Informatics and the Electronic Medical Record for Syndromic Surveillance in Companion Animals: Development, Application and Utility

Anholt, Rae

http://hdl.handle.net/11023/453

Downloaded from PRISM Repository, University of Calgary
Informatics and the Electronic Medical Record for Syndromic Surveillance in Companion Animals: Development, Application and Utility

by

Rae Michele Anholt

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MEDICAL SCIENCES

CALGARY, ALBERTA

JANUARY, 2013

© RAE MICHELE ANHOLT 2013
Abstract

There is growing interest in companion animal surveillance to provide an early warning for emerging environmental health risks including zoonotic diseases and chemical contamination of food, water and air. The objectives of this thesis were to plan, implement and evaluate a companion animal surveillance system capable of using existing clinical data to detect emerging environmental risks to people and/or pets. Twelve companion animal practices participated in a sentinel veterinary practice syndromic surveillance network. Enteric diseases were targeted due to their anticipated prevalence and relevance to animal and public health. Customized data extraction software automatically extracted the required fields from the veterinary practice management software at each practice and exported them to a secure data warehouse (n = 447,388 records from January 1, 2007 to December 31, 2010; extraction 99.4% accurate). The records were in free-text with no diagnostic codes or standardized nomenclature. A categorization dictionary developed using commercially available text-mining software automatically classified and retrieved cases of enteric syndrome (n = 18,832; sensitivity, 87.6% and specificity, 99.3%). Using this data it was possible to identify clusters of enteric disease in space and time, describe these patterns by the host factors and observe patterns of antimicrobial use (AMU). There was often not enough specific information recorded in the medical record to describe the clusters by the proportion of vaccinated cases, the probable etiological agent or by environmental factors that may be predictive of increased risk. The lack of contextual information in the medical records limited the usefulness of this system to provide an early warning of environmental hazards. The system could be used to monitor temporal trends in AMU or conduct analytical studies to explore patterns of AMU and concomitant increases in antimicrobial resistance. The system may be useful for studies that support evidence-based veterinary medicine. Future companion animal surveillance systems should consider imposing structured clinical reports onto the submissions from the participating veterinarians. This would address many of the data and system quality issues encountered in this study, but may have an impact on the willingness of veterinarians to participate in community-based surveillance.
Preface

The following manuscript has been published. R. Michele Anholt obtained and analyzed the data, interpreted the results and wrote the paper. Her co-author and supervisor, Dr. Craig Stephen and co-author, Dr. Ray Copes contributed intellectual content and provided critical review of the paper. Written permission for reproduction of the article in its entirety has been obtained from the publisher and the co-authors.

Acknowledgements

I would like to thank the sentinel veterinary practice owners; this project would not have been possible without their generous participation. I am most grateful.

To my supervisor, Dr. Craig Stephen, who trusted that I could work independently but when I asked was always willing to provide insight, expertise and encouragement. His expectation for excellence challenged me and my thesis is all the better for it. Thank you.

I am very grateful to my excellent supervisory committee; Drs. John Berezowski, Carl Ribble and Margaret Russell. I relied on your support, enthusiasm for the project and attention to detail. Thank you.

I would like to acknowledge my generous financial support provided by the Centre for Coastal Health, the Faculty of Veterinary Medicine at the University of Calgary, Mitacs, Pfizer Animal Health and Merck Animal Health. In-kind support from Pet Lynx is also gratefully acknowledged.

To the faculty and staff in the Department of Medical Sciences, Faculty of Medicine and the Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, thank you for supporting me and my research.

Thank you to my lab mates Drs. Kate Sawford and Amanda Salb, and all of my fellow graduate students without whose encouragement, intelligent appraisals and humour would have made these past 5 years bleak indeed.

I would also like to thank my parents, siblings, extended family and friends for your love, support and understanding even when I was most neglectful. You are all so precious to me.

I am most grateful to my husband Craig and sons Bryce and Marshall. You had to make sacrifices but you were always encouraging, interested in my project and proud of my efforts. I love you dearly.
Dedication

To my husband Craig Dorin who supported my wish to do a PhD and then proceeded to do whatever was necessary for me to achieve it; I am so grateful to you.
# Table of Contents

Abstract ............................................................................................................................... ii  
Preface ............................................................................................................................... iii  
Acknowledgements ............................................................................................................. iv  
Dedication ............................................................................................................................ v  
Table of Contents ............................................................................................................... vi  
List of Tables ..................................................................................................................... xi  
List of Figures and Illustrations ........................................................................................ xiv  
List of Abbreviations ....................................................................................................... xvi

CHAPTER ONE: INTRODUCTION ................................................................................. 1  
1.1 Surveillance system goals ........................................................................................ 1  
   1.1.1 Animals as sentinels of diseases for environmental hazards ......................... 1  
   1.1.2 Companion animals as sentinels of environmental risks to people ............... 2  
      1.1.2.1 Companion animals as sentinels of endemic zoonotic disease ............... 5  
      1.1.2.2 Companion animals as sentinels of emerging zoonotic diseases ........... 6  
      1.1.2.3 Companion animals as sentinels of non-infectious diseases ................. 8  
   1.1.3 Companion animals as sentinels for other animals ......................................... 10  
1.2 Surveillance system objectives .............................................................................. 10  
   1.2.1 Provide early warning ..................................................................................... 10  
   1.2.2 Monitor endemic disease ................................................................................ 12  
   1.2.3 Evaluate medical and surgical management of animal diseases ................. 12  
   1.2.4 Investigate the epidemiology of a disease ...................................................... 13  
   1.2.5 The purpose and objectives of a pilot companion animal surveillance system 14  
1.3 Governance of Animal Disease Surveillance in Canada .................................... 14  
1.4 Develop case definitions ....................................................................................... 16  
1.5 Determine the type of system ............................................................................... 17
1.5.1 Approaches to animal disease surveillance .....................................................17
1.5.2 Syndromic surveillance ...................................................................................18
1.5.3 Sentinel surveillance .....................................................................................20
1.5.4 Develop data collection instruments ...............................................................21
   1.5.4.1 Informatics .............................................................................................21
   1.5.4.2 Knowledge management and text-mining to support analysis of
       surveillance data..............................................................................................22
1.6 Thesis objectives and structure ...........................................................................23

CHAPTER TWO: THE DEVELOPMENT AND IMPLEMENTATION OF DATA
EXTRACTION AND DATA WAREHOUSING TECHNOLOGIES FOR
COMPANION ANIMAL DISEASE SURVEILLANCE.........................................29
2.1 Introduction ..............................................................................................................29
   2.1.1 Companion animals as sentinels....................................................................29
   2.1.2 Informatics .....................................................................................................31
   2.1.3 Evaluation frameworks for surveillance systems ............................................32
2.2 Methods ...................................................................................................................34
   2.2.1 Sentinel recruitment ......................................................................................34
   2.2.2 Development of the data extraction tool .........................................................35
   2.2.3 Evaluation .......................................................................................................37
      2.2.3.1 User Experience .....................................................................................37
      2.2.3.2 System Quality ......................................................................................38
      2.2.3.3 Information Quality ...............................................................................39
2.3 Results ......................................................................................................................40
   2.3.1 Sentinel recruitment and user experience ......................................................40
   2.3.2 Development of the data extraction program, user experience and system quality
       ..............................................................................................................................43
   2.3.3 Data management and user experience ..........................................................45
   2.3.4 Information Quality ........................................................................................46
2.4 Discussion .................................................................................................................47
2.4.1 User experiences – sentinel practices ......................................................... 47
2.4.2 System quality and system user experiences ........................................... 47
2.4.3 Information quality .................................................................................. 49
  2.4.3.1 Data accuracy ............................................................................... 49
  2.4.3.2 Geographic representativeness ....................................................... 50
  2.4.3.3 Demographic representativeness ...................................................... 51
  2.4.3.4 Selection of the denominator .......................................................... 51
2.5 Conclusion .................................................................................................. 52

CHAPTER THREE: MINING FREE-TEXT MEDICAL RECORDS FOR COMPANION
ANIMAL ENTERIC SYNDROME SURVEILLANCE ........................................ 62
  3.1 Introduction .............................................................................................. 62
  3.2 Methods .................................................................................................... 65
    3.2.1 Data .................................................................................................. 65
    3.2.2 Development of the categorization dictionary in WordStat ............... 66
    3.2.3 Validation of the categorization dictionary ......................................... 68
  3.3 Results ........................................................................................................ 69
    3.3.1 Development of the categorization dictionary ................................. 69
    3.3.2 Validation of the categorization dictionary ....................................... 70
    3.3.3 Evaluation of the classification errors ............................................. 71
  3.4 Discussion .................................................................................................. 72
  3.5 Conclusion .................................................................................................. 74

CHAPTER FOUR: SPATIAL-TEMPORAL CLUSTERING OF COMPANION ANIMAL
ENTERIC SYNDROME: DETECTION AND INVESTIGATION ....................... 79
  4.1 Introduction .............................................................................................. 79
  4.2 Methods .................................................................................................... 82
    4.2.1 Study population and enteric syndrome data .................................... 82
    4.2.2 Time series ....................................................................................... 83
    4.2.3 Scan statistic for space-time clusters ............................................... 83
List of Tables

Table 1.1: The zoonotic capabilities of federally notifiable enteric pathogens (Public Health Agency of Canada) ............................................................................................................... 27

Table 2.1: Evaluation of the sentinel surveillance system........................................................................................................ 55

Table 2.2: List of variables extracted by the University of Calgary Data Extraction Program (UCDEP)................................................................. 56

Table 2.3: Data completeness in the extracted medical records, n = 296,087 records ............... 57

Table 2.4: Comparing the proportion of age groups of dogs and cats seen in Canada to those seen by the sentinel practices in 2008................................................................. 60

Table 2.5: Proportions of sterilized cats and dogs in Canada compared to those seen by the sentinel practices in 2008................................................................. 61

Table 3.1: Case definition applied to the sample of cases to determine if enteric syndrome positive or enteric syndrome negative. ................................................................. 76

Table 3.2: The accuracy of the text-miner, WordStat™, to identify cases of enteric syndrome compared to a manual review of the records as the external reference standard............... 77

Table 3.3: Sources of false positives and false negatives when classifying and retrieving cases of enteric syndrome with the text-miner................................................................. 78

Table 4.1: Expected values of the variables for the enteric syndrome cases calculated from the cases of enteric syndrome, January 1, 2007 to December 31, 2009 ( n = 13,063) ........... 95

Table 4.2: Significant (< 0.05) retrospective (R) space-time clusters of enteric syndrome ........ 98

Table 4.3: Significant (< 0.05) prospective (P) space-time clusters of enteric syndrome ........... 99
Table 4.4: Age standardization to compare the association of animals that had not been spayed or neutered between the etiologically described clusters and the clusters with expected numbers of etiologies................................................................. 101

Table 4.5: Clusters with statistically different (p < .05) proportion of feline cases when compared to the reference population................................................................. 102

Table 5.1: The accuracy of the text-miner, WordStat™ to identify cases that had diagnostic testing compared to a manual review of a sample of 500 records serving as the external standard. ................................................................. 117

Table 5.2: The accuracy of the text-miner, WordStat™ to identify cases with a stated etiological diagnosis compared to a manual review of a sample of 500 records serving as the external standard. ................................................................. 117

Table 5.3: The accuracy of the text-miner, WordStat™ to identify cases treated with an antihelmintic compared to a manual review of a sample of 500 records serving as the external standard. ................................................................. 118

Table 5.4: The accuracy of the text-miner, WordStat™ to identify cases treated with an antimicrobial compared to a manual review of a sample of 500 records serving as the external standard. ................................................................. 118

Table 5.5: Cases with a stated etiological diagnosis from the sample of 500 enteric positive cases .................................................................................................................. 119

Table 5.6: Distribution of antimicrobials used by the sentinel veterinary practices in the treatment of the 8441 enteric syndrome cases (n = 18,832) in 2007, 2008, 2009 and 2010.................................................................................................................. 121

Table 6.1: Scenarios presented to the workshop participants to focus a discussion on the essential elements of interdisciplinary collaboration for emerging and zoonotic diseases. 142
Table 6.2: Distribution of $771 million in Canadian Institute for Health Research (CIHR) awards and grants for interdisciplinary collaborations in 2008/2009 (331). ................. 144

Table 6.3: Summary of the steps to effective interdisciplinary collaboration as described in the business management literature (335) and the systems engineering literature (343). .. 145
List of Figures and Illustrations

Figure 1.1: The sentinel framework in context. Reprinted with permission from (7) ................. 26

Figure 1.2: Objectives and thesis structure ................................................................................. 28

Figure 2.1: Map of study area based on the first 3 digits of the owner’s home postal codes, the forward sortation area (FSA) ................................................................................. 54

Figure 2.2: Number of dog cases seen by the sentinel practices divided by the number of licensed dogs in each of the City of Calgary’s 35 Forward Sortation Units (FSA) ............. 58

Figure 2.3: Distribution of pet species in Canadian homes compared to the distribution of pet species (ie. number of dog cases divided by all cases) seen by the sentinel practices. ....... 59

Figure 4.1: Schematic of the methods used in this study (after (145)) ......................................... 94

Figure 4.2: Number of cases of enteric syndrome seen by the sentinel practices on each day of the week in 2007, 2008 and 2009 ............................................................... 96

Figure 4.3: Daily count and 7 day moving average of counts of the enteric syndrome cases seen by the sentinel veterinary practices in 2007, 2008 and 2009 ......................... 97

Figure 4.4: Map showing location of etiologically-defined clusters. ......................................... 100

Figure 4.5: Exposure information found in the records of cases from statistically significant time-space clusters of enteric syndrome ................................................................. 103

Figure 5.1: Flow diagram describing the laboratory diagnostics, percentages of cases with a diagnosis, and antihelmintic and antimicrobial management of enteric syndrome cases... 120

Figure 5.2: Time series of the proportion of enteric cases treated with any antimicrobial, nitroimidazole class and penicillin class ........................................................................ 122

Figure 5.3: Scattergrams of the numbers of cases of enteric syndrome treated with .......... 123
Figure 6.1: The elements of effective ID collaboration and KTE .............................................. 143
### List of Abbreviations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMR</td>
<td>Antimicrobial resistance</td>
</tr>
<tr>
<td>AMU</td>
<td>Antimicrobial use</td>
</tr>
<tr>
<td>BSE</td>
<td>Bovine spongiform encephalopathy</td>
</tr>
<tr>
<td>CAD</td>
<td>Canadian dollars</td>
</tr>
<tr>
<td>CFIA</td>
<td>Canadian Food Inspection Agency</td>
</tr>
<tr>
<td>CIHR</td>
<td>Canadian Institute for Health Research</td>
</tr>
<tr>
<td>CPV</td>
<td>Canine parvovirus</td>
</tr>
<tr>
<td>D &amp; M IS</td>
<td>DeLone and McLean model of information success</td>
</tr>
<tr>
<td>DSA</td>
<td>Data sharing agreement</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assays</td>
</tr>
<tr>
<td>EMR</td>
<td>Electronic medical record</td>
</tr>
<tr>
<td>EBVM</td>
<td>Evidence-based veterinary medicine</td>
</tr>
<tr>
<td>ETL</td>
<td>Extract, transform and load</td>
</tr>
<tr>
<td>FSA</td>
<td>Forward sortation area</td>
</tr>
<tr>
<td>GI</td>
<td>Gastrointestinal</td>
</tr>
<tr>
<td>GM</td>
<td>General Motors</td>
</tr>
<tr>
<td>ICD-9</td>
<td>International Classification of Diseases, 9th revision</td>
</tr>
<tr>
<td>ID</td>
<td>Interdisciplinary</td>
</tr>
<tr>
<td>IE</td>
<td>Information extraction</td>
</tr>
<tr>
<td>IT</td>
<td>Information technology</td>
</tr>
<tr>
<td>KTE</td>
<td>Knowledge transfer and exchange</td>
</tr>
<tr>
<td>LAS</td>
<td>London Ambulance Service</td>
</tr>
<tr>
<td>MRSA</td>
<td>Methacillin-resistant <em>Staphylococcus aureus</em></td>
</tr>
<tr>
<td>NLP</td>
<td>Natural language processing</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>ODBC</td>
<td>Open database connectivity</td>
</tr>
<tr>
<td>OIE</td>
<td>World Organization for Animal Health</td>
</tr>
<tr>
<td>OWOH</td>
<td>One World, One Health</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PHAC</td>
<td>Public Health Agency of Canada</td>
</tr>
<tr>
<td>RMAF</td>
<td>Results-based Management and Accountability</td>
</tr>
<tr>
<td>SARS</td>
<td>Severe acute respiratory syndrome</td>
</tr>
<tr>
<td>SNOMED-CT</td>
<td>Systemized nomenclature for medicine, clinical terms</td>
</tr>
<tr>
<td>SNOVET</td>
<td>Standard nomenclature of veterinary diseases and operations</td>
</tr>
<tr>
<td>SPS</td>
<td>Sanitary and Phytosanitary</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured query language</td>
</tr>
<tr>
<td>UCDEP</td>
<td>University of Calgary Data Extraction Program</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>US</td>
<td>United States</td>
</tr>
<tr>
<td>VMDB</td>
<td>Veterinary medical database</td>
</tr>
<tr>
<td>VPMS</td>
<td>Veterinary practice management software</td>
</tr>
<tr>
<td>WNV</td>
<td>West Nile Virus</td>
</tr>
<tr>
<td>WTO</td>
<td>World Trade Organization</td>
</tr>
</tbody>
</table>
Chapter One: Introduction

“While often viewed in terms of public health, the challenges of detecting natural and intentionally introduced disease outbreaks are equally shared by the plant and animal health communities” (1). This quote arising from the Forum on Microbial Threats (Institute of Medicine) addresses the epidemiological interdependency of people, animals and their shared environment. It suggests that risk arising in one domain may reflect risks in another and hence provide early warning. New surveillance strategies are required for the early detection and rapid response to emerging global health issues that arise at the human-animal-environment interface. This thesis describes the development and assessment of a new approach to companion animal surveillance that aimed to observe animals for signals relevant to risk management for both people and animals. I have described the planning of this surveillance system using steps demonstrated by Teutsch (2): describe the purpose, establish the objectives, develop case definitions, determine the type of system and develop data collection methods.

1.1 Surveillance system goals

Surveillance is an essential public service in animal health and veterinary public health. It is defined as “the ongoing systematic collection, analysis and interpretation of outcome-specific data essential to the planning, implementation and evaluation of public health practice, closely associated with the timely dissemination of these data to those who need to know. The final link in the surveillance chain is the application of these data to prevention and control” (3). This definition emphasizes that surveillance is outcome-orientated. When planning a new surveillance system the developers must first consider the purpose of the system and describe the desired goals or outcomes (2).

1.1.1 Animals as sentinels of diseases for environmental hazards

One goal of this animal surveillance system was to provide information that can link environmental factors to human health risks so that actions can be taken to improve the health and welfare of people. The National Research Council Committee on Animals as Monitors of
Environmental Hazards recommended that systems be established “for collecting data on parallel diseases of animals and humans living in the same environments” (4). Animal sentinels are monitored for disease and health to identify disease outbreaks, or to investigate the epidemiology of conditions of interest in people or other animals (5, 6). There is evidence supporting the use of companion animals as sentinels, as well as challenges and areas where there is currently insufficient knowledge to evaluate the usefulness of companion animal sentinel surveillance.

Ideally the selection of an animal sentinel population would target those animals with a higher probability of disease (6) but there are other epidemiological and pragmatic considerations. A framework for evaluating animals as sentinels has been developed by Halliday et al. (Figure 1.1) (7). The framework highlights some key features to consider when selecting a sentinel species for infectious disease risks; I have applied it to non-infectious environmental hazards as well. The framework assesses the ability of animals to serve as effective sentinels by using the characteristics of the sentinel species and their relationship with the target population, the nature of the environmental hazard and the availability of supporting diagnostic and epidemiological infrastructures (7).

1.1.2 Companion animals as sentinels of environmental risks to people

The nomination of pets to act as sentinels for human disease has been supported by numerous case reports and surveys. The Canary Database is an online database of animal-sentinel studies published in the biomedical literature (http://www.canarydatabase.org). A review of the Canary database identified 157 articles that linked human and companion animal sentinel health data (8) (personal communication, M. Scotch, Yale Centre for Medical Informatics, 2011). This discussion will begin by assessing companion animals as sentinels for human disease within the conceptual framework described above and will continue with the application of pets as sentinels for endemic zoonotic diseases, emerging zoonotic diseases and non-infectious diseases.
The framework developed by Halliday et al. (7) emphasizes that the relationship between the sentinel and the target populations should be scrutinized to ensure there is some degree of a spatial, epidemiological or behavioural association: they may live in close proximity, be part of the same food chain, share water supplies or have some other connections that will allow changes in the sentinels to signal changes in the exposure risks for the target populations. Household pets meet many of these requirements. Fifty-six percent of Canadian homes own at least one dog or cat (9). People enjoy an intimate relationship with their dogs and cats. Much has been written about the benefits of pet ownership and the human-animal bond to the quality of human life and health (10). Within that closely shared environment there is also the opportunity for shared risks.

The route of transmission for the pathogen should be similar for the sentinel and the target populations so that host responses to pathogens are easily interpreted. For non-infectious hazards, dogs and cats share the same trophic level as people so they may have similar mechanisms for absorbing and metabolizing disease causing agents, and provide an identifiable and measureable response (11-13). For example, a pet dog clinically ill with lead poisoning led to the testing of 2 asymptomatic children in the same household. Both were found to have lead intoxication and the source was determined to be contamination of the yard with lead based paint chips following an exterior home renovation (14). The route of exposure was the same for both species and the test results in the dog were easily interpreted and acted upon.

Companion animals are smaller and have faster metabolic rates than people. This may result in increased sensitivity to exposure and a shorter latency to clinical disease (15, 16). This has been demonstrated through experimental inhalation studies of *Yersinia pestis* in cats (17) and observational studies of asbestos-associated mesothelioma in dogs (18).

There are also behaviours and characteristics of dogs and cats that are distinct from their human owners which can increase their exposures to environmental hazards and enhance their ability to provide a signal of a hazard circulating in the human environment. Carnivore pets that spend time outdoors have a greater risk of being exposed to environmental hazards compared to
people (19). Pets may be exposed to infectious or non-infectious hazards by drinking environmental sources of water, by increased contact with contaminated soil, contact with livestock and livestock manure, and through scavenging behaviours including coprophagia. They also have a greater risk of contact with wildlife as predators, prey or contact through a shared ecosystem. Owners may also increase their pets’ exposure to foodborne pathogens by feeding raw diets. Evidence of exposure in the pet may then signal a potential increased risk of exposure for people (20).

Sentinel animals should display a measurable response to the presence of the hazard (sensitivity) but that response should not be present when the hazard is absent (specificity) (7). For the information provided by the sentinels’ signals to be meaningful to the target population it must be possible to interpret the signal based on an understanding of the disease’s epidemiology. It may be possible to more accurately describe the environmental exposure history for pets because there are fewer confounding variables such as working in industry, travel to other geographic locations or smoking that often complicate the analysis of human exposure data (20).

The sentinel species should be observable, amenable to being sampled and accessible for data or sample collection. A recent survey found that dogs and cats are in 70% of Canadian pet-owning homes (9) so they are widespread and accessible. In this same survey 78% of the dogs and 50% of cats had visited a veterinarian in the previous 12 months providing an opportunity for health and disease data collection.

Species other than dogs and cats are found in 30% of the Canadian pet-owning homes (9). These include ferrets, rabbits, rodents, birds, reptiles, amphibians, fish and others and will be referred to herein as specialty and exotic pets. The risks and pathways of zoonotic disease transmission to the owners of these pets will vary depending on the species of the pet, whether they were captive bred or wild caught, where they originated, husbandry conditions and proximity to wildlife (21). Using American data, it is predicted that a significant proportion of specialty and exotic pets (2% of the birds, 30% of the reptiles and amphibians, and 15% of the freshwater fish) are illegally acquired native wildlife and legally and illegally imported exotic
wildlife (21). The legal, intentional import of exotic animals to the USA from 2000 to 2004 was estimated to be 38 million amphibians, reptiles, birds and mammals. Of these, 90% were intended for the pet market (22). Illegal imports are more difficult to quantify but the US Fish and Wildlife Service seizes $10 million in illegal wildlife imports annually, a number they expect “only scratches the surface” (23).

There is less opportunity to collect veterinary surveillance data from specialty and exotic pets than from dogs and cats. An American pet owner survey determined that the percentages of specialty and exotic pets that were seen by a veterinarian in the previous year were: ferrets (38.5%), guinea pigs (18.4%), rabbits (16.4%), other rodents (15.8%), birds (13.9%), fish (1.9%), and reptiles (0.0%) (24). With limited observation there is less opportunity to detect a response to an environmental hazard in these pets limiting their usefulness as sentinel species.

1.1.2.1 Companion animals as sentinels of endemic zoonotic disease

It is important to consider how companion animals respond to a pathogen of interest and the ability to detect that response (7). Dogs and cats have similar susceptibilities as people to many endemic infectious diseases. A database of infectious disease pathogens compiled from references of domestic mammal diseases found that 261 out of 374 pathogens that infect humans also infect dogs and/or cats (25). Many of these diseases are usually self-limiting or easily treated such as dermatophytoses and campylobacteriosis (21). In children and immunocompromised individuals, more serious consequences are possible (26). Other zoonotic diseases such as salmonellosis are potentially severe even in immunocompetent people.

While most cases of human salmonellosis are believed to be foodborne, an unknown proportion may result from contact with pets; so pets may be useful sentinels for human salmonellosis (27). Supporting evidence has been provided by studies that found Salmonella spp. in 21% of the samples of commercially available raw dog food (28) and 4% in commercial pig ear dog treats (29). Another study collected fecal samples from asymptomatic dogs and found that dogs fed raw diets were significantly more likely to shed Salmonella than the dogs that did
not eat raw meat (Odds ratio, 22.7; 95% confidence interval 3.1-58.8) (30). Cross-sectional studies that sampled the feces of asymptomatic dogs (27) and group-housed kittens (31) demonstrated a *Salmonella* spp. carriage rate of 23.2% and 51.4% respectively. Investigations of human salmonellosis outbreaks have been attributed to pets or pet products including outbreaks associated with feeding frozen rodents to snakes (32), the importation of tropical fish (33), and the care of diarrheic pets at 3 veterinary facilities and a humane society (34). In each of these outbreaks, investigations into the animal sources were initiated after the human cases had been identified. Dogs and cats are being exposed and are susceptible to *Salmonella* but it may not be possible to detect a change in the prevalence or incidence of *Salmonella* spp. by monitoring the pet population. A high proportion of asymptomatic carriers and limited laboratory confirmation of infection in symptomatic cases will prevent detection. There is no reported evidence of the utility of companion animal surveillance to provide warning of increased human risk for *Salmonella* spp. or other endemic zoonotic diseases.

1.1.2.2 Companion animals as sentinels of emerging zoonotic diseases

Pets may provide a warning of an emerging infectious environmental hazard (20). While the evolutionary process that leads to emergence is a complex interaction between the pathogen, host and environment and is unique for each organism, human contact with animals or animal products has been common to many recently emerged human diseases (35, 36). A literature survey determined that 73% of the emerging and re-emerging human diseases were zoonotic (37). A joint consultation of the World Health Organization, the Food and Agriculture Organization and the World Organization for Animal Health defined an emerging zoonosis as “a pathogen that is newly recognized or newly evolved, or that has occurred previously but shows an increase in incidence or expansion in geographical, host or vector range” (38). Examples of recently emerged zoonotic diseases include severe acute respiratory syndrome, West Nile virus, highly pathogenic avian influenza, variant Creutzfeldt-Jakob disease, feline and canine H5N1 and multi-drug resistant *Salmonella* (39-42). Ecological and social factors that contribute to increased human and animal contact and enhanced disease spread include: growth in both the
human and domestic animal populations, living in closely shared environments with animals, the encroachment of people and domestic animals into wildlife habitats and globalization (43-46).

Few pathogens of domestic carnivores are considered to be emerging with the exception of antimicrobial resistant (AMR) organisms which are causing significant concern among public health agencies (25, 47-49). AMR is a complex problem with drug-resistant bacteria acquiring resistance following exposure to antimicrobials or from the transfer of resistance genes. Dogs and cats represent potential sources of spread of AMR due to the use of antimicrobials in these species and their close contact with the environment and people. For example methicillin-resistant *Staphylococcus aureus* (MRSA) is a clinically significant opportunistic bacterium in human medicine. Investigations have demonstrated concurrent MRSA colonization in people and pets showing that interspecies transmission may be possible (50, 51). Furthermore, the human patient could not be verified as the source of the infection in all of the pets suggesting there may be an environmental source of AMR bacteria in some pets. (51).

To better understand the epidemiology of AMR in companion animals and their owners, data on AMR and antimicrobial use (AMU) in these species are required (52). Monnet et al. (53) demonstrated that studies of variations in human AMU and concomitant variations in AMR over time and place provided the best evidence of an association between the suspected cause (AMU) and the observed effect (AMR). AMR data from companion animals have mostly been obtained by retrospective surveys and studies (47, 52). The collection of AMU data in animals has not been legislated in Canada and collecting AMU data at the pet level has been a challenge (54). AMU data in pets have been collected from veterinary teaching hospital records. These retrospective cross-sectional studies described AMU and compared AMU patterns with prudent use guidelines (55-59), or with the AMR profiles of selected bacteria (60). The ongoing collection and evaluation of the occurrence and risk factors for AMR in companion animals is necessary but the feasibility of such projects has not been determined (52).

Multi-host pathogens that also infect wildlife are twice as likely as other pathogens to become emerging diseases of public health interest (25). There are numerous examples where wildlife,
including imported wild caught animals, have been associated with disease emergence where the infected wildlife either became pets or had associated with pets. Examples are monkey pox (61), Lyssavirus (62), tularemia (63), and leptospirosis (64). These infections were documented in developed nations. Disease emergence by this mechanism is anticipated to become more commonplace because of the increase in pet ownership in the developing nations where these new owners are also interested in specialty and exotic animals as pets (65). The increased risks associated with ownership of wild species would justify their increased surveillance. An additional challenge is that the specialty and exotic species may be illegally obtained and they are less likely to receive veterinary care than dogs and cats (21). It may be possible to improve their visibility by targeting veterinary practices that specialize in these species but there has been no documented evidence of the usefulness of this approach (66).

1.1.2.3 Companion animals as sentinels of non-infectious diseases

Most of the toxicity testing in the past 70 years has relied on exposing laboratory animals to high levels of toxins and then observing adverse health effects (67). The results were extrapolated back to people at lower exposure rates. Progress has been slow (67) and laboratory studies are not capable of providing an early warning of a current environmental threat (15). There have been both historical (canary in the coalmine) and recent examples that support the use of animals as sentinels for environmental toxin exposures.

A recent example was the addition of melamine and cyanuric acid to a wheat gluten protein supplement intended for pet food that was imported into Canada and the United States from China. It resulted in the illness and death of many dogs and cats in the spring of 2007 (68). The number of affected dogs and cats is unknown because there is no central pet disease reporting agency in Canada or the United States but is estimated to be in the thousands (68, 69). Canada’s regulatory agency responsible for food safety recognized the importance of this toxic event and had begun screening for melamine in imported foods because “it could have just as easily have been baby food” (personal communication, C. Argue, Canadian Food Inspection Agency, February 2008). This warning was not heeded by the Chinese government and in the fall of 2008
an estimated 294,000 babies were affected by melamine contaminated baby formula in China (70).

