Developing and Implementing Techniques to Harvest Surveillance Information from Existing Veterinary Diagnostic Laboratory Data

By

Fernanda C. Dórea

A Thesis

Submitted to the Graduate Faculty
in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

Department of Health Management
Faculty of Veterinary Medicine
University of Prince Edward Island

Condition for the use of the thesis

Chairman of the Department of Health Management

Faculty of Veterinary Medicine

University of Prince Edward Island

Charlottetown, P. E. I.

Canada C1A 4P3

Permission to use postgraduate thesis

Title of thesis: Developing and Implementing Techniques to Harvest Surveillance Information from Existing Veterinary Diagnostic Laboratory Data

Name of Author: Fernanda C. Dórea

Department: Health Management

Degree: Doctor of Philosophy Year: 2013

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from the University of Prince Edward Island, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the professor or professors who supervised my thesis work, or, in their absence, by the Chairman of the Department or the Dean of the Faculty in which my thesis work was done. It is understood any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Prince Edward Island in any scholarly use which may be made of any material in my thesis.

Signature:

Address: Department of Health Management

Faculty of Veterinary Medicine

University of Prince Edward Island

550 University Avenue

Charlottetown, PE C1A 4P3

CANADA.

Date: April 08,2013

University of Prince Edward Island

Faculty of Veterinary Medicine

Charlottetown

CERTIFICATION OF THESIS WORK

We, the undersigned, certify that **Fernanda C. Dórea**, a candidate for the degree of **Doctor of Philosophy**, has presented his thesis with the following title: "**Developing and Implementing Techniques to Harvest Surveillance Information from Existing Veterinary Diagnostic Laboratory Data**" and that the thesis is acceptable in form and content, and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate through an oral examination held on March 27,2013.

Examiners

Dr. Dan Hurnik (chair) _____

Dr. Crawford Revie _____

Dr. Ian Gardner _____

Dr. C. Anne Muckle _____

Dr. David Pearl _____

Date: March 27,2013

I am thankful,

To God, my source of energy;
To my mother, my best friend and source of motivation, laughter and sanity,
(and even the right dose of insanity);
To my brother, my source of inspiration;
To Holger, my source of happiness;

this accomplishment is dedicated to:

my father;
the man who gave his children only one thing above his means: education
the man who I have watched study every single day of my life
the man responsible for my English education
the man who gave me wings, and made sure I used them.

Abstract

Syndromic surveillance is a tool for continuous, automated extraction of surveillance information from health data sources. The research documented in this dissertation aimed at exploring informatics and data mining tools in order to develop and implement techniques to harvest additional surveillance information from existing diagnostic laboratory data. Data concerning laboratory test requests for diagnosis in cattle were provided by the Animal Health Laboratory (AHL), at the University of Guelph, Ontario. A thorough review of the initiatives of syndromic surveillance in animal health was conducted. Documented difficulties regarding the acquisition of clinical data, and especially sustainability of systems based on voluntary participation of veterinarians or data providers in scattered locations, resulted in the choice of using laboratory data in this research. Automated methods to classify laboratory submission data into clinical syndromes were investigated. One of the challenges of working with laboratory data was determining how to transform diagnostic data into epidemiological information. The most time-consuming step of classification was the creation of a dictionary of keywords relevant to each classification task, and the definition of the relationship between these words, their co-occurrences and the target syndromic group. Once defined, however, these relationships were easily translated into a set of rules that achieved high classification performance. After classification was performed, the data were reduced to multiple time-series registering daily (or weekly) submissions to the different syndromes monitored. Retrospective evaluation of the time-series representing daily counts for each syndromic group were carried out in order to identify temporal effects present, and define methods to model or remove them on-line. A method is presented for automated removal of excessive noise and historical outbreaks in historical data, in order to construct baselines of normal behaviour. These baselines could be used as training data for the algorithms implemented in the next stages. Lastly, the prospective phases of system development were carried out, that is, the analyses which scan the time series in an on-line process, one day at a time, in order to detect temporal aberrations in comparison to a baseline of historical

data. Several aberration detection algorithms were evaluated. Upon the conclusion that no single algorithm was superior in all outbreak scenarios, a scoring system to combine algorithms was developed. All steps were set up using open source software, and delivered to the data provider as a simple desktop application scheduled to run daily in an automated manner. Fast development and simple maintenance is expected to lead to incorporation of this system into the routine of the data, becoming an indispensable tool for diagnosticians and epidemiologists, and encouraging further technical development.

