viable and dead
bacteria [15]. Other factors such as the presence of yet unidentified helicobacter species may account for false positive results when testing for *Helicobacter pylori* [10]. Positive results using some PCR primers but not others indicate that this may be occurring [48].

In light of this, a number of approaches have been taken to determine the viability of *H. pylori* in water. For example, although non-culturable, *H. pylori* was deemed viable after immersion in natural freshwater as determined by a LIVE/DEAD BacLight assay that distinguishes between living and dead bacteria on the basis of intactness of cell membranes [49, 50]. Another approach is to culture for a limited period of time, which prevents complete overgrowth of competing bacteria, followed up by PCR [14]. Based on these and other findings, it seems likely that *H. pylori* passes through a viable but non-culturable state for a brief period before proceeding to cell death [51]. As demonstrated in previous studies, it is likely that the speed of the conversion process is affected by the aqueous environment and the presence of other organisms in the water [23, 49].

However, the question remains as to whether this non-culturable form is able to colonize. The ability of the coccoid form to colonize has been tested using animal models, though results have been contradictory. Successful colonization was observed in mice after intragastric colonization with the coccoid form, however colonization was unsuccessful in piglets [52, 53]. Experiments with other bacteria indicate that they can maintain their pathogenicity in the VNBC state and are able to multiply after a dormant phase thereby making colonization possible [54-57].

In summary, the presence of *H. pylori* has been detected in the environment with various tests; however, it has not been convincingly cultured from naturally occurring samples. Isolation from the environment by means of culture would demonstrate that viable *H. pylori* is present outside the human stomach and further strengthen the theory of waterborne transmission.

### 2.3 The epidemiology of transmission of *H. pylori* in water

The epidemiology of *H. pylori* transmission has been studied extensively and numerous risk factors have been identified. This section reviews studies that have examined water source as a factor in *H. pylori* transmission in addition to other factors that may be important in the water transmission pathway.
2.3.1 Water transmission factors

Water source has been implicated as a factor for the transmission *H. pylori* bacteria in epidemiological studies in a variety of settings, mostly in developing countries. In different settings, water from different sources (i.e. municipal versus private) is variable in quality because of the diversity in environmental influences and water distribution systems. For example, a study of Peruvian children found that those relying on municipal water were significantly more likely to have a positive urea breath test (UBT) than those using community well water (OR = 11.4, p = 0.02) [58]. In Lima, Peru, municipal water is taken from the Rimac River, which is heavily contaminated by industrial and domestic pollution. A study in Germany found that children drinking from non-municipal sources were at an increased risk of infection compared to those relying on municipal water supplies (OR = 2.8, 95% CI: 1.0 – 8.2; p < 0.05) [59]. In Germany, the municipal water supply is generally thought of as the cleaner water supply and, although private wells exist for economic reasons, there is a ban on drinking untreated water [60]. Other surveys indicate that the prevalence of infection is higher among people drinking untreated water or water from shallow wells, and among people lacking tap water or an internal water supply in the home [58, 61-66].

Findings from other studies strengthen the possibility that water plays a role in *H. pylori* transmission. In a rural area of The Gambia, infants enrolled in a cohort study at a time when the deep bore hole water supply was disrupted had a median age of first positive urea breath test (UBT) of eight weeks, while those recruited when the water supply was re-established had median first positive UBT at 28 weeks [36]. A cross-sectional survey in Kazakhstan found that people with a high clean water index (CWI) (based on frequency of boiling water, behaviours regarding storing and reusing water, and frequency of bathing) were less likely to have a positive UBT than those with a medium (OR = 1.9, 95% CI: 1.3-3.1) or low CWI (OR = 5.1, 95% CI: 1.8-20.5) after adjustment for age and socio-economic status [64]. The importance of water storage has been demonstrated in a study in rural Bolivia, where a lower seroconversion rate was observed for children whose families used a narrow-mouthed water vessel with a lid compared to those using the traditional, open-mouthed water vessel (OR = 0.4; 95% CI: 0.2-0.9) [67]. Frequency of swimming in streams and other waterways susceptible to contamination are also noted transmission factors for infection [62, 68].
The presence of *H. pylori* DNA as detected by PCR in water close to human-impacted areas may be an indication of faecal contamination and suggests the importance of faecal-oral transmission [69]. Evidence of this can be found in relationships between infection and the presence and type of sewage system. People with indoor bathrooms are generally less likely to be infected and the presence of a flush toilet is protective of *H. pylori* infection [63, 64, 66, 70, 71]. Though these findings may suggest that transmission occurs by direct faecal-oral contamination, a closer look may be needed to investigate the interaction between sewage disposal and water, a feasible pathway for transmission that has not been thoroughly investigated in the literature.

Some studies suggest that other vehicles may also be involved in water-related *H. pylori* transmission. For example, the consumption of uncooked vegetables has been identified as a risk factor in different settings [68, 72]. Bacterial contamination may have originated from irrigation water containing faecal matter as *H. pylori* has been detected in water used for agricultural purposes in at least a few settings [30, 72]. The role of food in *H. pylori* transmission is discussed in more detail elsewhere [73].
Zoonotic transmission may occur with animals acting as carriers for *H. pylori*. A high prevalence of infection among shepherds compared to farmers offers evidence that sheep may pass on the bacteria [74]. The presence of pets has also been identified as a factor associated
with transmission of infection in some situations but not in others [59, 75-77]. In some cases, the presence of pets is an indicator of elevated social status, while in others, it may exacerbate lower levels of hygiene. High levels of \textit{H. pylori} antibodies found in workers exposed to fresh cut animal parts relative to those in clerical positions without animal exposures further suggests that \textit{H. pylori} infection may be zoonotic [78]. While direct contact with animals remains one possible mode of transmission, indirect transmission may occur when animals contaminate susceptible surface water or groundwater sources used for drinking water, which is more likely to occur in areas of high-density agricultural activity.

2.3.2 Evidence against waterborne transmission

Epidemiological evidence against common-source transmission such as drinking water comes from studies finding a negative association between \textit{H. pylori} infection and hepatitis A virus (HAV), which is used as a marker of exposure to faecal matter because it is transmitted via faecal-oral pathways [79-81]. However, low rates of coinfection do not eliminate the possibility of a common-source transmission as hepatitis A is more contagious than \textit{H. pylori}, hepatitis A virus infection lasts a relatively short period of time, multiple routes of transmission may exist for each organism and the distribution of the two microorganisms in the population may be different [82]. Absence of differences in \textit{H. pylori} prevalence in workers occupationally exposed and not exposed to sewage and low rates of seroconversion among travelers to developing countries have led some to reject the notion of common-source pathways [83-86]. However, this ignores the likelihood that people take special precautions to avoid potentially high-risk exposures while at work or traveling.

2.4 Difficulties in investigating waterborne transmission of \textit{H. pylori}

The principal challenge with conducting environmental sampling is the inability to demonstrate the existence of viable \textit{H. pylori} in water that is able to then colonize upon entering the stomach or duodenum. Despite the detection of nonviable \textit{H. pylori} using other methods, the concept of waterborne transmission is likely to remain in question until it is cultured from natural sources. However, we have presented microbiological and epidemiological evidence to suggest that water is a feasible transmission route for \textit{H. pylori} infection. In the absence of methods to culture \textit{H. pylori} from natural sources, this theory warrants further investigation from a public health perspective.
From an epidemiological perspective, many studies investigating *H. pylori* transmission face similar limitations to one another. In general, *H. pylori* transmission has been difficult to investigate because of the high proportion of asymptomatic chronic infection and the absence of a known indicator of recent *H. pylori* acquisition. Studies of children, especially in developing countries, may be biased because some non-invasive diagnostic tests are less accurate in children than adults and validation of diagnostic tests in the study population may be less common [87, 88]. The urea breath test is more reliable in children older than six years [89]. Its use in studies of infants has yielded high rates of apparent spontaneous clearance, which is unlikely [90]. Therefore, studies that use only a single urea breath test measurement to determine *H. pylori* carrier status may be inaccurate and in these situations, non-differential misclassification is likely to bias risk estimates towards the null.

In studies where drinking water is the exposure of interest, it is difficult to track all drinking water sources for an individual as they may rely on several sources in a single day and even water from a single source may be quite variable in quality over time. Other problems arise when understanding the importance of factors that may be relevant to more than one transmission pathway. For example, clustering patterns in families may be an indication of either oral-oral transmission via direct contact or the sharing of drinking glasses, or of fecal-oral transmission through a shared exposure to a specific water source.

### 2.5 Waterborne *H. pylori* transmission model

The plausibility of waterborne *H. pylori* transmission has been demonstrated by various microbiological and epidemiological studies. Based on this evidence, a conceptual model of waterborne transmission is presented in Figure 2.2. The most important assumption of this model is that humans and animals are long-term hosts and water is a relatively short-term reservoir in which *H. pylori* may remain for a period of time before it is spread to susceptible humans through direct consumption of water containing the bacteria or indirectly through the consumption of contaminated food. A human carrier will spread the bacteria by shedding it in their faeces. At this point, the bacteria may come into contact with a susceptible person and infect him or her through direct faecal-oral transmission or the faeces may enter bodies of water. This can occur if people defecate directly into, or if sewage effluent has contact with water used for drinking. Animal carriers can also contaminate water supplies by defecating directly into
surface water or if their faeces penetrate into groundwater that is either unconfined or that has a high water table. Soil type is also an important factor because it affects penetrability of pathogens into groundwater. Heavy rainfall events may play an important role as they may facilitate the spread of manure that contains the bacteria.

Once the bacterium enters water, it may remain there until it is ingested by a person as drinking water, during recreational activities, or using food as a vehicle. As microbiological studies have shown, *H. pylori* may have a limited survival time in water and thus colonization (i.e. successful transmission) may be dependent upon the amount of time between introduction of the bacteria in water and ingestion by a susceptible person. As temperature has been as an important variable with respect to survival in water, seasonal cycles may also be possible. Other factors, such as pH and the presence of specific microorganisms or suspended materials may also affect waterborne transmission. The absence of the bacteria in the original water source (i.e. river or well) may not prevent infection if it is delivered through a pipe or stored in a container where *H. pylori* exists in a biofilm. Treatment of water is probably helpful to minimize the risk of waterborne transmission. Table 2.1 outlines important components of the waterborne *H. pylori* transmission cycle. Other possible transmission routes may involve direct person-to-person transmission through oral-oral routes [12, 91] and through direct exposure to animals.
Table 2.1 Important factors in hypothesized waterborne *H. pylori* transmission cycle

**Entry into water**
- Proximity to human and animal impacted areas
- Penetrability of soil
- Precipitation
- Susceptibility of water to contamination
- Presence of *H. pylori* in biofilm in water pipe or container

**Ingestion**
- Drinking water
- Consumption of raw vegetables
- Swimming or bathing

**Survival**
- Length of time in water
- Water temperature
- Water pH
- Presence of other microorganisms or suspended sediments
- Water treatment

**Infection**
- Viability of *H. pylori*
- Susceptibility of host

This model incorporates many of the factors related to water systems and sources for *H. pylori* transmission. For example, transmission via the waterborne pathway explains the higher prevalence and lower age of acquisition of *H. pylori* in developing countries. This is because the possibility for contamination of water is greater in areas where sewer systems and water treatment facilities are less developed. In addition, the time between *H. pylori* being introduced into water and ingestion by a susceptible individual may be shorter because water treatment, a time-consuming and risk-reducing step, is less common. The absence of indoor plumbing, which is seen more often in the developing world, requires that people rely on closest available sources of water for drinking, regardless of contamination. A lack of running water also means that people bathe in open water, where ingestion may occur. The model also supports geographic clusters of infected individuals as the transmission cycle may be contained within susceptible water systems. Thus, the presence of different environmental influences, such as livestock, around water systems explains why infection is more common in certain areas. However, it is
important to recognize that even if the most frequent route of transmission involves water, other routes are plausible and are likely responsible for at least a small proportion of cases as can be seen in the model.

2.6 Conclusion

The details of *H. pylori* transmission remain unclear, hampering efforts to reduce its burden of illness. Microbiological and epidemiological evidence have repeatedly indicated the possibility that water may be a reservoir in faecal-oral *H. pylori* transmission. The proposed model indicates that the risk of *H. pylori* infection is greater in areas where faecal contamination of water intended for human consumption is more likely and explains why people depending on more susceptible water sources may be at higher risk of infection.

Future studies should employ a combination of methodologies from microbiology and epidemiology. As microbiological methods develop further, it may enhance our ability to understand the physical nature and important characteristics of the bacterium as it enters and colonizes the human stomach. Phenotyping and genotyping of strains isolated in humans and water will further clarify the steps involved in waterborne transmission. To date, many epidemiological studies investigating the relationship between drinking water and *H. pylori* infection have simply looked at water source but not the additional factors that may contribute to the presence of the bacterium in water and hence increase the risk of transmission. Researchers in future studies need to examine exposures to a number of known factors hypothesized to affect water quality including: water source (e.g. surface water and groundwater); type of groundwater source (e.g. confined, semi-confined, and unconfined); the presence of water treatment; type of sewage disposal system; surrounding land use (residential versus agricultural and specific agricultural); and soil type around groundwater sources. By examining these possible types of exposure, the role of water in the *H. pylori* transmission may be clarified and lead to the development of interventions that can reduce the burden of *H. pylori*-related disease in the population. In summary, this article highlights important areas to address in future research, including investigation and specification of the mechanisms and processes by which water-related characteristics affect and influence the transmission of *H. pylori* infection.
2.7 References


33. Karita M, Teramukai S and Matsumoto S. Risk of *Helicobacter pylori* transmission from drinking well water is higher than that from infected intrafamilial members in Japan. Dig Dis Sci 2003;48:1062-7


35. Park SR, Mackay WG and Reid DC. Helicobacter sp recovered from drinking water biofilm sampled from a water distribution system. Water Research 2001;35:1624-1626


82. Sundaram KK, Mendall MA. Helicobacter pylori and hepatitis A: a common route but uncommon mode of infection? Helicobacter 1998;3:140-1


86. Potasman I, Yitzhak A. Helicobacter pylori serostatus in backpackers following travel to tropical countries. American Journal of Tropical Medicine & Hygiene 1998;58:305-8


CHAPTER 3: CONSTRUCTION AND CHARACTERISTICS OF AN ENVIRONMENTAL DATABASE FOR POPULATION HEALTH RESEARCH

3.1 Background

The Langley Water and Health Study is a population-based epidemiologic study with the primary aim of investigating the relationship between adverse health outcomes affecting the gastrointestinal (GI) system and environmental factors hypothesized to impact water quality. The study is set in the Township of Langley, British Columbia (BC), located in the Fraser Valley outside of the Greater Vancouver area. The Township has a current population of approximately 100,000 people and a land area of 303 square kilometers [1]. It is an ideal place to examine the relationship between health and water quality related factors because it includes areas of urban and rural living that brings about variability in the environmental variables to be studied. In addition, the Township’s Water Resource Management Strategy, established to determine the most effective approach for managing and protecting its water supply, has developed an abundance of data describing water-related characteristics.

In population-based studies, the cohort or sampling frame can be defined as all people living within a set of geopolitical boundaries [2]. This study design minimizes selection bias, a common challenge in observational research; however, large study populations often make it difficult to collect data at the level of the individual for all study members. To address this challenge, characteristics are often assigned to multiple individuals within a group defined by a similar set of characteristics. For environmental studies, this can include groups of individuals residing within large area boundaries, as people residing within these boundaries tend to share similar exposure characteristics. Depending on the size of the area, this method of assigning environmental variables may limit the study to be more ecologic in nature or lead to misclassification that can affect the validity of results [3].

---

2 This manuscript is being prepared for submission to the International Journal of Health Geographics. The authors of this manuscript are Bellack NR, Koehoorn MW, MacNab YC, Morshed MG, Teschke K, Atwater JW, Schreier H, Elmieh N, and Isaac-Renton JI.
The use of geographic information systems technology in epidemiological studies has increased greatly as the technology has advanced. It is able to incorporate spatial data by applying latitude and longitude coordinates to categories of variables in addition to carrying out every activity of traditional databases such as expanding, updating and editing data [4]. Data can therefore be added to the database based on common physical location (i.e. shared geographic coordinates) or by deterministic linkage that generates links based on full agreement of a single or group of identifiers [5]. The storage and presentation of data are known as a geographic information system (GIS) and contains a database (geodatabase) and corresponding map with the features described in the database.

The objectives of this paper are to describe the construction and characteristics of the environmental GIS database for the Langley Water and Health Study and to characterize the distribution of environmental factors related to the water system and source that will ultimately be assigned to Township residents based on their household residence. The distribution of risk factors will be described for the Township over time at the smallest relevant geographic level, defined in this study as each parcel of land with a street address in the Township during the study period. The discussion will focus on the primary water factors of interest for the hypothesized role in the transmission of *H. pylori* infection, the accuracy and reliability of data, the potential for exposure misclassification, and how these issues might affect the results of the epidemiological study. Note that only a subset of the environmental water database constructed for the larger Langley Water and Health Study was used in this work investigating the association with *H. pylori* infection, in particular those characteristics hypothesized to be related to water’s role as a reservoir in *Helicobacter pylori* transmission. However, for posterity, the construction of the full environmental water database is described below.

### 3.2 Methods

#### 3.2.1 Overview

The study population consisted of all people living in the Township of Langley from 1995 to 2003 inclusive, as defined by their postal code of residence recorded in the provincial Ministry of Health Client Registry. The Registry is comprehensive for most BC residents within the provincial universal health care system and is updated as new address information is provided to
the Ministry. For the epidemiological analyses, health outcomes will come from records of hospitalizations and physician visits also maintained by the Ministry and from a serological database maintained by the BC Centre for Disease Control.