The melamine and cyanuric acid contamination of pet food was detected by veterinary practitioners concerned about the number and unusual presentation of patients with acute renal failure. Following their reports to veterinary associations and to the pet food manufacturer, the suspected pet food was recalled and studies were instituted to identify the toxin (71). Acute toxicities that affect large numbers of animals are rare events. Questions remain about the usefulness of companion animal surveillance for the detection of acute non-infectious disease hazards beyond investigating concerns of astute practitioners reporting unusual events.

Stahl (72) argues that using domestic animals for sentinel surveillance of non-infectious diseases is an imprudent use of research dollars. Acute toxicities such as melamine and cyanuric acid poisoning, have well described endpoints and the biological process is well understood in people and animals (73, 74). This is not the case for chronic and more complex endpoints such as cancer. Stahl is skeptical that surveillance findings in animals can be interpreted for much of the human risk associated with environmental toxicities (72). Rabinowitz et al. (20) recognized that it is often not possible to understand what the relevance of a toxicological event in animals, is to people. They observed that there has been no ongoing effort to bring human health and veterinary health data or the two professions together so as to link surveillance data from both sectors. The authors see promise in the use of informatics to gather both human and animal data, monitor and analyze the outcomes in order to identify environmental factors implicated in their diseases (20). The appeal of an integrated surveillance system is that perhaps, a warning of poor health in the pet population will not be disregarded.

There is optimism that a new approach for toxicity testing presented by the National Research Council will provide more relevant data on which to make ecological health assessments for chronic outcomes. This approach combines laboratory-based studies using human cells and cell-lines with population-based exposure and surveillance data (67). In this context there are still challenges to using animal data as sentinels for human risk. Assessing toxin exposure often relies
on measuring appropriate biomarkers of exposure and effect. However there may be limited knowledge of the animal biomarkers that could be accurately interpreted as evidence of human risk. An understanding of the differences in susceptibility between people and animals would also be required (75). If this knowledge did exist, appropriate samples would need to be collected, tested and interpreted by an informed and astute practitioner and a willing animal owner.

1.1.3 Companion animals as sentinels for other animals

A second goal is to monitor health events or evaluate health interventions in companion animals to improve the health of other domestic and wild animals. For example, canine parvovirus which affects both domestic and wild animals has continued to evolve since it first emerged. Monitoring susceptible populations of dogs to identify changes in the virus may be a method for signaling increased risk to both domestic and wild canines (76). Responses to the detection of new viral strains could include new diagnostic methods and vaccination programs.

1.2 Surveillance system objectives

Surveillance system objectives should describe how the data will be used to meet public or animal health goals. Objectives must be clearly defined early, during the planning and development phases of a new system (2, 77). Companion animal surveillance data may be used in any or all of the following ways.

1.2.1 Provide early warning

The threat of bioterrorism, the emergence of infectious diseases and concerns about environmental contamination have stimulated interest in using animals for early warning of increased human risk (78). Using additional data from non-traditional sources may provide an earlier signal in an outbreak, leading to an more rapid response, limiting morbidity and mortality in both animal and human populations (78, 79). For example, the timely submission of dead birds is recognized as a spatially and temporally sensitive method for the early detection of
increased risk of human exposure to West Nile Virus (WNV) (80). Bird surveillance is a key component of the Public Health Agency of Canada’s (PHAC) WNV surveillance plan (81).

Once it has been determined that a contagious disease outbreak is occurring, rapid implementation of mitigation strategies is necessary to control the spread of the disease. Public health measures to control outbreaks are aimed at containing the first cases of an outbreak through isolation, quarantine, vaccination or in the case of WNV, mosquito control and avoidance (81, 82). In veterinary medicine, slaughter of positive cases and contacts is also frequently employed (83). The severe acute respiratory syndrome (SARS) epidemic provides an example of the importance of a timely response for limiting the number of infected individuals and the size of the outbreak. SARS emerged from a wildlife reservoir (likely bats) through civet cats as the intermediate host in Guangdong province of China in 2002 (84). In Canada there were 252 cases of SARS and 43 deaths (85) with an estimated cost of $1.5 billion CAD to the Canadian economy (86). A study reported that a delay of one week in detecting and controlling the SARS outbreak, would have nearly tripled the number of cases and extended the outbreak by 4 weeks (87).

In order for animal data to be capable of providing an early warning, the signal must be recognized and communicated to those who need to know. In New York City in the summer of 1999, a physician reported 2 unusual cases of neurological illness to public health which led to an epidemiological and laboratory investigation (88). The initial diagnosis of St. Louis encephalitis was incorrect but the decision to initiate mosquito control and inform the public to protect themselves from mosquito bites was appropriate (89). However, 59 patients were hospitalized and 7 patients died in this outbreak (90). The New York City Department of Health was unaware that a group of veterinarians and wildlife workers had identified a large avian die-off in New York City several weeks earlier. Following media reports of the human encephalitis cases the bird deaths were also determined to be caused by encephalitis and this led ultimately to the identification of WNV (88). Investigation of the unusual disease cluster in people was not integrated with information from the disease outbreak in birds. Had this happened, WNV may
have been identified earlier. Furthermore, diagnostic investigations to determine the etiology of the bird deaths was not instituted until after the human cases were reported. For a signal from animals to provide early warning of human exposure, the signal must be detected and there must be enough information with which to make a decision (79). For these reasons it is not known whether companion animal surveillance data will be useful for early warning of an emerging disease risk.

1.2.2 Monitor endemic disease

Animal surveillance has provided the data necessary to monitor the distribution, spread and changes in endemic disease (91). This surveillance can result in improved animal health outcomes and if the disease is zoonotic there may be measurable benefits for human health as well (92). A *Leptospira* spp. study using companion animal data from the National Companion Animal Surveillance Program is illustrative (93). Despite the availability of vaccines containing *Leptospira canicola* and *icterohemorrhagiae* serovars since the 1960’s, the number of cases of canine leptospirosis diagnosed at US and Canadian veterinary teaching hospitals had increased since 1983 (94, 95). Using the results of Leptospira microscopic agglutination tests from January 2002 through December 2004 (n = 23,005 serological tests) it was determined that the *L. canicola* and *icterohemorrhagiae* serovars were the lowest proportions of the seropositive tests (94). These findings suggested that the most commonly available leptospirosis vaccines were of questionable efficacy.

The leptospirosis study described above had several requirements: i) an etiologically precise diagnosis to the subspecies level; ii) data that could be collected from a large corporate network of veterinary hospitals and laboratories (93); and the information could be interpreted (2). This requirement may not be met for many animal species or pathogens.

1.2.3 Evaluate medical and surgical management of animal diseases

Animal surveillance data may be useful for evidence-based practice in veterinary medicine (EBVM). There are currently limited veterinary resources available for EBVM compared to
human medicine and there have been few published rigorous, randomized clinical trials with adequate power to demonstrate a difference (96). Medical record surveillance data may provide a large numbers of cases using alternative therapies and it could be possible to conduct observational studies to provide evidence of therapeutic efficacy (66, 97). In veterinary medicine most clinical studies use retrospectively collected clinical data. The School of Veterinary Medicine and Science at the University of Nottingham has created the Centre for Evidence-based Veterinary Medicine (http://www.nottingham.ac.uk/vet/research/population/areasofexpertise.aspx), to collect prospective animal surveillance data for evidence-based research. Outcomes from this project have yet to be reported.

1.2.4 Investigate the epidemiology of a disease

Animal surveillance data can been used to understand the natural history of disease and the management factors that are important in disease distribution and spread (98). This knowledge can guide control strategies to improve the health of the affected populations or generate hypotheses for future research (83).

From April 2000 to December 2003 a survey of foxhounds owned by hunt club members, other breeds of pet and shelter dogs and wild canines in Canada and the USA were tested for antibodies to *Leishmania* spp. (99). Persons associated with dogs in the study were also invited to submit serum samples. Antibodies to *Leishmania* spp. were confirmed in sera samples from foxhounds from 58 hunt clubs in 18 states and 2 Canadian provinces. There were 2 positive cases from wild canines and no positive cases in other breeds of dogs or people. These findings led to an investigation of the management of foxhounds to identify factors in that may favor *Leishmania* transmission. The authors of this study concluded that *Leishmania* was being transmitted dog to dog in the US and Canada. However, if North American sandfly species were to become competent hosts, new enzootic cycles of *Leishmania* could be initiated similar to those seen in other geographic locations. This would greatly increase the probability of human exposure and has been identified as an area for future research (99).
In this example, comprehensive companion animal data collected in a cross-sectional survey provided enough detailed information to support a qualitative epidemiological description of leishmaniasis. It is not known if it is possible to collect enough detailed information to support qualitative or quantitative epidemiological assessments of a disease by monitoring routinely collected companion animal health record data.

1.2.5 The purpose and objectives of a pilot companion animal surveillance system

To realize the potential of an animal population to serve as sentinels for infectious and non-infectious environmental hazards the population must be under observation (6, 7). There is currently no ongoing disease surveillance in companion animals in Canada (69). As a developmental project, the objectives and purpose of the surveillance system described in this thesis was to evaluate the utility of companion animal clinical data to provide an early warning, monitor endemic disease or enable epidemiological investigations of companion animal diseases for the benefit of human and animal health.

1.3 Governance of Animal Disease Surveillance in Canada

The Honorable Marc Lalonde in his influential report, New Perspectives on the Health of Canadians, emphasized the importance of understanding the legal framework under which the public health system must function (100). The obligations and limitations of veterinarians are regulated under Canadian federal and provincial laws. In Canada, the Constitution Act of 1867 established the framework within which the management of animal health, food safety and zoonotic diseases operates (101). The Act provides parliament with the power to regulate inter-provincial and international “trade and commerce”. These broad federal powers must also accommodate Section 92(13) which gives the provinces exclusive power over property and civil rights which has come to mean intra-provincial trade. Health is not addressed specifically in the Constitution Act so power with regards to public health is divided between both levels of government. This has resulted in fragmentation across and within levels of government (102). The jurisdictions of relevance to veterinary public health have been divided between the federal
and provincial statutes of agriculture, veterinary medicine, Canadian Food Inspection Agency (CFIA), health, quarantine and emergency measures, environment, natural resources, and the Canadian Charter of Rights and Freedoms (102). Alberta provincial public health authorities also have the authority to “make regulations respecting the prevention, investigation and suppression among animals of infectious disease communicable to humans” (103).

The control and prevention of animal diseases in Canada has been largely influenced by the Sanitary and Phytosanitary (SPS) Agreement of the World Trade Organization (WTO) (104, 105). The WTO designated the World Organization for Animal Health (OIE) as the international organization responsible for drafting standards and guidelines on risk analysis in animal health and zoonoses (104). Canada’s surveillance and publicly funded diagnostic laboratory capabilities have historically been directed towards diseases on the OIE list (106); predominantly diseases that affect livestock and livestock products that are important to trade. The federally reportable diseases relevant to companion animals are limited to rabies and brucellosis (107). There are additional zoonotic pathogens that must be reported by laboratories or are annually notifiable by the public health agencies and laboratories (69). In the province of Alberta, Salmonella spp., Lyme disease and its vector and avian chlamydiosis are also listed as notifiable (108).

In response to the risk to public health from emerging and re-emerging disease, the OIE has changed its disease reporting requirements: “The OIE will continue to support and urge Member Countries to make progress on timely and accurate reporting of zoonoses and sharing information on emerging and re-emerging zoonotic diseases, realizing that many of these new diseases are not associated with animal trade or traditional listed diseases” (109). These changes have placed an obligation on disease control organizations to develop surveillance systems for new pathogens, some which may never have been seen before, and in species where there is limited experience in surveillance such as companion animals.

As there is no legislated mandate for companion animal surveillance in Canada or Alberta, planning and implementing such a system requires an interdisciplinary, cross-sectoral approach. Potential stakeholders include public health, public animal health, academia, private practitioners
and industry (66). With these multiple perspectives there will be several distinct objectives for a pet surveillance system; from providing an early warning of environmental hazards to people to collecting the data to inform future evidence-based management of pet diseases.

1.4 Develop case definitions

The decision to use enteric disease as the model for this pilot surveillance project was based on several criteria. Gastrointestinal (GI) disease is a common problem in companion animals. A study found 13.9% of new insurance claims for dogs in the UK were due to GI problems (110). However there are no evidence-based practice guidelines for the management of diarrhea in companion animals reported in the refereed literature. Similarly, diarrhea from acute GI infections is one of the most commonly diagnosed illnesses in people in the developed world (111). Of the eleven federally reportable enteric, food and waterborne pathogens, six are capable of causing disease in pets (Table 1.1). *Toxoplasma* and some helminths are not reportable but they are also zoonotic enteric pathogens. Most of our understanding of pet transmitted zoonotic enteric infections is from surveys that measured the prevalence of enteric zoonotic agents in companion animal species (31, 112-115). There have also been reports of epidemiological investigations of the risk factors for human enteric disease that considered pet exposure (116-119). However, the role that companion animals play in the burden of human gastrointestinal illness in Canada is largely unknown.

Health surveillance requires clear case definitions that are often a combination of clinical and laboratory criteria (2). It is often necessary for case definitions to evolve over time. For example during an outbreak when there is a need for improved timeliness, the case definition may include less specific and more timely criteria with which to identify cases (2). To be effective, a case definition will need to be modified when consistent and detailed case definitions cannot be assembled from the data. The use of medical records for surveillance is a secondary use of data that has been collected for another purpose, and they often do not provide all of the required information demanded by the project (120, 121). Using medical records for surveillance may
necessitate a less specific case definition; one defined by clinical signs without a laboratory confirmed diagnosis.

1.5 Determine the type of system
1.5.1 Approaches to animal disease surveillance

Surveillance systems are often described by their methods of data collection (91). Passive surveillance is the examination and reporting of clinically affected animals (83). It often begins with an animal owner concerned with a change in their animal’s health or behaviour and the decision to seek veterinary advice. The veterinarian then decides if a submission to a diagnostic laboratory is justified and the owner must agree to pay the costs. The test results provide a basis for treatment decisions and are the foundation of many animal disease surveillance and control programs including most of the reportable and notifiable diseases (91). The financial cost of diagnostic testing, poor timeliness for return of test results, and the failure to find a significant diagnosis have been found to impede laboratory submissions and may result in underreporting of disease (122). To ensure compliance for the detection and management of zoonotic pathogens such as *Mycobacterium bovis*, testing may be mandatory by legislation, offered free of charge to animal owners and owners compensated for destroyed animals (123). Passive surveillance cannot provide estimates of disease frequency but it can identify a change in pattern of disease that may indicate further investigation or the implementation of active data collection (83, 91). Passive surveillance has also been credited with identifying new and emerging diseases; the first cases of bovine spongiform encephalopathy (BSE) were identified by passive surveillance (83, 91).

Active surveillance uses a formal sampling process to select individuals from a population that are then tested for a specific disease. Testing for known diseases makes active surveillance unsuitable for identifying new, previously unknown diseases. It can provide valid population-based disease prevalence estimates, providing a random or representative sample of the population has been collected. Active surveillance is often costly so it tends to be conducted in the short-term. To improve its efficiency, active surveillance may be targeted to populations
more likely to be positive for the health related event. An example includes testing all fallen bovine stock for BSE (119).

Regardless of how the data are collected, laboratory-based disease surveillance may be less effective when the objective of the surveillance system is to provide an early warning of an emerging infectious disease (124). The submission biases described above, poor timeliness from submission to diagnosis, and the challenge of recognizing a disease previously not described in the region serviced by the laboratory will all contribute to delays in recognizing an emerging infectious disease.

1.5.2 Syndromic surveillance

In the past decade, the demands for improved bioterrorism preparedness and emergency response required that many health departments move to strengthen their public health surveillance capacity (125). To achieve an early warning, surveillance research has focused on detecting changes in the patterns of health related events before a diagnosis has been established and to no longer rely entirely on laboratory-based data (126). Commonly used data include emergency department chief complaints (127, 128), laboratory test orders (129, 130), and the purchase of over the counter medications (131). The data are subjective and non-specific clinical symptoms that are often grouped together by a specific set of clinical features (syndromes) such as fever, gastrointestinal, respiratory, flu-like illness, neurological, hemorrhagic, dermatologic, coma, and sudden death (126). The appeal of early detection has led to the application of syndromic surveillance to endemic and emerging disease outbreak detection (80, 127, 132).

Prospective syndromic surveillance studies are necessary to demonstrate that, in the field, the system will lead to earlier detection than existing methods (133). Early event detection uses any of a number of statistical algorithms applied to the syndromic data to provide a warning of a change in the normal levels of the health data (134). Further research is required to understand when the alerting algorithms may be useful (135). Localized outbreaks involving large numbers of cases or easily diagnosed outbreaks are likely to be detected by clinicians before a signal is
detected in the statistical algorithm (135). The algorithm may not have the power to detect small, diffuse outbreaks, and in this situation, the astute clinician may prove more timely (135). A review of surveillance systems for emerging zoonoses showed that the systems were rarely evaluated for their ability to detect both known and unknown pathogens (136).

Human practice-based experience has demonstrated that syndromic surveillance had greater utility when it was complemented with other information sources and used for monitoring public health events other than for the early detection of small outbreaks (124, 137, 138). Syndromic surveillance has evolved so that its objectives may now include measures of disease prevalence or incidence and to monitor disease trends that can enhance understanding of health events (139). A survey of public health officials found that outbreak detection had motivated their investment in syndromic surveillance but they found greater utility using syndromic surveillance for “situational awareness” especially for influenza surveillance (138). Situational awareness is described as disease detection, characterization, tracking, response and outcome. This is in contrast to a syndromic surveillance; a disease detection and alarm system (1). A review of syndromic surveillance in veterinary medicine found that the focus for some systems was early disease detection but most were used to add value to animal disease surveillance such as providing sentinel data for zoonotic disease, to improving coverage, linking animal data with public health or provide data for evidence-based practice (140). The Center for Disease Control guidelines also assess surveillance system usefulness beyond its ability for early outbreak detection to include “the impact or value added by its application” (141).

Retrospective detection of geographic clusters or temporal variations in the disease incidence can be used to establish the baseline model against which the prospective detection model is tested (142, 143). Detecting aberrations in spatial and/or temporal patterns within the data have been useful to generate a hypothesis for further investigation (144).

Once a signal in the data has been detected, the signal must be characterized often by using outbreak investigation protocols to determine if the signal represents an important health event that requires a response (145, 146). The signal is described by the patients involved, the
geographic location, time of occurrence, history and clinical features. This detailed information helps to develop a case definition which is used to focus the investigation and establish a diagnosis (125). Once an outbreak has been confirmed and its characteristics defined, a hypothesis can be generated about cause or risk factors associated with the outbreak. The final step in an early warning syndromic surveillance system will be to determine if the event is of animal or public health importance and to communicate those findings to those who need to know so an intervention can be initiated (18, 19). The objective of retrospective space-time data analysis may be an epidemiological investigation of a health event. Unexpected findings from cluster detection can, for example, lead to an understanding of the natural history of the disease or facilitate planning for its control and management (2).

1.5.3 Sentinel surveillance

Clinical practice is an important source of data for human syndromic surveillance systems and the main sources of data for veterinary syndromic surveillance (138, 140). Sentinel practice surveillance involves selecting private practices to monitor disease in the patients they treat and is useful when a population-based case reporting system is not feasible (83, 147). Inclusion in a surveillance network is often based on the ability or willingness of the practice to provide data rather than targeting areas with higher probability of the disease. The advantages of sentinel practice surveillance are: i) the clinical setting provides access to detailed demographic, geographic and clinical information (147); ii) it can provide signals of a change in the prevalence or incidence of the disease under surveillance (6); iii) many of the recently emerged infectious diseases were first recognized by hospital or clinical networks (148); and iv) an appropriately selected sentinel network of practices has the potential to be more cost effective for detecting change in the health status of a population than extensive cross-sectional surveys (6).

Sentinel practice surveillance does not offer complete case finding nor can it provide a population-based sample. Other disadvantages include: i) it requires a sufficient number of sentinel practices to be able to detect a change in the population (132); ii) it cannot be used to measure disease magnitude in the population; iii) it may not be clear if the practices are
representative of all of the practices in the region and achieving geographic or demographic representativeness of the cases can be problematic (149); and iv) without a population-based sample it is not possible to use a population-based denominator; the epidemiological ideal (150). When the denominator is not known, the crude numbers of cases can accurately reflect the epidemiology of the disease but with the assumption that the relation of the cases observed by the sentinels and the cases in the underlying population remains constant over time and region (150).

There are numerous examples of effective sentinel health networks in public health practice (149, 151) and for food animal health surveillance (6). There are companion animal sentinel surveillance networks that have been recently launched (152, 153) or are under development (154). To date there have been few study results from the networks published. An exception is Radford et al. (155) who reported antimicrobial prescribing patterns in companion animals using their sentinel data. Each of these surveillance systems employs different informatics methodologies to collect, store and analyze companion animal clinical data.

1.5.4 Develop data collection instruments

1.5.4.1 Informatics

If the change in behaviour or the symptoms and signs of the population are to provide the earliest warning of an infectious disease outbreak, then the syndromic data must be collected, analyzed and assessed quickly (156). For disease surveillance where the goal is the early detection of a health anomaly, the focus has been on pre-diagnostic data that is easily and electronically available (157). The adoption of electronic medical records by veterinary practitioners provides an opportunity to employ informatics for data collection, management and analysis (156). Informatics can take advantage of the uptake of the electronic medical record (EMR) by health care providers (138, 140). Public health informatics is defined as ”the systematic application of information and computer science and technology to public health practice, research and learning” (158). It is an interdisciplinary field with its core in systems design engineering.
Using informatics, it is possible to extract data from the EMR from multiple veterinary practices and integration the data into a single external data warehouse. The clinical data warehouse can function as the hub of a network of private practices for disease surveillance (159), for performing any number of population-based studies (160, 161) or to evaluate medical and surgical interventions to inform evidenced-based practice (162).

There are numerous examples of the use of informatics in the human health literature (159, 160, 163, 164). In the veterinary literature the utility of data extraction and warehousing has been shown with the National Companion Animal Surveillance Program. It is a partnership between Purdue University, Banfield Pet Hospitals™ and Antech Diagnostics™ (93). This system has informed evidence-based practice to improve pet animal health such as post market pharmaceutical safety reports for cat vaccines (165). This program has also provided surveillance data on exposures that may impact human health including leptospirosis in pets in the United States (94) and assessed the health of dogs and cats following exposure to industrial chemicals to investigate the potential of human health effects (166). Data warehousing from a network of veterinary hospitals does not currently exist in Canada.

1.5.4.2 Knowledge management and text-mining to support analysis of surveillance data

Medical record data are typically high volume and heterogeneous. There are challenges with preparing the medical record data for analysis, identifying the cases of interest and analyzing the data (167). Information technology is necessary to achieve these tasks. Knowledge management software such as data mining platforms (e.g., Knime, http://www.knime.org) can support the analysis process including data storage, transformation, analysis and reporting.

In Canada, companion animal veterinary medical records are documented in free-text with little structure and no standardized clinical coding or fixed vocabulary (140). Text-mining is an information technology used to convert free-text into computer understandable language. It automatically extracts information from textual data which can then be used to identify patterns in the surveillance data (168). Text mining been applied successfully in human surveillance
systems (169-174) and there was one report in the veterinary literature (175). Once the desired data is stored in a database, analytical processes are applied that will meet the objectives of the surveillance system. There have been no published accounts of veterinary syndromic surveillance systems that utilize automatic extraction of the electronic medical record, data warehousing and text-mining to collect syndromic data for disease detection and investigation.

1.6 Thesis objectives and structure

Problem Statement: There is currently insufficient knowledge regarding the potential utility and possible strategies for the design, implementation and evaluation of a companion animal surveillance system to provide an early warning of environmental hazards to people or animals or support evidence-based companion animal practice.

The structure of the thesis is presented in Figure 1.2. The key objectives of my research were:

1. To develop, deploy and evaluate a data extraction program that would accurately and completely extract, transform and load retrospective and prospective clinical data for data warehousing from sentinel veterinary practices.
2. Determine if warehoused veterinary data can provide relevant, valid, complete and representative data useful for epidemiological investigations of clinical diseases in pets.
3. Develop and evaluate a text-mining application for automatically classifying and retrieving cases of enteric syndrome from warehoused free-text companion animal medical records.
4. Determine if it is possible to distinguish an abnormal space-time pattern of enteric syndrome cases from a normal pattern.
5. Determine if the information provided by the medical records was sufficient to characterize an abnormal spatial-temporal pattern.
6. Using the information within the medical records, describe the patterns of medical management of enteric syndrome by the sentinel practices.
7. Gain an overview of essential elements for interdisciplinary collaboration and information linkages across human, animal and environmental health systems.

In Chapter 2, I address the first and second objectives by describing the process of recruiting sentinel companion animal practices, creation of a mutually acceptable data sharing agreement and development of the University of Calgary Data Extraction Program. The description of the system and measures for data completeness and representativeness are presented.

In Chapter 3, I focus on my third objective. Using a commercially available text-mining program I developed a categorization dictionary to extract the enteric syndrome cases from the warehoused data. Estimates of the diagnostic sensitivity and specificity of the text-mining processes to retrieve cases of enteric syndrome were determined.

In Chapter 4, I examine objectives four and five. By using time series analysis I investigated the temporal patterns of companion animals with enteric syndrome from the sentinel practice data. I analyzed the data using space-time methodologies to see if it was possible to detect statistically significant clusters of enteric syndrome in both a retrospective and a prospective study. I examined the medical records of the cases included in the clusters. My goal was to describe the clusters by demography, etiology, and possible risk factors to determine the biological and epidemiological significance of the cluster.

In Chapter 5, I characterize the medical management of enteric syndrome cases by the sentinel veterinarians (objective six). One of my goals was to test the ability of text-mining to accurately classify the cases by their management to determine if it is possible to automate this process. The other purpose of characterizing the cases was to describe the patterns of antimicrobial and antihelmintic use.

With multiple objectives, planning a companion animal surveillance system will require a collaborative effort from diverse disciplines and jurisdictions. While a systems approach under the term One Health is often advocated (176) there has been little attention directed at how best
to build interdisciplinary (ID) collaborations across species and sectors. A structure for ID collaboration and a method to effectively evaluate the outcomes of such collaborations is necessary. In Chapter 6, I addressed objective seven and gathered information relevant to interdisciplinary collaboration with a survey of the peer-reviewed literature and by scanning the websites of relevant organizations. A workshop involving people with professed experience in ID collaborations at the human-animal-environmental interface followed. My goal was to examine examples outside of human-veterinary-environmental health to gain insight from the experiences of other professions. Common themes are organized into a framework of the essential elements of interdisciplinary collaboration.

Resources invested in surveillance cannot be used elsewhere. Health managers need to ensure that a surveillance system is meeting its objectives efficiently and effectively but assessment of the systems attributes and usefulness will also depend upon the objectives of the stakeholders, the system type and its stage of implementation (177). Evaluation frameworks have been developed for public health (141), for the early detection of outbreaks (141), communicable disease (178) and informatics-based (179) surveillance systems. Conclusions and recommendations must be acted upon to meet the needs of the system users. There has been little critical evaluation of emerging zoonoses surveillance systems (136). In Chapter 7, I evaluate my surveillance system and provide applications where it will be most useful. I also present ideas for its improvement and areas for future research.
Figure 1.1: The sentinel framework in context. Reprinted with permission from (7).
Table 1.1: The zoonotic capabilities of federally notifiable enteric pathogens (Public Health Agency of Canada)

<table>
<thead>
<tr>
<th>Pathogens causing human gastrointestinal disease that are under national surveillance¹</th>
<th>Potential of causing disease in companion animals (180)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cryptosporidia</td>
<td>Yes</td>
</tr>
<tr>
<td>Cyclosporidia</td>
<td>No</td>
</tr>
<tr>
<td>Giardia</td>
<td>Yes</td>
</tr>
<tr>
<td><em>Campylobacter</em> spp.</td>
<td>Yes</td>
</tr>
<tr>
<td>Cholera</td>
<td>No</td>
</tr>
<tr>
<td><em>Clostridium difficile</em></td>
<td>Yes</td>
</tr>
<tr>
<td><em>Listeria</em></td>
<td>Rare</td>
</tr>
<tr>
<td><em>Salmonella</em> spp</td>
<td>Yes</td>
</tr>
<tr>
<td><em>Shigella</em></td>
<td>No</td>
</tr>
<tr>
<td><em>Verotoxigenic E. coli</em></td>
<td>Yes</td>
</tr>
<tr>
<td>Norovirus</td>
<td>No</td>
</tr>
</tbody>
</table>

Figure 1.2: Objectives and thesis structure

Objective #1 and
Objective #2

Objective #3

Objective #4

Objective #5

Objective #6

Clinic records

Data Warehouse

Cases of enteric syndrome

Detect clusters in time and space

Characterize the clusters

Case management Antihelminthic use Antimicrobial use

University of Calgary Data Extraction Program

Text-mining

Space-time analysis
Chapter Two: The development and implementation of data extraction and data warehousing technologies for companion animal disease surveillance

2.1 Introduction

2.1.1 Companion animals as sentinels

The National Research Council Committee on Animals as Monitors of Environmental Hazards recommended that systems be established “for collecting data on parallel diseases of animals and humans living in the same environments” (4). There are pets in approximately 60% of Canadian homes (9), and in our largely urban population, pets are the only animals with which many people interact. The nomination of pets to act as sentinels for human disease has been supported by numerous case reports and surveys. The Canary Database is an online database of animal-sentinel studies published in the biomedical literature (http://www.canarydatabase.org). A review of the Canary database identified 157 articles that linked human and companion animal sentinel health data (8). These studies included zoonotic diseases (181), environmental contamination causing chronic diseases (18), and antimicrobial resistance (31).

Domesticated carnivores have several attributes that support their use as sentinels for detecting circulating infectious hazards to which people could be exposed. Dogs and cats account for 70% of the animals in pet-owning homes so they are widespread and accessible (9). A review of the human pathogen database showed that almost half (43%) of human pathogens have carnivore hosts (25). People and their carnivore pets enjoy a close and intimate relationship within the same environment. People and their dogs and cats have similar susceptibilities to various pathogens and share the potential for exposure to these pathogens in their environment.

Establishing a link between chronic, non-infectious disease in people and environmental factors is difficult, and the use of non-human species that share our environment has been advocated to provide an early warning of increased human health risk (20). Carnivore companion
animals may meet the criteria as appropriate sentinels for environmental contamination: i) the shared environment allows for similar exposures (13); ii) as mammals that share the same trophic level as people, they may have similar mechanisms for absorbing and metabolizing contaminants (11, 12); and iii) their smaller size and faster metabolism may result in increased sensitivity to contaminants and a shorter latency to clinical disease (15, 18).

While dogs and cats are more popular pets, a broad range of other species are found in 30% of Canadian pet owning homes (9). These include native and exotic species of fish, birds, reptiles, amphibians, and small mammals including rodents. The potential for these specialty and exotic pets to introduce infectious diseases was of particular concern for some authors (21, 22, 26, 182). For example, an increase in reported human cases of *Salmonella paratyphi B* infections in Quebec from 2000 to 2003 was found to be associated with contact with imported tropical fish or their aquariums (33).