Acknowledgements

I thank OMAFRA for the funding and the Animal Health Laboratory (AHL) at the University of Guelph for the data. I would like to thank Dr. Bruce McNab and Dr. Beverly McEwen from these respective institutions for their help and support, which went far beyond the professional obligations. I am very grateful for their kindness, patience and mentorship.

I thank Dr. David Kelton, Dr. Ann Muckle and Dr. J.T. McClure for their support on my advisory committee.

I thank Dr. Javier Sanchez for his mentorship, and for his personal assistance during my moving to PEI, a place where I have been truly happy.

I thank Dr. Crawford Revie for his mentorship. Finishing my PhD would not feel as meaningful if it would dissolve the mentor-pupil bond we established, and I look forward to the chance of continuing to work under his mentorship after completing my PhD training.

I thank all those who have made my experience at the university pleasant, especially Leanne Newson and Rosemary McIver.

I thank Dr. Mary McNiven, Ruth Cox, the Jay family and Jennifer van der Lee for enriching my experience in Charlottetown.

I thank my family for standing the distance and yet making themselves close enough to keep me going.

I thank he who, along these 3 years, has been my boyfriend, fiancee and husband. Holger's understanding and support has been, at times, what enabled me to keep going on, and at all times the reason I was able to go on with a smile on my face.

List of abbreviations

AHL	Animal Health Laboratory
AUC	Area Under the Curve
CUSUM	Cumulative Sums
DOW	Day of Week
EWMA	Exponentially Weighted Moving Averages
HW	Holt-Winters Exponential Smoothing
LIMS	Laboratory Information Management System
OMAFRA	Ontario Ministry of Agriculture, Food and Rural Affairs
ROC	Receiver Operating Characteristic

Table of Contents

Condition for the use of the thesis ii

Permission to use postgraduate thesis iii

Certification of thesis work iv

Abstract vi

Acknowledgments viii

List of Abbreviations ix

List of Tables xiv

List of Figures xvi

1 Introduction and objectives 1

1.1 Introduction 2

1.2 Objectives 4

2 Veterinary syndromic surveillance: current initiatives and potential for development 8

2.1 Abstract 9

2.2 Introduction 10

2.3 Population coverage and timeliness in syndromic surveillance 12

2.4 Syndromic surveillance initiatives in veterinary medicine 14

2.4.1 Syndromic surveillance based on notification of atypical cases 18

2.4.2 Syndromic surveillance based on analysis of all clinical cases 19

2.4.3 Syndromic surveillance focusing on public health (animals as sentinels for human diseases) . . . 21

2.5 Data sources for syndromic surveillance in veterinary medicine . . . 24

2.5.1 Voluntary notification . . . 25

2.5.2 Clinical data . . . 26

2.5.3 Herd management data . . . 27

2.5.4 Laboratory data . . . 28

2.5.5 Others . . . 29

2.6 Implementation of disease aberration detection from animal health data . . . 31

2.6.1 Definition of events and syndromes . . . 31

2.6.2 Aberration detection algorithms . . . 35

2.6.3 Evaluation . . . 37

2.7 Discussion . . . 39

3 Exploratory analysis of methods for automated classification of laboratory test orders into syndromic groups in veterinary medicine 46

3.1 Abstract . . . 47

3.2 Background . . . 48

3.3 Methods . . . 51

3.3.1 Data source . . . 51

3.3.2 Data structure . . . 51

3.3.3 Syndrome definition . . . 52

3.3.4 Mapping of test codes . . . 54

3.3.5 Algorithms for automated syndrome classification . . . 55

3.3.6 Assessing algorithms performance . . . 56

3.4 Results . . . 57

3.5 Discussion . . . 63

3.6 Conclusion . . . 68

4 Retrospective time series analysis of veterinary laboratory data: preparing a historical baseline for cluster detection in syndromic surveillance 69