The environmental GIS database was constructed from existing data sources with the unit of observation being each unique street address (defined by land parcel identifiers) in the Township during the study period. Environmental water system and source variables were assigned to individuals for the epidemiological analyses based on their residential street address(es) while living in the Township, as recorded and updated in the Registry. In order to protect the privacy of individuals, the linkage of environmental water-related factors to individuals by street address was conducted at the Ministry and all personal information was removed from the research database released to the study team.

3.2.2 Environmental Database

The environmental GIS was constructed to describe water system and water source characteristics that have either been shown or hypothesized to be determinants of poor water quality and are therefore factors for pathogen transmission for gastrointestinal (GI) illness. These variables, listed in Table 3.1, describe various features of the drinking water supply, sewage disposal system, uses of surrounding land, precipitation, and hydrogeologic attributes for each parcel in the Township. Primary variables are those with complete data and for which there are clear a priori hypotheses with respect to transmission of pathogens related to gastrointestinal illness. Secondary variables are either incomplete for the study area or study period but were included for exploratory investigation based on previous evidence of an association with GI illness. Tertiary variables lack a priori hypotheses because they have a less direct relationship with water quality but it is plausible that they influence outcome through a mechanism such as facilitating the transport of pathogens. It is hypothesized that a higher incidence of GI illness is associated with drinking water from private wells (versus municipal water supply), with septic systems (versus municipal sewer system), living on parcels designated for agriculture use (versus residential), receiving a mixture of surface water and groundwater (versus groundwater only), receiving untreated water, and following periods of high precipitation.
Environmental water-related data originated from numerous sources including: municipal, provincial and federal governments; the local health authority; academic and community surveys; and environmental assessments commissioned by the Township. In addition to assigning variables to each parcel, temporal variations in these variables were also captured (e.g. a switch from private well water supply to the municipal water supply).
Table 3.1  Variable descriptions and origin

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CATEGORIES</th>
<th>DATA ORIGIN</th>
<th>LINKAGE METHOD</th>
<th>LEVEL OF SOURCE DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Drinking water supply variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drinking water supply (1)</td>
<td>Categorical (municipal; private)</td>
<td>Municipal government</td>
<td>Property identifier</td>
<td>Parcel level (polygon)</td>
</tr>
<tr>
<td>Type of private water supply (1) *</td>
<td>Categorical (community; single-user)</td>
<td>Local health authority</td>
<td>Address</td>
<td>Parcel level (polygon)</td>
</tr>
<tr>
<td>Drinking water source (1)</td>
<td>Categorical (mixture; groundwater only)</td>
<td>Municipal government</td>
<td>Water system identifier</td>
<td>Water system</td>
</tr>
<tr>
<td>Well construction (2)</td>
<td>Categorical (drilled; dug)</td>
<td>Provincial government</td>
<td>Spatial join</td>
<td>Geographic location (point)</td>
</tr>
<tr>
<td>Well depth (for groundwater sources) (2)</td>
<td>Continuous</td>
<td>Provincial government</td>
<td>Spatial join</td>
<td>Geographic location (point)</td>
</tr>
<tr>
<td>Disinfection (1)</td>
<td>Categorical (chlorination; UV; none)</td>
<td>Municipal government</td>
<td>Water system identifier</td>
<td>Water system</td>
</tr>
<tr>
<td>Aquifer (3)</td>
<td>Categorical (14 major aquifers)</td>
<td>Environmental assessment</td>
<td>Spatial join</td>
<td>Geographic location (polygon)</td>
</tr>
<tr>
<td>Aquifer type (3)</td>
<td>Categorical (unconfined, shallow confined, deep confined)</td>
<td>Environmental assessment</td>
<td>Spatial join</td>
<td>Geographic location (polygon)</td>
</tr>
<tr>
<td><strong>Sewage disposal variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of sewage disposal system (1)</td>
<td>Categorical (sewer; septic)</td>
<td>Municipal government</td>
<td>Property identifier</td>
<td>Parcel level (polygon)</td>
</tr>
<tr>
<td>Age of septic system ** (2)</td>
<td>Date (MM/DD/YYYY)</td>
<td>Local health authority</td>
<td>Address</td>
<td>Parcel level (polygon)</td>
</tr>
<tr>
<td><strong>Land use variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General land use (1)</td>
<td>Categorical (residential; agricultural; other)</td>
<td>Municipal government</td>
<td>Property identifier</td>
<td>Parcel level (polygon)</td>
</tr>
<tr>
<td>Agriculture land use (1)</td>
<td>Categorical (crops; beef &amp; dairy; poultry; mixed; other)</td>
<td>Municipal government</td>
<td>Property identifier</td>
<td>Parcel level (polygon)</td>
</tr>
<tr>
<td><strong>Precipitation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily precipitation (1)</td>
<td>Continuous (mm)</td>
<td>Federal government</td>
<td>Spatial join</td>
<td>Geographic location (raster)</td>
</tr>
</tbody>
</table>
Table 3.1 Variable descriptions and origin (continued)

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CATEGORIES</th>
<th>DATA ORIGIN</th>
<th>LINKAGE METHOD</th>
<th>SPATIAL RESOLUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Other environmental variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation (3)</td>
<td>Discrete (2m contours)</td>
<td>Environmental assessment</td>
<td>Spatial join</td>
<td>Geographic location (polyline)</td>
</tr>
<tr>
<td>Surficial soil texture (3)</td>
<td>Categorical (2 major categories and 20 subcategories)</td>
<td>Environmental assessment</td>
<td>Spatial join</td>
<td>Geographic location (polygon)</td>
</tr>
<tr>
<td>Slope angle (3)</td>
<td>Continuous</td>
<td>Web: Geobase.ca</td>
<td>Spatial join</td>
<td>Geographic location (raster)</td>
</tr>
<tr>
<td>Water table elevation (3)</td>
<td>Continuous (Discrete, by 10m contours)</td>
<td>Environmental assessment</td>
<td>Spatial join</td>
<td>Geographic location (polyline)</td>
</tr>
<tr>
<td>Watershed (3)</td>
<td>Categorical (11 different watersheds)</td>
<td>Environmental assessment</td>
<td>Spatial join</td>
<td>Geographic location (polygon)</td>
</tr>
<tr>
<td><strong>Water quality variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coliforms – total and fecal (municipal) (2)</td>
<td>Discrete (cfu/ml)</td>
<td>Municipal government</td>
<td>Spatial join</td>
<td>Geographic location (point)</td>
</tr>
<tr>
<td>Coliforms – total and fecal (community) (2)</td>
<td>Discrete (cfu/ml)</td>
<td>Local health authority</td>
<td>Address</td>
<td>Parcel level</td>
</tr>
<tr>
<td>Nitrate levels (municipal) (2)</td>
<td>Continuous (mg/L)</td>
<td>Municipal government</td>
<td>Spatial join</td>
<td>Geographic location (point)</td>
</tr>
<tr>
<td>Nitrate levels (private) (2)</td>
<td>Continuous (mg/L)</td>
<td>University survey</td>
<td>Address</td>
<td>Parcel level</td>
</tr>
<tr>
<td>Turbidity (municipal) (2)</td>
<td>Continuous (NTU)</td>
<td>Municipal government</td>
<td>Spatial join</td>
<td>Geographic location (point)</td>
</tr>
</tbody>
</table>

* Nested in drinking water supply variable; ** Nested in type of sewage disposal variable
Data were provided to the study team in several formats: geographic information system files with data assigned by individual or sets of latitude and longitude coordinates; electronic data files with data recorded by a unique property identifier or street address; and hard copies of paper files with data recorded by street address. All hard copy records were entered into a study database using double entry techniques to reduce data entry errors.

### 3.2.2.1 Variable descriptions

The variables in the database focus on the drinking water system characteristics and environmental factors of the residential environment. The variables are described in greater detail below including variable category definition, data origin and a brief explanation of each variable as a risk factor of GI illness.

**Drinking water supply.** There were five distinct municipal water systems in the Township serving regions characterized by higher density (i.e. urban) living. Parcels that were connected to one of these systems were classified as having a municipal water supply. Otherwise, parcels were classified as private water users of either a single-user private system or multi-user private water system. Since the inception of the Safe Drinking Water Regulation in 1992 (which has since been revised as the Drinking Water Protection Act), all water systems with greater than one connection to a well that are not run by the municipality require a permit and records are maintained at the local health authority [6]. These multi-user private systems are known as community water systems.

Taxation records from the municipality indicate whether a parcel is connected to a municipal water system. Parcels not connected a municipal water supply were designated as private water systems and parcels identified from local health authority records were designated as community water systems.

Individuals relying on private water systems are expected to be at an increased risk of exposure to pathogens and gastrointestinal illness because they lack the safety measures put in place to protect municipal water supplies. Epidemiological studies have identified a higher risk of illness in private water supplies due to the limited resources available for maintaining and monitoring
these systems [7, 8]. As such, it is more likely that pathogens or microbial agents are able to pass through undetected and cause illness in the individuals relying on that water.

Drinking water source. There are two types of water sources in the Township. Parcels receive either a mixture of groundwater and surface water or water that comes from groundwater sources only. The different municipal water supply systems in the different parts of the Township are connected to municipal wells that pump water into reservoirs and ultimately, to those parcels connected to the municipal water system. In the western part of the Township, the water supply is supplemented with surface water from reservoirs located in the nearby coastal mountains. The proportion of surface water that enters the system is dependent upon demand for water at that time. Other municipal water systems and all private water systems are recipients of groundwater only. Parcels connected to the West System were classified as receiving water from mixed sources and all others will be categorized as relying on groundwater only.

A higher risk of gastrointestinal infections and illnesses is expected among those relying on surface waters because they are more susceptible to contamination [9]. Surface waters are thus expected to carry a greater concentration of pathogens and other potentially harmful agents.

Water treatment. At the beginning of the study period (1995), municipal water derived from groundwater sources was not treated. Chlorination units were installed in Township reservoirs at different points in the years that followed. The surface water that supplements the groundwater supply in the west water system is chlorinated prior to mixing. Township records were used to identify date of installation of the water treatment units. Water treatment was assigned to all parcels connected to the municipal water system starting on the date of installation of the chlorination units. Files from the local health authority, recorded by address, were used to assign treatment to community water systems based on the permit issuing date. Single-user private systems were assigned as having no water treatment.

Treating water, in this case primarily by means of chlorination, is intended to diminish the risk of gastrointestinal illness by deactivating pathogens present [10]. Thus, the absence of water treatment is a determinant of poor water quality and individuals relying on such water are at a higher risk of pathogen exposure/transmission and subsequent infection and illness.
**Sewage disposal system.** Parcels in the Township are either connected to the municipal sewer system or must rely on a private septic system to dispose of their sewage. Taxation records maintained by the municipality were used to identify the parcels connected to the sewer system. All other parcels were assumed to have a septic system.

The presence of a septic system on an individual’s property is a potential determinant of poor water quality because septic system effluent has the potential to leak into water sources, especially if not properly maintained [11]. Therefore, groundwater may contain pathogens that are shed in the feces of those using nearby septic systems.

**Land use.** There are diverse land uses in the Township as a result of the mixture of living environments (i.e. urban and rural) and the presence of the Agricultural Land Reserve. Parcels of land are designated for use as residential, agricultural, commercial, industrial, public and transportation, or communication. BC Assessment, the provincial crown corporation responsible for property assessments assigns each parcel a three-digit classification that categorizes the parcel into the groups mentioned above based on the first digit. The specific three-digit land use codes are also used to categorize the specific type of agricultural activity occurring on the parcel and distinguishes between crop farming and animal farming that may be associated with transmission of pathogens into the water supply. Some addresses have more than a single land use code associated with them because of multiple activities occurring in that space. For example, a property may have a residential structure on a parcel that is also designated for agricultural activities.

Proximity to agricultural activity is a determinant of poor water quality because the manure or sewage sludge used for fertilizer and the waste produced from livestock farming have the potential to enter waterways and spread pathogens [12].

**Precipitation.** In the study area, periods of high precipitation occur in the winter season and periods of low precipitation predominate in the summer. Daily precipitation data is collected at climate stations operated by Environment Canada and are stored in databases. Data were
downloaded from the National Climate Data Archive for the stations in and surrounding the Township.

Periods of high precipitation make it possible for wash events to occur and for surface pathogens to wash into water sources and increase the risk transmission of pathogens and of gastrointestinal illness [13, 14]. For pathogens with a defined incubation period, it is possible to investigate the relationship between periods of heavy precipitation and the incidence of gastrointestinal illness.

*Well characteristics.* Information on wells location (i.e. latitude and longitude) construction date, well depth, and type of well are reported by well drillers to the BC Ministry of Water, Land and Air Protection (now the Ministry of Environment), the provincial ministry responsible for environment preservation and maintenance. Parcels with wells located on them were assigned the information associated with that well.

Wells that draw water from sources closer to the surface are more likely to have poor water quality because of susceptibility to contamination. Therefore, dug wells are more susceptible than drilled wells, and shallow drilled wells are more susceptible than deep drilled wells.

*Water quality data.* Laboratory measurements of drinking water in the Township are made weekly for municipal and community water supplies. The Township is responsible for municipal water systems and community systems are generally maintained by a representative individual or group. Municipal systems have total and fecal coliform counts and turbidity measurements taken weekly and nitrates tested bi-annually. During the study period, municipal water quality measurements were taken at 32 different sites located in different areas of the Township. At community systems, coliform counts are measured weekly and nitrate and turbidity once every three years. Water quality parameters for seventy single-user private systems were taken at different intervals during the study period as part of academic survey and work done by community environmental organization.

These direct measures of water quality indicate different types of pollution. For example, total and fecal coliform counts are indicators of poor water quality and fecal contamination [15]. The presence of nitrates suggests possible contamination by animal or human waste [16]. High levels
of turbidity are indicative of particle contamination and disinfection may be less effective as a result [17].

*Other environmental variables.* Other factors such as soil type, aquifer characteristics and elevation can influence whether a pathogen that is introduced at the surface can be transported to groundwater. Different soil types will facilitate different types of pathogen transfer based on the ability of the soil and sediment to pack together. Aquifers that are unconfined or are shallow and confined may be more susceptible to surface contamination. Elevation influences the potential for flooding and run-off that may have an effect on pathogen transfer. Information on these factors were made available based on the results of an environmental assessment commission by the Township.

### 3.2.2.2 Geographic information systems

The flexibility of the GIS was instrumental in the construction of the database to assign risk factor categories to parcels where a deterministic linkage by street address was not possible. In GIS terminology, files containing data according to latitude and longitude are called shapefiles, or data layers, and the mapping of the exposures from one shapefile to another can be carried out by joining these data spatially based on shared physical location [18]. Different types of shapefiles exist for the display of different kinds of spatial data (Figure 3.1). Data can also be stored in raster format, where values for a particular characteristic are stored in a grid of evenly divided cells [19].
Figure 3.1  GIS shapefile description

Polyline shapefile
A set of lines that represent geographic features such as roads or contour lines

Point shapefile
A set of points that represent geographic features such as fire hydrants or wells

Polygon shapefile
A set of lines that close upon themselves to make a closed shape that represent geographic features such as lakes or properties

A complete list of street addresses for the study period was obtained from municipal tax records representing different cross-sections of time throughout the study period. Tax records were in the form of GIS polygon shapefiles and each street address was mapped to a parcel of land defined geographically and represented graphically by the legal property lines. Parcels of land are generally associated with a single address although changes in address (such as house number) for a given parcel were observed between municipal databases. A very low percentage (1-2%) of parcels in each database was not associated with an address: examples are parks or areas adjacent to major roadways. A polygon shapefile was constructed with all parcels of land and relevant street addresses so that data could be linked deterministically where needed to the environmental database by using a property identifying field such as street address fields (street name, house number and, where applicable, unit number) or by using the spatial join function in GIS software. This GIS with variables linked to unique street addresses could then be used to assign time-varying water system and water source variables to study subjects based on their street address as recorded and updated in the Ministry of Health Client Registry. The following sections describe how environmental data were linked to street addresses.
3.2.2.2.1 Variables linked by spatial join

Numerous variables were added to the database using the GIS spatial join feature. Figures 3.2A, 3.2B and 3.2C demonstrate how data in various GIS file formats were added to the database.

Data on variables such as surface elevation and water table elevation were in the form of polyline shapefiles. These variables were added to the database using the “Spatial Join” function in the Analysis Tools:Overlay feature of the ArcToolbox. Parcels were assigned the polyline value that was closest to the geometric centre. For example, parcels were labeled as contour line closest to the geometric centre (Figure 3.2A).

Variables were added from polygon shapefiles to the parcels using a similar mechanism. For example, soil type data were added from the polygon shapefile to the parcels. Similar to spatial joins from polyline shapefiles, when adding information from polygon shapefiles, the value assigned to the parcel is the value closest to the geometric centre of the parcel (Figure 3.2B). Other data in the form of polygon shapefiles were added to the database in this manner included watershed, and aquifer information.

Daily precipitation and slope data were added from raster files. Daily precipitation data were downloaded from the Environment Canada archive for each of seven climate stations active during the study period and located in or on the periphery of the Township [20]. For each day of the study, a spatial interpolation was conducted to create daily precipitation raster files. Slope angle was added from a digital elevation map (DEM), a raster containing a series of elevation values from which slope is derived. Each address was assigned a daily precipitation value and a slope value using a procedure analogous to that displayed in Figure 3.2B using the procedure described above.