The sentinel value of pets cannot be realized unless there is a sufficient level of ongoing observation (6). In Canada, the federally legislated disease reporting requirements for companion animal veterinarians and diagnostic laboratories is limited to two reportable diseases (rabies and brucellosis) and one immediately notifiable disease (psittacosis) (107). In Alberta, *Salmonella* spp. and Lyme disease and its vectors are also listed as reportable diseases (108). Despite the limited scope of official surveillance of pets, there are opportunities to access more frequent and detailed medical information from this target population. An online Ipsos-Reid™ survey conducted in 2008 and targeting the urban pet owner, determined that 78% of dogs and 50% of cats had been examined by a veterinarian in the previous 12 months (9). Under the regulations of the Alberta Veterinary Medical Association (183) the findings from these examinations must be documented in the examining veterinarian’s medical records making them a potential source of surveillance data.

This potential has been illustrated by the Veterinary Medical Database (VMDB) (98). In 1964 the National Cancer Institute started the VMDB which abstracts and codes the medical records from 26 veterinary colleges in North America. More than 100 publications on a broad range of
epidemiological studies have used this database (https://cvmsecure.missouri.edu/VMDB/Publications/BrowseAll.aspx). However, the system is passive, lacks timeliness because many colleges are currently several years behind in their submissions and the data may be subject to substantial referral bias (184). To better connect trends in pet population health to changing patterns of toxic and infectious environmental hazard risk to people, a more timely system that better reflects the pet population is required. This includes a means to directly access pet medical records in primary care veterinary facilities. Animal health records are not abstracted at general veterinary practices and the volume of data in pet records precludes the manual review of paper-based records. However, the adoption of electronic medical records (EMR) by veterinary practices has made access to a huge volume of veterinary clinical data more feasible.

2.1.2 Informatics

The collection, storage and analysis of EMR’s for public health purposes can be facilitated by informatics (185, 186). Public health informatics is defined as ”the systematic application of information and computer science and technology to public health practice, research and learning”(158). It is an interdisciplinary field with its core in systems design engineering.

There are many challenges to accessing and extracting the data from EMR’s. First, under Alberta provincial legislation, the Personal Information Protection Act, veterinarians have a duty to their clients to ensure the privacy of their patients’ medical records (187). Second, the purpose of the EMR is to document and support patient care, assure continuity of care and assist with the financial management of the practice. These databases are not intended to be used for outcome analysis or to support population level studies (160, 162, 164, 186). In order to use the EMR for surveillance, the data must be moved out of the veterinary practice management system. And third, there has been little uptake of standardized codes or fixed vocabulary in veterinary records and EMR’s are predominantly in free text with minimal structure.
With the use of informatics it is possible to extract data from the EMR at practices, transform it into a functional format and load it for export to an external data warehouse. Extract, transform and load (ETL) computer technologies enable the integration of data from multiple veterinary practices into a single external data warehouse where subsequent analyses looking for patterns and trends can occur (107, 160). The clinical data warehouse can function as the hub of a network of private practices for disease surveillance (159), for performing any number of population-based studies (160, 161), or to evaluate medical and surgical interventions to inform evidenced-based practice (162).

There are numerous examples of similar applications of ETL technology in the human health literature (159, 160, 163, 164). In the veterinary literature the utility of data extraction and warehousing has been demonstrated. The National Companion Animal Surveillance Program is a partnership between Purdue University, Banfield Pet Hospitals™ and Antech Diagnostics™ (93). There were more than 500 Banfield Pet Hospitals across the United States in 2004. The hospitals used proprietary software called PetWare™ and SNOVET™ disease coding. The medical records were uploaded nightly to a data warehouse. These records and any necessary laboratory test results from Antech Diagnostics™ were then transferred to Purdue University. This system has informed evidence-based practice to improve pet animal health such as post market pharmaceutical safety reports for cat vaccines (165). This program has also provided surveillance data on exposures that may impact human health. These include leptospirosis in pets in the United States (94) and assessed the health of dogs and cats following exposure to industrial chemicals in order to investigate the potential of human health effects (166). Data warehousing from a network of veterinary hospitals does not currently exist in Canada.

2.1.3 Evaluation frameworks for surveillance systems

Animal health data that were obtained to document the results of a patient visit in the medical record may not provide the information needed for disease surveillance. An evaluation of the data is necessary to understand the limits of the analysis of the data. As well, an assessment of the strengths and weaknesses of the data collection methods can inform their potential for use in
a surveillance system. Evaluation frameworks have been developed for public health (77), for the early detection of outbreaks (141) and communicable disease (178) surveillance systems. Measures such as timeliness, data quality and system usefulness are elements of these frameworks and are also relevant to surveillance systems that employ informatics for data collection. However, these evaluations focused on outbreak detection (141) or the prevention of adverse health events (77) and were not applicable. Therefore a more general evaluation framework was required. Buckeridge (179) described an evaluation framework for automated surveillance systems that utilized the DeLone and McLean model of information system success (D & M IS Success Model) (188), the framework for evaluating surveillance systems for early detection of outbreaks (141), and a public health logic framework structured on the D & M IS Success Model (189). Based on the conclusions of these frameworks, Buckeridge (179) defines four dimensions as potential foci to measure the information system success of an automated surveillance system: information quality, system quality, user experience and system benefits. These dimensions are interrelated and the “selection of the success dimensions and their measures should be contingent on the objectives and context of the empirical investigation” (190).

The purpose of this study was to determine if the application of medical informatics to the veterinary management programs at sentinel veterinary practices was able to retrieve and store complete companion animal health records in a form that allowed for future classification and analyses of disease trends. The objectives were to:

1. Develop a sentinel practice network and a mutually acceptable data sharing agreement with the prospective veterinary practices.
2. Develop and deploy a data extraction program that was acceptable to the practices and would accurately and completely extract, transform and load the retrospective and prospective clinical data for data warehousing.
3. Evaluate the users’ experiences.
4. Evaluate the system quality.
5. Evaluate the quality of the information. Demonstrate that the warehoused data provided relevant, valid, complete and representative data which could be used for epidemiological investigations of clinical diseases in pets.

2.2 Methods

2.2.1 Sentinel recruitment

Only practices with completely computerized medical record systems and using the same veterinary practice management software (VPMS) were eligible to participate in the study. To help protect the identity of the participating practices, the trade name of the VPMS is withheld. The study area was the city of Calgary, Alberta, Canada and the neighboring communities that are situated between 10 and 40 kilometers of the Calgary city limits. The communities lie northwest (Cochrane), north (Airdrie), east (Chestermere and Strathmore) and south (Okotoks) of the city. We used the 2009 Alberta Veterinary Medical Association Directory to identify the companion animal and mixed food animal/companion animal practices from the study area (Figure 2.1).

The owner/manager of each practice in the study area was initially contacted by telephone. The objectives and proposed methods of the study were described. The reasons expressed by the practice owner/manager for participating or not participating were documented in writing. If the practice was eligible and the owner/manager expressed an interest in the project, the project proposal was presented in a face to face meeting. The author responded to questions about the study from the owner/manager and written consent in the form of a data sharing agreement (DSA) was obtained.

The DSA was developed by the University of Calgary’s legal counsel. The DSA described the project and defined the data that would be extracted from the VPMS. It stated that the project would protect the identity of the clients, patients, practices and veterinarians and defined the ownership of the EMR and the extracted data. It ensured that the results of the study could be published but the data could not be used beyond the scope of the study described or be
transferred to a third party. The author of this thesis (Anholt) was responsible for maintaining the confidentiality of the data. The mutually acceptable data sharing agreement was signed by each practice’s managing partner(s) and the author (Anholt) prior to any access to the practice’s computer systems.

2.2.2 Development of the data extraction tool

The first five practices that agreed to participate in this study were recruited to develop and pilot the customized data extraction program called the University of Calgary Data Extraction Program (UCDEP). A computer software developer with extensive experience in veterinary database development (Business Infusions, http://www.businessinfusions.com/) was contracted to create the data extraction software. The desired data from the VPMS database included the date the animal was seen, the patient identification number used by the practice, the first three digits of the owner’s home postal code, the animals’ species, breed, sex, date of birth, and the medical records including the prescription drugs that had been administered and dispensed. The practices used similar numerical patient codes, so to further distinguish the records, a unique numerical code was assigned to each practice and to the records that originated from that practice. To preserve the confidentiality of the medical record, none of the fields that were intended for recording financial information or information that would identify the owner or pet were extracted.

The data extraction software developer built the UCDEP by initially searching the computer code on a pilot practice’s computer. He identified database tables that had promising names such as “PatientID” and then programmed the extraction software to extract those database tables. The UCDEP converted the extracted database tables into a common spreadsheet file format (comma separated values or .csv files) and the results were reviewed. If the extracted data contained confidential elements then the data extraction software was modified so as to not extract that particular database table. If desired data were missing then a piece of unique text would be selected from the missing data. The programmer then wrote a search program to scan the code of the practice computer in order to match that piece of unique text. The search program reported
where in the VPMS code this text could be found so that the corresponding data table could be included in the subsequent version of the data extraction program. Every data extraction resulted in the creation of several .csv files to capture all of the required data. Each .csv file represented a different database table from the VPMS. This was an iterative process. The computer code that queried the VPMS database for the desired information was designed so that it could be edited and modified quickly and easily during the development.

The data used in this study covered the period from January 1, 2007 to December 31, 2010. Prior to 2007, the management programs at the participating practices were typically an older technology and data extraction was more difficult. The UCDEP extracted retrospective data beginning January 1, 2007. If the practice opened or began using computerized records after this date, the retrospective began the first day that the practice installed and used the VPMS. The UCDEP was then configured to collect the prospective data so that every day it automatically extracted the previous day’s records. The scheduled time for the daily data extraction was chosen by the clinic manager. The UCDEP encrypted the daily files and they were emailed to the author (Anholt) daily until the end of the study. The retrospective and the prospective data were warehoused on a secure computer at the University of Calgary. As well, each practice’s files were stored on the hard drives of the respective practice computers so the practices could review their practice’s extracted data.

Development of the UCDEP at each of the 5 pilot practices continued until it was extracting only the required data fields and was accurately identifying and emailing the daily clinical records. It was then installed at the remaining practices that had signed the data sharing agreement. Retrospective data was collected at all of these practices and if the program was installed before December 31, 2010, then prospective data was collected as well.

The authors warehoused, managed and analyzed the data on a password protected computer, a Dell Alienware™ M17X with Intel Core i7 740 Quad Core Processor, 2 hard drives (a 500 GBSATA-HDD and a 256 GB solid state hard drive) and 6 GB of RAM. We used Microsoft
Office Excel 2007™ (Microsoft Corporation, Redmond, Washington) and Konstanz Information Miner 2.2.2 (Knime, http://www.knime.org) for data management and analyses.

2.2.3 Evaluation

The framework for evaluating automated surveillance systems was used for our evaluation (179). The evaluation focused on (a) the experiences of the users working with the system, (b) the quality of the system, and (c) the quality of the information used in the system.

2.2.3.1 User Experience

To assess the experience of the veterinary practices involved in this project we documented:

- The responses of the owners/managers when they were asked to participate.
- The number of practices that did or did not sign the DSA to become sentinel practices.
- The number of sentinel practices that remained involved throughout the data collection and data verification periods as well as the number of practices that dropped out of the project.
- Key informant interviews and open ended questions were used to determine if there were any concerns about the system.

To assess the experience of the system users we documented:

- The time invested in developing the program in the 5 pilot practices.
- The ease of use and time to perform tasks by the system users.
- When the author needed further investigations to verify or understand the data.
2.2.3.2 System Quality

In evaluating the quality of the data extraction program and data warehouse we considered if the system had the potential for a universal application and if the system remained stable throughout the study period.

Another important criterion for success was the ability of the UCDEP to extract complete and accurate records from the 5 practices in which the UCDEP was developed and piloted. This was measured by comparing the extracted files to the medical records on each of the practices’ computers using a one sample comparison of proportion (Stata/IC 10.0™, StatCorp College Station, TX). A sample size was calculated to detect a 0.01 difference in proportion between the extracted records and the practices’ records at the 5% significance level (2-sided) with 90% power. The proportion of successfully extracted medical records (expected proportion = 0.99) was compared to the value of 1, which represented the original medical records stored on sentinel practices’ computers. The calculated sample size was 163 records (Stata/IC 10.0™, StatCorp College Station, TX).

To acquire a random sample, dates from the study period were randomly generated with Konstanz Information Miner 2.2.2 (Knime, http://www.knime.org). Sufficient days of records were reviewed to meet the calculated sample size. For each randomly selected day, we confirmed that all of the medical records on the practices’ computers were also present in the extracted files. A complete and accurately extracted medical record was defined as having no missing data fields (for example we ensured that the prescription data was extracted) and that the text in the extracted records exactly matched that on the practices’ records. The complete and accurate extraction of an animal’s medical record was recorded as a success. The same sampling protocol and sample size was repeated when new versions of the program were installed in order to ensure that data extraction was still accurate and complete.

At the end of the development period the UCDEP was installed at the remaining practices that had signed the data sharing agreement. Before each practice was confirmed as a participating
practice, we verified that the data extraction by the UCDEP was >99% complete and accurate by comparing one or two days of records to the extracted files to ensure that there were no missing records and that all of the desired fields in the records were complete and accurate.

The final quality measure of the UCDEP and data warehouse was to verify that the data extraction process had not compromised the confidentiality of the medical records. We reviewed the sample of extracted medical records to determine if the files included any financial information or the names or contact information of the owners, veterinarians or veterinary staff.

2.2.3.3 Information Quality

Data completeness was measured by determining the proportion of complete and construct valid entries in each of the following data fields: owner’s home postal code, pet species, age and sex.

The first three digits of the Canadian postal code are called the forward sortation area (FSA). The FSA has the alphanumeric format of A0A and does not include the letters D, F, I, O, Q, or U (Canada Post, http://www.canadapost.ca/business/tools/pg/manual/PGaddress-e.asp#1382487). A postal code record in this format was classified as valid. The entire study area included 50 FSA’s (Figure 2.1). The FSA’s of the owner’s home postal code were used to determine which records were within the study area.

A construct valid ‘Species’ entry was recorded as avian, canine, chinchilla, feline, ferret, gerbil, guinea pig, hamster, mouse, rabbit, or reptile/amphibian. A construct valid ‘Age of Pet’ entry was between 0.001 and 25 years for dogs and .001 and 30 years for cats based on documented maximum life expectancies (77, 191, 192). A construct valid ‘Sex’ entry was recorded as female, female spayed, spayed, male, male neutered or neutered.

To assess the representativeness of the data collected from our set of volunteer sentinel practices, the geographic location of the pet, based on the FSA of the owner’s home address, the distribution of pet species, the age distribution of dogs and cats, and the proportion of cats and
dogs that had been surgically sterilized were compared to other sources of information about the underlying pet population. Geographic distribution data was provided by the City of Calgary dog license data (personal communication, D. Taylor, City of Calgary 2011) and the demographic data was provided by the Ipsos-Reid survey previously described (9). The information provided by the dog license and the survey was only relevant to pets living within the City of Calgary boundaries (35 FSA’s). There was no information available with which to compare to the pets living in the neighboring communities and rural areas and they were not included in this analysis.

The 2010 City of Calgary dog license data were available and the FSA’s of the licensed dog owners’ home addresses were retrieved. The 2009 sentinel data were weighed against the 2010 dog license data by dividing the number of dog records by the number of licensed dogs in each city FSA. To investigate the impact of the geographic location of the sentinel practices on the geographic distribution of the records, we analyzed the catchment areas of each of the sentinel practices located in Calgary. We determined the total number of Calgary FSA’s from which they drew their clientele and the number of FSA’s that represented approximately 75% of their clientele. These results were described using summary statistics.

The 2008 sentinel demographic data were compared to the Ipsos-Reid data which were collected in 2008 (2) using a 2 sample test of proportion (Stata/IC 10.0™, StatCorp College Station, TX).

2.3 Results

2.3.1 Sentinel recruitment and user experience

In August 2009, 29 of the 96 companion animal and mixed animal practices in the study area had completely computerized medical records. Twenty of these were using VPMS and were considered eligible as sentinel practices. The remaining 9 practices were using a variety of DOS-based, proprietary, and other structured query language (SQL) database programs and these were excluded from the project. Of the 20 small animal and mixed animal practices in the study area
using VPMS, only one practice was not comfortable providing access to their data and refused to participate. The remaining 19 practices were agreeable for a willing-to-participate rate of 95%.

There were 7 VPMS practices that had agreed to participate in the project but did not become sentinels. The reasons were: i) technical difficulties in getting the UCDEP to function in their practices within the financial and time constraints of this project (2 practices); or ii) because it was impossible to obtain the signatures of all of the practice co-owners on the data sharing agreement within the time constraints of this project (5 practices). Twelve practices participated in the study providing a participation rate of 60% (Table 2.1).

Eight practices provided data beginning January 1, 2007. Four practices had either computerized their record-keeping system or began operations after this date. The medical records were available for extraction from these 4 practices in each of: February 2007, August 2007, January 2008 and July 2009.

The sentinel practices in this study were exclusively veterinary-owned. They varied from single practitioner (n=2) to multi-practitioner (n=3) and multi-hospital businesses (n=7). Three practices limited their practice to dogs and cats with fewer than 10 visits by other species in a year. Eight practices treated other pet species routinely and one practice treated all species of companion animals, horses and food animals. Seven practices had regular office hours, defined as being open from approximately 8am to 5pm, weekdays with one or two late (about 8pm) evenings and Saturdays. Five practices offered 24 hour care, emergency call services or extended hours with late opening 7 days per week. The practices did provide care to stray pets but these animals could not be distinguished from pets with owners in the extracted database. Each managed their practice, patients and medical records in a manner that best suited their independent needs. There was no fixed vocabulary and no practice used drop-down pick lists or diagnostic codes. All had in-clinic facilities for basic parasitology, hematology, chemistry, and some limited serological testing.
The sentinel practice owners as well as many of the non-eligible practice owners contacted saw value in the project for both population studies and to enable research to support evidence-based companion animal practice. Some of the practitioners had an interest in research and were keen to be involved. Despite their interest and willingness, before the practice owners/managers would agree to participate they needed to be reassured that: i) the confidentiality of the patient records would be protected; ii) the use of their data would be restricted to this project; iii) the project would not impact the running of their practice; iv) there would be no financial costs or staff time expenditures and, perhaps the greatest concern; and v) the UCDEP would not crash the VPMS application or their computers’ operating systems.

Practitioner concerns about confidentiality were addressed by the project’s ethics approval from the University of Calgary Conjoint Research Ethics Board and by our reassurances that fields containing private information would not be extracted. Our responsibility for preserving the confidentiality of the records was also written into the DSA. The development of a mutually acceptable DSA was a lengthy process that consumed most of a year before all of the participating practices had signed. Clearly defined limits of data use and of data ownership were necessary for its approval (Appendix A).

The project did not burden the practices with additional work. We targeted the data that veterinarians and their staff generated in the routine care of their patients and we did not request any additional information. The UCDEP worked as a background application on their computer hard drives so we could reassure the practice managers that the data extraction program did not require any input from their staff nor would any costs be imposed upon the practice.

It was very important to the practices to protect the operation of their IT systems. We explained that the UCDEP accessed the data within the VPMS through a standard interface called the Open Database Connectivity (ODBC) which was designed to be independent of the database system. By using the ODBC it was impossible for the UCDEP to crash their VPMS software. If it did encounter an error, it was not a critical part of the computer’s operating system so the computer itself was not threatened.
These features satisfied the practice managers that the project would not compromise the running of their practice and we were able to proceed.

2.3.2 Development of the data extraction program, user experience and system quality

During the development of the UCDEP, we found that if practices were using a different version of the VPMS or used different templates within the same version, the desired data elements were traced to different locations within each practice’s computer code. Therefore, the development process needed to be repeated at each of the 5 pilot practices. An additional challenge was that new updates of the VPMS program installed at a practice would affect the function of the UCDEP. Close monitoring of the extracted data was required to detect and respond to the change with further modifications to the UCDEP. Eighteen versions of the UCDEP were developed and trialed over 6 months before the program was accurately extracting data from the 5 pilot practices. A total of 795 records at the 5 pilot practices were reviewed. The proportion of the records that were completely and accurately extracted from the practices’ records was calculated to be 0.994 (95% confidence interval, 0.988 - 0.999) using a one sample test of a proportion (Stata/IC 10.0™, StatCorp College Station, TX).

VPMS could be accessed by the staff and data entered during extraction. The UCDEP did utilize some of the computer’s processing capacity when it was extracting data. However, the staff did not report that their computers were performing more slowly at these times and there were no complaints about any reduction in their computers’ or VPMS’s ability to function. During the data collection period, there were no complaints from the practice managers about the UCDEP on their computers or any indication that the personal visits to the practice to manage or validate the data extraction were problematic for the practices. One veterinarian remarked as the program was uninstalled from their computer that “he hadn’t even notice it was there” (Table 2.1).

There was some desired data that could not be extracted. Reports from private laboratory services and specialist reports were filed as a hard copy at the practice or were scanned into
VPMS and usually filed under a tab that also contained digital images and financial information. The reports could not be extracted because they were in the same database table within the computer code and as a result it was linked to the financial information. Orders for laboratory diagnostics, recommendations for a specialist referral and the veterinarian’s interpretation of a report or a radiographic image may be found within the medical record and these were extracted. This information provided some insight into the unseen report but to view the complete laboratory or specialist report required a visit to the practice (Table 2.1).

The fields that contained owner’s names, contact information and financial information were not extracted. However, within the text of the medical record the names of owners, pets, veterinarians and staff could be found. This was an undesirable outcome when we evaluated the system’s ability to maintain the confidentiality of the record (Table 2.1).

Retrospective data extraction began January 1, 2007 at 8 practices. Three practices began operation and one established computerized records after this date. For these practices the extraction began on the first day that computerized records became available. The extracted data included the information recorded in the appointment schedule, the medical notes (history, clinical exam, interpretations of diagnostic tests, clinical assessment, differential diagnoses, and treatment), the in-clinic laboratory results and all prescription information. These data were contained in 4 or 5 .csv files. The number of files varied depending upon how the practice used their VPMS. The retrospective data extraction took between 30 minutes and 3 hours depending upon the size of the files and the speed of the practice computer.

At 4 practices the program was installed in January 2011 at the end of the study period, so the retrospective data were collected and the program was then uninstalled. Data were collected prospectively from 8 practices. The prospectively collected data was intended to only include data from the previous 24 hours so the daily files of emailed data were of a modest size (usually less than 100KB) and the process was completed within 5 minutes. At 3 practices the daily extraction was not limited to the previous 24 hours and instead included data from the date that the UCDEP had been installed. This problem was the result of variable date formats and the
UCDEP was not able to distinguish recent data. The daily prospective data extraction and transfer at these practices was longer, about 20 minutes.

In some practices configuration of the practices’ computers to establish the file transfer by email required several visits to the practice before it was working properly. In the practices that had well-developed information technology (IT) security, support from their IT specialist was required to set up the email service. In addition, the encryption of the files prevented internet security software from scanning the files for viruses. Therefore, to shield its network, the University of Calgary’s internet protection blocked all of the attachments. A Gmail™ (Google Canada, Montreal) account was then used to receive the files but attachments were still occasionally blocked for the same reason. All of the files were stored on the computer hard drive at the practices and so the blocked files could be retrieved during visits to the practice (Table 2.1).

2.3.3 Data management and user experience

Data cleaning began with the spreadsheet files from January 1, 2007 to December 31, 2009. In these 3 years of data there were 1,523,222 lines of data from the 12 sentinel practices. There were numerous duplicate files. In addition, within the text of the medical note, were templates that described normal findings in a clinical exam, information for owners and communications with owners for business purposes (e.g., vaccine reminders). These data were not considered useful for documenting a medical case. Using Microsoft Excel’s™ conditional formatting function, the irrelevant data was tagged using text matching, and then filtered and deleted from the spreadsheet. Matching and deleting all of the extraneous data rows in this large and complex data set was not feasible. Variable formats, especially of dates, were another issue that required formatting changes so that the file formats were consistent for all of the practices’ records.

The files were combined and the rows sorted and linked so that there was one row of data for each record (an animal with a unique identifier) every day. The unique identifier was the combination of the numerical code I assigned to the practice and the identification number of the
patient provided in the clinic records. Using the unique identifier, it was possible to search the data warehouse for more information in an individual animal’s record. The linkage also pulled the information recorded in the appointment schedule, the medical notes and all prescription information for each record into one variable, ‘Note’. The 3 years of data from the 12 practices resulted in 308,820 records. Three months was invested in establishing the protocols for and completing the data cleaning (Table 2.1).

2.3.4 Information Quality

The percentages of fields in the medical record that had been completed with a construct valid entry are shown in Table 2.3. Records were removed from the study if they were outside the study area or had an invalid or incomplete FSA (3.88%) or if the species field was invalid or incomplete (0.25%). Records with invalid or incomplete age of the pet (0.15%) or invalid or incomplete sex (0.9%) were not removed from the study but only from the analyses specific to those parameters. The final number of records stored in the data warehouse was 296,087.

An estimated 90% of dogs were licensed in the Calgary in 2010 (personal communication, D. Taylor, City of Calgary 2011). The number of dog records divided by the number of dogs in each FSA was greater in the south and northwest regions of the city than in the northeast and west regions of the city (Figure 2.2). The 9 sentinel practices that were located within the City of Calgary saw patients from 31 to 35 of Calgary’s 35 FSA’s. However most pet owners visited practices closer to their homes. Four of the 5 practices that limited services to regular office hours saw approximately 75% of their clients from an average of 3.75 (s = 0.96) FSA’s around the practice. The exception was one practice with regular office hours but was centrally located and drew its clientele from 8 closely positioned FSA’s. The 4 practices that offered extended hours had approximately 75% of their clients travel from an average of 7.75 (s = 1.5) FSA’s surrounding each practice.

The sentinel practices saw more dogs, fewer cats and a much smaller proportion of other pets (p < 0.000) than was described by Perrin(9) (Figure 2.3). The dogs seen by the sentinel practices
were more \( (p < 0.000) \) frequently puppies and more likely to be spayed or neutered than those seen in the general pet population (Tables 2.3 and 2.4). Participating practices saw more \( (p < 0.000) \) kittens, more geriatric cats and more spayed and neutered cats than are found in the general pet population (Tables 2.3 and 2.4). It was not possible to identify the records that represented stray dogs or cats (Table 2.1).

2.4 Discussion

2.4.1 User experiences – sentinel practices

The ability to recruit and maintain private practice veterinarians was fundamental to the success of this project. The sentinel network was built on previous relationships and trust between the author and the practices. Nevertheless, face to face meetings and clear explanations of the provisions to protect their time, IT systems, and the confidentiality of the medical records were necessary to move the project forward. Once the project was underway, the sentinels did not express any concerns about their participation. However, the practice managers/owners may have been more comfortable expressing negative opinions of the project if they had been given the opportunity to provide anonymous feedback. This limitation is an area for future research.

2.4.2 System quality and system user experiences

Various programming languages are used in relational database management systems (193). Even widely used programming languages such as structured query language (SQL) may not be consistent between different database products because of an incompletely specified standard (personal communication, K. McLean, Business Infusions). It should be possible to extract data from any veterinary management program. However, this would have required considerable data extraction software development and IT support from a skilled and committed software programmer throughout the project. To simplify the development process and to limit the costs of data extraction, the recruitment of practices was limited to those using VPMS. This was a potential source of selection bias. In addition, different versions of, or inconsistent use of the VPMS program compromised the capability of the UCDEP. With the data extraction method
used in this project, it was not possible to create a generic data extraction tool that was effective in all of the veterinary practices that used a VPMS. The variability in EMR systems limited the universality and stability of the UCDEP.

The UCDEP was customized to extract data in the 5 pilot practices and worked effectively in 12 practices. The UCDEP had many positive attributes. At the practices where it was effective, it was simple to install and configure. The retrospective and the prospective data collections were simple, fast, complete and accurate. The data extraction and data warehouse provided access to the information rich text contained within the medical record. The records for all prescription medications that were either administered or dispensed were obtained and could be connected to the clinical data. Patients were identified by unique numerical codes so the notes from any one day could be linked to previous or subsequent clinical and diagnostic exams so the records of an individual patient could be examined.

There were also some important limitations of the UCDEP including the inability to access reports from laboratory and pathology services. Relying on veterinarian notes within the medical record about an unseen report may result in the misclassification of a case. The importance of this misclassification is not known at this time. It was possible to retrieve the required documents at the practice from which it originated but this had a negative impact on the time and effort required by the system users. Moving the storage of scanned documents to another folder in VPMS where they would not be linked to private data and therefore accessible for extraction is an option that should be explored.

There were numerous frustrations associated with establishing the file transfer by email. Recommendations for future development should consider file transfer protocols (FTP) and other available technologies to transfer data from the sentinel practices to the data warehouse (194).

Data extraction and warehousing is only the first step in knowledge discovery in databases (164). The next step is to use data mining software and applications for discovering patterns in the data. Establishing protocols for data cleaning and standardizing the data fields so that the
large and complex databases from the 12 sentinel practices could be integrated into a single
database in the data warehouse and in a form suitable for further analysis was time consuming.
However, in my opinion, the potential advantages of the timely and effective access to 4 years of
data from 12 practices outweighed the disadvantages from the time and effort in data
management.

The use of routinely collected clinical data ensured that participation did not incur an
additional burden on the time or resources of veterinarians. However, as a consequence, the
confidentiality of records became a concern because patients, owners and veterinarians’ names
were found in the text fields of medical records. This is an important issue and an area of
ongoing research in the field of health informatics (195, 196). Automated de-identification or
‘scrubbing’ of text is challenging and was beyond the scope of this project. Instead this problem
was managed by limiting access to the data warehouse to one author who was responsible for
protecting its confidentiality. However, if the purpose of a data warehouse is to build a database
from numerous data sources that can then be utilized by epidemiologists, clinicians and practice
managers to help answer a diverse number of research questions, future veterinary data
warehouse developments must solve this confidentiality issue.

2.4.3 Information quality

It is important to understand the validity of the data for epidemiological inference (197).
Validity includes the accuracy of the data, whether the system is representative of the underlying
pet population and the selection of a denominator that is appropriate for the data when the
measures of disease are not population based (150).

2.4.3.1 Data accuracy

Data accuracy is the completeness of the data fields, the correctness of the data entries and the
absence of diagnostic errors (198). The UCDEP provided relatively complete data. Most (>97%)
of the fields contained complete data entries, had construct validity and the distributions of the
pets home addresses, species, ages and sex were plausible. The correctness of the data entries and diagnostic accuracy can only be established by comparing the data to ‘true’ values. This was not possible within the limits of this research project but is an area for future research.

2.4.3.2 Geographic representativeness

The ratio of dog records seen by the sentinel practices to the licensed dog population was not equally distributed across the city. The impact of this geographic bias was difficult to quantify but may be significant as the geographic differences may affect the species, age and sex profiles of the pets seen. For example, dog owners are more likely to live in owner-occupied single family homes whereas cat ownership is greater in rented apartments (24). Geographic location may also impact veterinary consultation rates. Not all companion animals have access to or receive the same level of veterinary care. A companion animal will be examined by a veterinarian if: i) it has an owner or is under the care of the humane society; ii) the symptoms displayed by the pet are severe enough that the owner believes a trip to the vet is justified; iii) the owner values the pet enough to rationalize the veterinary expenditure; and iv) the owner has the financial means to pay for the visit (199, 200). Consultation patterns are therefore determined in part by the socio-economic status of the owner. It is possible that pets in some areas of Calgary were less likely to have access to veterinary care and have fewer resources invested in the animal’s care. Therefore, the geographic locations of the pets seen by this sentinel system may not accurately reflect the underlying pet population and may limit the representativeness of the data.