4.1 Abstract . . . 70

4.2 Introduction . . . 71
4.3 Methods . . . 72
4.3.1 Data source . . . 72
4.3.2 Case definition and syndromic groups . . . 73
4.3.3 Data characterization . . . 75
4.3.4 Aberration removal . . . 76
4.4 Results . . . 78
4.4.1 Case definition and syndromic groups . . . 78
4.4.2 Data characterization . . . 78
4.4.2.1 Time series for syndromes monitored daily . . . 79
4.4.2.2 Time series for syndromes monitored weekly . . . 80
4.4.3 Aberration removal . . . 81
4.4.3.1 Time series monitored daily . . . 81
4.4.3.2 Time series monitored weekly . . . 83
4.5 Discussion . . . 84
4.6 Conclusion . . . 88

5 Syndromic surveillance using veterinary laboratory data: data pre-processing and algorithm performance evaluation 90

5.1 Abstract . . . 91
5.2 Introduction . . . 91
5.3 Methods . . . 93
5.3.1 Data source . . . 94
5.3.1.1 Simulated data . . . 97
5.3.2 Detection based on removal of temporal effects and use of control charts . . . 100
5.3.2.1 Exploratory analysis of pre-processing methods . . . 100
5.3.2.2 Control charts . . . 101
5.3.3 Detection using Holt-Winters exponential smoothing . . . 103
5.3.4 Performance assessment . . . 104
5.4 Results . . . 105
5.4.1 Preprocessing to remove the DOW effect . . . 105

5.4.2 Qualitative evaluation of detection algorithms . . . 106
5.4.3 Evaluation using simulated data . . . 108
5.5 Discussion . . . 112

6 Syndromic surveillance using veterinary laboratory data: algorithm combination and customization of alerts 118

6.1 Abstract . . . 119
6.2 Introduction . . . 120
6.3 Methods . . . 122
6.3.1 Data source . . . 122
6.3.2 Simulated data . . . 123
6.3.3 Algorithms for aberration detection . . . 123
6.3.4 Correcting temporal aberrations . . . 124
6.3.5 Combining algorithms: Scoring system . . . 125
6.3.6 System reports . . . 127
6.4 Results . . . 128
6.4.1 Correcting temporal aberrations . . . 128
6.4.2 Scoring system . . . 130
6.4.2.1 Detection using real data: qualitative analysis . . . 130
6.4.2.2 Detection using simulated data: quantitative analysis . . . 132
6.4.3 System reports . . . 135
6.5 Discussion . . . 135
6.6 Conclusion . . . 141

7 Conclusions and future directions 143

7.1 System delivery to AHL and OMAFRA . . . 147
7.2 System limitations and potential for further development . . . 150
7.3 Conclusion . . . 154

Bibliography 171

List of Tables

2.1 Published initiatives in veterinary syndromic surveillance 16

2.2 Peer reviewed publications investigating the potential of different datasets in implementing veterinary syndromic surveillance systems. 17

3.1 Sample of the laboratory submission data used in this work. 53

3.2 Syndromic groups 58

3.3 Instances and syndromic groups in the 'unmapped' subset of the data 61

3.4 Performance measures for the algorithms implemented. 61

4.1 Syndromic groups 74

5.1 Performance evaluation of different detection algorithms. 110

6.1 Detection limits for each of the three algorithms implemented 126

List of Figures

1.1 Overview of the process of developing a syndromic surveillance system 7

2.1 Disease continuum in a population. 13

2.2 Process of monitoring disease trends and detecting clusters using animal health data sources. 32

3.1 Number of syndromes identified in each case using information from individual test requests. 59

3.2 Percentage of test requests classified by direct mapping and automated classification. 60

3.3 Daily counts of cases allocated to Bovine Viral Diarrhea and Mastitis syndromes. . . 67

4.1 Time series of daily and weekly counts for a few syndromic groups 79

4.2 Day-of-week and month effects exemplified using the mastitis series 80

4.3 Poisson model predictions and the original BLV time series used to fit the model. . . 81

4.4 Raw data of daily counts for the BLV series, superimposed by a baseline constructed after removal of temporal aberrations. 83