Environmental data in the form of point shapefiles were added to the parcels. For example, a point shapefile of the wells in the Township was created using the latitude and longitude coordinates for each well contained in the database maintained by the Ministry of Water, Land and Air Protection. Well information, including well depth, well construction type and well installation date, were assigned to a parcel if located within parcel boundaries (Figure 3.2C).
Municipal water quality data measurements of total and fecal coliforms and turbidity were also added from a point shapefile. Each address classified as a recipient of water from the municipal water supply was assigned the water quality measurement value taken for that week at the measurement site located closest to it and the distance from the site was also included.

Figure 3.2A  Spatial join of data from surface elevation layer to parcels layer

Parcels 1, 2 and 3 were assigned a surface elevation of 10 metres and Parcels 4, 5 and 6 are assigned a surface elevation of 20 metres.
Figure 3.2B  Spatial join of data from soil type layer to parcels layer

Soil type A is assigned to Parcels 1, 2, 3 and 6 and Soil type B is assigned to Parcels 4 and 5.

Figure 3.2C  Spatial join of data from water wells layer to parcels layer

Well A is assigned to Parcel 4, Well B to Parcel 5, Well C to Parcel 6, Well D to Parcel 2, and Well E to Parcel 3.
3.2.2.2 Variables linked by property identifier

Some variables were linked to a property identifier such as street address or Property Roll Number, a unique, ten-digit number assigned by BC Assessment, the provincial corporation responsible for real estate assessments for taxation purposes.

Property Roll Number was used to assign variables describing water supply, sewage disposal system and land use and to trace changes in these dynamic variables throughout the study period. Roll Number was used instead of street address because it is not susceptible to changes in street name or house number that occurred and would have resulted in incomplete exposures for some parcels over the course of the study period. These data were maintained in Township databases as they affect tax payments of property owners. From each of eleven Township databases representing a different period of time, variables were added to describe water supply, sewage disposal and land use. Data from the earliest database (April 1999) were assigned back to the beginning of the study period (January 1995). For the other databases, data were assigned from the month and year it represented to the month and year of the next available database.

Data on private sewage treatment systems, such as date of installation and type of system (conventional versus alternate) were available in hard copy records maintained by the regional health authority responsible for the Township. This information was entered into an electronic database and linked to the main database deterministically based on street address.

Community water system characteristics and water quality measurements were linked deterministically by addresses identified from the files of the local health authority. Private water system measurements were linked to the addresses from which samples were taken in the surveys done throughout the study period.

3.2.2.3 Variables assigned according to water system

A few key characteristics, such as water source (groundwater versus mixture of groundwater and surface water) and water treatment (chlorination versus none), were shared by all parcels connected to the municipal water supply within a water system. Municipal water users within each water system district were assigned the same source and treatment characteristics. For example, all residents using the municipal water supply in the West System were recipients of a
municipal water supply that consisted of a mixture of local groundwater and surface water that originated from surface water reservoirs in the coastal mountains. All other municipal water users and private water users relied on groundwater originating from either private or municipal wells.

Chlorination units, installed for disinfecting water, were introduced in the municipal water systems in different regions of the Township at different times throughout the study period. All residents living in a region were assigned as having chlorinated water starting from the installation dates obtained from Township records.

### 3.3 Distribution of variables

The complete GIS database contained a total of 29,861 unique street addresses representing 29,652 unique parcels in the Township. There were 191 parcels that had a change in unit number or a change in street name at least once during the study period: this explains the discrepancy between the number of street addresses and the number of parcels. The results shown below focus on the exposures at the parcel level as these describe what is occurring in the physical space. Conducting the analysis at the individual address level would incorrectly overweight parcels that underwent a change in address.

Figure 3.3 demonstrates how parcels with addresses were created and deleted from the municipal databases over time. A total of 27,100 parcels existed for the duration of the study period. Of the 27,279 parcels that were present at the beginning of the study, 179 ceased to exist at one point during the study period and an additional 2,373 were created. A total of 29,652 unique parcels existed at one point or another during the study period. Reasons for a parcel ceasing to exist would be the defining of new property lines that occurs most often if a new subdivision is built, and new parcels are therefore created. Fifteen of these newly created parcels were temporary and did not exist at the end of the study period: this was likely a result of short-term land use assignment that occurred during development. The greatest increase in the number of parcels, an indication of development activity, occurred towards the end of the study period.
Table 3.2 presents a cross-tabulation of three key primary variables. Of the 29,652 parcels, 19,659 (66.3%) were recorded as municipal water users only and 9,153 (30.9%) were recorded as private water users only. A switch from private to municipal water use was recorded in 680 parcels and 24 of these were recorded as switching back to a private water supply. Conversely, 160 parcels first recorded as municipal water users had a record of switching back to private water supply and 17 of these returned back to municipal water supply. During the study a total of 20,499 (69.1%) parcels had at one point been connected to the municipal water supply.
Table 3.2  Distribution of key primary water variables at beginning and end of study period

<table>
<thead>
<tr>
<th>START</th>
<th>WATER SUPPLY</th>
<th>WATER SOURCE</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Municipal</td>
<td>Private</td>
<td>Groundwater</td>
</tr>
<tr>
<td>SEWAGE DISPOSAL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Municipal</td>
<td>12843 (47.1%)</td>
<td>31 (0.1%)</td>
<td>2988 (11.0%)</td>
</tr>
<tr>
<td>Private</td>
<td>5241 (19.2%)</td>
<td>9164 (33.6%)</td>
<td>9569 (35.1%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WATER SOURCE</th>
<th>Groundwater</th>
<th>Mixture</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groundwater</td>
<td>3362 (12.3%)</td>
<td>9195 (33.7%)</td>
<td>12257 (46.0%)</td>
</tr>
<tr>
<td>Mixture</td>
<td>14722 (54.0%)</td>
<td>0 (0%)</td>
<td>14722 (54.0%)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>18084 (66.3%)</td>
<td>9195 (33.7%)</td>
<td>27279 (100%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>END</th>
<th>WATER SUPPLY</th>
<th>WATER SOURCE</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Municipal</td>
<td>Private</td>
<td>Groundwater</td>
</tr>
<tr>
<td>SEWAGE DISPOSAL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Municipal</td>
<td>14818 (50.3%)</td>
<td>44 (0.1%)</td>
<td>3143 (10.7%)</td>
</tr>
<tr>
<td>Private</td>
<td>5439 (18.5%)</td>
<td>9157 (31.1%)</td>
<td>9604 (32.6%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WATER SOURCE</th>
<th>Groundwater</th>
<th>Mixture</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groundwater</td>
<td>3546 (12.0%)</td>
<td>9201 (12.0%)</td>
<td>12747 (43.3%)</td>
</tr>
<tr>
<td>Mixture</td>
<td>16711 (56.7%)</td>
<td>0 (0%)</td>
<td>16711 (56.7%)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>20257 (68.8%)</td>
<td>9201 (12.0%)</td>
<td>29458 (100%)</td>
</tr>
</tbody>
</table>

For sewage disposal systems, 14,492 parcels (48.9%) were recorded as being connected to the municipal sewer system only, while 14,465 (49.4%) were recorded as using private sewage systems only. Of the 417 parcels initially recorded as users of a private system that had a documented change to the municipal sewer system, 21 returned back to private system use. Eighty-eight parcels were recorded sewer system users at the beginning of the study switched back to private systems according to the municipal records. A total of 14,997 (50.6%) parcels were connected to the municipal sewer system at one point during the study period.

During the study period, there was a slight increase in the proportion of parcels relying on the municipal water supply and sewer system (increases of 2.5% and 3.3% respectively). While the majority of the Township was either using both or neither municipal services (> 80%), there was a substantial proportion of parcels that were connected to the municipal water supply but had private sewage systems (~19%). There were few parcels that relied on a private water supply
and were connected to the municipal sewer system. The increase in the proportion of parcels receiving water from a mixture of surface and ground water sources (43.3% to 46.0%) coincides with the increase in the number accessing municipal water in the West System.

At the beginning of the study period, no parcels were recorded as receiving chlorinated water while over two thirds (20,387) were by the end of the study. All single-user private wells were recorded as having no treatment.

Well data were available for 4,360 parcels, the majority of which had either one well (3,637) or two wells (557) within their boundaries. Another 635 parcels were associated with community water systems and were therefore classified as being connected to a well. Therefore, well data were available for a total of 4,995 parcels (16.9%). Well data were available for 4,404 of the 9,153 parcels in our database that were never connected to the municipal water system (48.1%).

Detailed private sewage system data were available for only 5,124 of the 14,655 parcels (35.0%) never connected to the municipal sewer system.

During the study period, 1014 parcels (3.4%) had a change in land use. A breakdown of the land uses at the beginning and at the end of the study can be seen in Figure 3.4. The land uses in the “Other” category included commercial, industrial and public use. There was a slight increase in the number of parcels with a residential use code from the beginning of the study (24,259 or 87.8% of the total) to the end (26,168 or 88.9%) while the number and proportion of agricultural parcels decreased (2,252 or 8.1% at beginning versus 2,050 or 7.0% at end).
3.4 Discussion

This paper describes how GIS has been used to construct an environmental database on water system and source characteristics and to assign these variables to individuals based on their addresses. These environmental factors have either been shown to be or are hypothesized to be factors associated with pathogen transmission for gastrointestinal illness. Mapping software allowed the linkage of geographic data in different formats from a variety of data sources. The resulting data illustrate the temporal and spatial variability in these characteristics at the street address level in the Township of Langley.

The results documented here reflect changes that occurred in the Township during the study period and therefore have inherent validity. The increase in the overall number of parcels and those coded for residential use coupled with the decrease in the number of agricultural parcels demonstrate the crowding of agricultural lands due to increased urban development. This development is also the driving force behind the changes in distribution of the water supply and sewage systems. Tracking changes over time revealed high stability as most parcels showed no change in water supply (97%), sewage disposal (98%) and land use (97%) from 1999 to 2003. This suggests that the missing data from 1995 to 1998 should not create substantial
misclassification. Municipal water supply chlorination commenced during the study period and was started over concern of the safety of the water supply that emerged in recent years [21].

GIS has been used in a variety of contexts in the field of environmental epidemiology as it allows researchers to enhance their analysis by investigating the relationship between location and health. For example, geocoding or address matching is carried out to pinpoint the location of a study participant in order to assign surrogate exposures based on proximity to a pollutant source. This has been done to investigate the association between traffic density and childhood leukemia, crop production in mothers’ area of residence and birth weight, and mothers’ exposure to pesticides and fetal death [22-24]. These studies have used aggregated environmental variables assigned to individuals and were therefore susceptible to ecologic fallacy.

Other studies have used GIS to assign exposures at the individual level in order to avoid this bias. For example, a GIS-based model of chemical transport in groundwater was used to estimate levels of exposure to tricholoethylene in individuals in order to investigate effects on neurobehavioral deficits [25]. In another study, potential exposures to land use contaminants and wastewater was estimated by analyzing historical land use in areas with groundwater wells identified as recharge areas where precipitation is likely to infiltrate and enter groundwater sources [26].

GIS is often used in air pollution epidemiology to assess exposure to pollutants being studied. As intra-urban variability in pollutants has been identified as a source of exposure misclassification, more advanced methods have been developed [27]. Land use regression is an alternative approach that starts by assigning environmental variables to numerous sampling sites dispersed throughout a city. Multivariable regression techniques are then used to identify the variables that account for the most variability in exposure. These variables are then used to create exposure surfaces for the entire city so that exposures can be assigned to all individuals residing there [28, 29].

There were certain challenges and limitations faced in the construction of this GIS. The quality and nature of source data were not always ideal. Generally, it is common that datasets get combined from a variety of sources. As a result, there may be issues related to the accuracy and
completeness of the different datasets [19]. This can make it difficult to analyze variables of interest and may lead to imputation of missing values. Some data did not allow for tracking of changes that may have occurred. Characteristics such as soil type and elevation are relatively static for the duration of the study period and therefore do not present a potential for bias. However, data describing well characteristics are applicable at the date of installation and it was not possible to capture modifications that may have been made to the well potentially due to water quality concerns (e.g. increased depth).

Detail was insufficient for precise assignment of characteristics in other cases. The polygon shapefile of aquifer information contained only a two-dimensional representation of the subterranean area covered by the aquifer. Parcels were therefore assigned all aquifers that were located underneath it. If depth data were available for all wells and aquifers, using the two in combination would have enabled the determination of the aquifer source. This will blur the association between aquifer as a water source of pathogen transmission and gastrointestinal illness.

The method used to add spatial data may have misrepresented characteristics of the parcel. When assigning data to parcels from line or polygon shapefiles, characteristics located at the geometric center of the parcel were assigned to that parcel (Figures 3.2A and 3.2B). In some cases, more than one characteristic may have been present but not been assigned to that parcel (e.g. a small area of Parcel 5 in Figure 3.2B actually contained some soil type A although B was assigned as it was the soil type at the geometric centre of the parcel). As polygons cover large areas, this occurred very infrequently and, if anything, introduced a non-systematic bias into our data that would attenuate risk estimates towards the null epidemiological analyses.

Assigning water treatment start dates can be difficult because chlorination units are located at reservoirs within a water system. Treatment at a parcel was assigned the day of chlorination unit installation for that system, ignoring the time to distribute chlorinated water throughout the system. In addition, parcels located closer to the treatment unit may have received more highly chlorinated water. Further, all surface water is chlorinated prior to mixture with groundwater and therefore all parcels in the West System were receiving at least partially treated water before chlorination units were installed in the Township. To account for these issues, chlorination
residuals data from water quality measurements were included to indicate the amount of chlorination present in different areas of the municipal water supply.

Data on water treatment of single-user water systems (i.e. private wells) were unavailable and all such systems were assigned as having no water treatment, although Township surveys have indicated that a small proportion of them have some form of treatment (~5-10%) (personal communication, Hans Schreier, May 2005). This limitation is a conservative one that will bias results in the opposite direction of that hypothesized (i.e. private water users and those without water treatment are expected to have a higher frequency of GI illness).

Accuracy assessments conducted by different stakeholders provide some insight into the quality of the environmental data. A comparison of reported versus true locations of wells in the Township indicated that approximately 10% of wells had locations accurate within 10 metres, 89% had locations accurate within 50 metres and the small remainder were either inaccurate or accurate only within 200 metres [30]. In the epidemiological study, a discrepancy of 50 metres can lead to substantial misclassification as characteristics of a certain well may be assigned to an incorrect household. Another survey done by the Ministry of Agriculture, Fish and Food indicated some discordance between designated land uses and those observed by surveyors. These data will be linked to the database in order to assess misclassification as part of further work beyond the scope of the current dissertation.

Data quality concerns arose from the municipal databases because it was not known whether the address changes represented real changes or administrative errors. All addresses found in these databases were included in the database sent to the Ministry of Health in order to maximize the linkage between health records and environmental data for Township residents. Switches observed in the use of municipal water and sewer services for some parcels were another source of concern. This would be expected to occur very rarely in the event of a service malfunction. As part of the larger Langley Water and Health Study, a survey will be conducted on a subsample of Township residents to assess agreement between administrative and self-reported data and to determine if changes in water supply and sewage disposal systems occur frequently.
Finally, missing data appears to be a substantive issue for some of the variables in the GIS. The lack of data for wells was expected because of the passive reporting system in place for conveying this information to the Ministry of Environment. Incomplete information within each well’s record made it impossible to determine the wells that were used for drinking water. However, during the study period and for a significant period prior to it, private sewage system installation required a permit from the local health authority and very little missing data were expected. Of the 9,531 parcels never connected to the sewer system, 77.3% (7,403) had residential land uses for which we would expect there to be a septic system. It is possible that a larger than expected proportion of private sewage systems were installed when an installation permit wasn’t required. A look at land use codes for some of those missing sewage system data indicated that they were manufactured homes inside manufactured home parks and therefore may be using a central or shared system that we were unable to capture in our database.

In general, the primary variables of water supply, sewage disposal system, water source and land use appear to be well characterized with little missing data. These factors hypothesized to be related to pathogen transmission were stable between municipal databases and qualitative examinations of the distribution of these variables by people familiar with the area suggest that they are accurate (i.e. workers at the municipality and researchers that have conducted environmental assessments). The daily precipitation variables are based on data collected by Environment Canada, the federal ministry responsible for the environment, that has a plethora of experience and uses the most up-to-date equipment for their collection of climate data.

Missing data of some secondary variables will make it difficult to use them in the analysis. One way to address this is to limit the epidemiological analysis to individuals without missing data, a method that has the potential to introduce bias into our results [31].

The construction of this GIS database and its use in the Langley Water and Health Study represent an effort to conduct population-health research using an enriched environmental characteristics database. The database is able to incorporate spatial and temporal changes throughout the Township during the study period. The use of existing administrative data for environmental variables allows for efficient research to be conducted in real world settings and large epidemiological studies where primary data collection at the individual level is not feasible.
This study may serve as a model for future population-based studies that aim to examine the role of an individuals’ environment and selected health outcomes. Specifically, this GIS database will be used to examine the association between the previously described water system and water source characteristics hypothesized to be related to pathogen transmission and *H. pylori* infection. The analysis will focus on the primary variables in order to determine whether there is an environmental component to *H. pylori* transmission specifically related to water.