An evaluation of the geographic locations of the participating practices and their catchment areas was used to explain the variable distribution of records. Veterinary practices do not have easily defined catchment populations. Numerous pet practices are located in every region of the city and only 38% of the pet owners consider the location of the practice as the most important attribute when choosing a new veterinarian (24). The FSA’s of the pet owners’ home addresses were used to determine that practice’s catchment area. Analysis showed that while some pet owners were willing to travel long distances to see a preferred vet, most visited a veterinarian
closer to home. Therefore, an equal geographic distribution of sentinel practices would be necessary to obtain a more equal geographic distribution of records.

2.4.3.3 Demographic representativeness

There was a significant sampling bias towards dogs, against cats and a large bias against other species with this sentinel practice network. As well, stray pets that were captured using this sampling strategy could not be identified as stray. The sentinel practices saw proportionately more of the very young, senior and spayed/neutered pets than exist in the underlying population which reflects the increased veterinary care of the sentinel pets.

Our interpretations of the impact of the sampling biases from this system are that: (i) disease outbreaks that have a broad geographic distribution and involve dogs and/or cats could be detected by this sentinel network, (ii) immature and senior pets are potentially more vulnerable to infectious diseases and environmental risks, so a bias towards this segment of the population may increase the opportunities to detect disease outbreaks, and (iii) outbreaks that are focused in poorly covered areas of the study site, or involve populations of companion animals that are less likely to receive veterinary care, may be overlooked by this sentinel system.

2.4.3.4 Selection of the denominator

Determining an appropriate denominator is necessary for a valid estimate of the frequency and distribution of the disease in the population (150). The ideal denominator would be the total population. A reasonable estimation of the total dog population in each FSA was available through the dog license data; however, we do not have all of the disease frequency information from each FSA for a corresponding numerator. Using the population in the catchment area of the sentinel practices was not possible because pet owners were free to choose from multiple practices in the same area and this would render any estimate unstable.

The number of participating practices was not constant throughout the data collection period. To control for the variation in coverage by the sentinel network it is possible to use the number
of sentinel practices per unit of time as the denominator (201). However, the practices in this sentinel network varied considerably in size, the number of veterinarians working per shift and hours of operation. Therefore the utilization rates are not consistent and simply dividing the number of target cases by 8, 10 or 12 sentinel practices would result in a biased estimate.

As an alternative, dividing the number of target cases by the total number of consultations (records) may represent a population based estimate. Day of the week and seasonal variation in consultation rates are common in veterinary practice (202). If the average number of consultations varies over time and region, there is sufficient information contained within this data to be able to filter the records by month, week or day and by FSA and in so doing, unmask the effects of variable consultation rates. Consultation rates can also be filtered by species, breed, age and sex to further describe the frequency and distribution of the target disease.

Using crude case numbers in a time series approach may reflect trends in disease incidence if the proportion of cases seen by sentinels to the cases in the general population remains constant. However, unpredictable shifts in care utilization are common in veterinary medicine (203). This vulnerability may be managed by using a network model which uses the ratio between a pair of time-series models, each monitoring a different clinical syndrome to remove the effect of the baseline shifts (204). Another alternative for using only case numbers is the space-time permutation scan statistic (205).

2.5 Conclusion

This study created and evaluated data extraction software and a data warehouse using private companion animal veterinary practices in Alberta, Canada. Many of our development challenges (development of mutually agreeable data sharing agreements, variable EMR systems, de-identification of text, access to diagnostic procedure results, “dirty data” and data quality issues) were similar to those experienced in the development of a human health sentinel system using the EMR as the data source (159). The use of electronic medical records by private veterinary practices provides an opportunity to collect timely and complete data from animal patients.
without increasing the workload of veterinary practitioners or their staff. Challenges include creating and maintaining an effective data extraction and file transfer program to a central database, accessing all relevant files within the medical record and removing all identifying information from the text. A future evaluation of the data will determine if it is possible to detect trends or patterns of infectious disease within the pet population that are relevant to human health or inform evidence-based practice to provide benefits to pet populations.
Figure 2.1: Map of study area based on the first 3 digits of the owner’s home postal codes, the forward sortation area (FSA).
Table 2.1: Evaluation of the sentinel surveillance system.

<table>
<thead>
<tr>
<th>Positive Attributes</th>
<th>Negative Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User Experience:</strong></td>
<td><strong>Sentinel Practices</strong></td>
</tr>
<tr>
<td>95% willing-to-participate</td>
<td>None reported</td>
</tr>
<tr>
<td>60% participation</td>
<td></td>
</tr>
<tr>
<td>0% drop outs</td>
<td></td>
</tr>
<tr>
<td>There were no complaints from participants.</td>
<td></td>
</tr>
<tr>
<td><strong>User Experience:</strong></td>
<td><strong>System Users</strong></td>
</tr>
<tr>
<td>Simple to format and install at each practice</td>
<td>A long development period</td>
</tr>
<tr>
<td>Data extraction was quick and easy</td>
<td>Outside lab work results, specialist reports, and e-files blocked by internet security needed to be retrieved at the practices</td>
</tr>
<tr>
<td><strong>System Quality</strong></td>
<td></td>
</tr>
<tr>
<td>UCDEP was fully automated and reliable</td>
<td>Data extraction program needed to be customized to the practice VPMS and was vulnerable to system upgrades.</td>
</tr>
<tr>
<td>Complete and accurate data extraction</td>
<td>File transfer by email was challenging</td>
</tr>
<tr>
<td>Data warehouse functioned well and was secure</td>
<td>Identifiers could be found in the text of some medical record.</td>
</tr>
<tr>
<td><strong>Information Quality</strong></td>
<td></td>
</tr>
<tr>
<td>&gt;97% complete and legitimate data entries.</td>
<td>Geographic regions of the city were not equally represented in the data</td>
</tr>
<tr>
<td>Records from each of the 50 study area FSA’s.</td>
<td>Distributions of demographic characteristics were not equal to the underlying population.</td>
</tr>
<tr>
<td>All species, age groups and sex of pets represented.</td>
<td>Stray pet population not represented.</td>
</tr>
</tbody>
</table>
Table 2.2: List of variables extracted by the University of Calgary Data Extraction Program (UCDEP)

<table>
<thead>
<tr>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinic identification number assigned by the researcher</td>
</tr>
<tr>
<td>Date animal seen</td>
</tr>
<tr>
<td>Animal identification number assigned by the practice</td>
</tr>
<tr>
<td>First 3 digits of the owner’s home postal code (forward sortation area or FSA)</td>
</tr>
<tr>
<td>Species</td>
</tr>
<tr>
<td>Breed</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>Date of birth</td>
</tr>
<tr>
<td>Appointment notes</td>
</tr>
<tr>
<td>Medical notes</td>
</tr>
<tr>
<td>Pharmacy prescriptions</td>
</tr>
<tr>
<td>In-house hematology and serum chemistry results</td>
</tr>
</tbody>
</table>
Table 2.3: Data completeness in the extracted medical records, n = 296,087 records

<table>
<thead>
<tr>
<th>Data Entry</th>
<th>Complete and legitimate data entry (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Owner’s Home Postal</td>
<td>97.72</td>
</tr>
<tr>
<td>Age of Pet</td>
<td>99.85</td>
</tr>
<tr>
<td>Species</td>
<td>99.75</td>
</tr>
<tr>
<td>Sex</td>
<td>99.10</td>
</tr>
</tbody>
</table>
Figure 2.2: Number of dog cases seen by the sentinel practices divided by the number of licensed dogs in each of the City of Calgary’s 35 Forward Sortation Units (FSA).
Figure 2.3: Distribution of pet species in Canadian homes compared to the distribution of pet species (ie. number of dog cases divided by all cases) seen by the sentinel practices.

Other species includes fish, birds, reptiles, amphibians and small mammals.
Table 2.4: Comparing the proportion of age groups of dogs and cats seen in Canada to those seen by the sentinel practices in 2008.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Ipsos-Reid Survey Dogs</th>
<th>Sentinel Dogs</th>
<th>Ipsos-Reid Survey Cats*</th>
<th>Sentinel Cats</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 1</td>
<td>.07</td>
<td>.20</td>
<td>.08</td>
<td>.16</td>
</tr>
<tr>
<td>1 - 3</td>
<td>.29</td>
<td>.24</td>
<td>.33</td>
<td>.19</td>
</tr>
<tr>
<td>4 - 7</td>
<td>.30</td>
<td>.23</td>
<td>.35</td>
<td>.18</td>
</tr>
<tr>
<td>8 - 10</td>
<td>.11</td>
<td>.11**</td>
<td>.11</td>
<td>.09**</td>
</tr>
<tr>
<td>10 +</td>
<td>.22</td>
<td>.23**</td>
<td>.24</td>
<td>.38</td>
</tr>
</tbody>
</table>

* As reported in Perrin (9), sum of proportions = 1.11.

** Comparisons by age group (within species) were not statistically different using a 2 sample comparison of proportion, p > 0.05 (Stata/IC 10.0™, StatCorp College Station, TX).

All other pair-wise comparisons were significantly different, p < 0.0000.
Table 2.5: Proportions of sterilized cats and dogs in Canada compared to those seen by the sentinel practices in 2008.

<table>
<thead>
<tr>
<th></th>
<th>Proportion of Sterilized Dogs</th>
<th>Proportion of Sterilized Cats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ipsos-Reid Survey</td>
<td>0.67</td>
<td>0.76</td>
</tr>
<tr>
<td>Sentinel Practices</td>
<td>0.85</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Proportions are significantly different, p < .0000
Chapter Three: Mining free-text medical records for companion animal enteric syndrome surveillance

3.1 Introduction

Sentinel surveillance uses a subset of the target population or a different species of animal to reflect the population of interest (83). Sentinels can serve to track endemic disease rates in the population or as an early warning to the presence of disease. They must be susceptible to the disease of interest and have a greater probability of being exposed to the pathogen, or a greater likelihood of an observed and measureable response to the pathogen than other populations (6, 19). In our largely urban population, companion animals may provide an early warning of zoonotic and environmental human health risks (20, 72, 206). Their smaller size and closer relationship to the environment can result in an early and increased exposure to some infectious or environmental agents (15, 207). As well, metabolic differences may result in shorter induction and latency periods allowing the effects of exposures to manifest in animals more rapidly than in people (208). There have been documented examples that pets can serve as sentinels for zoonotic diseases (e.g., visceral leishmaniasis (99)), and environmental diseases (e.g., mesothelioma (18)).

Gastrointestinal (GI) disease is a common problem in companion animals. A study found 13.9% of new insurance claims for dogs in the UK were due to GI problems (110). Similarly, diarrhea from acute GI infections is also one of the most commonly diagnosed illnesses in people in the developed world (111). The estimated burden of acute GI in people is 1.17 episodes per person year which corresponds to millions of cases across Canada every month (209). The costs of these human infections are measured in terms of loss of productivity, expenditures associated with hospitalization and deaths. For example, in Canada in 2004, there were 729 hospitalizations and 5 deaths due to Salmonella spp. (210).

Sargeant et al. (209) estimated that only 24% of human GI cases sought medical care and of these about 21% submitted stool samples for microbiological diagnosis. Tens of thousands of human GI infections caused by federally notifiable pathogens are reported each year in Canada (211). The sources for community acquired diarrhea in people include animals, food, water and
other people. Of the 11 reportable enteric, food and waterborne pathogens (http://www.phac-aspc.gc.ca/bid-bmi/dsd-dsm/duns-eng.php), 6 are capable of causing disease in pets and may have the potential to be transmitted from animal to people (Table 1.1). *Toxoplasma* and some helminths are not reportable but they are also zoonotic enteric pathogens.

Source attribution of reported infectious enteric diseases is a key objective of GI surveillance in Canada (212). Ravel (213) examined 30 years of human GI outbreak data in Canada and determined that for 16.8% a food source was not identified in the record. For another 72.8% of the cases the food source was suspected but not confirmed or the records did not indicate if the recorded food was the confirmed source of the infection. An etiological agent was identified in 50.3% of the cases but confirmed in only 9.8% of the cases. In an analysis of food-borne disease outbreaks in the United States, Jones, Imhoff (214) found similar limitations with GI disease outbreak data. Although many outbreaks were not traced to food and pets are acknowledged sources of enteric infection for people, the attributable fraction of exposure to pets in the human cases of GI illness was not investigated or discussed in many publications reviewing the epidemiology of acute human GI infections (215, 216). Furthermore, the majority of human cases are sporadic and not associated with an outbreak. These single cases or small clusters involving 2 or 3 cases may not undergo an epidemiologic investigation or be reported (214).

Most of our understanding of pet transmitted zoonotic enteric infections is from surveys that measured the prevalence of enteric zoonotic agents in companion animal species (31, 112-115). There have also been epidemiological investigations of the risk factors for human GI disease that considered pet exposure (116-119). However, the role that companion animals play in the burden of human GI illness in Canada is largely unknown.

To link patterns of animal disease with human illness requires new integrated capacities to gather and assimilate information from diverse sources and disciplines (217). However there are few disease reporting requirements for companion animals, or a method for the routine collection of animal data for the tens of millions of pets across Canada. There is no validated surveillance
system tracking disease in companion animals and that can link to human health outcomes in Canada.

The uptake of electronic medical records (EMR) by private practice veterinarians provides an opportunity for pet disease surveillance in the clinical setting but it is limited by the lack of standardized medical record software systems. Further, the medical records are documented in free-text with little structure and with no standardized clinical coding or fixed vocabulary. Informatics or the application of computer technologies to public health has been used to overcome these same barriers in human health (160, 164, 170, 218). It has become a priority of public health agencies around the world to use these tools to develop automated surveillance systems to enhance the early detection of public health threats (158, 171, 219). The use of informatics has also been advocated for surveillance in veterinary medicine (162, 185, 186).

Exploiting the information in free-text medical records for surveillance purposes requires text-mining. Text-mining is a component of natural language processing (NLP) and its goal is to convert free-text into computer understandable language. The process includes information extraction (IE) which is extracting predefined types of information from text. This is accomplished by applying statistical algorithms (trained or rule-based) or by recognizing the target using word and phrase matching and rules (untrained or linguistic approach) (218, 220, 221). Text mining software that employs a linguistic approach requires the user to have knowledge of the meaning, context and relevance of the words that are used to convey meaning in the text. Therefore, linguistics-based IE is well suited to users from the field to which it is applied (221). IE is followed by analysis of the structured data that has been retrieved to look for clusters and relationships between the concepts.

Text mining has been applied successfully in human surveillance systems (169-174). Many published applications of text-mining for veterinary surveillance have been for monitoring web-based electronic information (222, 223). Lam (175) successfully used text-mining to identify and extract syndromes from the free-text electronic medical records of racehorses to determine why they had been retired.
Text-mining can be valuable for case detection in surveillance if it can be performed with sufficient accuracy. Previous work has demonstrated that the performance of a text miner can be validated using established methods for measuring the accuracy of a diagnostic test (169, 170, 172). Reliable estimates of the diagnostic sensitivity and specificity of the test are necessary for calculating sample sizes and adjusting for misclassification when interpreting the surveillance data (224). Evaluating a test’s performance should consider criteria from the early stages of its development and assess its performance during implementation (225).

The use of text-mining to retrieve targeted electronic medical records is relatively new in veterinary epidemiology. The records collected may be used to detect patterns of disease in animals and identify changes that may signify an increased disease risk to animals and people. The objectives of this research were to:

1. To develop a categorization dictionary to text-mine and automatically classify and retrieve cases of enteric syndrome from warehoused free-text companion animal medical records.
2. To estimate the diagnostic sensitivity and specificity of text-mining processes for retrieving cases of our target syndrome (enteric syndrome) from companion animal EMR’s.

3.2 Methods

3.2.1 Data

There were 296,086 electronic medical records extracted from the veterinary management programs at twelve sentinel companion animal practices from January 1, 2007 to December 31, 2009. The study area included 6 communities in the province of Alberta, Canada: Calgary, Cochrane, Airdrie, Chestermere, Strathmore and Okotoks. All of the records were stored in a data warehouse at the University of Calgary. The features of the sentinel practices, data extraction and data warehouse were described in Chapter 2.
The appointment schedule, medical notes (history, clinical exam, interpretations of diagnostic tests, assessment, differential diagnoses, and treatment) and prescription data for each case were combined into one variable named ‘Note’, in the data file. All of the records were in free-text. There was no standardized diagnostic coding, fixed vocabulary or pick lists used by the sentinel practices.

3.2.2 Development of the categorization dictionary in WordStat

To reduce computer processing time, a random subset of 25,000 records was used to develop the categorization dictionary. The 25,000 records were stored on the base product QDAMiner 3.1™ (Provalis Research, Montreal, QC). The text analysis software, WordStat6™ (Provalis Research, Montreal, QC) was used for the automatic categorization of the text.

WordStat6™ is a linguistics-based program, so the text, in the form of individual words or compound phrases, was extracted and organized into categories and together these categories formed the categorization dictionary that was used for the identification and retrieval of cases. First, words that have little semantic value such as pronouns and conjunctions (n = 561 words) were automatically removed from the analysis by using an existing exclusion dictionary in WordStat™. Without further processing of the text, WordStat™ then performed a frequency analysis of all of the remaining words contained in the variable ‘Note’. WordStat™ identified 38,480 unique words and counted the number of occurrences of each word.

The list of unique words generated by WordStat™ was manually reviewed to identify words consistent with our target (enteric syndrome) and they were moved into the categorization dictionary. Misspellings and abbreviations resulted in many variant forms of a word. This was particularly true of the word ‘diarrhea’. The wildcard form of the word ‘diar*’ treated all instances of words beginning with these letters as diarrhea and captured many misspellings and abbreviations. Other unique spellings and abbreviations such as ‘dairhea,’ d’hea and D+ needed to remain in the dictionary as specific words. Words that were synonyms for diarrhea such as ‘cowpie’ and words that supported a diagnosis of enteric disease (gastroenteritis, colitis, Parvo+)
were also part of the inclusion dictionary. All of these words were placed into a category called ‘Diarrhea’. The word ‘vomiting’ and all of its variations were also included in this category. WordStat™ can also scan the text for the most frequently used phrases and those that were synonyms for diarrhea such as ‘watery stool’, and ‘bloody bowel movement’ were added to the inclusion dictionary in a category called ‘Other Diarrhea’. Both ‘Diarrhea’ and ‘Other Diarrhea’ were grouped into an ‘Enteric Signs’ category within the categorization dictionary.

The initial categorization dictionary was used to retrieve cases with enteric signs from the subset of 25,000 records. WordStat displayed the words and phrases that had been retrieved by the dictionary in context so they could be examined in the record from which they were retrieved. By scrutinizing the words and phrases in context it was possible to evaluate their accuracy for identifying and retrieving cases of enteric syndrome. From this process, it was determined that ‘vomiting’ and its variations were too non-specific a clinical sign for the accurate retrieval of enteric cases and they were removed from the inclusion dictionary. As well, retrieving words and phrases in isolation from other information resulted in classification errors. Modifiers such as negatives (e.g., no diarrhea), warnings (e.g., watch for diarrhea) and history (e.g., had diarrhea last month) were all falsely identified as cases of enteric disease.

Under a new category (No Diarrhea) within the categorization dictionary, rules were added that defined the conditions under which words and phrases should be categorized. The rules were complex expressions that include Boolean logic (AND, OR, NOT) or other modifiers (NO, IF, WATCH) and proximity operators (NEAR, BEFORE, AFTER) that specified where and how closely words should be associated within the sentence or paragraph. For example the rule NO BEFORE #DIARRHEA/C 5 specified that paragraphs with the word ‘no’ within 5 words of any of the expressions included in the Diarrhea category should be included in the ‘No Diarrhea’ category. Therefore the phrases ‘no diarrhea’ and ‘no coughing, vomiting or diarrhea’ were not falsely identified as enteric syndrome. The category ‘Enteric Signs’ was used in combination with the category ‘No Diarrhea’ to retrieve the targeted cases.
The dictionary was refined through an iterative process and deployed with the same subset of 25,000 records, the results were assessed by reviewing the keywords in context and then further modifications were made to the dictionary. To optimize the dictionary this process was repeated until additional adjustments to the dictionary did not improve the accuracy of the retrieved cases.

3.2.3 Validation of the categorization dictionary

The principles and methods for validating diagnostic assays for infectious diseases described in the OIE Terrestrial Manual 2010 (225) directed this evaluation. We studied the text-miner’s categorization dictionary for repeatability, diagnostic sensitivity and diagnostic specificity.

A sample of 2500 records was randomly selected from the entire file of 296,086 extracted medical records. The assumptions and the sample size calculation were as follows (226):

\[ n = \frac{Z^2_\alpha \cdot pq}{L^2} \]

- \( \alpha \) Significance level = 0.05
- \( Z_\alpha \) Under the normal distribution, the 1-\( \alpha/2 \) percentile = 1.96
- \( p \) \textit{A priori} estimate of the proportion, conservatively = 0.5
- \( q \) 1 – \( p \) = 0.5
- \( L \) precision of the estimate, allowable error = 0.1
- \( n \) number of target cases required in the sample = 96

From the process of developing the categorization dictionary, I estimated that there was a 4% prevalence of enteric positive cases.

Sample size = 96/0.04 = 2400 non-cases (controls) + 96 target cases = 2500 cases

To assess repeatability, the same categorization dictionary was applied to the same random subset of 2500 records on three separate occasions. The cases identified and extracted in each of the replicates were compared to determine the variation in results.
A case definition for enteric syndrome was developed in consultation with an experienced, board certified veterinarian (Table 3.1). Two final year veterinary students were recruited to manually review and determine the status of each of the records in the 2500 record sample. Each reviewed the records independently (blinded to the results of the other reviewer) and applied the case definition to identify each record as enteric syndrome positive or enteric syndrome negative. Their results were compared and the kappa coefficient was calculated to measure the agreement between the two reviewers. Where they differed they sought a consensus. A third reviewer, an experienced veterinary clinician, also manually reviewed the classification of the 2500 cases by the student reviewers and the text-miner to verify their findings and address any discrepancies. The agreed upon status of each case as target positive or target negative became the external reference standard. There was no method that would unequivocally classify cases as enteric syndrome positive or negative so the term external reference standard was used rather than “gold standard” (225).

The dichotomous test results were cross tabulated in a 2 X 2 table. The sensitivity (Se) and the specificity (Sp) of the text-miner’s identification and retrieval of target cases were then calculated using the cases classified by the external standard. The 95% confidence intervals (CI) for the Se and Sp were also calculated (Exact method, Stata/IC 10.0™, StatCorp College Station, TX) (227, 228). The cases that were improperly classified (false positives and false negatives) were reviewed to determine why they had been misclassified.

3.3 Results

3.3.1 Development of the categorization dictionary

The limited number of etiological diagnoses and the ambiguity of the language used by veterinarians in the documentation of medical records hindered the identification of etiologically-defined cases of infectious enteric disease. The frequency analysis in WordStat indicated that words that may have been documenting an infectious etiology (e.g., giardia, parvo, coccidia or enteric parasites) occurred in 3.6% of the cases. However, they were not specific to a case of enteric disease; rather they could be indicative of preventive measures such as vaccines,
differential diagnoses or a warning to owners. The category of words that described the clinical
signs of enteric disease such as diarrhea and runny stool occurred with a frequency of 6.4%.
Therefore, to maximize the accuracy of the text-miner’s ability to retrieve cases, it was necessary
to identify cases by the symptoms of enteric diseases rather than an etiological diagnosis. The
target for our study was therefore enteric syndrome.

We optimized the categorization dictionary as described above in about 30 iterations. Not all
of the modifications made to the categorization dictionary improved the text-miner’s accuracy in
identifying and retrieving cases of enteric syndrome. Reversing the unsuccessful changes and
exploring new alternatives resulted in the large number of iterations. The development of the
categorization dictionary was a lengthy process but once completed the text-miner could reliably
scan 25,000 records in less than two minutes.

3.3.2 Validation of the categorization dictionary

The proportion of enteric syndrome positive cases classified by the text-miner from the
sample of 2500 records was 0.048 (n = 121). When the categorization dictionary was applied at
three different times to the same sample of 2500 records, the retrieved cases were identical each
time. Therefore the repeatability of the test was 100%.

Compared to the external reference standard, the text-miner retrieved cases with enteric signs
with a sensitivity of 87.6% (95%CI, 80.4% - 92.9%) and a specificity of 99.3% (95%CI, 98.9% -
99.6%) (Table 3.2). The 2 reviewers’ agreement was 98.8% and the Kappa coefficient was 0.845
± 0.02.

There was one case where the two student reviewers could not find a consensus. In the case
history it was uncertain whether the patient had had diarrhea and the diagnosis was not explicit
but the clinician treated the case in a manner consistent with enteric syndrome. The experienced
clinician as the third reviewer determined this was an enteric syndrome positive case. There were
2 cases that the text-miner identified as enteric syndrome positive and both student reviewers had
not. When assessed by all three reviewers it was agreed that these 2 cases were enteric syndrome positive. In both of these records the symptoms of enteric syndrome were only briefly described and the records were 350 to 400 words long compared to most which were about 150 words or less. In addition, one of the overlooked records was recorded in all capital letters so it was more difficult to quickly scan the text to identify the key words. None of the remaining cases were reclassified when the third reviewer, the experienced clinician, rechecked the results.

3.3.3 Evaluation of the classification errors

The reasons that the text-miner misclassified cases were common to both the false positives and the false negatives (Table 3.3). Some of these errors were addressed by adding rules to the dictionary. For example, many prescriptions of non-steroidal anti-inflammatory drugs included the warning ‘If vomiting or diarrhea, please discontinue this medication and call the office.’ The rule ‘IF BEFORE DIARRHEA/C4’ correctly moved this phrase into the NO DIARRHEA category. However, when numerous rules were developed to try and capture all of the semantic variations of a phrase, the sensitivity or the specificity of the text-miner would improve but at the expense of the other. For example, to categorize the phrase ‘Radiographs, lateral and V/D (i.e., ventral-dorsal) – cardiomegaly evident’ as NO DIARRHEA, the rule V/D NEAR RAD* C4 would be effective. However, this same rule would improperly categorize ‘Puppy presented with V/D (i.e., vomiting/diarrhea). Radiographs – no significant findings’ as NO DIARRHEA. Therefore, the dictionary was constructed to optimize both sensitivity and specificity.

Templates for documenting a clinical case in the veterinary management software were used by 10 of the 12 practices and they provided some organization to the otherwise unstructured narrative. Templates (n = 4) that arranged the clinician’s records into history, vital signs, exam of each system (i.e., integument, cardiovascular, etc.), assessment, options to owner, and treatment were useful for text-mining because the results of the clinical exam were often well described, coherent and complete. Six of the practices used more explicit templates where the default was the normal findings of an exam of a body system. If there were abnormal findings then the default needed to be replaced with the clinician’s findings. These records resulted in redundant
information. The greater disadvantage of the explicit template was contradictory statements that resulted from the veterinarian not deleting the default description such as: “The loops of intestines had normal sizes, location, and consistency. There is no history of vomiting or diarrhea. Jack has had diarrhea for 2 days.” The text-miner misclassified this case as enteric syndrome negative.

3.4 Discussion

This study demonstrates that text-mining retrieved enteric syndrome cases from electronic veterinary medical records with high sensitivity and specificity when compared to human reviewers. The text-miner was able to scan large files to identify the target syndrome, even if referred to only briefly. Using text-mining to scan warehoused records may be advantageous for identifying an abnormal increase in the number of cases of pets presenting to their veterinarians with enteric syndrome.

A study that used natural language processing to identify adverse events (172), used the records from 30 human patients that had had a central venous catheter placed to optimize their pool of words and phrases. WordStat™ quickly analyzed the content of our 25,000 records which was helpful when the categorization dictionary was applied repeatedly. The categorization dictionary had high measures of sensitivity and specificity but we do not know if the same results could have been achieved with a smaller subset or improved with a larger one.

There were challenges as a result of the linguistic variations within the clinical narrative and therefore the text-miner misclassified some records. Considering the sources of the false positive and false negative cases, improving the sensitivity and specificity of the text-miner in this context would be difficult. Improvements in the clinical documentation of cases would improve the accuracy of case retrieval using text-mining. The text-miner was most efficient when the diagnosis was explicitly stated with the supporting signs and symptoms. Further study would be required to determine if other syndromes, such as seizure, with fewer semantic variations, abbreviations, synonyms and acronyms would present fewer challenges for text-mining. Our
results were consistent with other studies that used text-mining to automatically identify adverse events (172) and fever (170) from human medical records.

Measuring the precision of this test requires evaluation of the text-miner across several locations (225). This was not done in our study and therefore we cannot comment on how well the text-miner would perform in other settings. It is expected that the categorization dictionary created for our text-miner would require modifications to capture the language used by the participating veterinarians in different regions or countries. Temporal changes in language usage within the same geographic area would also impact the performance of the text-miner so adaptations to capture evolving practice standards and language may be necessary for future applications.

Some authors consider the use of clinical terminology standards as essential to a veterinary informatics infrastructure (162, 185). We do not know if there has been any uptake of such standards in private veterinary practices in Canada but a standardized diagnostic coding system such as Systemized Nomenclature for Medicine – Clinical Terms (SNOMED-CT) may remove some of the ambiguity found in the medical records. While promising, the reliability of SNOMED-CT coding is also imperfect. A study comparing the agreement among 3 physicians experienced in biomedical informatics and using 2 SNOMED-CT terminology browsers found the inter-coder and intra-coder agreement ranged from 44% to 55% (229).

Furthermore, studies in human medicine have contradicted the assumption that standardized clinical coding would improve case detection. Two studies compared case detection using both structured and non-structured data. Using a sample of cases, both the structured data (ICD-9 diagnostic codes) and text-processing of the free-text documentation of the medical records were used to detect cases of GI syndrome (230, 231). Both studies concluded that text-processing showed improved sensitivity although decreased specificity in case retrieval as compared with using ICD-9 codes. South (2007) showed that using both case detection models together performed better than either used in isolation. No studies comparing SNOMED record retrieval
to text-mining were found but another study did show that SNOMED indexed cases were more accurately identified than those coded using ICD-9 (232).

Case ascertainment is a source of bias and arises from how the population is sampled and the cases are assessed. Our study could not target all of the pets in the study area with clinical signs consistent with enteric syndrome because it is likely that not all pets experiencing the symptom of diarrhea will be presented to the veterinarian (199, 200). A proportion of owners will choose to manage the problem without a consultation. This proportion is unknown for animals but it is estimated that only 12-24% of people with enteric illness in Canada seek medical care (209, 233, 234). If the characteristics of the pets seen by the veterinarians were systematically different from those managed at home then could be a source of bias. Second, this text-mining system will only capture those cases that present to one of the sentinel practices. Compared to what is known about the underlying pet population, the pets seen by the sentinel practices in this study saw more dogs, younger pets and relatively more pets from some regions of Calgary (Chapter 2).