4.5 Raw data of weekly counts for the systemic syndrome series superimposed by a baseline constructed after removal of temporal aberrations. 84

5.1 Syndromic groups used to exemplify the time series used in this work. 96

5.2 Synthetic outbreak simulation process. 98

5.3 Comparative performance of pre-processing methods. 106

5.4 ROC curves documenting the performance of several aberration detection algorithms evaluated. 109

6.1 Schematic representation of the on-line process of aberration detection and correction. 125

6.2 Alarm detection and automated correction in the BLV series. 129

6.3 Aberration detection using three algorithms combined using the scoring system and applied to real data. 131

6.4 Sensitivity of detection and false alarm rates when the combined algorithms are applied to the simulated mastitis series. 133

6.5 Sensitivity of detection and false alarms rate in the BLV series 134

6.6 Example page of a daily report sent to analysts in case of alarm. 136

7.1 Overview of system implementation. 149

Chapter 1

Introduction and objectives

1.1 Introduction

At the turn of the millennium Doherr and Audigè (2001) [1] pointed out the changing demands in animal health surveillance, as disease control and eradication around the globe have increased the necessity to deal with rare events and provide evidence of disease freedom. The authors highlighted the role of early detection of emerging (or re-emerging) diseases, calling attention to the need of developing and implementing "scientifically based approaches that use the resources (and data) available." Over the intervening decade technological infrastructures and the health data available within them have developed rapidly, disease prevalence has reduced for some diseases in many areas, and general awareness of the need for early disease detection has grown as a result of publicity around major disease outbreaks such as pandemic influenza as well as the concern regarding bioterrorism events.

These significant changes have caused surveillance to shift, in the last decade in particular, towards systems capable of early detection of disease [2]. Modern bio-surveillance systems are designed to take advantage of data assumed to contain signatures of healthcare-seeking behaviours, which are not as specific as diagnosis, but allow for more rapid detection, and can be aggregated into syndromes. Surveillance based on these types of data is therefore referred to as *syndromic surveillance* [3]. Other definitions of syndromic surveillance exist, all of which emphasize the use of prediagnostic data [4].

Syndromic surveillance has been used not only for early detection of diseases, but also for real-time monitoring of outbreaks (situational awareness [5]), monitoring of disease trends, and to provide reassurance of disease freedom [6, 7]. These goals can be summarized as continuous analysis of health data to provide immediate feedback [6].

Syndromic surveillance is therefore a tool for continuous, automated extraction of

surveillance information from health data sources, as timely as the rate of gathering data in electronic format. The data source scanned should in turn be acquired continuously, in an automated routine, and be stored electronically. In order to improve the time to output in comparison to traditional surveillance, two directions are possible [8]: improve the collection of traditional health data, or look for data sources that are already collected frequently and electronically, but have not been traditionally used for disease surveillance. The role of laboratory data as a source for syndromic surveillance development has been recognized in public health due to its high population coverage [9], and electronic recording [10]. The scarcity of centralized electronic collections of clinical data in veterinary medicine further strengthens the motivation for the use of laboratory data in animal health surveillance.

While laboratory results are only available late in the disease continuum, laboratory test requests are a type of syndromic data. They are timelier than test results, and can be grouped into syndromes according to the nature of the disease and/or symptoms observed by the veterinarian [11, 12, 10]. Stone (2007)[13] investigated the potential of using laboratory test requests for syndromic surveillance in veterinary medicine and reviewed the potential biases associated with these type of data. The author also pointed out the variability in the submission rates year to year, and misclassification bias (veterinarian not submitting the right sample or requesting the correct test), but concluded that the data are suitable for syndromic surveillance. Similarly, Shaffer (2007) [14] assessed the potential of microbiology test submissions for syndromic surveillance in companion animals assuming that the consistency of test orders over time allows for the use of these data in prospective monitoring, and that increases in the number of test orders can be used as indicators of an increase in disease burden. The availability of historic data is another advantage of laboratory data in veterinary medicine over other types of data, since some estimation of a baseline of disease burden is needed in syndromic surveillance to compensate for the lack of denominator data. Lastly, laboratory test requests screen a larger

proportion of the animal population than sick animals, as animals can be tested for different purposes. Zhang et al. (2005) [15] reported that four different purposes were recorded as reasons for test requests in their laboratory data: diagnostic, export testing, government monitoring, and industry monitoring.