Overall, this study has the potential to fill important gaps that will increase our knowledge of the etiology of GI illness. The variety of factors and the study setting may allow for findings to be extrapolated to other settings and can be valuable to water and land management policy both inside and outside of British Columbia and Canada. The results may indicate potential targets for prevention and lead to the development and implementation of strategies to combat both endemic and epidemic GI illness.
3.5 References


5. Li B, Quan H, Fong A and Lu M. Assessing record linkage between health care and Vital Statistics databases using deterministic methods. BMC Health Serv Res 2006;6:48


CHAPTER 4: THE PREVALENCE AND INCIDENCE OF HELICOBACTER PYLORI-RELATED ILLNESSES IN THE TOWNSHIP OF LANGLEY, BRITISH COLUMBIA

4.1 Background

The bacterium *Helicobacter pylori* (*H. pylori*) was first isolated in 1982 by Warren and Marshall from the gastric mucosa of patients with chronic gastritis [1]. Subsequent investigation demonstrated that *H. pylori* was associated with gastritis, peptic ulcer disease and gastric cancer [2-7]. The etiology of the bacteria in the development of these diseases has since been described [8-10].

With improvements in hygiene over the last half century, the incidence and prevalence of *H. pylori* infection and related diseases are now decreasing, especially in the developed world [11-13]. However, these illnesses continue to affect a substantial proportion of the population and represent a burden on the health-care system. This burden is attributed to the resources used for diagnosis, symptom management, and treatment. Primary prevention is hindered by our lack of knowledge of how the bacterium enters the stomach of susceptible individuals; however, a number of environmental factors including water are hypothesized to play a role in transmission [14]. There are currently limited data on the burden of *H. pylori*-related diseases in Canada.

The objectives of this study are to describe the incidence and active prevalence of *H. pylori*-related outcomes using outpatient and inpatient health data for a cohort of residents from the Township of Langley, British Columbia (BC) from 1995 to 2003. This cohort will be used for ongoing epidemiological studies to examine the relationship between gastrointestinal health outcomes and exposure to a variety of water system and environmental characteristics. Specifically, we will describe the prevalence and incidence of gastric cancer, peptic ulcers and gastritis and the economic burden on the BC health care system associated with physician contacts. Demographic characteristics of residents at the time of their incident diagnosis will be presented.

---

3 This manuscript is being prepared for submission to the journal *Helicobacter*. The authors of this manuscript are Bellack NR, Koehoorn MW, MacNab YC, and Morshed MG.
4.2 Methods

The cohort was defined as all people residing in the Township of Langley at any point in the study period according to their six-digit postal code of residence. Postal code data were taken from the Client Registry, the provincial Ministry of Health’s repository for demographic information of people covered by its universal health insurance plan. The Registry is comprehensive for the provincial population with a few exceptions for groups covered under federal funding. For each eligible individual identified, a research database was created for the individuals’ date of birth (month and year), gender, and socio-economic status. As socio-economic status data were not available for individuals, the mean household income status for the census tract (i.e. neighbourhood) in which an individual resided was assigned as a surrogate measure based on data from the Census of Canada.

*H. pylori*-related diagnoses were identified by linking cohort members to their outpatient medical services in the Ministry of Health’s Medical Services Billings file and to records of their hospitalizations in the Hospital Discharge Records File for the study follow-up period.

Cohort identification and medical record extraction were confidentially conducted by the Centre for Health Services and Policy Research (CHSPR), a research organization located at the University of British Columbia. This precaution was taken to ensure that the privacy of cohort members was maintained. All personally-identifiable information was removed by the CHSPR programmer before the linked research database was released to the research team.

Health data obtained from the Medical Services Billings files were restricted to general practitioner and specialist visits and excluded other health care professionals for which diagnosing *H. pylori*-related illnesses was considered unreliable such as chiropractors, naturopaths, massage or physical therapists.

Contacts with the health care system were coded using International Classification of Diseases codes, 9th revision (ICD9), published by the World Health Organization [15]. Visits to physicians, including specialists, are recorded in the Medical Services Billings File using three-
digit ICD9 codes while hospitalizations are recorded to the first decimal, thereby providing increased specificity of the diagnosis [16].

All physician visits and hospitalizations identified by their ICD9 code for malignant neoplasm of the stomach (ICD9 code: 151), gastric ulcer (531), duodenal ulcer (532), peptic ulcer, site unspecified (533), gastrojejunal ulcer (534) and gastritis and duodenitis (535) were used for analysis. We attempted to quantify the number of individuals in the cohort with an ICD9 code specific for *H. pylori* infection (041.86) but were unable to find any as coding in the health care utilization records are not coded to the fifth digit required for its specific description.

Individuals were counted as a prevalent case in each year if they had at least one of the diagnoses of interest with a physician visit date or hospitalization admission date in that year. This definition was developed to identify prevalent cases with active symptoms and management of disease. If an individual had any diagnosis indicating peptic ulcer disease (ICD9 codes 531, 532, 533 and 534), they were counted as a prevalent case of general peptic ulcer. The total number of hospitalizations and physician visits per year for each diagnosis and the number of unique individuals accounted for in these visits were also tabulated to investigate the health care burden of these conditions as well as the total cost of visits coded for each outcome. All cohort members were followed from date of first residence in Langley (dynamic cohort from 1995 to 2003) until the end of follow-up (2003) or until they left the Township (as recorded by yearly postal code data in the Client Registry), whichever occurred first.

Incident cases of gastric cancer were defined as individuals who were diagnosis-free for a minimum period of three years prior to their incident diagnosis. To ensure that they had not been previously diagnosed, individuals were excluded for the following reasons: if they had a gastric cancer diagnosis in the three-year window from January 1st 1992 to December 31st 1994; if they were not living in Langley at the time of the diagnosis, or; if they were not registered with the BC Medical Services Plan for a period of three years before the incident diagnosis. As a result of the universal health care coverage system in BC, all health care utilization records for the cohort over the entire study period were included in the research database for a comprehensive profile of pre-existing outcomes from 1992 to 1994 regardless of where they were living in the province. Yearly postal code data were used to indicate on-going residency in the Township.
To calculate the incidence of peptic ulcer disease, each of the site-specific diagnoses of gastric (531), duodenal (532) and gastrojejunal (534) ulcers were examined with the less specific peptic ulcer (533) diagnosis. For example, for individuals with a diagnosis of gastric ulcer (531), records were also identified for that individual indicating peptic ulcer (533) as it may be used to indicate the former. The earliest date of either diagnosis was taken as the incident date, provided that there was a minimum three-year diagnosis-free period before (dating back to the earliest availability of data). The same procedure was followed for duodenal and gastrojejunal ulcers. The incidence of peptic ulcers for individuals diagnosed only with the more general peptic ulcer ICD9 code (533) was also calculated. The burden on the pediatric population (< 16 years) in the Township was also tabulated.

To determine the reliability of coding for incident peptic ulcer cases, an algorithm was developed to identify whether the records for each case (fee item or procedural code) indicated an appropriate diagnostic procedure such as radiology, endoscopy, surgery or laboratory work on the day of diagnosis or within one year prior.

The incidence of gastritis was calculated using the same methods as for the other two health outcomes. An algorithm was used to detect whether cases had a procedural code for endoscopy on the day of diagnosis or within one year prior. The incidence of gastritis in the pediatric population was also tabulated. In addition to quantifying the burden of illness in this age group, the objective was to explore the possibility of a seasonal influence on the number of diagnosed cases. The pediatric population was also chosen because children are less likely to be exposed to non-\emph{H. pylori} risk factors for gastritis (i.e. non-steroidal anti-inflammatory drugs, alcohol and cigarette smoke) and that childhood is a potentially high risk period for incident acquisition of the bacteria [17, 18].

Prevalence and incident rates were calculated per 100,000 people (i.e. residents of the Township of Langley) per year. The prevalence and incidence over the course of the study period was investigated to determine if a trend developed.
Statistics on age, gender and socio-economic status on date of initial diagnosis were calculated. Corresponding demographic summaries were generated for individuals living in the Township during the mid-year of the study period (1999). Age at time of diagnoses was calculated from the month and year of birth by assigning the 15th of the month to all cohort members.

### 4.3 Results

The cohort consisted of 144,274 people who resided in the Township at a point during the study period. The population of the Township according to the Client Registry was calculated as 82,485 people in 1995 and grew to 93,900 in 2003.

#### 4.3.1 Prevalence of *H. pylori* illnesses

Table 4.1 shows the prevalence of gastric cancer, peptic ulcer and gastritis per 100,000 people per year as well as the total number of hospitalizations and physician claims recorded in each year for these diagnoses and the costs of these physician visits. The active prevalence of gastric cancer, according to our definition, during the study period ranged from 13 to 27 cases per 100,000 people per year. A total of 116 individuals had diagnoses of gastric cancer during the study period accounting for a total of 147 hospitalizations. Gastric cancer was considered the illness most responsible for the patient’s stay in hospital for 98 of those admissions, which accounted for 984 days in hospital. The cost of the 217 physician service claims with a gastric cancer diagnoses totaled $1.47 million over the study period.
Table 4.1 Prevalence, health care utilization and costs of gastric cancer, peptic ulcer disease and gastritis

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gastric cancer</td>
<td>Prevalent cases per 100,000 people</td>
<td>21</td>
<td>17</td>
<td>21</td>
<td>13</td>
<td>27</td>
<td>19</td>
<td>16</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Total number of hospitalizations</td>
<td>18</td>
<td>11</td>
<td>17</td>
<td>12</td>
<td>25</td>
<td>14</td>
<td>8</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Total number of physician claims</td>
<td>57</td>
<td>15</td>
<td>13</td>
<td>8</td>
<td>21</td>
<td>24</td>
<td>30</td>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Cost of physician claims ($CDN)</td>
<td>362,893</td>
<td>74,398</td>
<td>36,994</td>
<td>41,550</td>
<td>270,719</td>
<td>122,011</td>
<td>150,770</td>
<td>287,591</td>
<td>122,404</td>
</tr>
<tr>
<td>Peptic ulcer disease</td>
<td>Prevalent cases per 100,000 people</td>
<td>570</td>
<td>530</td>
<td>462</td>
<td>443</td>
<td>461</td>
<td>383</td>
<td>401</td>
<td>418</td>
<td>382</td>
</tr>
<tr>
<td></td>
<td>Total number of hospitalizations</td>
<td>72</td>
<td>58</td>
<td>50</td>
<td>45</td>
<td>66</td>
<td>43</td>
<td>54</td>
<td>61</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Total number of physician claims</td>
<td>642</td>
<td>591</td>
<td>555</td>
<td>514</td>
<td>549</td>
<td>437</td>
<td>442</td>
<td>462</td>
<td>414</td>
</tr>
<tr>
<td></td>
<td>Cost of physician claims ($CDN)</td>
<td>2,386,027</td>
<td>2,095,193</td>
<td>2,079,804</td>
<td>1,698,204</td>
<td>2,063,461</td>
<td>1,543,289</td>
<td>1,687,589</td>
<td>1,985,354</td>
<td>1,538,127</td>
</tr>
<tr>
<td>Gastritis and duodenitis</td>
<td>Prevalent cases per 100,000 people</td>
<td>1,941</td>
<td>1,793</td>
<td>1,835</td>
<td>1,658</td>
<td>1,769</td>
<td>1,675</td>
<td>1,622</td>
<td>1,770</td>
<td>1,931</td>
</tr>
<tr>
<td></td>
<td>Total number of hospitalizations</td>
<td>113</td>
<td>119</td>
<td>101</td>
<td>104</td>
<td>127</td>
<td>100</td>
<td>79</td>
<td>107</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>Total number of physician claims</td>
<td>2,306</td>
<td>2,018</td>
<td>2,173</td>
<td>1,733</td>
<td>1,895</td>
<td>1,769</td>
<td>1,718</td>
<td>1,922</td>
<td>2,068</td>
</tr>
<tr>
<td></td>
<td>Cost of physician claims ($CDN)</td>
<td>7,646,650</td>
<td>6,892,115</td>
<td>7,790,557</td>
<td>5,561,486</td>
<td>6,173,366</td>
<td>5,797,623</td>
<td>5,704,295</td>
<td>6,541,559</td>
<td>6,774,858</td>
</tr>
</tbody>
</table>
The prevalence of ulcer diagnoses ranged from 383 to 570 cases per 100,000 people per year. Among 2,846 individuals diagnosed with an ulcer during the study period, there were 482 hospitalizations of which 305 included ulcer as the primary diagnosis. These visits accounted for 1,455 days in hospital. The cost of the 4,606 physician services billings for ulcers totaled $17.08 million over the study period. The annual prevalence of peptic ulcer disease shows a decrease over the course of the study period (p < 0.001). The different prevalence between the different age groups suggests a relationship between age and peptic ulcer disease (Figure 4.1). There is a downward trend in the prevalence of cases in the 30-49 and 50-69 years age-groups. The increase in prevalence for both the 30-49 and 70+ age groups in 1999 is prominent. The variability in the 70+ age group is likely a result of the smaller population.

Figure 4.1  Annual active prevalence of peptic ulcer disease by age

The prevalence of diagnosed gastritis and duodenitis ranged from 1,622 to 1,941 cases per 100,000 people per year. The 10,604 individuals with a gastritis diagnosis over the course of the study period accounted for 957 hospitalizations, in 559 of which it was the primary diagnosis. These 559 hospitalizations accounted for 941 total days in hospital, indicating that at least some of these individuals were both admitted to and released from hospital within one day. The cost of the 17,602 physician services billings for gastritis and duodenitis totaled $58.88 million over the study period. Figure 4.2 demonstrates the trend in the annual prevalence of gastritis by age group. A slight and steady decrease can be seen for the youngest two age groups whereas no
pattern for the oldest two groups can be observed. The prevalence in the 30-49 year age group is steady with the exception of a spike for 1999.

Figure 4.2 Annual active prevalence of gastritis by age

4.3.2 Incidence

The incidence of gastric cancer ranged from 9 to 19 cases per 100,000 per year and a total of 107 new cases were found in the nine year follow-up period (Table 4.2). Over the course of the study period, 2,393 new cases of peptic ulcer disease were diagnosed. There was a decrease in the number of incident cases from the beginning to the end of the study period (p < 0.001). This was impacted most by the decrease in the number of newly diagnosed cases of duodenal ulcer (532) and peptic ulcer (533) and the overall number of incident cases in the 17-30, 30-49 and 50-69 year-old age groups (Figure 4.3). There was no apparent trend in the pediatric age groups, however, the rates for gastrojejunal ulcers (534) were much higher than for the other ulcer types. Nearly ninety-one percent (2173) of the 2,393 ulcer cases were substantiated by the presence of a fee-item code for a procedure that would be used to diagnose the outcome, with approximately 93% of gastric and duodenal ulcer cases (531 and 532), 90% of peptic ulcer cases (533) and 85% of gastrojejunal cases (534) identified by the algorithm.
Figure 4.3  Annual incidence of peptic ulcer disease by age

Figure 4.4  Annual incidence of gastritis by age
Table 4.2  Incidence of gastric cancer, peptic ulcer disease and gastritis

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gastric cancer</td>
<td>Incidence per 100,000 people</td>
<td>19</td>
<td>11</td>
<td>14</td>
<td>11</td>
<td>17</td>
<td>11</td>
<td>9</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Pediatric cases per 100,000 pediatric population *</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>13</td>
<td>4</td>
<td>0</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Gastric ulcer</td>
<td>Incidence per 100,000 people</td>
<td>69</td>
<td>60</td>
<td>64</td>
<td>54</td>
<td>56</td>
<td>38</td>
<td>41</td>
<td>56</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Pediatric cases per 100,000 pediatric population *</td>
<td>13</td>
<td>18</td>
<td>9</td>
<td>9</td>
<td>18</td>
<td>22</td>
<td>13</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Duodenal ulcer</td>
<td>Incidence per 100,000 people</td>
<td>150</td>
<td>141</td>
<td>118</td>
<td>99</td>
<td>99</td>
<td>83</td>
<td>72</td>
<td>75</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Pediatric cases per 100,000 pediatric population *</td>
<td>13</td>
<td>18</td>
<td>9</td>
<td>9</td>
<td>18</td>
<td>22</td>
<td>13</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Gastrojejunal ulcer</td>
<td>Incidence per 100,000 people</td>
<td>48</td>
<td>43</td>
<td>39</td>
<td>37</td>
<td>51</td>
<td>42</td>
<td>59</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Pediatric cases per 100,000 pediatric population *</td>
<td>76</td>
<td>61</td>
<td>43</td>
<td>26</td>
<td>35</td>
<td>18</td>
<td>49</td>
<td>27</td>
<td>23</td>
</tr>
<tr>
<td>Peptic ulcer</td>
<td>Incidence per 100,000 people</td>
<td>144</td>
<td>126</td>
<td>95</td>
<td>116</td>
<td>100</td>
<td>86</td>
<td>78</td>
<td>76</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>Pediatric cases per 100,000 pediatric population *</td>
<td>27</td>
<td>4</td>
<td>43</td>
<td>26</td>
<td>4</td>
<td>22</td>
<td>9</td>
<td>9</td>
<td>28</td>
</tr>
<tr>
<td>All ulcers</td>
<td>Incidence per 100,000 people</td>
<td>412</td>
<td>370</td>
<td>315</td>
<td>306</td>
<td>306</td>
<td>249</td>
<td>251</td>
<td>258</td>
<td>231</td>
</tr>
<tr>
<td></td>
<td>Pediatric cases per 100,000 pediatric population *</td>
<td>130</td>
<td>83</td>
<td>95</td>
<td>70</td>
<td>70</td>
<td>67</td>
<td>75</td>
<td>45</td>
<td>65</td>
</tr>
<tr>
<td>Gastritis and duodenitis</td>
<td>Incidence per 100,000 people</td>
<td>1,437</td>
<td>1,200</td>
<td>1,196</td>
<td>1,023</td>
<td>1,078</td>
<td>1,012</td>
<td>913</td>
<td>974</td>
<td>995</td>
</tr>
<tr>
<td></td>
<td>Pediatric cases per 100,000 pediatric population *</td>
<td>1,274</td>
<td>1,037</td>
<td>1,011</td>
<td>787</td>
<td>859</td>
<td>818</td>
<td>756</td>
<td>796</td>
<td>754</td>
</tr>
</tbody>
</table>