The data quality issues may not substantively limit the usefulness of the data for syndromic surveillance. The goal of syndromic surveillance is to recognize changes in patient consultation patterns by continuously monitoring data that describes clinical features, such as diarrhea, before there is confirmed diagnosis (235). A statistical deviation from a baseline rate in a data time series can act as a trigger for an epidemiological investigation to determine the reason for the deviation and its value as an early warning signal (141, 146). Such signals require additional investigation to determine the etiology or origin of the increased rates of enteric disease. Once the significance to animal health or public health has been established, an appropriate response can then be initiated.

### 3.5 Conclusion

We have shown the use of text-mining on data warehoused from the EMR of private veterinary practices, provided for automated, timely and complete enteric syndrome case detection for syndromic surveillance in companion animals. The ability of this system to
distinguish an abnormal pattern of diarrhea cases from a normal pattern in the population of pets presenting to sentinel veterinarians has yet to be determined. However, development of a validated text mining tool serves as a foundation for the epidemiological investigation of changing companion animal environmental health outcomes in a community setting.
Table 3.1: Case definition applied to the sample of cases to determine if enteric syndrome positive or enteric syndrome negative.

<table>
<thead>
<tr>
<th>Description of feces</th>
<th>Inclusion as a case</th>
<th>Exclusion as a case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjectives describing feces</td>
<td>Diarrhea - all linguistic and spelling variations</td>
<td>Negatives</td>
</tr>
<tr>
<td></td>
<td>Adjectives describing feces</td>
<td>• No diarrhea</td>
</tr>
<tr>
<td></td>
<td>• Soft</td>
<td>• Normal BM</td>
</tr>
<tr>
<td></td>
<td>• Watery</td>
<td>• Formed BM</td>
</tr>
<tr>
<td></td>
<td>• Cowpie</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Runny</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Mucous</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Bloody</td>
<td></td>
</tr>
<tr>
<td>Duration</td>
<td>Acute (&lt; 7 days)</td>
<td>Warnings/Instructions</td>
</tr>
<tr>
<td></td>
<td>Chronic (&gt; 7 days)</td>
<td>• Watch for…</td>
</tr>
<tr>
<td></td>
<td>Intermittent or occasional</td>
<td>• If…</td>
</tr>
<tr>
<td></td>
<td>Resolving/improving</td>
<td></td>
</tr>
<tr>
<td>Concurrent conditions</td>
<td>Dietary indiscretion</td>
<td>Prophylactic</td>
</tr>
<tr>
<td></td>
<td>Stress</td>
<td>• To prevent (re-) occurrence</td>
</tr>
<tr>
<td></td>
<td>Medication</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metabolic diseases</td>
<td></td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Enteritis</td>
<td>Constipation</td>
</tr>
<tr>
<td></td>
<td>Gastro-enteritis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Colitis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parvo/giardia/ova positive</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.2: The accuracy of the text-miner, WordStat™, to identify cases of enteric syndrome compared to a manual review of the records as the external reference standard.

<table>
<thead>
<tr>
<th></th>
<th>External Standard +</th>
<th>External Standard −</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>WordStat +</td>
<td>106</td>
<td>17</td>
<td>125</td>
</tr>
<tr>
<td>WordStat −</td>
<td>15</td>
<td>2362</td>
<td>2372</td>
</tr>
<tr>
<td>Total</td>
<td>121</td>
<td>2379</td>
<td>2500</td>
</tr>
</tbody>
</table>

Se = 87.6%  
(95% CI, 80.4% to 92.9%)  
Sp = 99.3%  
(95% CI, 98.9% to 99.6%)
Table 3.3: Sources of false positives and false negatives when classifying and retrieving cases of enteric syndrome with the text-miner.

<table>
<thead>
<tr>
<th>Cause of Error</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spelling</td>
<td>Not captured by wildcard (diar*) or identified in the frequency analysis</td>
</tr>
<tr>
<td>Vague wording</td>
<td>Feed bland diet until the stool is normal.</td>
</tr>
<tr>
<td>Poor diagnostic precision</td>
<td>The kitten has a poor, dull coat – worried about gastrointestinal parasites.</td>
</tr>
<tr>
<td>Contradictory statements</td>
<td>Has had bloody diarrhea for 3 days. No diarrhea since noon.</td>
</tr>
<tr>
<td>Homonyms</td>
<td>V/D = vomiting/diarrhea and ventral/dorsal (radiographs)</td>
</tr>
<tr>
<td>Warnings</td>
<td>Diarrhea is a possible side-effect of this medication.</td>
</tr>
<tr>
<td>Ongoing case</td>
<td>Explicit diagnosis made in previous day’s record but not this day.</td>
</tr>
<tr>
<td>History</td>
<td>Has diarrhea when eats table food.</td>
</tr>
<tr>
<td>Resolved</td>
<td>Was sick with diarrhea last month.</td>
</tr>
<tr>
<td>Preventive</td>
<td>Use ‘medication’ to prevent diarrhea.</td>
</tr>
<tr>
<td>Spacing</td>
<td>Vomit/Diarrhea; NoCoughing for 3 days.</td>
</tr>
</tbody>
</table>
Chapter Four: Spatial-temporal clustering of companion animal enteric syndrome: detection and investigation

4.1 Introduction

Syndromic surveillance was developed for the early detection of a bioterrorist event (128, 236). To identify clusters of illness as early as possible, syndromic surveillance uses data that continuously monitors the change of behaviour or the symptoms and signs of the population before there is a confirmed diagnosis (235). Commonly used data include emergency department chief complaints (127, 128), laboratory test orders (129, 130) and the purchase of over the counter medications (131). Without a specific diagnosis, syndromic surveillance targets illnesses that have been grouped together by a specific set of clinical features such as fever, gastrointestinal, respiratory, flu-like illness, neurological, hemorrhagic, dermatologic, coma, and sudden death (126). The appeal of early detection has led to the application of syndromic surveillance to endemic and emerging disease outbreak detection (80, 127, 132) and to veterinary disease surveillance (140).

While the objective of an automated, prospective syndromic surveillance systems has often been to provide an early warning of a public health threat, a survey of public health departments found that syndromic surveillance data had greater utility when it was used to complement other information sources for monitoring public health events (138). The survey respondents reported that the data were more useful when used to track and characterize infectious and non-infectious disease outbreaks and trends for improved “situational awareness” (138, 139). Examples included monitoring seasonal influenza and monitoring impacts after natural disasters (139).

Syndromic surveillance also can provide data for retrospective studies. The purpose may be to gain insight into the suitability of the data for early disease detection in a prospective study (140). The retrospective detection of geographic clusters or temporal variations in the disease incidence can be used to establish baseline parameters for a prospective detection model (142, 143). Retrospective surveillance data has also been useful for epidemiological studies such as the effect of heat waves (237), vaccine adverse effects (238), to augment laboratory data for
foodborne disease surveillance (239, 240), and emerging viral zoonoses risk assessments (241). The evolution of syndromic surveillance from early disease detection to its use in other research initiatives and to support planning and policy development have resulted in its continued growth in veterinary medicine (140).

Obtaining syndromic data and continuously monitoring a large volume of pre-diagnostic data for syndromic surveillance must consider the timeliness, cost and effort of acquiring the data. The uptake of computerized medical records by companion animal practitioners provides an opportunity to access data for pet disease surveillance (see Chapter 2). Automated data extraction and case detection from electronic medical records (EMR) necessitates the use of informatics which is the application of information and computer technologies to managing, processing and analysing data for disease surveillance (See Chapters 2 and 3) (242). Human syndromic surveillance systems have been developed that automatically extract data and use text-mining to retrieve the cases consistent with the syndromic category or categories of interest (169-172, 231). There have been no published accounts of veterinary syndromic surveillance systems that utilize electronic medical records and text-mining to collect syndromic data for disease detection and investigation.

Once collected, syndromic cases subsequently undergo any of a number of described analytical processes to detect and distinguish an abnormal disease pattern from a normal one (134, 235). Spatial-temporal methods have been used to identify day and place aberrations in disease frequency from an expected baseline (243). Kuldorff et al. (205) have developed a space-time permutation scan statistic that uses case counts from one data stream to identify disease clusters and does not require information about the underlying population at risk. Without denominator data, this method relies on historical information to serve as the control. Therefore it is well suited to data where the catchment area of the practice is undefined and the population at risk is unknown (a characteristic of veterinary practices). It does require a large volume of data (205, 244). The space-time permutation scan statistic has been applied in human (127, 245, 246) and veterinary (247) surveillance systems.
The goal of space-time analysis is to detect a disease cluster that indicates there is an unusual aggregation of disease occurrences. Suspected clusters should then be investigated to determine whether they are valid. In veterinary medicine this may be accomplished using protocols that are similar to the frameworks described for outbreak investigations (145, 146). Additional data is sought for the descriptive epidemiological characterization of the cluster. The characterization includes a description of the place and dates of the clusters, the animals involved (species, age, sex, breed), and their clinical histories and findings. If syndromic data are used for disease event detection, the system users will need to find additional information on the etiological diagnoses associated with the cluster (145, 146). The distribution of the cluster characteristics is compared to an expected distribution for the population at risk over the same time and space (248). If the cluster is determined to be unexpected or unusual, then a hypothesis can be formulated as to its cause. In a prospective surveillance project with the objective of an early warning, the final step will be to determine if it is of animal or public health importance and to communicate those findings to those who need to know so an intervention can be instituted (145, 146). The objective of retrospective space-time analysis may be an epidemiological investigation of a health event. Unexpected findings from cluster evaluations can, for example, lead to an understanding of the natural history of the disease or facilitate planning for its control and management (2).

In this chapter I present a retrospective and a prospective analysis of companion animal enteric syndrome data using retrospectively collected EMR’s extracted from sentinel veterinary practices. My research had the following objectives:

1. Determine if time series analysis of the EMR’s from companion animal practices is useful for optimizing user-controlled settings in the scan statistic software.
2. In a retrospective and prospective study, determine if it was possible to identify statistically significant spatial-temporal clusters of enteric syndrome in the study area.
3. In the retrospective space-time analyses determine if there was sufficient information to interpret the relationship between the cluster and the time and place of a statistical increase in disease frequency.

4. In a prospective space-time analysis determine if there was sufficient information in EMR’s to determine the biological and epidemiological significance of the clusters to human or animal health.

4.2 Methods

4.2.1 Study population and enteric syndrome data

The EMR’s were extracted from 12 sentinel veterinary practices in the city of Calgary, Alberta and the surrounding communities of Cochrane, Airdrie, Chestermere, Strathmore and Okotoks (n = 296,087 records (January 1, 2007 to December 32, 2009) and n = 151,571 records (January 1, 2010 to December 31, 2010)). The records from each practice were aggregated into a single file with a total of 447,658 companion animal records from January 1, 2007 to December 31, 2010 and stored in a data warehouse at the University of Calgary (see Chapter 2). I used text-mining technology and a syndromic case definition (Table 3.1) to identify and retrieve enteric syndrome positive cases from the warehoused medical records (see Chapter 3). The enteric syndrome cases from January 1, 2007 to December 31, 2009 (n = 13,063) were used in the time series examination, the retrospective space-time analysis and was the reference population against which the cases within each cluster were compared. A prospective space-time analysis used the retrospectively collected enteric syndrome cases from January 1, 2010 to December 31, 2010 (n = 5,769). Data were stored and managed in Microsoft Office Excel 2007 ™ (Microsoft, Redmond, WA) and Konstanz Information Miner 2.2.2 (Knime, http://www.knime.org). Figure 4.1 provides an overview of the study’s methods.

The date that the animal presented to the veterinarian served as the time unit for the temporal analysis. The cases were not limited to the first presentation so an animal that presented multiple days for the same episode of enteric syndrome were identified as a ‘case’ for each of those days. The first 3 digits of the owner’s home postal code (the forward sortation area or FSA) provided
the spatial data. The animal’s demographic data (species, age, sex and breed) and the medical notes recorded by the veterinarian or his/her staff were used to further characterize the cases.

**4.2.2 Time series**

The total number of enteric syndrome cases recorded for each day of the week was counted and the daily and the 7-day moving average of cases were plotted against time (Stata/IC 10.0™, StatCorp College Station, TX). The results from these analyses were used to determine the appropriate user-controlled settings in the scan statistic software.

The time series data was normalized by dividing the number of enteric syndrome cases for each month by all cases presenting to the sentinel practices for each month. Linear regression was used to model the relationship between the proportion of enteric cases seen by the sentinel practices and the time (Stata/IC 10.0™, StatCorp College Station, TX).

**4.2.3 Scan statistic for space-time clusters**

4.2.3.1 Retrospective space-time analysis

An open source scan statistic software, SaTScan™, version 9.1.1 (Kulldorff Information Management Services, http://www.satscan.org), and the retrospective space-time permutation model (205) were used to detect clusters. This analysis required two data files; the geographic coordinates (latitude and longitude) for the centrum of each FSA (249) and the number of enteric cases for each FSA on each day of the study. The time aggregation parameter was based on the time series analyses (see results below) and was set at 7 days. The maximum temporal window size was also informed by the time series analysis. The maximum temporal window default in SaTScan™ is 50% of the study period or 1.5 years in this study. This would have effectively masked the seasonal trend so the maximum temporal window parameter was set at 100 days, approximately the length of the observed seasonal peak of enteric cases (see results below). In SaTScan™ the default maximum spatial window is 50% of the population contained within a spatial scanning window. There was not a known biologically relevant means to determine the
optimal maximum spatial window size, therefore multiple scans using a maximum spatial window of 10, 20, 30, 40 and 50% of the defined population (all cases of enteric syndrome) were run. The outputs from these scans were assessed for consistency based on the total number of clusters detected, the time frames and the FSA’s contained in each cluster. The default maximum spatial window size of 50% of the population was assessed to be appropriate (see results below). Statistical significance was evaluated using Monte Carlo re-sampling (999 repetitions) (205).

4.2.3.2 Prospective space-time analysis

Fifty-two prospective time-space permutation analyses were performed, one for each week from January 1, 2010 to Dec 31, 2010. The 3 years of retrospective data were used in the analysis as the initial reference baseline. The baseline subsequently increased in length by one week as the prospective analysis moved forward one week. The same time aggregation and maximum temporal window parameter settings (7 days and 100 days respectively) used in the retrospective space-time analysis were also used in the prospective analysis. In the prospective study the maximum spatial window size could not be defined by the percent of the population; this parameter required an additional grid file to define the circles used by the scan statistic. This grid file was not available so as an alternative, using the results from the retrospective analysis, the circle with the largest radius (75km) was identified and became the maximum spatial window size for the prospective study.

4.2.4 Evaluating the cluster signals

The cases identified in the statistically significant (p < 0.05), retrospective and prospective clusters were investigated using the information available within the EMR’s stored in the data warehouse. All of the cases contained within a cluster were reviewed. If necessary the records from previous or subsequent days (and outside the time frame of the cluster) were reviewed to further characterize a case included in a cluster. For example, if an animal’s record indicated that a fecal sample had been submitted to a laboratory for diagnostic testing, subsequent records were searched to find the laboratory results.
Each cluster was described by the following variables:

1. The median age of all animals, the proportions of each species, and the proportion of intact animals (not spayed or neutered) were calculated (Stata/IC 10.0™, StatCorp College Station, TX). The cases were also sorted by breed. Proportions were calculated as follows;

\[
\text{Proportion of cases} = \frac{\text{Number of cases in cluster, } i \text{ with documented variable}}{\text{All cases within the cluster, } i}
\]

2. Information from the animal’s medical record about preventive measures or the potential source of the etiologic agent.

3. An estimate of the severity of each case was determined using the recorded clinical descriptions of the animal’s enteric signs and symptoms. It included any of: hematochezia; ii) if the animal was admitted for intravenous fluid therapy; and/or iii) if the animal died or was euthanized.

4. The etiological diagnosis (when available).

Each of these variables was compared to an expected value (Table 4.1). The expected values for; i) median age; ii) proportion of cases represented by each species; iii) proportion of animals intact or spayed/neutered; and iv) the proportion of each breed were calculated using Knime 2.2.2 and all of the enteric syndrome positive cases from January 1, 2007 to December 31, 2009 (n = 13,063 cases). The proportions of cases describing the history of vaccinations, exposure history, disease severity and the etiologic diagnosis were assessed by reading the medical notes. To establish a reference population against which the cases within a cluster could be compared, the medical notes from a random sample of 500 cases (from the enteric syndrome positive cases from January 1, 2007 to December 31, 2009) were reviewed. The sample size was calculated using (226):
\[ n = \frac{Z_{\alpha}^2 \cdot pq}{L^2} \]

- \( \alpha \)  
  Significance level = 0.05
- \( Z_{\alpha} \)  
  Under the normal distribution, the 1-\( \alpha \)/2 percentile = 1.96
- \( p \)  
  \( A \) priori estimate of the proportion, conservatively = 0.5
- \( q \)  
  1 – \( p \) = 0.5
- \( L \)  
  precision of the estimate, allowable error = 0.10
- \( n \)  
  number of target cases required in the sample = 96

To reach the target of 96 positive cases in the sample required an estimate of the proportion of cases that would be positive for each variable. This was unknown and would differ for each variable so a proportion of 0.25 was selected.

\[ \frac{96}{0.25} = 384 \text{ non-cases (controls)} + 96 \text{ target cases} = 480 \]

I elected to use a sample of 500 cases.

Median ages of the cases in a cluster were compared to the expected values using the Wilcoxon-Mann-Whitney test. Proportions were compared using a 2 sample comparison of proportions (Stata/IC 10.0™, StatCorp College Station, TX).

To assess the potential usefulness of the data for further epidemiological investigations, selected variables were evaluated with additional statistical testing. Statistically significant clusters were divided into 2 subgroups: i) clusters in which there was a higher than expected proportion of cases (\( p < .05 \)) with a positive canine parvovirus (CPV) diagnosis; and ii) clusters that had an expected proportion of cases (\( p > .05 \)) with a positive CPV diagnosis. The measure of effect for the sex variable compared each of the subgroups to the reference population and was reported as an odds ratio and 95% confidence interval (CI). If the expected median age and the median age of the animals in the subgroups differed, direct age standardization methods (250)
were used to remove the effect of confounding by age and enable comparison of the sex characteristics of the two groups.

After the characteristics of the cluster were described and it was determined that the findings were unexpected, the information was evaluated for its ability to develop a hypothesis as to the cause of the outbreak, to assess the possible risk factors for enteric syndrome in the cluster, and/or to inform a response by animal health or public health officials.

4.3 Results
4.3.1 Time series

The number of enteric cases presenting to the sentinel veterinary practices on Sundays was less than those presenting the remainder of the week (Figure 4.2). This pattern was a reflection of the veterinary practice hours. The daily and 7 day moving average of counts of enteric cases were plotted against time. The numbers of enteric syndrome cases presenting to the sentinel veterinarians increased in the late summer and fall for each of the 3 years of the retrospective study, a time frame of approximately 100 days (Figure 4.3). Linear regression showed a statistically significant (p = .004), small upward (β coefficient = .0002) trend in the proportion of enteric cases to all cases seen by the sentinel practices.

4.3.2 Space-time analysis

A 7-day aggregation setting was used in SaTScan™ to adjust for the day of the week effect. Time aggregation also served to reduce the computing time. Repeated scans of the retrospective data using variable maximum spatial window sizes had identical results when the maximum spatial window sizes were 50, 40, and 30% of the population. The scan using the 20% maximum window size identified the same six clusters as the scans using the larger spatial windows. The 20% maximum spatial window size differed in 2 clusters; the numbers of FSA’s in the cluster were reduced and the timeframes were 1 to 2 weeks shorter. The scan using 10% of the study population as the maximum window size identified 5 of the same 6 clusters as the previous scans
as well as 2 additional small clusters. SaTScan™ results are generally sensitive to the parameter settings but using variable maximum spatial window sizes had little effect on the output. Therefore the default maximum spatial window of 50% of the population at risk was used. There were 6 significant (p < 0.05) clusters identified in the retrospective space-time analysis (named R1 through R6) and 6 significant clusters identified in the prospective space-time analysis (named P1 through P6).

4.3.3 Evaluating the cluster signals

The proportions of cases that had hematochezia or were placed on intravenous fluids or euthanized were not statistically different from the reference population in any of the clusters. There were no unassisted deaths in any of the reviewed records.

Four clusters (R1, R5, R6 and P3) had a larger proportion (p < 0.05) of cases positive for CPV than the reference population (Tables 4.2 and 4.3). These 4 clusters are herein referred to as being etiologically-described. These clusters were all located in northeast Calgary or in the rural areas east of the city (Figure 4.4) and overlapped the seasonal increase in enteric syndrome identified in the time series (Figure 4.3, Tables 4.2 and 4.3). A larger than expected proportion of cases positive for *Giardia* spp. was found in addition to CPV in one cluster and greater numbers than expected of coccidia and hookworms was found in addition to CPV in another (Table 4.2). Eight clusters (R2, R3, R4, P1, P2, P4, P5 and P6) had expected (p > 0.05) values for the percentage of cases with a positive etiological diagnosis (Table 4.2, Table 4.3).

The odds ratio for an animal in an etiologically described cluster not being spayed or neutered was 4.75 (95% CI, 3.29 to 6.86) compared to the animals in the reference population. The odds ratio for an animal in the remaining 8 clusters (had expected numbers of an etiological diagnoses) being intact was 1.31 (95% CI, 1.02 to 1.69) compared to the reference population. The animals in the etiologically described clusters were also younger (p < .05) than the reference median age (Table 4.2). Age standardization was used to remove the effect of age and compare the association between surgical sterilization and whether the clusters had higher than expected
numbers of cases with confirmed etiologies or not. The age-adjusted proportion of animals that had not been spayed or neutered in the etiologically-described clusters (0.49) was statistically different (p < 0.05) than the age adjusted proportion of intact animals in the remaining 8 clusters (0.34) (Table 4.4).

In 2 clusters (R1 and R6), the dog breed distribution was unusual. In cluster R1, 41% of the cases (n = 61) and 52% of the positive CPV cases (n = 19) were shepherd crosses. Based upon the recorded ages, 12 CPV-positive puppies were from one of 4 litters. Similarly, in cluster R6, 33% of the cases (n =26) were Rottweiler crosses and 8 of the 9 positive *Giardia* spp. cases were identified as Rottweiler cross littermates. Clusters R3, R4, P3 and P6 had a higher than expected proportion of cat enteric cases (Table 4.5). In cluster P3, cats accounted for 5 of the 7 coccidia positive cases in the cluster. There was no other information found in the records to explain the unusual increase in feline enteric cases in these clusters.

The vaccination history was inconsistently recorded in the records from the clusters (0.26 ± 0.14, standard deviation). Most (0.86 ± 0.03, standard deviation) of the records in each cluster had no exposure history documented (Figure 4.5). When present, the information on potential exposures or risk behaviours included; no known dietary indiscretion, known dietary indiscretion, staying at a boarding kennel, a recent purchase from a breeder/pet store/humane society, visits to parks/lakes/rivers, a hunter/scavenger, fed raw diet, a co-morbidity, medications and stress. In cluster P3, the 2 puppies diagnosed with hookworms had been imported from the Dominican Republic.

### 4.4 Discussion

This study demonstrates that the extracted electronic medical records from sentinel veterinary practices could be useful for monitoring trends in enteric syndrome and for identifying clusters of enteric syndrome in retrospective and the prospective space time analysis. There was no alternative data source against which to measure the sensitivity and specificity of the system’s
ability to detect a signal (141). In this study the investigation was limited to a review of the clinical records.

Describing the clusters by disease severity was not useful for determining if the cluster was unexpected or unusual. Increased numbers of etiologic diagnoses were useful for determining that 4 clusters were unusual. All of the etiologically-described clusters were contained within or overlapped the late summer and fall seasonal peak of enteric syndrome in companion animals that was identified in the time series of enteric syndrome cases (Figure 4.3 and Table 4.3). Seasonal variation in disease occurrence is epidemiologically significant because cycles in environmental changes can influence exposures, physiology and behaviours (250). CPV, *Giardia* spp., coccidia and hookworm are endemic pathogens of companion animals and outbreaks are possible. They are of animal health importance and are capable of causing disease in pets and spreading through direct fecal-oral contact or environmental contamination (251).

The records from the etiologically-described clusters identified in this study provided some information from which risk factor hypotheses could be developed. Dog breeders were linked to two clusters of CPV or *Giardia* spp. The etiologically-described clusters were all located in northeast Calgary and the rural areas east of the city (Figure 4). The animals in these clusters were younger and had greater odds of not having been surgically sterilized when compared to the reference population. After age standardization, the difference in the proportions of intact animals, 0.49 versus 0.34, suggests that being intact is associated with cases from etiologically-described clusters. One possible hypothesis is that the pets in these clusters received less preventive veterinary care than the pets in other regions of the study area. The vaccination history would have provided some insight into this hypothesis but this information was not routinely recorded in the enteric syndrome records. Further investigations such as surveys of the pet owners and of the veterinarians practicing in this area of Calgary would help to understand the context these findings.

The retrospective findings provide mixed results regarding the utility of the outputs of this type of surveillance. Text-mining based syndromic surveillance data could yield information for
a targeted educational campaign. Detection of CPV and *Giardia* spp. clusters, for example, might inspire an information campaign directed towards dog owners and dog breeders in these regions about the importance of CPV vaccination or *Giardia* spp. prevention. However, the lack of detailed etiological data would likely be an impediment to using these results as signals of public health risk. For example, some clusters of giardiasis were found and some strains of *Giardia* spp. are potentially zoonotic (180), but without a strain specific diagnosis or evidence of human transmission it is unlikely that a response by public health officials would be triggered. In the 8 clusters that had expected numbers of etiological diagnoses, there was insufficient diagnostic or exposure information to determine if the increased numbers of enteric syndrome cases were biologically or epidemiologically related and therefore the importance of these clusters to animal or human health was unknown.

Detecting significant disease clusters prospectively is more difficult because the analysis must include the last event in the series (205). Despite this limitation, 6 significant disease clusters were detected in this study. There was one cluster with greater than expected diagnoses of CPV, coccidia and hookworms. The actions taken after detecting clusters such as these could include issuing an animal health alert to veterinarians and the public of the outbreak and information on the appropriate preventive measures. The remaining 5 prospective clusters were all limited by the lack of etiologic diagnoses or exposure information. Further investigations would be required to determine the etiologic and epidemiologic links among the patients to be able to understand if the increased number of cases of diarrhea was significant. There is no mandate or provisions for companion animal disease outbreak investigations in the province of Alberta and it is unlikely that any action would be taken by the public sector upon cluster detection. Application of this surveillance system would need to be led by the private or academic sector because they are the 2 sectors most likely to benefit from or use the outputs. Action following the prospective detection of a cluster of enteric disease could include phone calls to the veterinary practices to seek additional information about risk behaviours and exposures. Given that few enteric syndrome cases are routinely subjected to etiological testing (see Chapter 5) communication of
an unusual number of cases of diarrhea could be used to motivate veterinarians in the cluster area to increase diagnostic testing.

The results of this study were similar to those of Balter et al. (252). These authors reported on a human syndromic surveillance system using emergency department chief complaint data for gastrointestinal syndrome. Their system was capable of detecting seasonal outbreaks of diarrheal illness due to norovirus. However, an outbreak that occurred outside of the seasonal peak of diarrheal disease that followed a widespread power outage had insufficient information recorded in the records to characterize that cluster by etiology or exposures.

This was the first reported study evaluating space-time cluster detection in veterinary syndromic surveillance using companion animal records and informatics. The study demonstrated that cluster detection is possible but the importance of the cluster to animal or public health was compromised by the lack of data. To overcome this limitation, adding a data stream with the fecal examination and fecal culture results from the laboratories outside the veterinary practices could be investigated. It is also possible that if this system was adapted for syndromic surveillance of alternative syndromes such as respiratory or urinary tract diseases, the information biases affecting the data would be different from those identified in this study. Future work could explore the implications for cluster detection with alternative analytical decisions such as excluding patient repeat visits or employing another detection algorithm.

The use of sentinel veterinarians and the associated selection bias may have impacted this study’s results. In order for a case of enteric syndrome to be captured by this syndromic surveillance system, an animal exhibiting clinical signs of enteric disease needed to have an owner, their pet’s symptoms needed to be severe enough that the owner could justify a visit to the veterinarian and the animal had to be seen by a sentinel practice. Previous work (see Chapter 2) demonstrated that the data collected by this system was not geographically or demographically representative of the underlying companion animal population. Pets in northeast Calgary were underrepresented in the data but enteric syndrome signals were detected in this region. The system also collected more information on dogs and on animals younger than
what is known about the pet population in Calgary (Chapter 2). This bias may have enhanced the ability of the system to detect the etiologically defined clusters of enteric disease. It is also possible that the limitations found in the medical records in this study (few diagnostic tests performed and limited exposure information) are not consistent across the study area. Therefore the results of this study cannot be generalized to all of the pet population in Calgary and area.