1.2 Objectives

The research documented in this dissertation aimed to explore informatics and data mining techniques, in order to develop and implement techniques to harvest additional surveillance information from existing diagnostic laboratory data. Data were provided by the Animal Health Laboratory (AHL), at the University of Guelph. The AHL is the primary laboratory of choice for veterinary practitioners submitting samples for diagnosis in food animals in the province of Ontario, Canada. The AHL has a laboratory information management system (LIMS) that is primarily used for reporting the results of diagnostic tests. However, it can also be utilized as a data retrieval tool for surveillance purposes. Surveillance information harvested should be delivered to the Ontario Ministry of Agriculture Food and Rural Affairs (OMAFRA) in order to support disease surveillance programs.

A thorough literature review of syndromic surveillance initiatives in veterinary medicine was performed and is documented in Chapter 2 of this dissertation. It provides important background for this work and sets the research in the context of related research activities. Based on this review the steps necessary to set up a syndromic surveillance system based on the data available were determined. An inventory of the methods to be explored was elaborated. Subsequently, the research necessary to adapt these methods to the specific characteristics of animal health data – the data at hand in particular – was outlined and carried out.

The substantive research objectives of the thesis, which are documented in the

following chapters of this dissertation, were:

1. Conduct exploratory analyses of machine learning and rule-based methods capable of recognizing medically relevant information from laboratory submissions, and classify them into syndromic groups (Chapter 3).

2. Evaluate, retrospectively, three years of historical data available at AHL, in order to assess the potential of those data for development of a syndromic surveillance system (Chapter 4). In particular, this assessment should:

   - determine the temporal effects present in the data, which could affect the performance of algorithms capable of detecting temporal aberrations (signals that must be investigated as they can indicate the presence of a disease outbreak);
   - evaluate statistical approaches to remove these temporal effects;
   - clean historical data from any outbreak signatures and excessive noise in order to establish the baseline behaviour of the data for aberration detection in real time.

3. Evaluate the performance of different algorithms capable of monitoring the data prospectively in near-real time (i.e. on a daily and weekly basis) in order to detect aberrations (Chapter 5). Specifically:

   - identify pre-processing methods that are effective in removing or dealing with temporal effects in the data;
   - explore methods that combine these pre-processing steps with detection algorithms, in light of the data available while keeping in mind the importance of having a detection process interpretable by the analysts;
   - identify the temporal aberration detection algorithms that can provide high sensitivity and specificity for this specific monitoring system.

4. Implement real-time monitoring using the algorithm(s) that showed best performance in the previous step (Chapter 6).

These objectives represent the natural steps necessary in order to construct a syndromic surveillance system from a new data source. Figure 1.1 represents these steps schematically.