* Pediatric cases are those aged 16 years and less at time of diagnosis

There were 8,736 incident cases of gastritis and duodenitis diagnosed during the study period with annual incidence ranging from 913 to 1,437 per 100,000 with a substantial proportion in pediatric patients (Table 4.2). A decrease was observed over the course of the study period for most age groups (Figure 4.4). A prominent peak was one again observed in 1999 for the 30-49 and 70+ age groups. Over the course of the study period, the months with the greatest number of cases diagnosed were January, February, March and April and with the fewest cases in August, September and October (Figure 4.5).
Table 4.3 shows a comparison of the demographic characteristics of incident cases to the Township of Langley population. Incident cases of gastric cancer and peptic ulcer disease (except gastrojejunal ulcers) were older than the cohort. For gastric cancer cases, the gender distribution was different than the cohort with a higher proportion of male cases (p < 0.001). Gastric cancer cases were also more likely to be in the lower census socio-economic category compared to other Township residents (p < 0.001). The socio-economic distributions for the other illnesses were similar to the cohort.
Table 4.3  Demographic characteristics of incident cases of *H. pylori*-related illnesses

<table>
<thead>
<tr>
<th>Variable</th>
<th>Gastric cancer</th>
<th>All peptic ulcers</th>
<th>Gastric ulcer (531)</th>
<th>Duodenal ulcer (532)</th>
<th>Peptic ulcer (533)</th>
<th>Gastrojejunal ulcer (534)</th>
<th>Cohort*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>68.2 (14.8)</td>
<td>48.7 (20.9)</td>
<td>53.0 (20.2)</td>
<td>51.3 (19.0)</td>
<td>48.6 (20.0)</td>
<td>38.6 (23.8)</td>
<td>38.8 (23.6)</td>
</tr>
<tr>
<td>Median</td>
<td>71.1</td>
<td>48.8</td>
<td>53.8</td>
<td>50.0</td>
<td>48.7</td>
<td>39.7</td>
<td>39.4</td>
</tr>
<tr>
<td>Female</td>
<td>40 (37.4%)</td>
<td>1214 (50.7%)</td>
<td>225 (52.4%)</td>
<td>375 (46.5%)</td>
<td>419 (53.9%)</td>
<td>195 (51.5%)</td>
<td>4775 (54.7%)</td>
</tr>
<tr>
<td>Male</td>
<td>67 (62.6%)</td>
<td>1175 (49.1%)</td>
<td>201 (46.9%)</td>
<td>431 (53.4%)</td>
<td>359 (46.1%)</td>
<td>184 (48.5%)</td>
<td>3949 (45.2%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>4 (0.2%)</td>
<td>3 (0.7%)</td>
<td>1 (0.1%)</td>
<td>0</td>
<td>0</td>
<td>12 (0.1%)</td>
</tr>
<tr>
<td>Socio-economic status **</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower 50th percentile</td>
<td>42 (39.3%)</td>
<td>672 (28.1%)</td>
<td>122 (28.4%)</td>
<td>253 (31.4%)</td>
<td>218 (28.0%)</td>
<td>79 (20.8%)</td>
<td>2722 (31.2%)</td>
</tr>
<tr>
<td>Upper 50th percentile</td>
<td>60 (56.1%)</td>
<td>1640 (68.5%)</td>
<td>289 (67.4%)</td>
<td>521 (64.6%)</td>
<td>541 (69.5%)</td>
<td>289 (76.3%)</td>
<td>5738 (65.7%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>5 (4.7%)</td>
<td>81 (3.4%)</td>
<td>18 (4.2%)</td>
<td>33 (4.1%)</td>
<td>19 (2.5%)</td>
<td>11 (2.9%)</td>
<td>276 (3.2%)</td>
</tr>
</tbody>
</table>

* Based on Township population at mid-point of study; ** Percentiles based on population in Canadian Census
4.4 Discussion

The prevalence and incidence of \textit{H. pylori}-related illnesses have been measured using population-based surveys and self-reported health outcomes such as the National Population Health Survey [19]. However, this is to our knowledge the first study to use administrative health data records in a Canadian context. This study showed that some outcomes related to \textit{H. pylori} are decreasing, though they still present a substantial burden to the health care system.

The estimated annual average active prevalence of 20.1 people per 100,000 and annual average incidence of 13.4 per 100,000 for gastric cancer are similar to those reported by the National Cancer Institute in the United States of 0.02\% for prevalence and 0.01\% for incidence. Administrative data are good for capturing cases of gastric cancer because patients require a high level of care and have frequent contact with the medical system.

When compared with other population-based studies examining the active prevalence of peptic ulcer disease conducted in similar settings, the estimate for the current study was lower. On average, 450 per 100,000 individuals in the Township presented to their physician with a diagnosis of peptic ulcer in a given year (range 382 to 570 per 100,000). The Canadian National Population Health Survey showed that 3 to 5\% of adults reported having current peptic ulcers that had lasted or were expected to last six months and were diagnosed by a health professional [19]. In the United States, Europe and Australia, prevalence estimates based mostly on survey data ranged from 2-6\% [20-24]. The low estimates generated in our study using health records may be explained by peoples’ tendency to manage ulcer symptoms on their own without frequent consultation of a physician or that this method captured only the more serious cases [25]. In addition, active prevalence was calculated rather than cumulative prevalence, and individuals were included only if they were actively seeking medical attention in a given year.

A total of 2,846 people had a diagnosis of peptic ulcer over the entire nine-year period. As the Township averaged approximately 90,000 residents during the study period, this represents a cumulative prevalence of over 3\%. Expanding the time period in which an individual would count as a prevalent case for a given year would likely increase our estimates.
Two other studies were identified that used physician service claims data to estimate the prevalence of peptic ulcer disease. From the General Practice Research Database in the United Kingdom, estimates decreased from 330 to 150 per 100,000 men and from 180 to 90 per 100,000 women over the course of a four-year period from 1994 to 1998 [25]. These estimates were somewhat lower than those reported here (range of 320 to 574 per 100,000 for males and 310 to 565 per 100,000 for females); the difference could be attributed to the lack of data on visits to specialists in the database from the UK. In the United States, approximately 1,500 per 100,000 people visited a general practitioner or specialist for a peptic ulcer in 1995 [12], an estimate over three times higher than ours. Other reasons for the discrepancy in estimates may be attributable to different patterns of health care utilization or different burden of peptic ulcer disease in study populations.

The downward trend in the prevalence of peptic ulcer disease observed in this study is consistent with other results. The decrease is likely related to the availability of proton pump inhibitors, antibiotic treatment that eradicates *H. pylori*, and improvements in sanitation [25]. The high prevalence estimate in the US study may be in part attributable to a hesitation among U.S. physicians to treat peptic ulcer disease with antibiotics [12].

The average annual prevalence of gastritis, according to our definition was 1,777 per 100,000 people per year, slightly higher than the prevalence of 1% reported in the US [26]. Assessing the true prevalence of gastritis in a population is difficult because people are likely to self treat. No previous studies were found to help estimate the underreporting of gastritis in the population.

This paper demonstrates the use of administrative data to estimate prevalence, incidence and costs of disease using the BC Linked Health Database (a collection of health service records), census statistics, and other population health data [16, 27, 28]. In the single-payer system that exists in British Columbia, nearly every individual’s health records are available and linkable to health records across databases using a Personal Health Number (PHN). Other advantages include the longitudinal nature of the database that allows for the determination of an individual’s health status throughout the study period. The complete population ascertainment and objective data sources of this study design are also served to minimize the effect of
sampling, selection and reporting biases that may affect the validity of survey-based research to estimate disease rates [29].

The largest drawback with using administrative data is that it is not collected for research purposes and data may be recorded in an inconsistent or inexact fashion. In this case, there is likely a conservative bias in the estimation of health outcomes and misclassification of cases by year and for the study area. However, life-threatening illnesses such as gastric cancer and serious cases of peptic ulcer disease are likely to be captured. The validity and reliability of administrative data for peptic ulcer disease and gastritis have been investigated by comparing ICD9 codes from administrative data to survey data or information taken from medical charts or files. Peptic ulcer disease and gastritis generally exhibit low sensitivity (~50%), and high specificity (> 90%) and positive predictive values that are very high for site-specific codes [30-33]. In these studies, the gold standard is either information taken from medical charts or the presence of appropriate diagnostic procedure codes.

Each of the specific codes for ulcer diagnoses (531, 532 and 534) were grouped with the more general code for peptic ulcers (533) because we felt that the latter might be used by physicians interchangeably with the others. To assess the integrity of the peptic ulcer disease and gastritis diagnoses, an algorithm was developed and found that 90% and 75% of cases, respectively had a diagnostic procedure on the date of diagnosis or within one year prior. However, without access to a gold standard, such as a medical chart, it is difficult to assess the sensitivity and specificity of the outcome definitions used here.

Although it has been seen that *H. pylori*-related diagnoses present a substantial burden to the health care system, it is not known what proportion of these outcomes can truly be attributed to *H. pylori* infection. For example, it is well known that non-steroidal anti-inflammatory drugs, in addition to *H. pylori*, also cause peptic ulcer disease [34]. However, very high proportions of peptic ulcer disease are associated with *H. pylori*, especially duodenal ulcers and pediatric cases where nearly all cases are *H. pylori*-associated [35-40]. The attributable risk of gastric cancer related to infection with *H. pylori* has been estimated to as high as ninety percent [10]. Gastritis is very highly correlated with *H. pylori* infection although other causes have been documented [40-42].
For gastric cancer in the Township of Langley, the $1.47 million amounts to a mean annual cost of over $163,000 per year. Based on an estimated average population of 90,000 people per year in Township during the study period, extrapolation would suggest that gastric cancer costs BC nearly eight million dollars annually and Canada nearly 60 million. For peptic ulcer disease, the estimates come in at approximately $90 million in the province and $688 million in the country. Likewise for gastritis, the costs are $313 million and $2.371 billion respectively. Even if only a small percentage of the costs associated with these outcomes are directly related to *H. pylori*, it still represents millions of dollars in health care expenditures per year in the study area alone.

The definition of disease incidence used a three-year diagnosis-free window (1992 to 1994) to identify incident cases. This approach may misclassify prevalent cases as incident cases in the earlier years. Incident cases diagnosed towards the end of our study period are more likely to be truly incident because of the greater number of years of diagnosis-free time leading up to the initial diagnosis. Cases diagnosed towards the beginning of our study period are more susceptible to incidence-prevalence bias and may in part account for the higher numbers seen earlier, though these findings are consistent with others demonstrating downward trends over time.

There may be other reasons for the patterns in the prevalence and incidence in *H. pylori*-related illnesses in the Township. Changes that occurred in the distribution of the Township’s water system may have played a role as environmental routes involving water have been proposed for *H. pylori* transmission based on epidemiological and microbiological evidence [14]. For example, the chlorination of the municipal water supply that began during the study period may have had an impact [43, 44]. From the beginning to the end of the study period, there was a slight increase in the number of people using the municipal water supply rather than private wells, a potential source of transmission for *H. pylori* infection identified from epidemiological studies [45-47]. Other factors such as precipitation that can lead to wash events may also influence *H. pylori* transmission. More gastritis cases were diagnosed in the rainy winter months than in the dry summer months. There was a significant range in annual precipitation in the study area that may be associated with the small spikes in yearly burden of disease seen for gastritis, for which symptoms would appear more quickly than for peptic ulcer disease. On an
ecological scale, these results support the possibility of an environmental impact for gastritis in the younger population.

It appears that illnesses related to infection with *H. pylori* continue to represent a substantial burden to the health care system. These estimates suggest that the cost to the BC Health Care System due to these outcomes in Langley alone is on the order of millions of dollars per year. While efforts to create a vaccine continue, the search for a mode of transmission into susceptible individuals is especially important in order to minimize the burden of these diseases in the population.
4.5 References


16. Hu W. Diagnostic Codes in MSP Claim Data. Program Monitoring and Information Management Branch, Resource Management Division, Medical Services Plan, BC Ministry of Health: Victoria, Canada, 1996


CHAPTER 5: THE INFLUENCE OF DRINKING WATER AND RESIDENTIAL ENVIRONMENT ON *HELCOBACTER PYLORI* INFECTION

5.1 Introduction

*Helicobacter pylori* (*H. pylori*) is a spiral-shaped bacterium that colonizes the human stomach and duodenum. It has been identified as playing a role in the development of gastritis, peptic ulcer disease, and gastric cancer. Its role in the pathogenesis of the preceding diseases means that treatment with antibiotics can be used to combat infection and eliminate related symptoms. However, as an estimated half of the world’s population are carriers of *H. pylori*, the bacteria still represents a huge burden to individuals and to the health care system [1]. This is due to the costs of treatment, the health problems encountered prior to diagnosis or in undiagnosed individuals, and the progression to more serious illnesses.

Ongoing improvements in living conditions over the past century have resulted in a decline in *H. pylori* infection. This has produced a cohort effect where a greater percentage of older individuals are carriers and are more likely to have acquired the bacteria in their early years than individuals born more recently [2, 3]. As a result, greater proportions of young adults remain susceptible and recent exposures may be more relevant in this group for the determination of transmission pathways.

Theories of *H. pylori* transmission involving waterborne pathways have been proposed [4, 5]. Previous studies have uncovered numerous transmission factors for *H. pylori* infection in individuals, especially those related to drinking water such as water source (groundwater or surface water), water supply (municipal or private) and water treatment [6-12]. Other transmission factors related to the environment such as sewage disposal system and exposure to animals may also influence exposure to the bacteria, either directly or through tainted drinking water [13, 14]. Most of these studies have been set in developing countries and have focused on

---

4 This manuscript is being prepared for submission to the *American Journal of Epidemiology*. The authors of this manuscript are Bellack NR, Koehoorn MW, MacNab YC, Schreier H, Morshed MG, Teschke K, and Isaac-Renton JI.
a small proportion of the population in a limited geographic area. In Canada, the setting for *H. pylori* transmission studies has been native communities that are often located in rural environments and have distinct social organization [15-17].

This study will use a population-based approach to examine the association between environmental and water system characteristics of an individual’s residence and *H. pylori* infection. Specifically, this investigation will focus on water supply, water source, water treatment, sewage disposal and land use as they have been identified as having a more direct and explainable relationship for *H. pylori* transmission [5]. Surface elevation and soil type will be investigated as exploratory variables as a relationship with *H. pylori* infection is less direct but may be important with respect to transfer of bacteria into drinking water sources. The study employs a nested case-control design and is set in the Township of Langley, in the province of British Columbia (BC), Canada, from April 1997 to December 2003.

### 5.2 Methods

#### 5.2.1 Sampling frame for cases and controls

The sampling frame consisted of all residents of the Township of Langley during the study period. Residence was defined by having a six-digit postal code in Township boundaries according to the address in the Client Registry, the BC Ministry of Health’s repository for demographic information of people covered by the provincial universal health insurance plan. The Registry also provides a source of information on change of address indicating movement within the study area or loss to follow-up (movement outside of the study area). The Registry covers nearly all of the BC population except for a few groups covered under a federal plan, such as aboriginals and members of the armed forces. For all individuals identified as Township residents, the Client Registry was used to populate a database that consisted of date of birth (month and year), gender, and socio-economic status (SES) defined by neighborhood-level census data. As socio-economic status data were not available for individuals, the mean household income for the census tract (i.e. neighbourhood) in which an individual resided was assigned as a surrogate measure based on data from the Census of Canada.
5.2.2 Case Identification

Carrier status for *H. pylori* was determined from a database of serological results maintained by the BC Centre for Disease Control (BC CDC). When an individual presents to their physician with gastrointestinal discomfort and *H. pylori* is suspected, a blood sample is taken and sent to the BC CDC for analysis. Blood samples are analyzed using an Immulite 2000 IgG test (Siemens Medical Solutions Diagnostics, Los Angeles, California) that has a sensitivity and specificity of 91% and 100%, respectively [18]. All serology results are recorded in a database at the BC CDC with the phlebotomy date. The serology data was merged with the Client Registry data at the individual level by the Ministry.

All individuals with a positive serology result were considered cases and the phlebotomy date was considered diagnosis date. Cases that had not resided in the Township for a minimum of three years were not considered for analysis. Results were available from April 1st, 1997 to December 31st, 2003.

5.2.3 Case-control matching

Each individual case of *H. pylori* was matched to two randomly selected controls in order to increase efficiency of the study design. Matching was conducted based on gender and date of birth (month and year), two of the most common confounders in epidemiological studies. Controls had to reside in the Township for a minimum of three years to be considered eligible.

5.2.4 Water system and related environmental variables

The environmental database was constructed with unique street address in the Township as the unit of observation and water system and environmental characteristics linked back to individuals by residential street address. Variables were added from numerous data sources including municipal, provincial and federal governments; the local health authority; academic and community surveys; and environmental assessments commissioned by the Township. These variables included water supply (municipal versus private); sewage disposal system (sewer versus septic); water source (groundwater versus mixture of groundwater and surface water); land use (residential versus other); water treatment (chlorination versus none); surface elevation; and surficial soil type.
Environmental data were available as geographic information system (GIS) files and electronic or hard-copy records from which electronic databases were constructed. The building blocks of the database came from municipal tax records and consisted of each unique address that existed during the study period, a unique property identifier, and a graphical representation of the address defined by property lines (i.e. a parcel of land) in GIS format. Variables across different data sources and coding schemes were added to the database by deterministically linking with address or unique property identifier fields, or through a spatial join using GIS software that assigned values based on shared location from the source data layer to the property data layer. The primary objective for the environmental database was to obtain risk factor variables for *H. pylori* infection for each unique street address in the Township of Langley over the study period to be linked back to individuals by residential address. The process of constructing the environmental database and the distribution of variables at the parcel level are discussed in more detail in Chapter 3.