4.5 Conclusion

The use of a syndromic case definition ensured that there were a sufficient number of enteric syndrome cases to identify space-time clusters. However, there were too few etiologic diagnoses to understand the relevance of most of the clusters and case assessment was limited to a few demographic parameters. Most of the case histories were only briefly documented especially in terms of the animals’ potential risk behaviours and risk exposures so an examination of the risk factors for these clusters using the records alone was compromised by a lack of data.
Figure 4.1: Schematic of the methods used in this study (after (145))
Table 4.1: Expected values of the variables for the enteric syndrome cases calculated from the cases of enteric syndrome, January 1, 2007 to December 31, 2009 (n = 13,063)

<table>
<thead>
<tr>
<th>Species, %</th>
<th>Median Age</th>
<th>Reproductive status, %</th>
<th>Each Breed, %</th>
<th>Diagnoses, %</th>
<th>Disease severity, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dog, 84.4</td>
<td>2.86 years</td>
<td>Intact, 22.2</td>
<td>4 to 6⁹</td>
<td>CPV³, 2.0</td>
<td>Hematochezia, 18.2</td>
</tr>
<tr>
<td>Cat, 14.8</td>
<td></td>
<td>Spay/Neuter, 77.8</td>
<td></td>
<td>Giardia spp., 1.8</td>
<td>Hospitalization, 4.6</td>
</tr>
<tr>
<td>Other⁸, 0.8</td>
<td></td>
<td></td>
<td></td>
<td>Bacterial o/g⁸, 1.4</td>
<td>Euthanasia, 1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Campylobacter⁹, 1.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Helminths, 0.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Coccidia, 0.4</td>
<td></td>
</tr>
</tbody>
</table>

a. Other = ferret, rabbit, rodent

b. Variable but the most common 15 breeds are represented 4 to 6% of the time
c. Canine parvovirus
d. Morphological diagnosis of bacterial overgrowth using light microscopy
e. Morphological diagnosis of Campylobacter spp., using light microscopy
Figure 4.2: Number of cases of enteric syndrome seen by the sentinel practices on each day of the week in 2007, 2008 and 2009
Figure 4.3: Daily count and 7 day moving average of counts of the enteric syndrome cases seen by the sentinel veterinary practices in 2007, 2008 and 2009
Table 4.2: Significant (< 0.05) retrospective (R) space-time clusters of enteric syndrome

<table>
<thead>
<tr>
<th>Cluster (p-value)</th>
<th>Number of cases (expected number)</th>
<th>Dates (M/D/Y)</th>
<th>FSA’s(^a)</th>
<th>Median Age (years)</th>
<th>Intact (%)</th>
<th>Diagnosis, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1 ((2.9E-15))</td>
<td>61 (15)</td>
<td>9/26/09 to 11/20/09</td>
<td>T3J T3N T1Y</td>
<td>0.46</td>
<td>53</td>
<td>CPV(^b), 31.1</td>
</tr>
<tr>
<td>R2 ((8.2E-7))</td>
<td>275 (181)</td>
<td>4/12/08 to 6/13/08</td>
<td>T2Y T2W T2X T2V</td>
<td>3.23</td>
<td>19</td>
<td>CPV, 1.5</td>
</tr>
<tr>
<td>R3 ((1.0E-6))</td>
<td>294 (197)</td>
<td>10/27/07 to 2/1/08</td>
<td>T2K T2L T2N T2R T2T T3A T3B T3C T3G T3H T3L</td>
<td>2.98</td>
<td>19</td>
<td>CPV, 2.0</td>
</tr>
<tr>
<td>R4 ((1.6E-6))</td>
<td>65 (26)</td>
<td>6/13/09 to 9/11/09</td>
<td>T1W T4C T3Z</td>
<td>2.32</td>
<td>12.3</td>
<td>-</td>
</tr>
<tr>
<td>R5 ((0.01))</td>
<td>43 (18)</td>
<td>9/12/09 to 10/23/09</td>
<td>T2A T2B</td>
<td>0.46</td>
<td>72</td>
<td>CPV, 27.9</td>
</tr>
<tr>
<td>R6 ((0.021))</td>
<td>26 (8)</td>
<td>11/29/08 to 1/9/09</td>
<td>T0J</td>
<td>0.2</td>
<td>67</td>
<td>CPV, 7.8 (Giardia, 34.6)</td>
</tr>
</tbody>
</table>

\(^{a}\) Forward sortation area

\(^{b}\) Canine parvovirus
Table 4.3: Significant (< 0.05) prospective (P) space-time clusters of enteric syndrome

<table>
<thead>
<tr>
<th>Cluster (p-value)</th>
<th>Number of cases (expected number)</th>
<th>Dates (M/D/Y)</th>
<th>FSA’sa</th>
<th>Median Age (years)</th>
<th>Intact (%)</th>
<th>Diagnosis, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 (.0055)</td>
<td>90 (48)</td>
<td>1/4/10 to 5/13/10</td>
<td>T1Y T3J T3N</td>
<td>1.5</td>
<td>45</td>
<td>CPV, 5.5</td>
</tr>
<tr>
<td>P2 (2.2E-7)</td>
<td>150 (74)</td>
<td>5/21/10 to 9/2/10</td>
<td>T1Y T2E T2K T3J</td>
<td>2.08</td>
<td>35</td>
<td>CPV, 2.0</td>
</tr>
<tr>
<td>P3 (0.0011)</td>
<td>85 (44)</td>
<td>7/24/10 to 10/29/10</td>
<td>T0J</td>
<td>0.50</td>
<td>55</td>
<td>CPV, 22.8 Coccidia, 9.7 Hookworm, 2.8</td>
</tr>
<tr>
<td>P4 (6.7E-9)</td>
<td>266 (199)</td>
<td>8/7/10 to 11/25/10</td>
<td>T1Y T2E T2K T2M T3J T3K T4A</td>
<td>2.54</td>
<td>37</td>
<td>CPV, 3.0</td>
</tr>
<tr>
<td>P5 (.015)</td>
<td>37 (11)</td>
<td>8/2/10 to 12/8/10</td>
<td>T4A T4B</td>
<td>1.99</td>
<td>35</td>
<td>CPV, 5.4</td>
</tr>
<tr>
<td>P6 (6.1E-7)</td>
<td>49 (21)</td>
<td>10/23/10 to 12/29/10</td>
<td>T1P</td>
<td>2.36</td>
<td>18</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 4.4: Map showing location of etiologically-defined clusters.

R = retrospective and P = prospective.
Table 4.4: Age standardization to compare the association of animals that had not been spayed or neutered between the etiologically described clusters and the clusters with expected numbers of etiologies.

<table>
<thead>
<tr>
<th>Age Group*</th>
<th>Number of cases</th>
<th>Age Distribution</th>
<th>Number Intact</th>
<th>Rate</th>
<th>Standard Population</th>
<th>Adjusted Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etiologically-described clusters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤6 mo</td>
<td>105</td>
<td>0.528</td>
<td>90</td>
<td>0.857</td>
<td>0.533^b</td>
<td>0.491</td>
</tr>
<tr>
<td>&gt;6 mo</td>
<td>94</td>
<td>0.472</td>
<td>24</td>
<td>0.255</td>
<td>0.135</td>
<td></td>
</tr>
<tr>
<td>Clusters with expected numbers of etiologies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤6 mo</td>
<td>231</td>
<td>0.223</td>
<td>139</td>
<td>0.602</td>
<td>0.533</td>
<td>0.340</td>
</tr>
<tr>
<td>&gt;6 mo</td>
<td>806</td>
<td>0.778</td>
<td>114</td>
<td>0.141</td>
<td>0.135</td>
<td></td>
</tr>
</tbody>
</table>

a. Less than or equal to 6 months of age and greater than 6 months of age
b. Calculated from the enteric syndrome positive cases January 1, 2007 to December 31, 2009
Table 4.5: Clusters with statistically different (p < .05) proportion of feline cases when compared to the reference population.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Reference</th>
<th>R3</th>
<th>R4</th>
<th>P3</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of Feline Cases</td>
<td>0.148</td>
<td>0.218</td>
<td>0.231</td>
<td>0.25</td>
<td>0.41</td>
</tr>
</tbody>
</table>

R = retrospective cluster and P = prospective cluster
Figure 4.5: Exposure information found in the records of cases from statistically significant time-space clusters of enteric syndrome.
Chapter Five: Using informatics and companion animal electronic medical records to describe the clinical management of enteric syndrome cases

5.1 Introduction

Diarrhea is a common clinical sign of enteric syndrome in companion animals. The pathophysiology of diarrhea is complex, poorly understood and can involve a wide array of infectious and non-infectious etiologies (253, 254). Clinical evaluation of ill animals directs the selection of diagnostic procedures such as parasite studies, microbiological examinations and/or toxin testing. Clinicians must weigh the cost of diagnostic procedures, the owner’s willingness to pay for them and the time spent waiting for a result against the likelihood that the results of a diagnostic test will affect their therapeutic recommendations. This cost: benefit analysis often results in empirical therapy with anthelmintics and antimicrobials being used to manage diarrheal illnesses in pets (255).

Infectious disease specialists advocate restricting antimicrobial use (AMU) to cases where there is evidence that AMU will result in improved clinical outcomes (254, 256, 257). Warnings against indiscriminate AMU in animals are increasing because the consequences of AMU include antimicrobial resistance (AMR) and decreased efficacy of important antimicrobials against significant animal and human pathogens (258, 259). Health Canada has categorized antimicrobial drugs based on their importance for treating serious bacterial infections in people and on the availability of alternatives should resistance develop to the antimicrobial of choice (260). The classes of antibiotics used in companion animal practice mirror those used in people (259). In their closely shared environment, pets may be a source of antimicrobial resistant enteric bacteria or resistance genes for their owners (47, 261).

Collecting AMU data at the patient level has been a challenge in human and veterinary medicine (53, 54). The collection of animal AMU data has not been legislated in Canada. Generally AMU data in animals is measured in terms of kilograms of antimicrobial active ingredients distributed for use in animals per year (54). There have been farm-based, time-
limited projects measuring AMU (262) and surveillance of AMU on sentinel farms representing specific livestock commodities (http://www.phac-aspc.gc.ca/cipars-picra/surv_farm-eng.php).

Surveillance of AMU by companion animal practitioners for the early detection of emerging AMR risks to animals and people is a priority for the Canadian Council of Chief Veterinary Officers (personal communication, S. Otto, Alberta Agriculture and Rural Development, 2012). AMU surveillance data provides the necessary exposure information to interpret AMR trends, identify potential problem areas in prescribing practice and provide evidence-based practice guidelines for practitioners (263-267). Monnet et al. (53) demonstrated that studies of variations in human AMU and the concomitant variations in AMR over time and place provided the best evidence of an association between the suspected cause (AMU) and the observed effect (AMR). Retrospective cross-sectional studies have described AMU in veterinary teaching hospital records and compared the AMU patterns with prudent use guidelines (55-59) or with the AMR profiles of selected bacteria (60). Each of these studies was able to identify the most frequently used antimicrobials, describe changing patterns of AMU and instances of inappropriate use or insufficient laboratory testing to justify their use. Another approach recruited general practice, companion animal veterinarians who were asked to complete a journal with a detailed description of 5 outpatients, their disease event, and the treatments administered one day per month for 12 months (268). There has been no studies reported that used systematic and continuous collection of AMU data from companion animal practices to monitor changes in AMU trends.

The uptake of the electronic medical record (EMR) by companion animal practitioners provides an opportunity for accessing AMU data. Informatics is “the application of information and computer science technology to public health practice, research and learning” (158). In Chapter 2, I showed how informatics can enable the collection, storage and analysis of EMR’s from participating companion animal practices. In Chapter 3, I demonstrated how text-mining can be used to identify and extract cases of enteric syndrome. This chapter explores text-mining for accessing and analysing the textual information in the medical records including orders for
diagnostic testing and AMU. Informatics has been applied elsewhere to text-based clinical records to describe disease-drug associations by physicians (269) but no similar references could be found in the veterinary literature.

Following my previous work, the objectives of this study were to:

1. Apply and evaluate text-mining technology to characterize the clinical management of enteric syndrome cases by veterinarians in a sentinel practice network.
2. Describe the diagnostic management of enteric syndrome in companion animals and the proportion of cases for which there was documented evidence of an infectious process.
3. Describe the use of antihelmintics and antimicrobials in the enteric syndrome cases.
4. Describe the temporal patterns of use for each antimicrobial class used in the treatment of enteric syndrome for a 4 year period (January 1, 2007 to December 31, 2010).

5.2 Methods

5.2.1 Study area and data

The electronic medical records were extracted from the veterinary management programs at twelve sentinel companion animal practices. The study area included 6 communities in the province of Alberta, Canada: Calgary, Cochrane, Airdrie, Chestermere, Strathmore and Okotoks. All records were stored in a data warehouse at the University of Calgary. The features of the sentinel practices, data extraction and data warehouse were described in Chapter 2. The appointment schedule, medical notes (history, clinical exam, interpretations of diagnostic tests, assessment, differential diagnoses, and treatment) and prescription data for each case were combined into one variable named ‘Note’, in the data file. All records were in free-text. Text-mining and a syndromic case definition were used to identify and retrieve the enteric syndrome positive cases from these warehoused medical records (see Chapter 3 for details). Data was stored and managed using Microsoft Office Excel 2007™ (Microsoft Corporation, Redmond, Washington) and Konstanz Information Miner 2.2.2 (Knime, http://www.knime.org).
5.2.2 Clinical management of enteric syndrome cases

Objectives 1, 2 and 3 were addressed using the enteric syndrome positive cases from January 1, 2007 to December 31, 2009 (n = 13,063). Text-mining software (QDAMiner3.1/WordStat6™, Provalis Research, Montreal, QC) was used to identify and retrieve cases that recorded the following variables:

- Cases for which diagnostic testing had been performed
- Cases where an etiological diagnosis had been made
- Cases treated with an antihelmintic
- Cases treated with an antimicrobial

For further explanation on the development and optimization of the text-miner’s categorization dictionaries see Chapter 3.

5.2.2.1 Diagnostic testing and etiologic diagnosis

Case definitions were developed to be used by the categorization dictionaries in WordStat™ and by the external reviewer. A case was classified as positive for diagnostic testing if any of the following diagnostic tests were recorded within the variable ‘Note’. A diagnostic test was a laboratory test that could either be performed in the practice by the animal health technologist or sent to an external veterinary laboratory. These were:

- Fecal flotations and fecal smears that provided a morphological diagnosis of helminths, protozoa or bacteria and included imprecise morphological diagnoses (made using light microscopy by the veterinarian or technician) of bacterial infections such as bacterial overgrowth and Campylobacter-type spp.
- Enzyme-linked immunosorbent (ELISA) assays to identify canine parvovirus or Giardia spp. infections.
The laboratory tests that were only performed by an external veterinary laboratory included:

- Fecal bacteria culture

A case was classified as positive for etiologic diagnosis if a positive outcome for any of the diagnostic tests described above was recorded in the variable ‘Note’.

5.2.2.2 Antihelmintic or antimicrobial use

Case definitions were developed to be used by the categorization dictionaries in WordStat™ and by the external reviewer. Positive antihelmintic or antimicrobial cases were defined as those enteric syndrome cases that were administered, dispensed or prescribed antihelmintics or antimicrobials for the management of the enteric signs. Cases that were on a scheduled deworming program were also classified as antihelmintic use positive. The antimicrobials used were classified by Health Canada’s categorization of antimicrobials based on their importance to human health (260).

5.2.2.3 Text-miner accuracy

To test the ability of the text-miner to accurately identify cases by the above variables, a sample of 500 enteric positive cases was randomly selected from the 13,063 enteric syndrome cases. The assumptions and sample size calculations were described in Chapter 4.

An experienced veterinarian clinician, blinded to the results of the text-miner, reviewed the information documented by the veterinarian and their staff in the extracted EMR of the sampled 500 cases and contained within the variable ‘Note’. The clinician reviewer classified each case as positive or negative for each of the variables listed above. This served as the external standard.
There was no method that would unequivocally classify cases as enteric syndrome positive or negative so the term external standard was used rather than “gold standard” (225).

The text-miner’s dichotomous test results were compared with the external standard and cross tabulated in 2 X 2 tables. The results for each of variables (diagnostic testing, etiological diagnosis made, antihelmintics and antimicrobials), were summarized as the sensitivity (Se) and the specificity (Sp) of the text-miner’s ability to accurately identify and retrieve the cases. The 95% confidence intervals for the Se and Sp were also calculated (Exact method, Stata/IC 10.0™, StatCorp College Station, TX) (227, 228). The cases that were improperly classified (false positives and false negatives) were reviewed to determine why they had been misclassified and if there was any opportunities to improve the text-mining results.

After assessing the ability of the text-miner to accurately classify cases by a diagnostic test performed and a diagnosis made, I determined that it was necessary to complete the description of the case management by manually reviewing the records. The same random sample of 500 enteric syndrome positive cases used to test the text-miner was manually reviewed and the cases (patients) were categorized by: i) no diagnostic testing performed, ii) diagnostic testing with a negative result or no result recorded; and iii) a positive diagnosis. Within each of the 3 categories the proportion of patients that were managed with antihelmintics and/or antimicrobials was determined.

In some cases, additional data to characterize a case were found in an animal’s records from the days previous or subsequent to the date that had been selected. For example, if an animal’s record indicated that a fecal sample had been submitted to a laboratory for diagnostic testing, subsequent records were searched to find the laboratory results. Access to results from an outside laboratory was not routinely available and therefore, in some cases, an order for a test or its result recorded in the medical record was used as a surrogate for the original laboratory report. Invoices that would provide an alternative source of information about the procedures and services provided were not accessible.
5.2.3 *Antihelmintic and antimicrobial use*

Objective 4 was addressed using the enteric syndrome positive cases from January 1, 2007 to December 31, 2010 (n = 18,832). The enteric syndrome positive cases that had been administered, dispensed or prescribed antihelmintics or antimicrobials were identified and extracted using the text-mining categorization dictionary described above. Using text-mining, the antimicrobial use positive cases were further classified by the class of antimicrobial with which they had been treated using Health Canada’s categorization of antimicrobial drugs based on importance to human medicine (260). Co-occurrences of antimicrobial use were identified by WordStat™ and the antimicrobials used in combination were described.

I examined the temporal trends of the Category I and Category II antimicrobials (260) for the 4 years of the study. For each month of the study, I determined the proportion of cases that had been treated with any antimicrobial and the proportions treated with each class of antimicrobial. The temporal trend for all antimicrobials combined and for each antimicrobial was examined by fitting a linear regression model to the data. The number of antimicrobial treated cases, normalized by the total number of enteric cases for each month, was the dependent variable and the month/year was the independent variable. If the antimicrobial use data fit the slope estimated by the linear regression (p < 0.05), the proportions of cases treated with this antimicrobial were plotted as a function of time (53, 270). Further exploratory data analysis included data smoothing by: i) pooling the number of cases treated with each class of antimicrobial in each quarter of each year; and ii) plotting the results in scatterplots with quadratic overlays (Stata/IC 10.0™, StatCorp College Station, TX).
5.3 Results

5.3.1 Text-mining

The estimates that measured the text-miner’s ability to distinguishing between cases that had: i) diagnostic testing performed; and ii) had a diagnosis made, were relatively low and there were wide confidence intervals around Se which indicated poor precision of the estimate (Tables 5.1 and 5.2). The primary reason the text-miner performed poorly when classifying and extracting these cases was that for some words, the context was relevant to the classification of the case. For example, the word “giardia” was associated with a diagnosis, a differential diagnosis, a past diagnosis, a diagnostic test, a vaccine, and a recommendation or a warning to owners. Despite repeated efforts, it was not possible to improve the performance of the text-miner for these variables and text-mining was not used to describe enteric syndrome cases by the diagnostic tests preformed or their etiological diagnosis.

In contrast, text-mining retrieved cases that had been treated with an antihelmintic and/or an antimicrobial with high Se and Sp when compared to a human reviewer (Tables 5.3 and 5.4). The text-miner misclassified cases if the name of the antihelmintic or antimicrobial was not provided or improperly spelled, if the record contained information about past treatment or future considerations for treatment or if the pet was receiving antimicrobials but they were being used to treat a co-morbidity (not dispensed for enteric syndrome). Given the high Se and Sp of the text-miner to classify cases by their antihelmintic and antimicrobial use, it was used for the remainder of the analysis.

5.3.2 Diagnostic testing, diagnoses and antimicrobial use

As the text-miner did not accurately retrieve cases that had laboratory testing performed or a diagnosis made, the results presented are from the manual review of the sample of 500 enteric syndrome positive cases only. The remaining enteric syndrome positive cases were not described by their diagnostic testing or etiological diagnosis. It was determined that 79 (15.8%) of the sample of 500 enteric syndrome positive cases were tested to identify an etiological diagnosis.
(Figure 5.1 and Table 5.5). Of the cases that underwent diagnostic testing, fecal examinations (smears and/or floats) were performed in 52% of the tests, 46% had ELISA assays to identify canine parvovirus or *Giardia* spp., and fecal cultures or PCR test were each ordered in 1% of the cases. Multiple testing using a combination of fecal exams and ELISA tests was documented in 31.6% of those tested. Thirty-four cases (43% of those tested, 6.8% of all cases) had a stated etiologic diagnosis in the EMR. Patients that had diagnostic procedures performed had more antimicrobials administered, dispensed or prescribed (68% when the animal had a positive diagnosis and 64% if a negative diagnosis) than patients that had no diagnostic testing performed (36%) (Figure 5.1). The proportion of cases treated with antihelmintics was greatest (32%) in the animals with a confirmed diagnosis (Figure 5.1).

Text-mining of the enteric syndrome cases from January 1, 2007 to December 31, 2010 determined that antihelmintics were administered, dispensed, or prescribed in 3015 (16%) of the enteric syndrome cases. Antimicrobials were administered, dispensed or prescribed in 8441 (44.8%) of the enteric syndrome cases. The distribution of the antimicrobial classes used in the management of enteric syndrome positive cases is summarized in Table 5.6. The distribution of antimicrobials classified by Health Canada’s importance to human health was: i) Category I, 79.2%; ii) Category II, 20%; iii) Category III, 0.8%; and iv) Category IV, none. Veterinarians prescribed more than one antimicrobial in 1250 (14.8%) of all cases treated with an antimicrobial. Nitroimidazole plus a penicillin was the most frequent treatment combination (n = 496) followed by nitroimidazole together with first and second generation cephalosporins (n= 105), nitroimidazole with fluorquinolones (n = 101), and penicillins in combination with fluorquinolones (n = 100).

### 5.3.3 Antimicrobial use temporal trends

The linear regression analyses of ‘all antimicrobials’ (n =8441), ‘nitroimidazole’ (n = 6820) and ‘penicillin’ (n = 1057) were significant (p < 0.05) and these variables were plotted against time (Figure 5.2). The graph and the slope coefficients (0.0002 to 0.0004) indicate a very small statistically significant, upward trend in the proportions of enteric cases treated with any
antimicrobial and treated with nitroimidazoles and penicillins. The regression analyses of the remaining antimicrobials were not statistically significant.

Smoothed scatterplots of the quarterly counts of cases treated with 3rd/4th generation cephalosporins and the penicillin β-lactamase inhibitor combinations showed patterns of antimicrobial use that were mirror images of each other (Figure 5.3). Scatterplots of the remaining antimicrobial class combinations did not show any recognizable patterns.

5.4 Discussion

The results from the text-mining methods used in this study varied depending on the variable of interest. The text-mining results were different because the documentation for antihelmintic or antimicrobial treatments was usually explicit and unambiguous; the meanings of the words did not depend upon the context in which they were used. However, the language used to communicate diagnostic procedures and diagnoses was highly context specific and the linguistic-based text-mining used in this study was unable to discriminate between the various meanings. It is possible that trained or rule-based text-mining would more accurately distinguish these cases and is an area for future study (218, 220).

Most cases of acute (less than 7 days) diarrhea are mild and self-limiting and supportive treatment without a diagnosis is considered appropriate (253) so it was not unexpected that less than 16% of the enteric syndrome cases in our study had diagnostic procedures performed. The recommended initial diagnostic approach to acute diarrhea is a fecal exam (271); more than half of the diagnostic procedures in our study were fecal flotation and/or fecal smears. In animals with severe disease (febrile, dehydrated, hemorrhagic or persistent diarrhea) further efforts at establishing an etiological diagnosis are warranted (253, 271). Animals in this study that were subjected to diagnostic laboratory tests were more likely to be given antimicrobials than those that were not tested regardless of the test results. This may indicate an assessment of more severe disease by the veterinarian although this judgement was not often explicitly stated in the medical
record. Despite efforts to identify an etiological agent, a positive diagnosis was established in less than half of the cases undergoing diagnostic testing.

A review by Weese (254) discusses the limitations of diagnostic tests for diarrhea. Laboratory tests for enteric parasites and bacteria have variable Se and Sp. Few papers in the peer-reviewed literature have validated molecular tests for enteric pathogens in the dog and cat. Not all species of a bacterium are pathogenic so a morphological diagnosis or diagnosis to the genus level renders the interpretation of laboratory results difficult. Additionally, there is a poor understanding of the normal intestinal microflora of the dog and cat so a precise diagnosis to the species level can still be problematic (254). It is a challenge for the veterinary clinician to determine the optimal, most cost-effective means for the diagnosis and clinical management of diarrhea in their patients.

Antimicrobials are recommended in the management of diarrhea in companion animals if there is a positive diagnosis of secondary bacterial overgrowth associated with inflammatory bowel disease or culture-confirmed primary bacterial infections of *Salmonella*, *Campylobacter*, *Clostridium* and enterotoxigenic *E. coli* (253, 255, 256, 272). Antimicrobials are also recommended if there is evidence of a breach in the mucosal integrity of the intestines (hemorrhagic diarrhea) or to manage the immunosuppressive effects of parvovirus (253, 255, 256, 271, 272). Other authors argue that while antimicrobials are commonly used in cases with a confirmed culture or if there is evidence of hematochezia, there is little objective information as to whether they are needed in all cases (254, 256).

Empirical antihelmintic and antimicrobial treatment was common in this study. The sample of 500 enteric syndrome positive cases showed that none of the 205 cases that received antimicrobials were culture positive for bacteria (Table 5.5 and Figure 5.1). Furthermore 13.4% of the cases received antihelmintics and 36.4% received antimicrobials without a recorded diagnosis (Figure 5.1). Empirical antimicrobial use may lead to treatment failures and antimicrobial resistance (254, 255, 271, 272). I found no post-prescription,
pharmacoepidemiological studies evaluating empirical antimicrobial management of diarrhea in pets in the refereed literature.

Using the data from the extracted medical records it was possible to detect changing trends in AMU. Despite increased AMR concerns (255, 257) there was evidence that nitroimidazole and penicillin use for the management of enteric syndrome in companion animals was increasing. Metronidazole (Class, Nitroimidazoles) was the most frequently prescribed antimicrobial and its use increased over the 4 years of the study. It is the drug of choice for anaerobic and microaerophilic bacteria (Bacteroides and Clostridia) and parasites (Giardiaspp.) in animals (255). In people it is important in the management of these pathogens and Helicobacter pylori (273, 274). There are few therapeutic alternatives for these infections in people and so it is classified as a Category I antimicrobial (260). Sensitivity testing for anaerobes is not routinely performed but treatment failures have been documented (274, 275) and the molecular basis for the resistance has been established (37). I found no papers documenting the transmission of metronidazole-resistant bacteria from pets to people. However, a fluorquinolone resistant strain of Campylobacter jejuni was shared between a 2 year old girl and her dog (276) and the transposable element associated with vancomycin resistance from Enterococcus faecalis isolates has been identified in dogs (277).

The increase in the number of cases treated with 3\textsuperscript{rd} and 4\textsuperscript{th} generation cephalosporins in early 2008 coincided with the Canadian approval on May 30, 2007 and subsequent distribution of Convenia\textsuperscript{TM} (Pfizer Animal Health, Kirkland QC) later in 2007 (278). Convenia\textsuperscript{TM} is the trade name for cefovecin, a third generation cephalosporin. The increase in its use corresponds to a decrease in the use of penicillin β-lactamase inhibitor combinations. The indications for use are similar for the 2 classes so it is possible that one class will be used as an alternative to the other. Starting in mid 2009, the relationship appears to be inverted and this trend continued until the end of 2010. The reason for this trend is unknown but suggests that it would be possible to use the methods described in this study to monitor changes in veterinary prescription behaviour following interventions to modify their use.
5.5 Conclusion

The use of text-mining to assess the laboratory procedures and diagnoses of enteric syndrome cases was impeded by the lack of standardized nomenclature or codes in the medical records. The language used to document antihelmintic and antimicrobial use was clear and text-mining was effective for documenting trends in use. My findings indicated that veterinarians commonly prescribed antimicrobials for enteric syndrome without any documentation that the animal’s diarrhea had an infectious etiology. The results from this study suggest that informatics and the EMR would be useful for monitoring trends in AMU. Temporal trends and regional differences could prompt further investigations to explore why the trends were developing. Confidential benchmarking by comparing AMU between veterinarians may serve to help veterinarians recognize problems and reduce AMU. Analytical studies to see if there is an association between AMU in companion animals with enteric syndrome and the development of AMR in fecal microorganisms are indicated and informatics could provide the exposure data necessary to interpret AMR results.
Table 5.1: The accuracy of the text-miner, WordStat™ to identify cases that had diagnostic testing compared to a manual review of a sample of 500 records serving as the external standard.

<table>
<thead>
<tr>
<th></th>
<th>External standard +</th>
<th>External standard -</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-miner +</td>
<td>66</td>
<td>34</td>
<td>100</td>
</tr>
<tr>
<td>Text-miner -</td>
<td>13</td>
<td>387</td>
<td>400</td>
</tr>
<tr>
<td>Sum</td>
<td>79</td>
<td>421</td>
<td>500</td>
</tr>
</tbody>
</table>

Se = 83.5%
(95%CI, 73.5% - 90.9%)

Sp = 91.9%
(95%CI, 88.9% - 94.3%)

Table 5.2: The accuracy of the text-miner, WordStat™ to identify cases with a stated etiological diagnosis compared to a manual review of a sample of 500 records serving as the external standard.

<table>
<thead>
<tr>
<th></th>
<th>External standard +</th>
<th>External standard -</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-miner +</td>
<td>23</td>
<td>21</td>
<td>44</td>
</tr>
<tr>
<td>Text-miner -</td>
<td>11</td>
<td>445</td>
<td>456</td>
</tr>
<tr>
<td>Sum</td>
<td>34</td>
<td>466</td>
<td>500</td>
</tr>
</tbody>
</table>

Se = 67.6%
(95%CI, 49.5% – 82.6%)

Sp = 95.5%
(95%CI, 93.2% - 97.2%)
Table 5.3: The accuracy of the text-miner, WordStat™ to identify cases treated with an antihelmintic compared to a manual review of a sample of 500 records serving as the external standard.

<table>
<thead>
<tr>
<th></th>
<th>External standard +</th>
<th>External standard -</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-miner +</td>
<td>76</td>
<td>4</td>
<td>80</td>
</tr>
<tr>
<td>Text-miner -</td>
<td>2</td>
<td>418</td>
<td>420</td>
</tr>
<tr>
<td>Sum</td>
<td>78</td>
<td>422</td>
<td>500</td>
</tr>
</tbody>
</table>

Se = 97.4%  
(95%CI, 91.0% – 99.7%)  
Sp = 99.0%  
(95%CI, 97.6% - 99.7%)

Table 5.4: The accuracy of the text-miner, WordStat™ to identify cases treated with an antimicrobial compared to a manual review of a sample of 500 records serving as the external standard.

<table>
<thead>
<tr>
<th></th>
<th>External standard +</th>
<th>External standard -</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text-miner +</td>
<td>203</td>
<td>18</td>
<td>221</td>
</tr>
<tr>
<td>Text-miner -</td>
<td>3</td>
<td>276</td>
<td>279</td>
</tr>
<tr>
<td>Sum</td>
<td>206</td>
<td>294</td>
<td>500</td>
</tr>
</tbody>
</table>

Se = 98.5%  
(95%CI, 95.8% - 99.7%)  
Sp = 93.8%  
(95%CI, 90.5% - 96.3%)
Table 5.5: Cases with a stated etiological diagnosis from the sample of 500 enteric positive cases

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Number of cases (% of 500 cases)</th>
<th>% of diagnosed cases</th>
<th>Diagnostic test</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>34 (6.8)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Helminths</td>
<td>2 (0.4)</td>
<td>5.9</td>
<td>Morphology</td>
</tr>
<tr>
<td>Coccidia</td>
<td>2 (0.4)</td>
<td>5.9</td>
<td>Morphology</td>
</tr>
<tr>
<td>Bacterial overgrowth</td>
<td>7 (1.4)</td>
<td>20.6</td>
<td>Morphology</td>
</tr>
<tr>
<td>Campylobacter-type</td>
<td>4 (0.8)</td>
<td>11.6</td>
<td>Morphology</td>
</tr>
<tr>
<td>Canine parvovirus</td>
<td>10 (2.0)</td>
<td>29.4</td>
<td>ELISA</td>
</tr>
<tr>
<td>Giardia spp.</td>
<td>9 (1.8)</td>
<td>26.5</td>
<td>Morphology or ELISA</td>
</tr>
</tbody>
</table>
Figure 5.1: Flow diagram describing the laboratory diagnostics, percentages of cases with a diagnosis, and antihelmintic and antimicrobial management of enteric syndrome cases.
Table 5.6: Distribution of antimicrobials used by the sentinel veterinary practices in the
treatment of the 8441 enteric syndrome cases (n = 18,832) in 2007, 2008, 2009 and 2010

<table>
<thead>
<tr>
<th>Health Canada Category¹</th>
<th>Antibiotic class</th>
<th>Number of cases (% of all cases)</th>
<th>Percent of antimicrobial treated cases, % (n = 8441)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Third/fourth generation cephalosporins</td>
<td>138 (0.7)</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>Fluorquinolones</td>
<td>270 (1.4)</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>Nitroimidazoles</td>
<td>6820 (36.2)</td>
<td>80.8</td>
</tr>
<tr>
<td></td>
<td>Penicillin β – lactam inhibitors</td>
<td>344 (1.8)</td>
<td>4.1</td>
</tr>
<tr>
<td><strong>Category II</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>First/second generation cephalosporins</td>
<td>473 (2.5)</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>Lincosamides</td>
<td>95 (0.5)</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>Macrolides</td>
<td>176 (0.9)</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>Penicillins</td>
<td>1057 (5.6)</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>Timethoprim-Sulpha</td>
<td>109 (0.6)</td>
<td>1.3</td>
</tr>
<tr>
<td><strong>Category III</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Choramphenicol</td>
<td>6 (0.0)</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Sulphonamides</td>
<td>61 (0.3)</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>Tetracycline</td>
<td>10 (0.1)</td>
<td>0.1</td>
</tr>
</tbody>
</table>

1. Antibiotics classified by the Health Canada’s Categorization of Antimicrobial Drugs Based on Importance in Human Medicine (260).
Figure 5.2: Time series of the proportion of enteric cases treated with any antimicrobial, nitroimidazole class and penicillin class.
Figure 5.3: Scattergrams of the numbers of cases of enteric syndrome treated with B-lactam inhibitors and cephalosporins in each quarter of the years.
Chapter Six: Strategies for collaboration in the interdisciplinary field of emerging zoonotic diseases surveillance

6.1 Introduction

Cross-jurisdictional surveillance has been advocated to investigate sources of human illness (source attribution) of infectious enteric disease (78, 212, 279) and to better understand movements of antimicrobial resistant pathogens and resistance gene determinants (52). Planning and implementing such a surveillance system requires input from multiple stakeholders including clinicians, public health, public animal health, laboratories, information technology specialists and others. Analysis and communication are part of surveillance; the surveillance data needs to be shared in a form that can ensure its use by the stakeholders. To fully benefit from the companion animal surveillance system described in this thesis, will require working across disciplines and sectors, and ID collaboration.