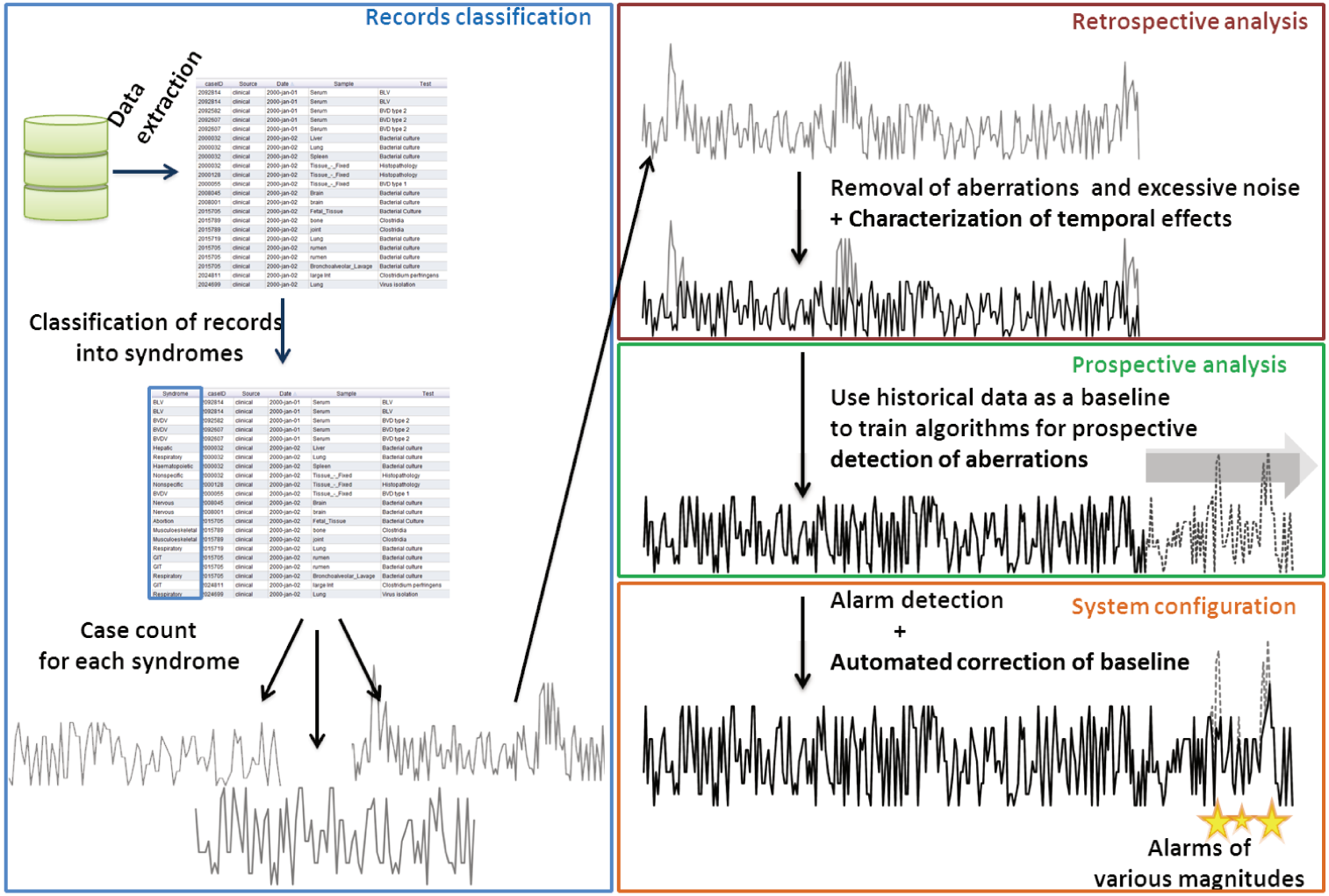


Figure 1.1: Overview of the process of developing a syndromic surveillance system from animal health data.

Chapter 2

Veterinary syndromic surveillance: current initiatives and potential for development

[0]Fernanda C. Dórea, Javier Sanchez and Crawford W. Revie. Preventive Veterinary Medicine 2011, 101: 1-17

2.1 Abstract

This paper reviews recent progress in the development of syndromic surveillance systems for veterinary medicine. Peer-reviewed and grey literature were searched in order to identify surveillance systems that explicitly address outbreak detection based on systematic monitoring of animal population data, in any phase of implementation. The review found that developments in veterinary syndromic surveillance are focused not only on animal health, but also on the use of animals as sentinels for public health, representing a further step towards *One Medicine*. The main sources of information are clinical data from practitioners and laboratory data, but a number of other sources are being explored. Due to limitations inherent in the way data on animal health is collected, the development of veterinary syndromic surveillance initially focused on animal health data collection strategies, analyzing historical data for their potential to support systematic monitoring, or solving problems of data classification and integration. Systems based on passive notification or data transfers are now dealing with sustainability issues. Given the ongoing barriers in availability of data, diagnostic laboratories appear to provide the most readily available data sources for syndromic surveillance in animal health. As the bottlenecks around data source availability are overcome, the next challenge is consolidating data standards for data classification, promoting the integration of different animal health surveillance systems, and also the integration to public health surveillance. Moreover, the outputs of systems for systematic monitoring of animal health data must be directly connected to real-time decision support systems which are increasingly being used for disease management and control.