Individuals were linked to their household variables by matching their residential address in the Client Registry to the residential address in the environmental database. This process was conducted by a programmer at the Ministry to ensure confidentiality of study members. The database that was released to the study team for analysis had all personally identifiable fields removed and replaced with an anonymous study identifier. A total of 90.8% of cohort members with a postal code indicating residence in the Township of Langley were subsequently linked to environmental factors for the study follow-up period.

Environmental variables based on the date of diagnosis were used for analysis. Although there is much uncertainty regarding the latency time between *H. pylori* acquisition and symptom development, symptoms may appear as a result of a recent infection indicating that current risk factor variables are relevant to determine transmission pathways [19]. Current environmental characteristics may be representative of those over longer periods of time when the bacteria may have been acquired. To determine the stability of exposures, the number of cases and controls with only a single category of exposure for the duration of the study period was tabulated. A high percentage of individuals with only one risk factor category for a given characteristics for the duration of the study period indicates that the categories used for analysis are representative of long-term influences of water quality. Stability in these characteristics would diminish the
effect of a potential incidence-prevalence bias. A low percentage suggests that misclassification is occurring, which would likely bias risk estimates. It is hypothesized that the characteristics on the date of diagnosis are a good representation of longer term characteristics for the current study population.

5.2.5 Analysis

Unadjusted odds ratios (ORs) were tabulated using the method of Mantel and Haenszel for 1:m matching of cases to controls [20]. Each case and matched controls are considered in their own stratum and the Mantel-Haenszel odds ratio is derived using the formula:

\[
OR = \frac{\sum [(M - m + 1) N_{1, m-1}]}{\sum [m N_{0, m}]}
\]

where M is the number of controls per cases; m is any number from 1 to M; \(N_{1, m-1}\) is the number exposed of cases with m-1 exposed controls; and \(N_{0, m}\) is the number of non-exposed cases with m of their matched controls exposed.

The cases and controls were split into four groups based on age at diagnosis: 20 years and under; 21 to 40 years; 41 to 60 years; and 61 years and older. This step was taken to further address issues related to relevant exposure periods that have been discussed in the literature. The cohort effect that has resulted from improvements in hygiene over the last century suggests that more recent exposures may be relevant for younger but not older individuals. Overall, the goal was to determine whether risk estimates were consistent between the different age groups. It was hypothesized that a stronger relationship between water source and water system characteristics and \(H. pylori\) infection would be observed among younger individuals.

Correlations between the study variables were investigated using Spearman correlation coefficients. Correlation between independent variables in a multiple regression model can lead to multicollinearity and make it difficult to disentangle the effects of those variables on outcome [21, 22]. Therefore, variables with a correlation coefficient > 0.9 were not considered for the same multiple logistic regression model. This was important because many regions within the Township share characteristics related to water supply and sewage disposal services. The presence of multicollinearity was assessed by examining standard errors and consistency of unadjusted versus adjusted risk estimates.
Variables found significant at the $p < 0.10$ level in Mantel-Haenszel analyses were used in the multiple regression models. Logistic regression was used to quantify the relationship between the water source and water system variables and *H. pylori* while controlling for other confounders and covariates. Uncertainty over period of acquisition and the lag between acquisition and the appearance of symptoms prevented a time-to-event analysis that may have provided beneficial information. Spatial analysis was not possible due to concerns over privacy of cohort members.

Model building was conducted using a backward stepwise approach starting with all variables identified as significant in the Mantel-Haenszel analyses. Interaction terms were tested in the model based on logical relationships between variables. The log-likelihood ratio test was used to assess the significance of the association of the fitted model and to select the most important exposure variables to arrive at a final parsimonious model where all variables were significant at $p < 0.05$. The regressions coefficients show a change in the odds of *H. pylori* infection for the exposure category under examination.

**5.3 Results**

There were 2,151 Township of Langley residents with a positive *H. pylori* serology test from April 1st, 1997 to December 31st, 2003. Water source and system data on the date of diagnosis were not available for 229 cases (10.6%) and therefore only 1,922 cases were included in the analysis. No significant differences were observed for gender ($p = 0.87$), age ($p = 0.41$) or socio-economic status ($p = 0.27$) between cases with and without exposure data. Four cases and nine controls matched to more than one address and therefore had multiple water source and water system characteristics associated with them for the date of diagnosis. These individuals were assigned the category hypothesized to be the higher risk for transmission of *H. pylori*. For example, if an individual was linked to both a municipal and a private water supply, the private water supply category was assigned. The distribution of water system and water source variables for cases and controls is shown in Table 5.1. As case-control matching was based on age and gender these variables are equally distributed.
Table 5.1  Proportional distribution of demographic and environmental variables for cases and controls

<table>
<thead>
<tr>
<th>Exposure variables</th>
<th>Exposure categories</th>
<th>Cases (Total = 1922)</th>
<th>Controls (Total = 3844)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Age</td>
<td>&lt; 20 years</td>
<td>38</td>
<td>2.0%</td>
</tr>
<tr>
<td></td>
<td>20 to 39 years</td>
<td>347</td>
<td>18.1%</td>
</tr>
<tr>
<td></td>
<td>40 to 59 years</td>
<td>838</td>
<td>43.6%</td>
</tr>
<tr>
<td></td>
<td>&gt; 60 years</td>
<td>699</td>
<td>36.4%</td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
<td>1000</td>
<td>52.0%</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>922</td>
<td>48.0%</td>
</tr>
<tr>
<td>Socioeconomic status</td>
<td>Upper 50%</td>
<td>1373</td>
<td>71.4%</td>
</tr>
<tr>
<td></td>
<td>Lower 50%</td>
<td>549</td>
<td>28.6%</td>
</tr>
<tr>
<td>Water supply</td>
<td>Municipal</td>
<td>1261</td>
<td>65.6%</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>661</td>
<td>34.4%</td>
</tr>
<tr>
<td>Sewage disposal</td>
<td>Sewer</td>
<td>946</td>
<td>49.2%</td>
</tr>
<tr>
<td></td>
<td>Septic</td>
<td>976</td>
<td>50.8%</td>
</tr>
<tr>
<td>Water source</td>
<td>Groundwater</td>
<td>950</td>
<td>49.4%</td>
</tr>
<tr>
<td></td>
<td>Mixed</td>
<td>972</td>
<td>50.6%</td>
</tr>
<tr>
<td>Water treatment</td>
<td>Chlorination</td>
<td>702</td>
<td>36.5%</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>1220</td>
<td>63.5%</td>
</tr>
<tr>
<td>Land use</td>
<td>Residential</td>
<td>1680</td>
<td>87.4%</td>
</tr>
<tr>
<td></td>
<td>Non-residential*</td>
<td>242</td>
<td>12.6%</td>
</tr>
<tr>
<td>Surficial soil</td>
<td>River deposits</td>
<td>624</td>
<td>32.5%</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>1298</td>
<td>67.5%</td>
</tr>
<tr>
<td>Elevation</td>
<td>≥ 50 metres</td>
<td>1030</td>
<td>53.6%</td>
</tr>
<tr>
<td></td>
<td>&lt; 50 metres</td>
<td>892</td>
<td>46.4%</td>
</tr>
</tbody>
</table>
The highest correlations of variables were for water supply with sewage disposal (0.68), water source (-0.75), and water treatment (0.55) and for water source with water treatment (-0.64) (Table 5.2).

Table 5.2 Correlations between water system and water source variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Water supply</th>
<th>Sewage Disposal</th>
<th>Water source</th>
<th>Water treatment</th>
<th>Land use</th>
<th>Surficial soil</th>
<th>Surface elevation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water supply</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sewage Disposal</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water source</td>
<td>-0.75</td>
<td>-0.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water treatment</td>
<td>0.55</td>
<td>0.29</td>
<td>-0.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land use</td>
<td>0.50</td>
<td>0.34</td>
<td>-0.37</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surficial soil</td>
<td>-0.29</td>
<td>-0.37</td>
<td>0.46</td>
<td>-0.40</td>
<td>-0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface elevation</td>
<td>-0.32</td>
<td>-0.13</td>
<td>0.57</td>
<td>-0.42</td>
<td>-0.20</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>Socio-economic status</td>
<td>0.01</td>
<td>-0.08</td>
<td>-0.12</td>
<td>0.08</td>
<td>-0.01</td>
<td>-0.05</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

Table 5.3 demonstrates the stability of water characteristics for cases and controls. For nearly all variables, between 91.5% and 94.7% didn’t change their characteristics over the course of the study period with the exception of exposure to treated groundwater. As chlorination was introduced during the study period, approximately half of cases and controls received untreated water only for the duration of the study period.

Table 5.3 Percentage of cases and controls with one exposure for duration of study period

<table>
<thead>
<tr>
<th>Variable</th>
<th>% with same exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases</td>
</tr>
<tr>
<td>Water supply</td>
<td>92.0</td>
</tr>
<tr>
<td>Sewage disposal</td>
<td>91.8</td>
</tr>
<tr>
<td>Water source</td>
<td>92.7</td>
</tr>
<tr>
<td>Water treatment</td>
<td>53.6</td>
</tr>
<tr>
<td>Land use</td>
<td>94.3</td>
</tr>
<tr>
<td>Surficial soil</td>
<td>94.6</td>
</tr>
<tr>
<td>Elevation</td>
<td>93.5</td>
</tr>
</tbody>
</table>
Odds ratios (ORs) and 95% confidence intervals in the two groups are presented in Table 5.4. Overall, there appeared to be an association between *H. pylori* infection and the variables describing water supply, water source, water treatment, land use, surface elevation and socioeconomic status. The odds ratios in the youngest group (≤ 20 years) were non-significant and had wide confidence intervals, likely the result of few cases (n = 38). In the 21-40 year age group, a higher odds of *H. pylori* infection was observed among individuals on a private water supply, receiving groundwater only, with non-residential land uses and living at elevations of ≥ 50 metres. In the next oldest age group (41-60 years), ORs for water source and land use were attenuated though remained significant while those for water supply and land use became non-significant. In the oldest age group (≥ 61 years), none of the ORs for the environmental variables reached significance. Socioeconomic status was significant for all age groups with the exception of the youngest. There was no association of *H. pylori* infection with type of sewage disposal system or surficial soil type.
Table 5.4  Unadjusted odds ratios and 95% confidence intervals for all age groups

<table>
<thead>
<tr>
<th>Variables</th>
<th>All subjects n = 5766 (1922 cases)</th>
<th>≤ 20 years n = 114 (38 cases)</th>
<th>21 to 40 years n = 1041 (347 cases)</th>
<th>41 to 60 years n = 2514 (838 cases)</th>
<th>≥ 61 years n = 2097 (699 cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water supply</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private (vs. municipal)</td>
<td>1.13 (1.01 - 1.27) *</td>
<td>0.94 (0.42 - 2.15)</td>
<td>1.72 (1.29 - 2.28) *</td>
<td>1.13 (0.95 - 1.35)</td>
<td>0.95 (0.78 - 1.14)</td>
</tr>
<tr>
<td>Sewage disposal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Septic (vs. sewer)</td>
<td>0.99 (0.89 - 1.11)</td>
<td>1.10 (0.52 - 2.33)</td>
<td>1.23 (0.95 - 1.61)</td>
<td>0.96 (0.81 - 1.14)</td>
<td>0.92 (0.77 - 1.10)</td>
</tr>
<tr>
<td>Water source</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed (vs. groundwater)</td>
<td>0.83 (0.74 - 0.92) *</td>
<td>1.09 (0.53 - 2.24)</td>
<td>0.54 (0.42 - 0.70) *</td>
<td>0.83 (0.70 - 0.98) *</td>
<td>1.00 (0.84 - 1.20)</td>
</tr>
<tr>
<td>Water treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None (vs. chlorination)</td>
<td>1.12 (1.00 - 1.26) *</td>
<td>1.00 (0.49 - 2.04)</td>
<td>1.24 (0.95 - 1.61)</td>
<td>1.22 (1.02 - 1.45) *</td>
<td>0.98 (0.81 - 1.18)</td>
</tr>
<tr>
<td>Land use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-residential (vs. residential)</td>
<td>1.17 (0.99 - 1.39) *</td>
<td>1.25 (0.41 - 3.82)</td>
<td>2.04 (1.38 - 3.01) *</td>
<td>0.98 (0.76 - 1.27)</td>
<td>1.10 (0.83 - 1.46)</td>
</tr>
<tr>
<td>Surficial soil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>River deposits (vs. other)</td>
<td>0.96 (0.85 - 1.08)</td>
<td>1.06 (0.47 - 2.39)</td>
<td>0.88 (0.67 - 1.16)</td>
<td>1.02 (0.85 - 1.21)</td>
<td>0.92 (0.76 - 1.12)</td>
</tr>
<tr>
<td>Elevation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 50 metres (vs. ≥ 50 metres)</td>
<td>0.85 (0.76 - 0.95) *</td>
<td>1.35 (0.63 - 2.91)</td>
<td>0.65 (0.50 - 0.84) *</td>
<td>0.84 (0.71 - 0.99) *</td>
<td>1.01 (0.85 - 1.22)</td>
</tr>
<tr>
<td>Socioeconomic status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower 50% (vs. upper 50%)</td>
<td>1.43 (1.26 - 1.62) *</td>
<td>1.36 (0.59 - 3.10)</td>
<td>1.63 (1.23 - 2.15) *</td>
<td>1.40 (1.14 - 1.70) *</td>
<td>1.38 (1.12 - 1.69) *</td>
</tr>
</tbody>
</table>

* p < 0.10
Using multiple logistic regressions, only the analysis of the entire study population and the analysis of the cases and controls aged 21-40 years revealed significant odds ratios other than socioeconomic status after adjustment for other variables (Table 5.5). In the overall group, a lower odds of H. pylori infection was observed among those on a mixed water source compared to groundwater only (OR = 0.86, 95% CI 0.63 – 0.96), adjusted for socioeconomic status. Similarly, in the 21-40 year age group, the same significant relationship was observed for mixed water source (OR = 0.63, 95% CI 0.47 – 0.83). However, individuals with non-residential land uses were also at a higher odds of H. pylori infection (OR = 1.58; 95% CI 1.03 – 2.42) in both the overall study population and the study population aged 21-40 years when compared to those with residential land use only. Interaction terms were non-significant, did not improve model fit and were not considered for the final models. Small standard errors and consistent regression coefficients suggested that multicollinearity did not occur in these models.

<table>
<thead>
<tr>
<th>Variables</th>
<th>All cases n = 5766 (1922 cases)</th>
<th>21 to 40 years n = 104 (347 cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Water source</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed (vs. groundwater)</td>
<td>0.86 (0.63 - 0.96)</td>
<td>0.63 (0.47 - 0.83)</td>
</tr>
<tr>
<td><strong>Land use</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-residential (vs. residential)</td>
<td>Not in model *</td>
<td>1.58 (1.03 - 2.42)</td>
</tr>
<tr>
<td><strong>Socioeconomic status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower 50% (vs. upper 50%)</td>
<td>1.40 (1.24 - 1.59)</td>
<td>1.57 (1.18 - 2.08)</td>
</tr>
</tbody>
</table>

Table 5.5 Adjusted odds ratios and 95% confidence intervals in parsimonious logistic regression models

Only socioeconomic status significant for the 41 to 60 year and > 60 year age groups; no variables significant in ≤ 20 year age group

* Land use not entered in model because unadjusted odds ratio not significant at p = 0.05

5.4 Discussion

This study employed a nested case-control design to examine the risk of Helicobacter pylori infection in relation to a number of characteristics describing residential water supply and related environmental water system characteristics among Township of Langley residents over a seven year period. The aim of this analysis was to investigate if an environmental pathway, mainly waterborne, is involved in the transmission of H. pylori in a North American setting. It is intended to complement previous studies that used a different methodological approach and were conducted mostly in developing countries. To our knowledge, this is the first to use
administrative data for a comprehensive environmental profile and a population-based study sample of cases and controls.

Of the variables most directly related to water quality, only water source remained a significant factor associated with of *H. pylori* infection. Surface water is generally more susceptible to contamination and *H. pylori* has been detected in surface water samples taken in the developed and developing world [13, 14, 23-25]. As a result, it was hypothesized that a higher odds of infection would be observed for individuals receiving the mixed water supply. However, in this study, the surface water originates from coastal mountain reservoirs, likely a cleaner source than those identified from other studies. In the final multivariable analysis, water treatment was not associated with *H. pylori* infection, although an elevated odds were observed with non-chlorination in the unadjusted bivariate analyses among those aged 21 to 60 years. Exposure to chlorinated water was assigned based on the date that chlorination units were installed in Township reservoirs. However, as all surface water is chlorinated prior to mixture with groundwater, this extra layer of treatment may account for the protective effect observed for exposure to surface water and may explain the absence of an effect for the implementation of chlorination units when adjusted for water source. Chlorination is one of the most common methods used in public health to remove microbial pathogens in water and was expected to decrease risk of *H. pylori* transmission as seen in this study as it has proven effective at inactivating *H. pylori* in water in experimental settings [26-30]. In BC, the susceptibility of surface water to pathogens has led to mandated water treatment of surface water sources by those responsible for the maintenance of water systems [31].