The disciplinary approach has been very successful and has made countless important discoveries that benefit human and animal health. We have extended human and animal lifespans and improved the productivity in food producing animals (280, 281). Vaccines, improved drug therapies and advances in sanitation have all made important contributions to the control of many infectious diseases (282). Despite these successes, infectious diseases remain a leading cause of disease and death worldwide. Emerging, re-emerging and increasingly resistant pathogens cause large scale social and ecological costs even in countries with adequate sanitation, vaccines and drug therapy. Greaves (283) concludes that a “new medical paradigm, rooted in an older and more global framework” is needed. Collaborative research involving multiple disciplines and spanning from the cellular and molecular to ecology, climatology, sociology and the political sciences, has been nominated as the next research paradigm necessary to confront the challenge of emerging infectious diseases (284).

An integrated systems approach is also needed to investigate and manage the complexity of non-infectious diseases, environmental contamination and the impacts of global climate change (285). A systems approach, whether under the term One Health, EcoHealth or Global Health is
increasingly promoted as a solution to many of the world’s health problems. For example, in the document *Contributing to One World, One Health™* (OWOH) (176) the authors advocate building OWOH teams on a foundation of interdisciplinary research capacity.

Taking a complex-systems view of health requires new integrated surveillance capacities, research efforts and project management models that can gather and assimilate information from diverse sources and disciplines. Understanding how to build capacity among diverse disciplines for an effective and scientifically rigorous collaborative response to a health problem underlies the systems view and will be predicated on effective interdisciplinary (ID) collaboration. However, there has been little attention directed towards the study of how to best build ID collaborations that cross species and systems.

The objective of our study was to gain an overview of the breadth, nature and status of research in ID collaboration by reviewing the literature from a number of sources and disciplines. Using a non-systematic review of the literature, a search of relevant websites and expert consultation we aimed to better understand the essential elements for ID collaboration as well as to identify gaps in the current research. We examined ID collaboration in health partnerships, health research knowledge transfer and exchange (KTE), business management and systems design engineering to identify the necessary features of effective ID collaboration that can support initiatives like the OWOH movement. A preliminary literature review indicated that these four areas had a significant amount of published experience in ID work. We highlight successes, failures, and perceived keys to effective implementation of inter-organizational learning. Structures, methods and outcomes for ID collaboration needed to strengthen linkages and information integration across human, animal and environmental health systems are nominated.

6.2 Methods

We undertook a scoping survey focused on identifying common themes, emerging issues, situations and potential pitfalls that may affect ID collaboration. Information was gathered by a
survey of peer-reviewed literature, by scanning the websites of relevant organizations and with a workshop involving people with professed experience in ID collaborations at the human-animal-environmental interface.

6.2.1 Literature review

Textbooks were initially selected to provide an overview of the subject area, to assist in refining the literature search strategy and to identify important resources from the reference sections.

Based on the preliminary review of the literature, a search of the University of Calgary’s Library Catalogue, PubMed, and Google Scholar was conducted using the following search terms: one medicine; one health; knowledge management; systems design engineering; knowledge transfer; knowledge exchange; multi-, inter-, transdisciplinarity; interdisciplinary health research; health partnerships; collaborations; teamwork; evidence based decision making; and results-based management. The search terms were used individually or with the Boolean logical operator ‘OR’. Articles from the literature search were diverse in their focus and design. The search results numbered in the many thousands and were not specific to the objectives of our study. The search was therefore modified with the key words: AND effective, veterinary, environmental OR ecosystem. For example the search term ‘knowledge transfer OR knowledge exchange resulted in a list of 6541 articles. Adding the term ‘AND effective’ resulted in a list of 438 titles that were of greater interest to this project. Achieving greater precision with this search of the descriptive literature was not feasible.

The criteria for inclusion were based on the objectives of the study and determined by the author and two collaborators with experience in animal health and public health surveillance (C. Stephen and R. Copes). The author was responsible for retrieving and reviewing the articles. No attempt was made to assess the quality of the papers. English language titles and abstracts retrieved were read and articles were selected if they met the following criteria: (i) had a holistic or interdisciplinary approach to business, systems design engineering, environmental, animal or
human health; (ii) described the participants’ experiences in ID collaboration or KTE; (iii) reviewed the factors that enabled or impeded ID collaboration or KTE; (iv) tried to understand the mechanisms behind effective or ineffective ID collaboration or KTE; or (v) included an evaluation of the efficiency or economic justification for ID collaboration or KTE. The findings were synthesized in a narrative form which was provided to the workshop participants described below.

Information was collected on the World Wide Web using the Google search engine with the search limited to pages from Canada using the following search terms: one medicine; one health; interdisciplinary health research AND veterinary OR ecosystem OR environment; collaborative health research AND veterinary OR ecosystem OR environment; population health AND collaboration OR veterinary OR environment health; knowledge exchange AND veterinary OR environment OR ecosystem. The results pages were searched until additional unique websites could not be identified. Websites for academic institutions, research networks, funding institutions, government agencies and professional organizations were scanned for activity in the fields of interdisciplinary research or collaboration. Journals, newspaper articles and private industry websites were not reviewed. The selection of websites was restricted to Canada because the national context could influence programs and activities and thus the experiences of experts working across species boundaries.

6.2.2 Expert consultation

In June 2009, 16 people (15 Canadian, 1 American) with documented experience in ID collaboration were invited to a workshop. A list of possible participants was developed from the literature, our involvement in past collaborations and professional networks. The candidates were asked to volunteer and if they were unable to participate were asked to nominate a replacement. There were five physicians, nine veterinarians and two biologists, working either exclusively or with multiple roles, in private practice, academia, and provincial and federal agencies. The objectives of the workshop were to: (i) have the participants share their experiences in and perspectives of ID collaboration; (ii) find the common barriers and enablers of ID collaboration;
and (iii) make recommendations to improve the structure, methods and outcome for ID collaboration. The intent of the expert consultations was to reach a consensus on what the participants identified as the key issues and themes for effective ID collaboration.

The workshop was conducted online using an internet collaboration tool (WebEx™, Cisco Systems Inc, San Jose, California). Computer mediated communication was selected for its efficiency and increased opportunity for participation. Research has shown that face-to-face communication to be more effective in building trust and relationships in a team which can result in improved outcomes (286). However, the use of digital technologies is increasing because effective communication and respect for the ability of the team members have been shown to result in a high quality performance in the absence of trusting relationships (287).

The workshop was organized so that each participant was involved in three different meetings held over 2 weeks in June 2009. Participants were asked to read the literature review on ID collaboration described above and consider whether these findings reflected their personal experience. The entire group discussed this in an online meeting room. They were then divided into 4 small multi-disciplinary groups and given an imaginary but plausible scenario (Table 6.1). Each small group then met together online. Participants were asked to bring their skills and expertise to the discussion and to approach the scenario with an ID perspective to determine what they viewed as the fundamental core attributes of ID collaboration. In the final meeting the whole group was asked to review a summary of the workshop discussions and comment on the findings. The meetings were all recorded and then transcribed.

6.3 Results
6.3.1 Literature review

There is significant attention paid to the terminology around knowledge and interdisciplinarity in the literature. A discipline is defined as a field of study. The professions of medicine, veterinary medicine and the environmental sciences are interdisciplinary at their core. For clarity, in the context of emerging infectious diseases and for this paper, interdisciplinarity
encompasses collaboration between the professions of medicine, veterinary medicine and the environmental scientists.

No references were found that specifically looked at ID collaboration between veterinary medicine, the environmental sciences and/or medicine. The search terms one medicine and one health were successful in providing titles and many of the references did promote veterinary and human medicine collaborations. Few documented any successes or positive outcomes from interdisciplinary collaborations. There were some notable exceptions such as Schelling (288) who described a combined vaccination program for nomadic pastoralists and their livestock in Chad. However, there were no references for ID projects involving veterinary medicine or the environmental sciences that described the experiences of team members, gave an evaluation of the processes involved in ID cooperation or identified best practices for collaboration. Therefore, references involving veterinary medicine and the environmental sciences were not included in our results.

The other search terms did yield titles on how best to achieve ID collaboration. Health partnerships, KTE, business management and systems design engineering are diverse subjects but they all focused on communication among people of different fields. The Canadian Institute for Health Research (CIHR), for example, has programs designed to support research networks and facilitate the active involvement of a number of disciplines in joint research activity (Canadian Academy of Health Sciences, 2005). These funding initiatives promote the “bench to bedside” approach to health research. Articles addressed the question of how universities can best implement, support and evaluate ID health research (289-292). Studies explored the best practices of a health care team of physicians, nurses, pharmacists, physiotherapists and others, with the goal of improving services in the health care setting (293, 294). Collaboration among health and welfare service providers were described in community level health service publications (295, 296). While all of the health publications portrayed these collaborations as valuable and did identify the features of a successful partnership, few provided information on how to achieve a successful collaboration. An exception was Lavis et al (297) who clearly
described the methods for the effective communication of research knowledge (KTE) to policy makers.

Business and systems engineering were selected as models to review because of their focus on knowledge use and integrative approaches. In the business management literature, the exploitation of knowledge assets was termed knowledge management. The business literature has many publications that focused on the ‘how’ of developing stable KTE networks for effective intra- and inter-organizational collaboration, as measured by improved profitability (298-300). The knowledge management literature provided information on how partnerships across organizations affect a business’s effectiveness, profitability and outcomes.

Systems engineering aims to provide holistic solutions to complex problems that transcend the traditional boundaries of engineering. Using the framework of systems theory, the systems engineer will view a problem as composed of systems that interact. These can be ecosystems, physiological systems, communication systems, transportation systems and so on (301). Systems engineers implement design processes to develop a new system or redesign an existing system to manage the problem. This requires an ID team effort throughout the development process involving complex management challenges. The system engineering literature discusses how to utilize the capabilities of the ID team and the need to measure its success by meeting the goals established for the new system (302).

Our review of the health, KTE, business and systems engineering literature found shared rewards, challenges, facilitators, and barriers to successful collaborations, regardless of the setting. The results of the literature review were grouped into 5 themes: people, the task, workplace and resources, information technology (IT) and evaluation of the results. Figure 6.1 summarizes the shared elements for effective ID collaboration.

There was general agreement among the experts consulted on the importance of all of the essential elements described in Figure 6.1 except for the use of formal agreements among
organizations. The conclusions of the expert consultations mirrored the findings of the literature search and both findings are discussed together.

6.3.2 People

Workshop participants and the literature emphasized the importance of highly qualified and passionate collaboration team members. Competent disciplinarians have strong standards of scholarship and large bodies of theoretical knowledge. Combined with broad interests and imagination they are better able to navigate the complex process of ID collaboration (303-305). But finding competent researchers experienced in interdisciplinary teamwork is a challenge (306).

Many post-secondary institutions have explored the question of how to develop interdisciplinarians (304, 307). Most post-secondary programs focus on developing excellent disciplinarians with an appreciation for life-long learning (303). Other institutions advocate problem-based teaching to encourage interdisciplinarity and allow students to examine complex problems (308, 309). ‘On the job’ experience was also a key contributor to creating ID personnel. All of the workshop participants had become interdisciplinarians through the opportunity to work on collaborative projects with experienced members providing mentorship and leadership. Many of the participants reported that participation in an ID project resulted in enthusiasm for participating in future projects. They had enjoyed the collaboration and also wished to reciprocate for past assistance. The first collaboration led to more collaborations and an ever expanding network of contacts. The experienced collaborator was also able to bring his/her previous relationships to the task which expanded the social networks of the less practiced participants.

In addition to the ‘hard’ skills of technical proficiency and sound professional knowledge, the interdisciplinarian must show aptitude for the ‘soft’ skills. Personal relationships are critical to knowledge creation and knowledge transfer. Within a diverse group, a culture of trust, reciprocity, and respect is required to enable effective collaboration. The major attributes of a
successful collaborative team are a willingness and ability to contribute, effective communication, commitment, readiness to seek consensus, compatibility, and an enthusiasm for sharing the credit (294, 296, 310-313). The workshop participants were unanimous in their belief that face to face meetings were essential to developing effective trusting relationships. Technology was important for facilitating ongoing communication and could help enhance relationships but at least initially, an opportunity to meet in person was inimitable.

The participants acknowledged that finding collaborators is difficult for a scientist that has not yet developed an interdisciplinary network. Workshop participants often meet potential collaborators through introductions by other colleagues or by understanding the structure of the agency or department with which they wished to work and then finding the appropriate person. Meetings and conferences with a social component involving food and drink were cited as useful for meeting people. For those who work in more isolation, communities of practice were valuable for first connections. A community of practice was defined as a social network that facilitates information exchange between individuals with the goal of improving the performance of their jobs. Some examples provided were Animal Determinants of Emerging Diseases Research Unit (314), Canadian Cooperative for Wildlife Health Centre (315), Veterinary Information Network (316), ProMED mail (317), professional organizations, and other networks and email distributions lists. The challenge was finding collaborators outside a discipline’s community of practice.

The importance of relationships was demonstrated by the experiences of the car manufacturing industry. To create a strong social network between firms, Toyota has had consultants and personnel work at suppliers’ locations for extended periods. Toyota personnel came to know what knowledge will be useful to the supplier, whom to contact, and where the learning capacity resided at the supplier. In contrast General Motors (GM) had not developed a stable network of supplier companies nor developed overlapping knowledge bases, social interactions or regular knowledge exchange networks. GM held a proprietary view of its
knowledge and as a result, suppliers viewed GM as less trustworthy. GM’s transaction costs were six times higher than those of Toyota (318, 319).

6.3.3 Task

The workshop participants felt that many complex tasks can benefit from ID collaboration. They stressed that the task needed to provoke passion and enthusiasm among the collaborators to be successful. They generally believed that the best ID collaborations were project driven and as a result their most successful collaborations had been reactive to a pre-existing issue. The business and systems design engineering literature emphasized the need for a shared vision or a defined problem and goal as being critical for effective ID collaboration. Systems design engineering inherently represents a target to be reached or a vision to be materialized (320).

Two examples illustrate this point. Letier (302) described goal-orientated design engineering using the example of the London Ambulance Service (LAS). The goal was that the ambulance must arrive at the incident site within 14 minutes. This goal provided a clear sense of direction and a commitment to the vision for the team members but did not represent an order or an instruction (321). Identifying the goals then naturally led to the repeated asking of “why”, “how” and “how else” (320). Characterizing the interaction of the LAS with its environment allowed the team to fundamentally change the functional and business practices of the LAS and thus meet the stated goal (302, 320).

The emergence of Cryptococcus neoformans var gattii (C. gattii) on Vancouver Island in 1999 was cited by the workshop participants as an excellent example of a successful interdisciplinary partnership that employed a systems approach to health. Its success was in large part because the problem stimulated passion and interest among a wide variety of disciplines. By investigating this emerging pathogen on an ecological scale, collaborating and sharing results, the researchers could gain an understanding about C. gattii on Vancouver Island as a dynamic, complex and large scale system (322-328).
6.3.4 Resources and workplace

Opportunities for ID collaboration now exist within and between academia, government and industry, but they are not without challenges that are created by institutional and workplace barriers (290).

Within academia, disciplines are divided physically and conceptually and it is difficult to bring down these walls (290). These divisions in space are reinforced by faculty positions, timetabling, curriculum obligations, faculty workload and budgets. The disciplinary divisions extend to the libraries as well which causes difficulties for researchers requiring information outside their specialization (329). Academic success depends on credit for grants, research, and publications in respected journals. Interdisciplinary collaborations can hinder a researcher’s academic success because ID research tends to take longer, credit for a grant may go to a collaborator (within or outside the university), the researcher may not receive adequate credit for an article with a long list of collaborators, and it may be difficult to find a journal interested in publishing an ID article at all (291, 304, 330). Funding sources including funding agencies, government and industry, favor pure disciplinary research. The contribution to ID research with a systems focus is small relative to the total investment in the Canadian health funding setting. The funding decisions of the Canadian Institute for Health Research were searched for those projects that the investigators had described as interdisciplinary or collaborative (331). The titles of the projects were scanned to identify those that had a cross-species or an environmental component. This search found that less than $3.8 million of $771 million (0.49%) in research funding was awarded to these types of interdisciplinary projects (Table 6.2). Those responsible for the public purse want to maximize the use of limited resources and they are not inclined to fund complex and lengthy ID collaboration projects (291, 304, 330).

The need for clear tasks, outcomes, milestones and timelines found within government and industry cannot be met with a complex interdisciplinary question. Legislation will define the extent of authority for government departments and these jurisdictions can be barriers to ID collaboration (304, 330, 332, 333). Industry is constrained by its need to generate a profit and
meet shareholders expectations. Corporations’ policies and directions may not reflect public concerns (330). Ownership of the intellectual and technological property when the collaboration crosses academic, industry, and/or government lines can be problematic (304).

Formality in structuring relationships (such as memoranda of understanding) was a strategy used by some of the workshop participants, especially those in government, to manage jurisdictional issues. In contrast, the private practitioners and those who worked in smaller agencies were wary of too much formality and the cumbersome reporting requirements they felt would result. They did recognize however that formality in relationships among several organizations did increase the credibility of the collaboration with government and funding agencies. The general feeling was that although a formal structure may enhance collaboration, it is not sufficient to motivate or sustain interdisciplinary work.

Participants at the workshop also concluded that passion and leadership could overcome most institutional and resource deficiencies, particularly in the short term. Projects that extend over a longer time period were more likely to founder with poor financial and institutional support.

6.3.5 Information technology

Throughout the literature on knowledge management in business, engineering, and health, information technology (IT) is regarded as an enabler of knowledge management. It can help bridge distances and build social capital but it cannot create knowledge (334-336). A systems approach to health requires information integration across systems to develop a network as opposed to a platform (312). Data on animal movement, uses and health or data on environmental determinants of health can potentially be rich sources of data for public health, but there were few published examples of the routine integration of diverse data sources into a single, coherent, collaborative information system. Data sharing often relies on informal agreements and personal relationships between agencies. For example, an unpublished report to the Canadian Food Inspection Agency (personal communication, S. Pollock et al., 2008) found
that a large amount of animal data is collected in Canada but it is typically collected and used for specific purposes and there is little sharing, integration or secondary use of the data.

Knowledge is a fluid mix of framed experiences, values, contextual information and expert insight that provides a framework for evaluating and incorporating new experiences and information (335). Technology will not replace the time and effort required of the human element in knowledge management but adequate technological support is critical for any organization to run smoothly, therefore a healthy IT network must be maintained (334-336).

Unmanaged knowledge networks become data dumps and that can limit their usefulness. Both information and knowledge can revert to data if not effectively managed and retrievable (298, 335). Institutions require data managers, often with training in the library sciences, to maintain their knowledge base so it remains current and relevant for their needs (321). Data bases such as the Cochrane Collection and the Canadian Institute for Health Information (CIHI) are examples of resources for assembling and disseminating evidence and such data bases are expanding.

6.3.6 Evaluation

Research to explore the effectiveness of ID work on the health of people, animal, and environments was not found in the literature but evaluation is necessary so that we can move beyond the belief that ID collaborations are simply a ‘good thing’ (295) to demonstrating that interdisciplinarity results in improved health outcomes. Most of the published research into partnerships and ID collaborations was focused on the experiences of the individual collaborators but there was a dearth of publications that provide the evidence of improved outcomes arising from such collaborations (295, 296, 303). A review of health and social care partnerships (296) found only 5 of the 36 papers reviewed had addressed outcome success determined by such measures as improved accessibility, efficiency, or health status of the people using the services. However, the review authors believed these 5 studies were unable to “establish a clear causal link between the partnership and its outcomes”.
During the workshop discussions, the participants divulged that their projects did not always meet their stated objectives but the participants still felt they were successful because of the development of new relationships, learning who would be useful contacts for future projects and an enhanced community of practice. However, they noted that governments and funding agencies will require additional, objective evidence of ID collaborations’ value and utility.

The problems confronting research into evaluating ID collaborations have been well described by Ansari et al. (295). One possible approach to the resolution of the measurement issue is to adapt a systematic evaluation framework. An example of such a framework is the Results-based Management and Accountability Framework (RMAF). This management framework for the Canadian government requires that public service managers will define strategic outcomes, focus on results, measure performance, and adjust policies, programs, and initiatives so that ongoing improvements can be made (337). The RMAF frequently uses such terms as timely, accurate, concise, clear goals, relevant and so on. While these are desirable adjectives for spending tax dollars, ID projects will not easily meet these requirements and this illustrates the challenge of evaluating ID collaborations. Existing evaluation tools will need to be adapted and may need to consider other relevant outcomes including change in process, awareness, knowledge, outcome or attitude within funding organizations academic research facilities, government decision maker and industry.

New approaches to evaluation have also been nominated in the literature. For example, in the 2005 joint report, *Building a Better Delivery System*, the Institute of Medicine and the National Academy of Engineering called for the two professions to work together to establish a systems framework in healthcare (338). Engineers use a systematic approach to the analysis, control and design of complex systems and they have been used effectively in the manufacturing and service industries. While there has been demonstrated utility in the health care system since the early 1990’s, there are limited applications reported in the literature (339-341).
6.3.7 Implementation of ID projects

The business management and systems engineering literature summarize the implementation of knowledge management or systems design as a series of steps or tasks (335, 342, 343). The concepts and themes in the steps are similar between these two disciplines, were consistent with feedback we received from workshop participants and are summarized in Table 6.3.

6.4 Discussion

There is a growing demand, both in Canada and internationally, for systems-based ID approaches to health research and management (176, 344). It is hypothesized that to manage complex health issues, there must be effective collaboration and exchange of knowledge among a diverse group of practitioners. However, in the context of One Health and emerging infectious diseases there is a dearth of information on the strategies to put ID collaboration into practice and no evidence of its effectiveness. We found that passionate individuals, with sound disciplinary knowledge, willingly assembled to solve a shared problem, are at the foundation of any actions taken to fulfill this demand. One of the objectives of this project was to identify the essential components of an interdisciplinary approach that could support initiatives like the One Health movement. In general, we found that: (i) professional social networks that provide informal and formal connections across disciplines need to be encouraged and supported; (ii) passionate ID leaders and advocates need to be cultivated and supported; (iii) there must be a focus on building a culture of trust and respect among disciplines such that when teams are assembled, there is willingness and confidence in each other to share information; (iv) ID teams need to have shared problems and visions for the tasks at hand; and (v) processes to allow for collaborative work in formal and informal settings need to be worked out in advance.

Education and experiential learning opportunities remain our best tools to develop the interdisciplinarian at this time. Educational institutions across the country train quality disciplinarians and embed in them the importance of lifelong learning. But there remains a diversity of opinions on the best way to instill the qualities necessary of team players and
leaders. Business schools have long recognized the value of teaching and rewarding leadership for the workplace and their experience could guide future leadership training in the health fields.

A ponderous institution is an obstacle to ID collaboration but the boundaries between disciplines and agencies continue to be broken down. This is achieved by the shared enthusiasm and commitment of the people willing to find workable solutions within the structure of the university, government agency or industry. Competent individuals, personal relationships and a task that incites enthusiasm among the participants were deemed most important for successful partnerships.

The authors believe that there is a potential role for a centralized agency, either government or not-for-profit, to provide the introductions and enable the face to face meetings of individuals from different disciplinary communities to create the professional social networks that we found to be instrumental in forging ID teams. Communities of practice have developed to meet the interests and needs of their members. They are a natural and spontaneous response to the necessity of sharing knowledge and solving complex problems (345). But the challenge is for the individual to make a connection outside the boundaries of their community of practice. Businesses wishing to access new resources and skills will appoint knowledge managers or directors of strategic alliances to find those with the required knowledge (335, 342, 346). Similarly, Lavis et al. (297) advocates the use of knowledge brokers to transfer research knowledge to policymakers. We suggest that collaboration across species and jurisdictional boundaries could also be facilitated by a third party. The issues and projects that become the focus of an interdisciplinary collaboration may still be mostly regional as will be the team created to deal with them. Regional knowledge brokers could play an important role in connecting regional players. A central knowledge broker could work with regional networks to develop an interdisciplinary team from a broader geographical area for problems of a national scale.

The final factor for successful collaboration is a supportive environment. Structural and cultural changes within an organization take time to develop and will evolve naturally if ID
collaboration is shown to be advantageous to that organization. We found similar experiences and views of the important elements to ID collaboration but there remains the major issue of how to document if the ID approach indeed resulted in more effective and efficient outcomes. We found no reports of instances where authors reported on the failure of ID collaboration; most likely reflecting a bias in submission of reports.

Our workshop participants attracted those people who were supportive in principle of ID collaboration and as volunteers represent a selection bias. As well, most of the participants were Canadian and only the websites of Canadian institutions were included in a search of the gray literature. However, the fact that the findings of the workshop and the gray literature reflected the findings in the literature suggests that the bias had little impact on the project and these findings can be generalized beyond the Canadian context.

Interdisciplinary collaborations involve a wide spectrum of disciplines with intersecting components. Developing standardized methods for their quality assessment will be challenging. Regardless of the measurement method, creation of a framework of ID evaluation to provide objective evidence of improved outcomes is essential and an obligation when spending public funds. Organizations and agencies wishing to develop ID capacity must be able to defend the decisions being made in ID development. Investing in research and engaging academic disciplinarians, communities and decision makers on how to evaluate the contributions of ID collaboration and developing ways to recognize success outside of the usual venues of publication must become a priority for institutions promoting ID programs.

6.5 Conclusion

A companion animal surveillance system that has the objective of providing an early warning of environmental risks to people will be more effective in meeting its objective if its developers collaborate across jurisdictional boundaries during its planning and implementation. It requires a skilled and committed public health officer who can articulate public health’s needs and who is receptive to the potential role of pet surveillance in meeting those needs. Further, the purpose of
the surveillance system must interest and inspire all of the stakeholders so they remain committed throughout the process. Interdisciplinary collaboration will help ensure that the surveillance findings will be useful in guiding changes in policy, practice or procedure.
Table 6.1: Scenarios presented to the workshop participants to focus a discussion on the essential elements of interdisciplinary collaboration for emerging and zoonotic diseases.

<table>
<thead>
<tr>
<th>Group</th>
<th>Topic</th>
<th>Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Food Animal</td>
<td>Following a water-borne outbreak of <em>E. coli</em>, there is public pressure for vaccinating cattle with an <em>E. coli</em> vaccine.</td>
</tr>
<tr>
<td>2</td>
<td>Companion Animal</td>
<td>There has been a human case of Leptospirosis. There was no travel history and their pet dogs were unvaccinated and seropositive for <em>Leptospira</em> sp.</td>
</tr>
<tr>
<td>3</td>
<td>Environmental</td>
<td>Dead ducks are found in small pond on an empty lot in a residential area where building materials have accumulated over the winter.</td>
</tr>
<tr>
<td>4</td>
<td>Wildlife</td>
<td>A northern Alberta native community are finding “tumors” in their country food. They are uncertain if the food is safe to eat.</td>
</tr>
</tbody>
</table>
Figure 6.1: The elements of effective ID collaboration and KTE
Table 6.2: Distribution of $771 million in Canadian Institute for Health Research (CIHR) awards and grants for interdisciplinary collaborations in 2008/2009 (331).

<table>
<thead>
<tr>
<th></th>
<th>Project described as ID or collaborative</th>
<th>Identified as having a cross species or environmental component</th>
</tr>
</thead>
<tbody>
<tr>
<td># Awards or grants</td>
<td>733</td>
<td>7</td>
</tr>
<tr>
<td>Value</td>
<td>$62,796,523</td>
<td>$3,792,790</td>
</tr>
<tr>
<td>Percent of total</td>
<td>8.1%</td>
<td>0.49%</td>
</tr>
</tbody>
</table>
Table 6.3: Summary of the steps to effective interdisciplinary collaboration as described in the business management literature (335) and the systems engineering literature (343).