Water supply (private versus municipal) was not related to outcome in any age group when other variables were considered. Individuals relying on a private well are thought to be more susceptible to water contamination based on the environment surrounding their well, especially if it is close to agricultural activity, or if the wells are too shallow [32, 33]. In Germany, non-municipal water supply was associated with an increased risk of *H. pylori* infection (OR = 2.8, 95% CI: 1.0 – 8.2) [9]. Municipal water users (versus community water supply) were at an increased risk of infection in Peru (OR = 11.4, p = 0.02) where the surface water was used for the municipal supply [11], highlighting the importance of the distinction between water source and water supply.
The presence of non-residential land uses, primarily composed of agricultural activities, was thought to pose substantial health risks because of the potential for pathogen contamination to water supplies [34]. Zoonotic transmission has been suggested because several domestic animals have been identified as potential carriers of *H. pylori* in both epidemiological and microbiological studies [9, 14, 35-37]. It is possible that direct exposure to animals or animal waste is the explanation for the increased odds ratios instead of less direct exposure to the bacteria via contamination of water supplies.

No significant relationship was found between *H. pylori* infection and sewage disposal although previous investigations demonstrated a difference in risk by type of sewage system [6, 12, 38]. As most were set in developing countries, it is likely that a lower standard of hygiene accounted for this difference.

Surface elevation and soil type were used in this analysis because it was hypothesized that they may play roles in transmission by facilitating the passage of waterborne agents. A relationship of *H. pylori* infection and surface elevation has been previously demonstrated, however, on a larger scale (i.e. larger differences in altitude) than those observed in this study [39, 40]. The significance of the elevation variable disappeared after adjustment for water source in the current study.

In the multivariable models, the water source and land use variables were all attenuated towards the null after adjustment for SES compared to the unadjusted odds ratios (Tables 5.4 and 5.5). The relationship between socio-economic status and *H. pylori* infection was maintained despite the aggregated socio-economic indicator based on census data for average household income in geographic census tracts that range in size from 2,500 to 8,000 people [41]. This type of variable assignment may be subject to ecological fallacy that would most likely attenuate the results toward the null.

The relationship between SES and *H. pylori* infection is consistent with many health outcomes analyzed in epidemiological research and has been criticized for drawing attention away from risk factors that may be modifiable [42]. In this case, SES may be representative of many things...
including lower levels of hygiene, occupation, country of birth or immigrant status, access to other resources or susceptibility to symptom development when acquisition of the bacteria occurs.

It is possible that people of lower SES lack the economic means to implement hygienic measures in the home and are therefore more exposed to a variety of bacterial agents [43]. The inverse association between \textit{H. pylori} prevalence and parental education levels indicates the possibility that people may lack the knowledge of behaviors that can be used to prevent infection [44].

Higher prevalence of infection of \textit{H. pylori} and earlier age of acquisition in developing countries may also account for the significance of the SES variable in this study. If immigrants acquire \textit{H. pylori} prior to arrival in Canada, the variables being investigated here are not relevant with respect to \textit{H. pylori} transmission [45, 46].

Individuals with occupations that require them to come in contact with human and animal waste are also more likely to be carriers [5, 42, 47]. Therefore, occupation, which is associated with SES, is another important factor.

As exposure to the bacteria through ingested water is a risk factor, the association between SES and \textit{H. pylori} infection may be explained by an inability to access other resources that would serve to prevent ingestion. For example, the use of bottled water instead of tap water has shown to decrease the risk of Cryptosporidium among children [48]. In the study area, individuals with more disposable income may be more likely to have a private water treatment system or purchase bottled water. Whether it is one of the factors associated with socioeconomic status that is acting alone or many factors acting together, it is likely that the result is either an increased exposure to the bacteria in water or the environment or perhaps an increased susceptibility to symptom development once the bacteria is acquired.

The issue of relevant exposure periods for \textit{H. pylori} acquisition is not conclusive in the literature. Both children and adults have been identified as susceptible but without follow-up studies it is impossible to know the exact time when infection occurred [49-51]. Two factors provide credibility to the results obtained from the current study. First, the stability of risk factors (Table
5.3) indicates that the variables analyzed were relevant for a substantial time period around the phlebotomy date. Second, the different results obtained for the separate age groups suggest that the variables analyzed were relevant only for the younger part of the population. This was expected based on the cohort effect that has been described previously. The significant ORs measured in the 21-40 years age group would only be irrelevant if *H. pylori* was exclusively acquired in childhood and if childhood risk factors were substantially different from current ones. If this were the case, the results would be attenuated towards the null hypothesis.

The assumption behind the acquisition of *H. pylori* is that it takes only a single exposure to the bacteria for acquisition to take place. It is hypothesized that, similar to other waterborne pathogens, *H. pylori* is not always present in the water supply. However, the bacteria is more likely present in water associated with determinants of poor water quality and the more of this that one drinks, the more likely it is that they will ingest it when present. This study compares the likelihood of that acquisition occurring between the different water system and water source factors hypothesized to play a role in transmission.

This study did have some limitations. The outcome measure used did not permit an analysis of the pediatric population (≤ 16 years) as there were few children in the database, likely because serology is not as sensitive a diagnostic tool in that group [52]. As childhood has been speculated to be an important period regarding *H. pylori* acquisition, this would have strengthened the study. Unfortunately, the small sample size in the youngest group (≤ 20 years) did not permit an informative analysis. The strongest associations were found for the second youngest age group.

It is difficult to distinguish between the effects of water source and of water treatment on outcome. This is mainly a result of the relationship between water-related variables in the study area and the assignment of exposures. The surface water that gets mixed with groundwater in part of the Township is treated prior to mixing. Therefore, all residents receiving water from mixed sources have their water treated whereas recipients of groundwater only may or may not have treated water.
Ideally, it would have been possible to examine the effect of other variables on *H. pylori* infection such as family size, birth order, occupation, country of origin, and SES status at the level of the individual rather than the neighborhood. Information on these variables was not available in the administrative data sources. However, this study did take advantage of the increased availability of data in GIS format with which to assign water source and water system characteristics to an individual address, an improved method of exposure assessment over previous works that have assigned everyone in a given region a single exposure value which may not be as precise [53]. For example, one study that examined the accuracy of postal code as a surrogate measure of water supply found that it was not good at classifying users of private water systems (63.5% predictive value) [54]. This study is part of a new wave of studies that takes advantage of available data in various formats and employs GIS technology for analysis of water-related illness [55-58].

The outcome measure used here relied on a diagnostic test that is only applied to individuals with clinical symptoms of *H. pylori* infection. Although the serology test has a high degree of accuracy, it is commonly believed that there are many carriers that do not present with any symptoms. Therefore, it is possible that some controls were indeed untested carriers of *H. pylori*. Misclassification of outcome would have made it difficult to identify significant water system and source variables involved in transmission.

The objective here was to describe and analyze risk associated with residential water system-related variables. As people consume water from a variety of sources, using their home water source as a surrogate for all water consumption may lead to errors [54]. Not accounting for exposures that people encounter at work or while traveling prevents us from being able to analyze what effect these exposures may have on outcome. In one recent study, approximately half of survey respondents reported their home water source as their only water source and 30% indicated that they did not rely on their home water source at all [54].

The potential exposure and outcome misclassifications described above are all biases that are believed to be non-directional. Non-directional misclassification bias produces measures of association (ORs) towards the null hypothesis making it difficult to detect significant risk factors and would have caused the ORs observed here to be underestimated [59].
The increased odds of *H. pylori* infection among recipients of groundwater only and those identified as living on parcels associated with non-residential activity is relevant to public health policy makers. First, individuals with these characteristics may be at increased risk of gastritis, peptic ulcer disease and gastric cancer as a result of increased contamination and transmission of the *H. pylori* bacteria via the water system. Individuals with *H. pylori* infection may be unaware that their symptoms are being caused by an infection and may be avoiding treatment by self-medicating. Communicating the effects of *H. pylori* infection and importance of seeking medical attention may prevent more serious illnesses later in life for carriers. Individuals with agricultural exposures may be a suitable group to target for public health communication.

We suspect that the decreased odds of *H. pylori* infection for those receiving the mixed water stems from the added layer of treatment (i.e. chlorination) that is applied to the surface water. While chlorination of the water supply may be controversial due to potentially harmful by-products, individuals should be informed of its benefits and other means of treatment should be publicized and made available to users of private water systems.

As sampling was based on outcome, this study employed a case-control methodology. A matching ratio of 1:2 for cases to controls was used to increase the statistical efficiency. Case-control studies are good for identifying potential factors as measured by the odds ratio; however, they are unable to provide an estimate of the risk associated with exposure variables. This is especially true in this case where there may be a high level of asymptomatic cases. However, the variables identified here provide the rationale for further investigation.

Although current methods are unable to culture *H. pylori* from the environment, it may be worthwhile to take water and agricultural samples and test them for *H. pylori* DNA. This will strengthen the possibility that there are potential environmental reservoirs for the bacteria. Further investigation may identify a suitable target for a public health intervention. It would be ideal to complement the epidemiological analysis with water sampling to determine whether the surrogates of *H. pylori* exposure used in this study are associated with the presence of the bacteria in the water supply.
Studies of this type are important to examine relationships between environment and health using observational data in a way that theoretical modeling cannot and they have important public health consequences. The methodology used here can provide the foundation and serve as a model for other studies investigating health and environmental relationships using a population-based approach. These results augment the body of literature suggesting that *H. pylori* transmission may have an environmental component related to water.
5.5 References


10. Karita M, Teramukai S and Matsumoto S. Risk of *Helicobacter pylori* transmission from drinking well water is higher than that from infected intrafamilial members in Japan. Dig Dis Sci 2003;48:1062-7


6 CHAPTER 6: CONCLUDING CHAPTER

6.1 Thesis Summary

This dissertation describes the burden of Helicobacter pylori infection on the population and examines the relationship between H. pylori infection and environmental factors, specifically related to water. A population-based approach was used to study residents in the Township Langley, British Columbia, (BC) from 1995 to 2003. Health outcomes were obtained from health care utilization records and a database of serological results, and environmental data originated from a variety of sources including municipal, provincial and federal government; the local health authority; environmental assessments commissioned by the Township; and surveys conducted by community groups and academic institutions. In addition to augmenting the existing literature on H. pylori by examining transmission and related illnesses, this work also makes a novel contribution to the field of environmental epidemiology through the methodology used to assign variables to individuals based on their place of residence.

6.1.2 Summary of results

6.1.2.1 Chapter 2: A conceptual model of water’s role as a reservoir in Helicobacter pylori transmission

A review of epidemiological and microbiological research that has investigated risk factors for H. pylori infection led to the development of a conceptual model of H. pylori transmission via environmental reservoirs and pathways, specifically involving water. Previous studies have shown different rates of infection between individuals relying on different types of water supply and that H. pylori has been detected in a variety of water samples, some intended for use as drinking water. Based on these findings, a higher proportion of H. pylori infection was expected among individuals relying on a private well-water system (versus municipal), whose water originated from surface water sources (versus groundwater), and whose water was untreated (versus treatment, mostly in the form of chlorination in the study area).

The conceptual model also identified other risk factors such as the use of a septic system (versus municipal sewer), especially among those using well water. The presence of agricultural and
other non-residential lands was thought to impact water quality by interaction with water systems. Surface elevation, slope, and soil type were thought to be important also because of their potential to facilitate the passage and survival of microbes along potential transmission routes.

The *H. pylori* transmission model was informative for the development of the geographic information system (GIS) discussed in Chapter 3.

### 6.1.2.2 Chapter 3: Construction and characteristics of an environmental database for population health research

In this chapter, the variety of data source and formats used to build the environmental database were described. The methods outlined the process for assigning environmental and water system characteristics to each address in the Township during the study period in assembling the GIS. These risk factors are determinants of water quality and are surrogates for exposure to the *H. pylori* bacterium itself and other waterborne pathogens. The adopted approach was taken to maximize the accuracy of assigning variables at the level of the individual as other methods used in large population-health studies introduce misclassification that affect validity of results [1].

The distribution of key variables, such as water supply, sewage disposal, water source, water treatment and land use were summarized and demonstrated that relatively little change occurred in the Township during the study period. There was an increase in the number of parcels connected to municipal water system resulting in an increase in the number of parcels receiving a mixture of surface water and groundwater. The number of parcels connected to the municipal sewer system and those designated for residential use also increased. The study period saw the widespread use of chlorination at reservoirs in the Township for groundwater treatment. Although risk factors were stable at each parcels, the variability in risk factors across parcels described in Chapter 3 demonstrate that the mixed urban-rural setting of the Township provide a good setting for studying the effect of these risk factors on *H. pylori* infection.
6.1.2.3 Chapter 4: The prevalence and incidence of Helicobacter pylori-related illnesses in the Township of Langley, British Columbia

In Chapter 4, the prevalence and incidence of \textit{H. pylori}-related illnesses were measured using administrative data in the form of health care utilization records. The health outcomes analyzed were those for which \textit{H. pylori} has a well-established role in their development. Specific definitions for these outcomes were based on International Classification of Disease (ICD) codes used in the administrative data.

The analysis demonstrated that gastric cancer, peptic ulcer disease and gastritis still represent a substantial burden on the population and to the health care system in British Columbia, although there was a decline observed in the prevalence of both peptic ulcer disease and gastritis. To address potential problems with the reliability of ICD coding, an algorithm was developed. For the incident cases of peptic ulcer disease and gastritis identified, over 90% and 75%, respectively, had a relevant diagnostic procedural code on the date of diagnosis or within one year prior.

This paper also demonstrated a seasonal trend in the incidence of gastritis, with more cases consistently diagnosed in the rainy winter months, which supports the theory on an ecological level that environmental factors influence \textit{H. pylori}-related disease.

6.1.2.4 Chapter 5: The influence of drinking water and residential environment on Helicobacter pylori infection

In Chapter 5, the relationship between \textit{H. pylori} infection and the risk factors described in Chapters 2 and 3 was investigated. A nested case-control design was used within a cohort of Township residents covering a period of 1997 to 2003. \textit{H. pylori} carrier status was derived from a database of serological results. The analysis contrasted results between different age groups in order to address issues related to relevant periods of acquisition and the appearance of symptoms.

Multiple regression analysis revealed an elevated risk of infection for individuals receiving groundwater (versus a mixture of groundwater and surface water) and living on land associated with non-residential activities (i.e. agricultural). The protective effect of receiving water from
mixed sources was a surprise and is likely a result of the additional treatment of the surface water prior to mixing with groundwater in part of the municipal water system. The association with non-residential land uses may be related to direct contact with agricultural exposures. Lower socioeconomic status was also a significant risk factor for infection.

No relationship was observed of *H. pylori* infection with sewage disposal system. In the North American context, this may not be as relevant as in the setting of developing countries, stressing the importance improved hygiene.

When assembled as a collected work, the preceding chapters:

1. Provide rationale for the investigation of a link between environmental water characteristics and *H. pylori* infection;
2. Demonstrate the concept of how different environmental risk factors interact to influence *H. pylori* transmission;
3. Outline the process for obtaining and combining environmental data from different sources to characterize variables at the individual level for large population-based health studies;
4. Summarize the distribution and variability of risk factors related to residential water system and environment among residents of the Township of Langley;
5. Describe the burden of *H. pylori*-related diseases in a North American population; and
6. Investigate environmental factors associated with *H. pylori* infection related to residential water and related characteristics.

### 6.2 Contributions

This work made various contributions to the fields of epidemiology and public health. The main foci of the papers that comprise this dissertation are the burden of *H. pylori* infection and the identification of important risk factors for infection. In broader terms, the methods used here provide a good example for other population-based environmental epidemiologic studies.

#### 6.2.1 *Helicobacter pylori* literature

The findings of this work with respect to *H. pylori* are broad. The theoretical model outlined in Chapter 2 is useful for researchers and policy makers. It reviews key findings, proposes a
relationship between the various risk factors and provides an explanation for the transmission cycle in the environment.

This seems to be the first study to use population-based sampling to investigate the association between \textit{H. pylori} infection and a range of environmental factors. Where previous studies have focused mainly on small cohorts of individuals, this study investigated all individuals in the study area who submitted a sample for serological testing over the course of the study period. The availability of administrative data was used to characterize important exposures in the study area and connect these to health outcomes at the individual level identified through health care utilization records. The results indicating an association of \textit{H. pylori} infection with water supply, water source and land use are among a few studies that have been done in a developed country and complement the results of previous investigations while informing on differences within the North American context. In addition, the significance of the water source variable but in the direction opposite of that commonly found in the literature shows how setting and context are important when considering influences of waterborne \textit{H. pylori} transmission.

This work described the burden of gastritis, peptic ulcer disease and gastric cancer, outcomes with a well established connection with \textit{H. pylori} infection, using administrative health records. Only two studies could be identified that used ICD9 codes from physician billings data to estimate the distribution of peptic ulcer disease in the population \cite{2, 3} and neither of these used a combination of inpatient and outpatient health care contacts. The reporting of the prevalence and incidence of these outcomes is a unique contribution to the literature and provides an important update on the status of these diseases in Canada.

\subsection*{6.2.2 Environmental epidemiology research}

Certain aspects of the methods used and the procedures developed in the completion of this work represent innovative contributions to epidemiologic research in general and in British Columbia in particular.

Data sources were identified and methods described that are necessary for the building of a GIS for exposure assignment in epidemiological studies. Data from municipal government was used for compiling the building blocks of address-level exposure data in the study and was
complemented by data from a variety of sources. The flexibility of GIS software with its ability to merge data using traditional methods, such as deterministic linkage, as well as geographic data using spatial joining functions, was essential for bringing together diverse data and producing a single exposure database. This database has already been used to investigate patterns in reportable enteric disease [4] and will also be used for a larger study looking at the relationship between exposures and acute gastrointestinal health outcomes called the Langley Water and Health Study.