<table>
<thead>
<tr>
<th>Step</th>
<th>Knowledge Management</th>
<th>Systems Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Start with a vision or a goal.</td>
<td>Formulation and analysis of the problem.</td>
</tr>
<tr>
<td>2.</td>
<td>Identify advocates of the vision.</td>
<td>Define what each design team does.</td>
</tr>
<tr>
<td>3.</td>
<td>Commit resources.</td>
<td>Define the overall development of project.</td>
</tr>
<tr>
<td>4.</td>
<td>Manage content.</td>
<td>The development process.</td>
</tr>
<tr>
<td></td>
<td>- Identify owners of knowledge and define their roles.</td>
<td>- Is organized around information transmission, reception, and utilization (347).</td>
</tr>
<tr>
<td></td>
<td>- Keep knowledge updated.</td>
<td>- Develop possible solutions and alternative hypothesis.</td>
</tr>
<tr>
<td></td>
<td>- Ensure knowledge is relevant and high quality.</td>
<td>- Analyze and evaluate the alternatives.</td>
</tr>
<tr>
<td></td>
<td>- Assign knowledge managers that have technical and professional skills as well as a sense of the cultural and political aspects of the knowledge applications.</td>
<td>- Select and model one alternative.</td>
</tr>
<tr>
<td></td>
<td>- Identify and define their roles.</td>
<td>- Implement and integrate the option in detailed design.</td>
</tr>
<tr>
<td></td>
<td>- Keep knowledge updated.</td>
<td>- Launch the system.</td>
</tr>
<tr>
<td>5.</td>
<td>Develop communities.</td>
<td>Develop teams.</td>
</tr>
<tr>
<td></td>
<td>- Develop a network of people, not a storehouse of information.</td>
<td>- Define how individual work stations are integrated into a suite that represents the whole system or operational architecture,</td>
</tr>
<tr>
<td></td>
<td>- Develop project leaders as coaches.</td>
<td>- Focus on people and knowledge behavior.</td>
</tr>
<tr>
<td></td>
<td>- Members must have a strong bond.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Trust, respect, common language, and a flattened company hierarchy all enable knowledge management.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Enhance richness of communication with face to face meetings.</td>
<td></td>
</tr>
<tr>
<td>Step</td>
<td>Knowledge Management</td>
<td>Systems Engineering</td>
</tr>
<tr>
<td>------</td>
<td>----------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>6.</td>
<td>Develop computer capabilities to store, transfer, and manage the knowledge.</td>
<td>Information technology is required to enable the process.</td>
</tr>
<tr>
<td>7.</td>
<td>Reward knowledge sharing.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Knowledge will not move without friction or motivation. It is seen as power and the owner may be concerned about what they may lose if they give it away. Knowledge sharers must be recognized and rewarded but not penalized for failures (348).</td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Use the experience of lessons learned to move forward.</td>
<td>Re-evaluate.</td>
</tr>
<tr>
<td>10.</td>
<td>Management</td>
<td>Management</td>
</tr>
<tr>
<td></td>
<td>• Knowledge management requires a cultural change from the individual to the collective to support a business need. It must be everyone’s responsibility but not added to existing tasks. Assign a knowledge manager to manage the company’s knowledge assets.</td>
<td>• Have an overall manager to utilize the capabilities of the contributors.</td>
</tr>
</tbody>
</table>
Chapter Seven: Conclusion

7.1 Overview

The early detection of environmental hazards for people, both infectious and non-infectious and the known and the unknown, is a challenge for public health officials (349). There has been a growing interest in the ability of companion animal disease surveillance to provide an early warning for human health (4, 206, 350). Animal health can benefit from collecting data on conditions encountered in clinical veterinary practice by reviewing and critically evaluating outcomes to provide relevant evidence for clinical decision making (351, 352). Currently, there are few reportable companion animal diseases and no routine collection of companion animal disease data in Canada. The objective of my thesis was to plan and implement a pilot companion animal surveillance system and evaluate its usefulness as an early warning for human disease and to inform evidence-based companion animal practice.

In Chapters 2 and 3, I demonstrated that surveillance using a network of general practice veterinarians was functional and acceptable to the practitioners. I presented the results of my research using a novel approach to data collection for companion animal surveillance by applying methods developed in human informatics and surveillance. There are no published examples of a veterinary surveillance system that used the automatic extraction of the electronic medical records (EMR) from the veterinary practice management system (VPMS) at sentinel veterinary practices and applied text-mining methodologies to identify and collect clinical cases of interest.

In Chapters 4 and 5, I analysed the surveillance data to describe patterns in my disease model (enteric syndrome) and looked for evidence of changes in the patterns. When possible, I interpreted the results of the analyses within the context of the clinical cases to better inform an appropriate response by public or animal health officials. Through my research presented in the these chapters I contributed to the understanding of the uses and limitations of unstructured, clinical data for companion animal disease surveillance, the technical limitations of my selected surveillance tools and the challenges posed by the human dimension of surveillance.
By definition, disease surveillance requires that if the data indicate that disease events have risen above an established threshold or are clustered in space and/or time that these findings will be communicated to those who need to know so that an action may be taken (3). If the purpose of the animal disease surveillance system is an early warning for public health, then the planning, implementation and communication of the surveillance results requires interdisciplinary (ID) collaboration. In Chapter 6, I describe the elements essential for successful ID collaboration. My research contributed to this field by applying the knowledge from the health, business and systems design engineering literature and a cross-sectoral workshop of experienced interdisciplinarians to better understand the ‘how’ of ID collaboration.

Surveillance systems need to be evaluated to ensure that they are meeting their stated purpose efficiently and effectively (177). Using the evaluation framework introduced in Chapter 2 and described by Buckeridge et al. (179) the attributes, limitations and potential benefits of this system will be described. Areas of future research that may improve its performance will be discussed.

### 7.2 Information quality

Information quality is a measure of the fitness for use of the information to the user (188). It includes desired characteristics such as representativeness, accuracy, completeness and timeliness (188). The type of surveillance system will have an impact of the quality of the information collected. The surveillance system described in this thesis was a sentinel practice-based system which are vulnerable to volunteer and case ascertainment bias (150, 177). Stärk and Nevel (353) described the selective participation and economic considerations that introduced bias into health data collected from a swine sentinel practice-based surveillance system. To be eligible for inclusion in this project, the clinics were required to have completely computerized medical records and be using the same VPMS. These conditions could have further contributed to the selection bias. Therefore, the cases of enteric syndrome seen by this surveillance system may not have been representative of the underlying population (see Chapter 2).
Early in the research project it was apparent that the medical records contained few etiological diagnoses for the cases of infectious enteric disease. This guided the decision to use a syndromic case definition - enteric syndrome (see Chapter 2). The warehoused medical records were my only source of information with which to describe the cases of enteric syndrome. Reports for outside laboratories and specialists were often not accessible (see Chapter 2). Without those report, I had to rely on the record of the veterinarian’s order for the laboratory procedure and documentation of the results in the medical record. I discovered this information was inconsistently recorded (see Chapter 5).

The demographic data in the medical records (species, age, sex, breed) and the data for the geographic location (the first 3 digits of the owner’s home postal code) were largely complete. Using this data it was possible to identify clusters of enteric disease in space and time, observe patterns of antimicrobial use and describe these patterns by the host factors (see Chapters 4 and 5). How accurately the true state of an animal’s health is reflected in the medical record will depend upon the veterinarian’s skill and experience, the willingness and financial capacity of the owner to pursue a precise diagnosis, the availability of reliable diagnostic tests and how completely procedures and results are documented (198, 354-356). There was often not sufficient specific information recorded in the medical record to describe the clusters by the immunity level in the population (the proportion of vaccinated cases), by the agent identified as a probable cause for the clinical signs of enteric syndrome or by the environmental factors that may be predictive of increased risk. Even clusters etiologically defined as canine parvovirus, *Giardia spp.* or others had inadequate detail to interpret the relationship between the disease and the time and place that a statistical increase in disease frequency had been documented. The unstructured data used in this surveillance system was lacking detail so that I could not understand the context of the signals. This constraint limited the usefulness of the system.

Informatics does provide the opportunity of including more than one data stream in a surveillance system. The addition of data from the veterinary laboratories would provide etiologically specific data from all of the veterinarians they service and this would serve to
improve the data quality. The feasibility of integrating data from the veterinary diagnostic laboratories in the Calgary area has not been assessed.

I described the management of cases including the veterinarians’ antihelmintic and antimicrobial prescribing behaviour (see Chapter 5). The clinical events and the veterinarian’s therapeutic reasoning and approach were not consistently recorded so the patient’s ‘story’ was not always clear. Determining if antimicrobial therapy is appropriate is best assessed when an etiological diagnosis has been established. In the absence of a diagnosis, the decision to dispense antimicrobials is subject to the veterinarian’s judgement. I found that I could not evaluate whether antibiotic use was appropriate in these situations due to a lack of contextual information recorded in the medical records. Therefore it was not possible to determine why an antimicrobial had been selected, and if the decision to prescribe an antibiotic was appropriate and aligned with guidelines on the prudent use of antimicrobials (357). I also found that the guidelines for appropriate use were subject to interpretation. One of the specific principles in the Canadian Veterinary Medical Association’s guidelines on the prudent use of antimicrobials is, “Antimicrobials should only be used therapeutically if a pathogen is demonstrated or anticipated to be present, based on clinical signs, history, necropsy examinations, laboratory data (including resistance testing), and if the pathogen is expected to respond to treatment” (357). To make any judgement about whether AMU was appropriate, clear definitions that define ‘appropriate use’ under different clinical situations would be necessary.

Timeliness is measured in intervals from the disease exposure, through data collection, analysis, interpretation and the intervention by public or public animal health (141). The timeliness of this system was not measured but the impacts that some of the system characteristics may have on timeliness can be considered. The automated retrieval of EMR from the VPMS and the subsequent text-mining to identify the cases of interest enhanced the timeliness of the system (see Chapters 2 and 3). Managing this large and complex data set (see Chapters 2 and 3) reduced its timeliness as did searching the records for all of the data relevant to the management of a case (see Chapters 4 and 5).
In conclusion, the information provided by this surveillance system was of sufficient quality to detect patterns or monitor trends but was often insufficient to understand the underlying epidemiology or the justification for a veterinarian’s therapeutic decisions.

7.3 System quality

System quality are measures of the information collection and processing system itself (188). These include reliability, flexibility and ease of use. The University of Calgary Data Extraction Program (UCDEP) developed in my research, used queries written against the SQL server on the VPMS as the data extraction method (see Chapter 2). I demonstrated that this approach was capable of completely and accurately extracting the available medical records. However, the SQL server is not consistent between database products and in the interest of limiting information technology costs and the time spent in development, it was necessary to restrict the participating practices to those all using the same veterinary management software product. Furthermore, different versions or different templates within the same software could render the data extraction program ineffective. The UCDEP was programmed to transfer the daily, prospective data extractions by email; a method frequently hampered by the email servers. Future research in this field should explore alternatives to accessing and processing the data in electronic medical records. Another companion animal surveillance system currently in development is using an XML schema to select and extract data of interest and is showing promise in overcoming some of the limitations of my approach (154). Savel and Foldy (156) emphasize the need for technology standards in order for information systems to communicate with one another and to improve the data transfer process.

The UCDEP did not extract the fields in the veterinary management software that contained confidential information. However, the names of the owners, pets, veterinarians and staff could be found within the text of the medical record. This is a serious breach of the confidentiality of the medical record and limits access to the records to only those researchers having signed confidentiality agreements with the participating practices. There has been progress in managing
this issue in human informatics (195, 196) and further research is required to investigate these methods for veterinary informatics.

The time and effort of developing the text-mining dictionary were compensated by the speed and efficiency with which it could scan the database and extract the cases of enteric syndrome. Text-mining consistently identified and collected cases of enteric syndrome with an accuracy approaching that of the human reviewers (see Chapter 3). In Chapter 5 I also used it as a tool to describe the cases and it performed effectively if the language used by the veterinarian was clear and unambiguous such as orders to administer or dispense antihelmintics or antimicrobials. It was not useful for selecting the enteric syndrome cases that had a laboratory confirmed diagnosis. The meaning of words such as ‘parvo’ was dependent upon the context in which they were used. This may not be a consistent finding if applied to another case definition. For example the record for a case of bacterial cystitis would not be confused with a laboratory test or a vaccine of the same name.

This system’s quality was satisfactory for meeting the objectives of this study. The limitations of the system can all be addressed with innovative information technology (IT) tools and I see enormous potential in the use of informatics for animal health surveillance.

### 7.4 User experience

User experience is the enjoyment felt when using the system. It is measured by the time spent accomplishing tasks, the ease of operation and the response to the use of the information provided. It is difficult to look at in isolation as it is the result of the interaction of the information quality and system quality described above (188).

The experiences of the participating sentinel veterinarians are a measure of user satisfaction. The ability to recruit and maintain private practice veterinarians was fundamental to the success of this project. No practices dropped out before the end of the project and I did not hear of any objections to the data collection method. The system met the expectations of the participants;
they did not incur any costs, the system did not impact the daily operation of their practices nor add any burden on their time.

In the preceding chapters I have described the difficulties of planning and implementing this system, true of any surveillance system development (153, 159, 358). Once development was complete the UCDEP was simple to install and configure at the sentinel practices and text-mining for the records of interest was quickly and easily completed. The unstructured clinical records used as the data source in this surveillance system resulted in a large and complex data set. I cleaned and processed the data as a retrospective dataset and so this task was completed once. I anticipate that if this system was used for prospective data collection and executed in real time, that data management would be very demanding. As the system user I was disappointed that the data volume did not equate to a more detailed characterization of enteric syndrome in the Calgary and area pet population. Information technology standards and data standards would improve the information quality, the performance of the system and the user experience.

7.5 System benefits, limitations and future directions

In my research I have described a surveillance system that applies informatics to the electronic medical record. The system that I have described is best implemented for retrospective data analysis to answer questions that are of veterinary clinical importance. A focus on retrospective data collection would neutralize three important limitations of the UCDEP: i) the UCDEP was vulnerable to software updates and so was not reliable in the prospective studies; ii) the daily prospective files were transferred by email, a clumsy and often ineffective method; and iii) one time data extraction is less complex and it may be cost-effective to customize the UCDEP for data extraction from additional management programs. This would improve the representativeness of the data.

I described the limitations of the enteric syndrome data. There was not enough actionable information for public health to respond to changes in enteric syndrome patterns in animals. Nor was there sufficient information to improve our understanding of the environmental factors that
may increase the risk of an animal presenting with enteric syndrome. However there is potential to use the data for evidence-based practice. This is the process of turning the data about past patients into knowledge so as to inform the treatment of current and future patients (352). The advantages of using medical record data bases for observational studies include larger sample size, lower costs and increased generalizability of the study when compared to other methods of data collection or randomized controlled trials (359). For example, I described the pattern of use of antihelmintics and antimicrobials for the management of enteric syndrome.

 Pharmacoepidemiological studies could critically assess whether these drugs or any of the other medications used to manage enteric syndrome cases are of benefit to the patient (359). Validity of the diagnostic information is cited as an important limitation for using retrospective clinical data and is relevant to the data from this system (360, 361). Information on confounders such as prescribing based on disease severity, the compliance of owners to give the medication or the use of over the counter medications will often be missing. Many of the records of patients with enteric syndrome have limited follow-up so an outcome may be misclassified. Confounders are managed with the methods common to all observational studies; restriction, stratification and regression. Proxy measures, such as the order for laboratory testing as a proxy for disease severity, are also used (361). Other clinically relevant questions may be answered with this data. While complex and perhaps imperfect, carefully planned and analyzed observational studies using this database could contribute useful information.

 In a population-based survey of people with episodes of acute diarrheal illness, 7.8% of the cases reported antimicrobial use (362). The authors described this proportion as “significant”. In Chapter 5 I reported that antimicrobials were administered or dispensed in 44.8% of the enteric syndrome cases. While it was not possible to assess the appropriateness of use, this was a considerable proportion of cases of which most were treated empirically based on clinical judgement and without laboratory confirmation.

 The antimicrobial use data from this study did provide information on temporal trends in antimicrobial use. These results could be used to investigate the causes for AMU trends, conduct
surveys to measure concomitant increases in antimicrobial resistance (AMR) and provide benchmarking data against which veterinarians could assess their AMU. The data can also be used to inspire research into the managerial and social perspectives of AMU. Medical guidelines are promoted as a means to improve the effectiveness of health care and reduce the wide practice variations for common clinical presentations (363, 364). I could find no papers in the refereed literature that describe practice guidelines for managing diarrheal disease in pets. Morley et al. (258) have drafted guidelines that promote stewardship and appropriate AMU in veterinary medicine. However, simply providing the information to medical practitioners about appropriate antimicrobial use has not always been successful (365). MacFarlane et al. found that understanding the ‘back-story’ to physicians’ AMU decisions and reducing the demand by patients was more effective at reducing use (366, 367). This signals the need to investigate how and why veterinarians use empirical antimicrobial therapy when treating enteric syndrome in pets.

My research has demonstrated that it was not often possible to identify the etiology of a cluster of diarrheal cases. For the data to be useful for outbreak investigations, to understand the epidemiology of a pathogen and to provide actionable information to public or animal health that identifying the etiology is often essential. Prospective surveillance can ensure that data collection is complete by responding to signals in real-time and advising veterinarians that further etiological investigations are necessary. If the objective of the surveillance is to assess AMU and AMR, thorough bacteriological studies including AMR profiles can be requested. However, I do not advocate the use of this surveillance system in its current form for prospective surveillance for the reasons described above.

In addition to exploring alternative data extraction technologies, I recommend adding structure to the medical records through the use of diagnostic coding and/or structured clinical reports to be completed by the participating veterinarians. This approach will enhance the patient records. The ambiguous clinical narrative can be transformed into a structured and standardized format with explicitly stated exposure history, diagnoses and rationale for treatment.
Furthermore, this system will help to address the complexity of the medical records and the issues of client confidentiality. The challenge lies in recruiting private practitioners to participate when there is an expectation of more work. Companion animal surveillance systems that require structured data input by the veterinarian have been developed and successfully implemented (152, 153, 368). In Alberta, the developers of a surveillance system that used some unstructured clinical data fields demonstrated its limitations to the participating swine veterinarians (personal communication, J. Keenliside, Alberta Agriculture and Rural Development, 2012). They were successful in convincing their participants that structured data was necessary in these fields. The requisite changes to their data collection process are currently in progress.

“When designing a surveillance system it is best to start with the end in mind (369)” I wanted to know if the routinely created companion animal patient data could: i) benefit human health and provide an early warning of environmental hazards for people; and ii) benefit animal health by monitoring disease and providing data to support population-based studies in the pet population. In its present form, the system I developed could support evidence-based companion animal practice but future research is necessary to improve its usefulness for both human and animal health.

I have described the technical limitations of my chosen informatics tools. Alternative approaches to data extraction and management are available and further research may prove these to be superior to my methods. Furthermore, technological innovation is rapid and ongoing and there is the expectation of improvements to informatics technologies (370). The development and application of universal technical standards to improve the integration of health care systems is a long term goal of database researchers (193, 370, 371).

Managing the limitations of the human dimension of companion animal surveillance may be more challenging. Research that explores ways to improve the quality of information in the medical record is necessary. Areas of research to consider are: i) to better understand the barriers to laboratory investigation in companion animal medicine which may present opportunities for promoting pursuit of etiological diagnoses by private practitioners; ii) validate the electronic
medical record by comparing the recorded content of the medical record against audio transcripts of patient visits; iii) investigate means of educating private practitioners about the data requirements of epidemiological studies and the potential benefit to their patients and their practice; and iv) determine if, when presented with sufficient evidence of improved outcomes with improved data quality, it is possible to document improvements in practitioners’ record-keeping behaviour.

In this research I have shown the need for structured data. Here again standards are seen as the solution to improving the interoperability of the EMR (364). Agreement on a set of data standards is a complex task that must meet the needs of a diverse group of stakeholders including the participant veterinary practitioners, academia, public animal health and public health. All of the recently developed companion animal surveillance systems (152-154, 368) have also included veterinary pharmaceutical industry partners in their projects. My research has shown that the shared goal is essential to effective interdisciplinary collaboration. The next step is to bring committed and talented people from all of the sectors together and with adequate resources to plan a system that can provide a basis for action.
References


42. Harder TC, Vahlenkamp TW. Influenza virus infections in dogs and cats. Veterinary immunology and immunopathology. 2010;134(1):54-60.


101. The Constitution Act, 1867, 30 and 31 Victoria c.3, UK


Syndromic surveillance in the Netherlands for the early detection of West Nile virus epidemics.

133. Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. Syndromic

134. Burkom H. Chapter 4; Alerting Algorithms for Biosurveillance. In: Lombardo JS,
Buckeridge DL, editors. Disease Surveillance: A Public Health Informatics Approach. New


Systematic review of surveillance systems for emerging zoonoses. Transboundary and emerging

137. Buckeridge DL. Outbreak detection through automated surveillance: a review of the

practice in the United States: findings from a survey of state, territorial, and selected local health

of syndromic surveillance. Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and

140. Dórea FC, Sanchez J, Revie CW. Veterinary syndromic surveillance: Current initiatives

for evaluating public health surveillance systems for early detection of outbreaks:
recommendations from the CDC Working Group. MMWRRecommendations and reports :
Morbidity and mortality weekly reportRecommendations and reports / Centers for Disease
151. Family E. Practice-Based Research Networks in the 21st Century.


164. Bhattacharyya SB, editor. Data warehousing and data mining—an overview. Indian Association of Medical Informatics; 2004; Calcutta.


330. Lomas J. Improving research dissemination and uptake in the health sector: beyond the sound of one hand clapping1997; C97-1(Journal, Electronic).


Appendix A

DATA TRANSFER AGREEMENT

This Agreement is made by and between

<table>
<thead>
<tr>
<th>Disclosing Party:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Contact information:</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

and

<table>
<thead>
<tr>
<th>The Governors of the University of Calgary (“Receiving Party”)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Contact information:</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

With respect to data that Disclosing Party will provide for (“the Study”), the Study protocol as approved by the Disclosing Party’s REB is attached hereto in Schedule “A”.

The Study will not commence until the Disclosing Party’s research ethics board (“REB”), if applicable, has (a) approved the Study protocol, and b) executed the Study informed consent form or waived the requirement to obtain consent.

The parties hereby agree as follows:

1. **Definition of Data Types; Transfer of Data.** (1) As used in this Agreement, the term "Chart Data" means medical data or information pertaining to specific cases that are collected by the Disclosing Party in the course of their practice and are part of the patient’s medical chart. “Abstracted Data” refers to anonymous information that has been abstracted from Chart Data
and is provided to the Receiving Party for the purpose of carrying out the Study in accordance with the Study protocol as approved by the Disclosing Party’s REB. “Summary Data” refers to the results of the analysis by the Receiving Party. (2) Disclosing Party shall collect and provide to Receiving Party the Chart Data in accordance with the Protocol. Disclosing Party retains the right to refuse the transfer of the Chart Data requested hereunder. For greater certainty, a client as referred to herein is an animal owner and a patient as referred to herein is an animal under veterinary care.

2. **Compliance with Law.** In the performance of the Study, the parties shall comply with all applicable laws, regulations, guidelines and policies (“Applicable Law”). Each party agrees to notify the others of changes to Applicable Law of which the party becomes aware that may necessitate an amendment to the terms of this Agreement. The parties shall promptly and in good faith take all necessary action to ensure that the terms of the Agreement remain compliant with Applicable Law.

3. **Non-Disclosure of Data.** (1) Without limiting the obligation set out in Section 2, the Receiving Party agrees that it shall, and shall require its directors, officers, employees, medical staff, research fellows, students, consultants, and advisors who need to know such Chart Data for the purposes of the Study (“Study Staff”) to:

   (a) confirm in writing that they are aware of the confidential nature of the Chart Data, their commitment to maintain the Chart Data in confidence, and not to disclose Chart Data except as permitted by this Agreement.

   (b) use Chart Data solely for the purposes of creating the Abstracted Data file and that the Abstracted Data are to be used for the Study, in compliance with:
(i) the approved Study protocol (attached as Schedule “A”) as it may be amended from time to time, provided that amendments are approved in writing by the Disclosing Party’s REB and made known to the Receiving Party;
(ii) Disclosing Party’s letter of approval of the Study protocol (and if applicable any written conditions imposed by the Disclosing Party’s or Receiving Party’s REB), attached hereto as Schedule “B”;
(iii) the informed consent form approved by the Disclosing Party’s REB or, if the requirement to obtain consent has been waived by the Disclosing Party’s REB, the waiver of consent given by the Disclosing Party’s REB, and attached hereto as Schedule “C”; and
(iv) any other conditions or restrictions imposed by Disclosing Party relating to the use, security, disclosure, return or disposal of the Data, as set out in this Agreement.

(c) not use the Chart Data to identify any individual clients or patients in full compliance with the privacy laws of Canada and the Province of Alberta;

(d) not transfer the Chart Data or Abstracted Data disclosed under this Agreement to any third parties, except as set out in Section 4 below, without prior written consent of the Disclosing Party and without obligating such third parties to comply with the terms and conditions hereof; and

(e) securely destroy the Chart Data as required by the Study protocol or by Disclosing Party and provide a written confirmation of the manner of destruction to Disclosing Party. The Abstracted Data may be retained indefinitely for the purposes of the Study.

(2) Receiving Party shall use appropriate safeguards to prevent any unauthorized use or disclosure of the Chart Data and Abstracted Data and shall promptly report to Disclosing Party any unauthorized use or disclosure of which Receiving Party becomes aware.
4. Permitted Disclosure of Data. The Receiving Party may disclose Chart Data and Abstracted Data:

   (a) to its Study Staff with a need to know, have signed the written confirmation identified in Paragraph 3(a) above, and who are obligated to maintain the confidential nature of such Data;

   (b) as otherwise permitted by the informed consent form approved by the Disclosing Party’s REB or the waiver of consent given by the Disclosing Party’s REB attached hereto as Schedule “C”; or

   (c) in order to comply with applicable laws or regulations or judicial process, or with a court or regulatory order, provided that the Receiving Party gives prior written notice of such intended disclosure to the Disclosing Party and that the Receiving Party takes all reasonable and lawful actions to obtain confidential treatment for such disclosure and, if possible, to minimize the extent of such disclosure.

5. Contact with Subjects/Individuals. The Receiving Party shall not make contact or attempt to make contact with a client or patient unless the Disclosing Party first obtains the client’s consent to being contacted.

6. Publication and Ownership. (1) The Receiving Party hereby acknowledges that all rights in and to the Chart Data shall be held by the Disclosing Party at all times for the duration of this Agreement and thereafter. (2) The Disclosing Party hereby agrees that the Receiving Party shall own and have the right to use the Abstracted Data and Summary Data as part of a publication or presentation of the results of the Study. The Receiving Party acknowledges that it shall be entitled to maintain a copy of the Abstracted Data and Summary Data but shall not be entitled to
use either for any purpose other than that described in Schedule “A” without the express written consent of the Disclosing Party. The Receiving Party’s rights of use of the Data during the Study and thereafter are identified in Schedules “A” and “D”. (3) Receiving Party shall not include any personally identifying information in any publication or presentation. (4) Disclosing Party’s and Disclosing Party’s investigator’s(s’) contribution to the Study shall be acknowledged appropriately in any such publication or presentation.

7. **Liability.** The Receiving Party assumes liability for any costs, suits, claims or penalties on account of wrongful disclosure of personal information which may arise as a result of its activities carried out in relation to the Study including such costs, suits, claims or penalties arise out of the negligent or intentional acts or omissions of another party or parties to which the Receiving Party granted access to and/or use of the Data. The Receiving Party shall not be liable for any wrongful disclosure of personal information arising from the forcible entry into the offices and/or files of the Receiving Party.

8. **Term, Termination and Survival.** This Agreement shall come into effect on the date of signature of the last party to sign this Agreement, and shall continue in effect until the completion or earlier termination of the Study. Any party may terminate this Agreement on thirty (30) days written notice to the other party. Any notice or other document to be given under this Agreement shall be in writing and shall be deemed to have been duly given if delivered by hand or courier, or sent by first class post, registered airmail or facsimile to the other parties at the addresses set out above or to such other address as a party may designate by written notice to the other. Any such notice or other document shall be deemed to have been received by the addressee on the day of delivery if delivered by hand or by courier, or five (5) working days following the date of dispatch if sent by post, or the day after the date of transmission if sent by facsimile. On termination, the terms in Schedule D, if applicable, concerning payment upon termination shall apply. The Receiving Party’s restrictions pertaining to the prohibited disclosure of the Chart Data shall survive the termination of this Agreement.

(1) No party shall be entitled to assign or transfer this Agreement or the rights and obligations hereunder to any third party without the prior written approval of the other party.

(2) This Agreement including the attached Schedules represents the entire understanding between the parties related to the Study and supersedes all previously or contemporaneously executed agreements related to the Study. In the event that there is a conflict between the terms of this Agreement and a Schedule, the terms of the Schedule shall govern.

(3) This Agreement shall not be amended, modified, varied or supplemented except in writing signed by both of the parties.

(4) No failure or delay on the part of any party hereto to exercise any right or remedy under this Agreement shall be construed or operate as a waiver thereof.

(5) The parties hereto are independent contractors. Nothing contained herein shall be deemed or construed to create between or among the parties hereto a partnership or joint venture or employment or principal-agent relationship. No party shall have the authority to act on behalf of any other party or to bind another party in any manner.

(6) Each party hereto acknowledges that it has been advised by the others to seek independent legal advice with respect to this Agreement and that it has not relied upon any of the other parties hereto for any advice, whether legal or otherwise, with respect to this Agreement.

(7) No party (“the first party”) shall use, or authorize others to use, the name, symbols, or marks of another party hereto or its staff without prior written approval from the party whose name, symbols or marks are to be used.

(8) This Agreement shall be governed by and construed in accordance with the laws of the Province of Alberta and the federal laws of Canada applicable therein.
Acknowledged and agreed by:

<table>
<thead>
<tr>
<th>Disclosing Party:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature:</td>
<td></td>
</tr>
<tr>
<td>Date:</td>
<td></td>
</tr>
<tr>
<td>Name and Title:</td>
<td></td>
</tr>
</tbody>
</table>

Acknowledged by:

| Signature: |     |
| Date:      |     |
| Title:     |     |

Receiving Party:

| Signature: |     |
| Date:      |     |
| Name and Title: |     |

Acknowledged by:

| Signature: |     |
| Date:      |     |
| Title:     |     |
Schedule “A”
Study protocol as approved by the Disclosing Party’s REB
Data Sharing Agreement Scope of Work Document

Schedule “B”
Disclosing Party’s Research Ethics Board approval letter; and written conditions, if any, imposed by the Disclosing Party’s or Receiving Party’s Research Ethics Board

Schedule “C”
Additional Terms of Use

1. Software Testing and Installation

a. Prior to the installation of the Software in the Disclosing Party’s computer system, the Disclosing Party shall conduct various testing, to its satisfaction, to ensure that disruption to the Disclosing Party’s computer system is minimized.

b. The Receiving Party hereby agrees to fully indemnify the Disclosing Party from and against any and all disruptions to the Disclosing Party’s computer system that the Disclosing Party may incur as a result of the installation of the Software on the Disclosing Party’s computer system. To minimize the potential for the loss of data (including without limitation the Data associated with this Agreement), the Disclosing Party agrees to conduct a backup of its computer system immediately prior to the installation of the Software and thereafter on a daily basis.

c. Installations of the Software by the Receiving Party shall be conducted at times when a representative of the Disclosing Party is available and present. To facilitate this requirement, the Receiving Party shall provide the Disclosing Party a schedule of the dates and times that it intends to install the Software at each of the Disclosing Party’s clinics (understanding that the Disclosing Party may need
to have some changes made to such schedule to satisfy its requirements).

d. Unless otherwise agreed by the Disclosing Party, the Software shall be removed from the Disclosing Party’s computer system within one year of the date of the Software was installed on the Disclosing Party’s computer system. The removal of the Software shall be supervised by a Disclosing Party representative. The costs associated with the removal of the Software shall be borne by the Receiving Party.

e. Should the Software become available for commercial use, the Disclosing Party shall at its written request be entitled to obtain a copy of the Software for each of its clinics. Prior to obtaining a copy of the Software, the parties agree to negotiate the terms of a license agreement, with such license agreement to address the applicable fees. Should the two parties wish to further develop the software to their mutual benefit, a partnership agreement would be negotiated.

f. In utilizing the Software to create the Abstracted Data file from the Disclosing Party’s computer system, the Receiving Party agrees to provide the Disclosing Party a copy of the Abstracted Data via email. Should the resulting files be too large to transmit via email, the Data shall be provided to the Disclosing Party on a mutually agreed upon storage device with all Data stored in a machine readable format. All copies of extracted Data shall be sent to the attention of “name” or other nominated individual in the Disclosing Party’s Surrogate Department. This is for assurance purposes only and all rights for use of the Abstracted Data lie with the Receiving Party.

2. Upon the completion of the Study, the Disclosing Party shall be provided a copy of the Study results on or before the release of such results to the general public. In receiving such Study results, the Disclosing Party shall be prohibited from disclosing such results
to anyone that is not reasonably approved by the Receiving Party.