The blueprint for this study may serve as an example for a number of other population-based studies. The availability of electronic administrative databases makes this type of research study possible and allows the investigation of research questions that would otherwise be difficult or very costly to consider. The key to conducting this research was the ability to link between databases based on a single unique identifier or combination of identifiers. This is the impetus for the development of the Population Health and Learning Observatory, a collaborative effort of many researchers in BC and Canada with the purpose of studying determinants of health in the population [5]. Observational epidemiological studies such as that described in this work are excellent means for evaluating risks at the population level and will likely be used with increasing frequency for risk assessment and regulatory purposes.

A new methodology was developed in the linking of individuals’ health records to the environmental variables of their place of residence. This procedure was influenced by privacy legislation in British Columbia. In order to protect privacy of the study population there was a need to rely on non-researchers to perform linkages at the address level. In working with the Health Information Management Division at the BC Ministry of Health, a process was developed for linking individuals via street address in the Client Registry, the Ministry’s repository for demographic information of people covered by the provincial, universal health insurance plan. For the 144,274 individuals identified from cohort extraction conducted by the Centre for Health Services and Policy Research, address information was extracted from the Client Registry. Originally, a linkage process was attempted based on full address in the exposure database and any of the four address fields in the Client Registry that resulted in a 75% match. Unmatched addresses were examined and it became clear that a data cleaning process was required to eliminate characters in the Client Registry address fields not present in the exposure database.
address fields. Once cleaned, a deterministic matching procedure resulted in the matching of 130,929 cohort members (90.8%). This surpassed the previously decided minimum goal of 90% that was decided upon to maintain a balance between the completeness of research and the devotion of time on behalf of data stewards at the Ministry. This methodology could be enhanced for future projects.

As a measure to determine the success of the linkage and potential sample selection bias, the number of cohort members linked per year and the mean number of person-days linked per cohort member was investigated over the study period (Table 6.1). The percentage of the estimated Township population (based on Census data) that was included in the study rose each year and was nearly 100% at the end of the study period. The lower values in the earlier part of the study may be a result of the lack of availability of municipal databases for these years meaning that some addresses were missing and could not be used to extract the study population from existing databases. The average of approximately 11 months of person-days linked per year is likely a result of many individuals linked for the full year and others who moved in or out of the Township during this time. These results are encouraging for the internal validity of studies using the linked data and indicate that future studies could benefit from employing these methods.

**Table 6.1  Cohort members and person-days linked per year of study**

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of cohort members linked*</th>
<th>Estimated population</th>
<th>% population linked</th>
<th>Person-days linked</th>
<th>Mean number of person-days linked per cohort member</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>65,845</td>
<td>80,000</td>
<td>82.30%</td>
<td>21,522,347</td>
<td>326.9</td>
</tr>
<tr>
<td>1996</td>
<td>70,910</td>
<td>80,179</td>
<td>88.40%</td>
<td>23,291,593</td>
<td>328.5</td>
</tr>
<tr>
<td>1997</td>
<td>74,731</td>
<td>81,500</td>
<td>91.70%</td>
<td>24,588,975</td>
<td>329.0</td>
</tr>
<tr>
<td>1998</td>
<td>78,117</td>
<td>82,250</td>
<td>95.00%</td>
<td>25,892,677</td>
<td>331.5</td>
</tr>
<tr>
<td>1999</td>
<td>80,722</td>
<td>83,750</td>
<td>96.40%</td>
<td>26,946,327</td>
<td>333.8</td>
</tr>
<tr>
<td>2000</td>
<td>82,634</td>
<td>85,000</td>
<td>97.20%</td>
<td>27,921,751</td>
<td>337.9</td>
</tr>
<tr>
<td>2001</td>
<td>85,257</td>
<td>86,896</td>
<td>98.10%</td>
<td>28,723,739</td>
<td>336.9</td>
</tr>
<tr>
<td>2002</td>
<td>87,747</td>
<td>89,375</td>
<td>98.20%</td>
<td>29,547,126</td>
<td>336.7</td>
</tr>
<tr>
<td>2003</td>
<td>90,279</td>
<td>91,000</td>
<td>99.20%</td>
<td>29,627,871</td>
<td>328.2</td>
</tr>
</tbody>
</table>

* Individual included if linked for at least one day in the year
6.2.3 Stakeholder involvement

Collaborating with the various stakeholders was an invaluable component of this research as they not only provided data but also knowledge of the study area and expertise of how existing environmental data was assembled. This collaboration was facilitated by Dr. Hans Schreier, committee member and professor of water resource management in the Institute for Resources, Environment and Sustainability in UBC, whose involvement with the Township over a long period of time has resulted in a trusting working relationship. For the larger Langley Water and Health Study, an advisory committee was formed to aid in the acquisition of data, the communication of proceedings between the study team and individuals within the participating groups, and in how to communicate results to the public. This committee could serve as a model for developing and maintaining good relationships for conducting environmental research with communities in the future.

6.3 Limitations and future work

There were a number of challenges faced during the completion of this work. These were primarily related to data access and quality, and epidemiological issues such as bias in the methodology that affect validity of study results, and interpretation of these results.

6.3.1 Data access

Data availability was an issue on many levels. Health care utilization records that were extracted by the Centre for Health Services and Policy Research (CHSPR) as part of the BC Linked Health Database required approval from the data steward, the BC Ministry of Health. The Data Access Request (DAR) was filed in September 2004 and no progress had been made on the approval process as of November 2005. This delay, while not unique to this project, and the concern about the timeliness in which approval could be obtained, almost led to the abandonment of the work. Intervention by individuals in the BC research community spurred action at the Ministry and the request was approved.

After this first approval step, the linkage of individuals to exposures in our database based on residential address was still pending and had to be conducted by a programmer at the Ministry of Health in Victoria due to privacy concerns. Privacy legislation also restricted the methods and
precluded the use of a spatial analysis of cases. All geographic indicators were removed in order to protect the privacy of cohort members. It would have been ideal to analyze the data in geographic format to permit the mapping of cases in search of clusters of *H. pylori* carriers. Hopefully, this analysis will be feasible in the future as valuable information can be produced from this type of research.

It was anticipated that more administrative data from the municipality would be available for study prior to 1999. The lack of data prior to 1999 may have led to a lower proportion of individuals linked in the early years of the study due to missing address data; however there was limited development in the Township during the early part of the study period according to people familiar with the area which would mean that few addresses would have ceased to exist. If changes with respect to water supply, sewage disposal system and land use did occur in this period, this may have led to misclassification of exposures for individuals. As the number of individuals relying on municipal services grew with increasing residential (i.e. urban) development, it is possible that other nearby parcels also switched over in the period prior to 1999. Assigning exposure from the April 1999 database would therefore have misclassified individuals as using municipal water supply and sewage disposal rather than private. This misclassification only would have affected a minimum number of parcels and would have biased the results in the direction of the null hypothesis provided it was not systematic. Further, the stability of exposures at the parcel level over time demonstrated in Chapter 3 indicates that misclassification did not appear to be an issue.

In Chapter 5, it was hoped that serology data would be available for at least the nine year period from 1995 to 2003. However, serology results were provided only for the time period from April 1997 to December 2003. In the multivariable analysis it would have been ideal to obtain a variable indicating household size as family size is associated with *H. pylori* infection. Extraction of this variable is possible through the BC Linked Health Database, however, as of this writing the data had not been received.

### 6.3.2 Data quality

Working with administrative data often means that there is an absence of a gold standard that can be used to assess validity of data. A survey is currently being conducted in the study area as part
of the larger Langley Water and Health Study to determine agreement between the administrative records used for exposure assessment and residents’ knowledge of the water system characteristics in their homes. Absence of agreement may indicate a data reliability issue but could also be attributed to a lack of resident knowledge about water system characteristics in their homes. This survey will also estimate the mobility of cohort members (frequency of moving) and how often the exposures change for an individual. This has the potential to provide insight into the bias introduced by using variables associated with the date of diagnosis as the risk factor of interest in the analysis for Chapter 5.

Another area of uncertainty lies in the addresses in the Client Registry that were used as residential addresses. According to the Client Registry Business Rules, there are fields for both home and mailing addresses in the Registry [6]. The only case where a mailing address might be different from a home address is if the individual had a work address as a mailing address. Cases where a person was linked to more than one address for a given period of time were observed, often with one as the home address and the other as the mailing address. The programmer at the Ministry added a field indicating address type and home address was used for analysis. There are legitimate reasons that an individual is linked to more than a single address for a given period of time. For example, a child of parents that no longer live together may multiple addresses on record.

6.3.3 Methodology and study design

The study design used here focused on residential environment and determinants of drinking water at place of residence. In one study that attempted to measure how much water people consume at home, they found that only approximately half of surveyed individuals indicated their home as their only water source and over thirty per cent reported not using their home water source at all [1]. This would likely lead to a non-directional misclassification in our study, unless residential water supply was correlated with the use of an alternate drinking water source. For example, if individuals relying on a private well are more likely to supplement their diet with bottled water due to concerns over the quality of their tap water. The results of the survey will shed some light on these questions in the future. It would be ideal to complement the epidemiological analysis with water sampling to determine whether these surrogates of *H. pylori* exposure are associated with the presence of the bacteria in the water supply.
This study focused mostly on adults, a population in which an ongoing risk of infection has been demonstrated. However it would have been preferable to have both a child and an adult population to investigate whether the measured risk estimates were consistent between the two groups. As an alternative, cases and matched controls were split into four groups based on age at diagnosis and noteworthy differences were observed. In the younger groups, it is more likely that variables on the date of diagnosis are either the relevant risk factors or are more representative of the longer-term (i.e. lifetime) risk factors than for the older group. To study children, a different diagnostic test such as the urea breath test or stool antigen test is recommended that has demonstrated higher sensitivity in that age group. However, the high specificity of the serologic tests suggests that younger participants who did test positive are unlikely to have been misclassified.

The use of administrative data was also used in the ascertainment of health outcomes in Chapter 4. The validity and reliability of coding in administrative data has been previously questioned. In this case, an algorithm was developed using fee items and procedural codes for that individual. However, without the presence of medical charts as a gold standard, it is not possible to verify the accuracy of the algorithm used.

6.3.4 Interpretation of results

The multiple regression analysis conducted in Chapter 5 demonstrated that water source and land use were associated with \( H. pylori \) infection in the 21-40 year age group and a lack of an association for the water treatment variable. In the study, water treatment referred to the practice of chlorination employed in the Township to the groundwater sources. As surface water entering the municipal system also gets treated prior to mixing with groundwater, it is hard to separate the effects of water treatment from water source.

Interpretation of the significance of the land use variable represents a source of uncertainty with regard to the possible mechanisms of transmission although the overall relationship between risk of \( H. pylori \) and proximity to agricultural land use seems valid. Either the proximity of agricultural lands to water sources results in contamination of drinking water or direct contact with agricultural exposures leads to infection.
These findings demonstrate association rather than causation as, at the very least, they are unable to satisfy the temporality criterion required to demonstrate causality. A cohort study is required to be able to demonstrate causal linkages between the presence of certain risk factors and new *H. pylori* infections.

### 6.4 Impact on Policy and Future Research

There are numerous uses and implications of this research for public health. First, the proposed theoretical model of *H. pylori* transmission (Figure 2.2) identified environmental pathways and indicated places that may be suitable targets for a public health intervention. For example, water treatment is identified as a means of minimizing the passage of the bacteria to susceptible individuals. As food has been identified as a possible vehicle for transmission, efforts directed at proper sanitation of fruits and vegetables may have the added benefit of disrupting the cycle. The interaction of septic systems and private wells is another potentially important phenomenon. The two must be kept separate by large areas to ensure that septic effluent does not interact with drinking water.

Individuals receiving groundwater only and living on agricultural lands appear to be at the highest risk of infection in addition to those of lower socioeconomic status. Most individuals on a private water supply are recipients of groundwater only and many lack a means of treating their water. It is likely that the chlorination of surface water prior to mixing provides an extra means of protection for individuals receiving a mixture of surface water and groundwater. Chlorination is a commonly-used measure to ensure delivery of potable water and has been identified in experimental studies but not yet demonstrated in an epidemiological study until now. The chlorination of water is somewhat controversial as by-products have been identified as potential carcinogens, however it has not been conclusively demonstrated that the association between drinking chlorinated water and cancer is causal \[7\]. Public health messages should emphasize the importance of chlorination as an effective means of decreasing the incidence of gastrointestinal illness. If chlorination is not acceptable to people, other means of treatment should be made available so that individuals have an opportunity to combat *H. pylori* infection in addition to other gastrointestinal ailments. The results indicate that efforts to combat acute and chronic gastrointestinal illness, originating through the ingestion of water, can be combined.
Successful primary prevention of *H. pylori* infection could substantially increase life expectancy and quality of life [8].

The management of water and surrounding land is an important priority for public health officials for a number of reasons. In this study, an association between land use and *H. pylori* was observed. The presence of agricultural land use for an individual’s residence in this study may be indicative of direct contact with agricultural exposures. As residential development increases and people live in areas formerly used for agricultural activities, public health officials should be vigilant of potential for increased exposure to harmful risk factors.

The results obtained here will be the basis for future research to be done in this area and bring about some experiences that can be used for an ideal study design to investigate *H. pylori* transmission. With a demonstrated link among adults in the Township between *H. pylori* infection and drinking water and environmental characteristics, focus can be directed towards children to see if these relationships hold true in the pediatric population. From an exposure perspective, following children would be an improvement because it would eliminate the uncertainty regarding when the bacteria was acquired. Combining secondary (i.e. administrative) data sources with primary data collection may be a cost-effective means of conducting research on this topic in the study area. Investigating patterns in past health care utilization among *H. pylori* carriers may reveal symptoms common among people with recent *H. pylori* acquisition that has been elusive to researchers to date. Testing water samples in the Township for *H. pylori* DNA may provide further evidence of water as a risk factor for *H. pylori* infection.

Population-based observational studies are relevant to policy makers because they provide a real world alternative for the evaluation of risks that cannot be achieved through theoretical modeling and experimental exposure studies involving humans or animals [9]. This study design can be used to evaluate the effects of one recent change in legislation. In the past, the Fraser Health Authority was responsible for the inspection of septic systems installed in the Township. As of 2005, the company installing the septic system now has this responsibility, a regulatory change that concerns many public health inspectors who have observed and failed several septic system installations. This methodology allows for the evaluation of GI illness for individuals with
systems installed before and after the regulatory change and results may impact future changes to policy.

Monitoring *H. pylori* infection and related outcomes is important. Although they appear to be in decline, a substantial burden on the population still exists given the high number of people affected, the pervasiveness of the symptoms and related illnesses, and the potential long term consequences of untreated infection. It is important to use the tools currently available to monitor the trend in *H. pylori* infection and related outcomes.

Knowledge translation is essential for researchers and policy makers to benefit on the experience gained and increased understanding resulting from research. Not only has a relevant public health matter been investigated but numerous lessons were learned in the development of this research that can be shared, especially with regard to the use of administrative data and population-based studies. The importance of *H. pylori* and its role in disease has been highlighted by the recent awarding of the Nobel Prize in medicine to Warren and Marshall, the physicians who first isolated it and proceeded to investigate its role in the diseases of the upper gastrointestinal system. Chapter 1, which outlines the conceptual model of *H. pylori* transmission, has already been published in the journal Epidemiology and Infection. Chapter 2, which outlines the compilation of the exposure GIS can be useful to others who wish to conduct population-based environmental research and will be submitted to the International Journal of Health Geographics. The description of the burden of *H. pylori*–related illnesses in Chapter 3 informs on the prevalence and incidence of these illnesses in a Canadian population and is being prepared for submission to the journal Helicobacter. Chapter 4 shows the relationship between *H. pylori* infection and a variety of environmental exposures and will be submitted to the American Journal of Epidemiology. In all, this work has a substantial impact on the development of research methods in population-based environmental health studies and evidence for further investigation of the link between water and environmental factors and *H. pylori* transmission based on real-life observations.
6.5 References


### APPENDICES

**Appendix A – Behavioural Research Ethics Board Certificate of Approval**

**The University of British Columbia**  
**Office of Research Services**  
**Behavioural Research Ethics Board**  
**Suite 102, 6190 Agronomy Road, Vancouver, B.C. V6T 1Z3**

**CERTIFICATE OF APPROVAL- MINIMAL RISK RENEWAL**

<table>
<thead>
<tr>
<th>PRINCIPAL INVESTIGATOR:</th>
<th>DEPARTMENT:</th>
<th>UBC BREB NUMBER:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judith L. Isaac-Renton</td>
<td>UBC/Medicine, Faculty of/Pathology &amp; Laboratory Medicine</td>
<td>H04-80300</td>
</tr>
</tbody>
</table>

**INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:**

**CO-INVESTIGATOR(S):**
- Mieke W. Koehoorn
- Neil Bellack
- Kay Teschke
- James W. Atwater
- Negar Elmieh
- Hanspeter E. Schreier
- Hadi Dowlatabadi
- Ying C MacNab

**SPONSORING AGENCIES:**
Canadian Institutes of Health Research (CIHR) - "The Impact of Water Quality Risk Factors on Endemic Gastrointestinal Illness in Rural-Urban Setting"

**PROJECT TITLE:**
The Impact of Water Quality Risk Factors on Endemic Gastrointestinal Illness in Rural-Urban Setting

**EXPIRY DATE OF THIS APPROVAL: June 18, 2008**

**APPROVAL DATE: June 18, 2007**

The Annual Renewal for Study have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

*Approval is issued on behalf of the Behavioural Research Ethics Board*