Keywords Syndromic surveillance; veterinary surveillance; animal health surveillance; emerging diseases; aberration detection; prospective monitoring.

2.2 Introduction

The evolution of disease control methods in veterinary medicine from campaigns and mass action to a new phase of surveillance and selective action was defined by Dr. Calvin Schwabe (1982) [16] as an epidemiological revolution, marked by the use of epidemiological intelligence and analysis key tools for diagnosis and decision making. The last decade has witnessed a further step in this revolution, with "epidemiological intelligence" being progressively improved through novel informatics and data mining techniques; these allow analysis to be carried out on an unprecedented quantity of data to identify novel and useful patterns in an automated manner [17].

In this new context, providing effective and comprehensive approaches for systematic information management and analysis plays a central role in achieving the goals of disease surveillance [18]. While the concepts behind integrating information from multiple sources are not novel [19], the past decade has seen an increase in research that is focused on developing, "the science and technologies needed for collecting, sharing, reporting, analyzing, and visualizing infectious disease data and for providing data and decision-making support for infectious disease" which Zeng et al. (2005) [18] defined as infectious disease informatics. This is an interdisciplinary field, taking advantage of a range of information technologies such as data sharing and security, geographic information systems (GIS), data mining and visualization, knowledge management, biostatistics and bioinformatics [17, 18].

The uptake of these approaches gained momentum when bioterrorist events, such as the anthrax attacks of 2001, and outbreaks of emerging infectious diseases, such as SARS [20] underlined the necessity to recognize patterns indicative of a possible introduction of human pathogens, natural or not, as early as possible. Using the tools provided by infectious disease informatics, real time surveillance systems were developed to make use of pre-diagnostic data already available and automatically

collected [21], such as sales of over-the-counter medicine, absences from work or school, patient's chief complaint upon emergency visit, or laboratory test orders [22, 23].

Due to the lack of specificity associated with pre-diagnostic data, this new type of surveillance targets general groups of diseases, or syndromes, and is therefore often referred to as "syndromic surveillance". The Centers for Disease Control (CDC, USA) has defined as syndromic surveillance those approaches which make use of "health-related data that precede diagnosis and signal with sufficient probability of a case or an outbreak to warrant further public health response" [3]. While less specific than confirmatory diagnosis, data used for syndromic surveillance are more timely [2], allowing for real-time or near-real-time analysis and interpretation of data [24]. The assumption is not that the data are representative of the disease burden in the population (and usually no attempt is made to estimate such parameters, as various biases are recognized to exist), but that they are sensitive to changes to the level of disease in the population, containing an early, though weak, signature of a disease outbreak [25]. While syndromic surveillance definitions focus on early detection of disease, Henning (2004) highlights the fact that with the continuous use of such systems longitudinal data are being accumulated, allowing for a broader achievement; "the use of existing health data in real time to provide immediate analysis and feedback to those charged with investigation and follow-up of potential outbreaks".

In veterinary medicine the development of systems for early detection of diseases followed a similar path to that taken in public health. Recent focus on the "One Medicine" concept has resulted in an increased awareness that the early detection of outbreaks in animal populations, whether zoonotic or not, can be of great public health importance.

While the past decade has seen a growth in the literature dealing with novel surveillance approaches, including a great increase in the use of cluster detection

techniques applied retrospectively to data, to the authors' knowledge there exists no systematic overview of the application of syndromic surveillance to veterinary medicine. This paper reviews the current progress towards developing syndromic surveillance in veterinary medicine, defining as such all those systems that explicitly address outbreak detection based on systematic monitoring of population data. While this review focuses on syndromic surveillance systems that are already operational or are in their implementation phase, we also review studies investigating the potential for early detection of disease using alternative types of data available in animal health, to help the reader gain a sense of potential future developments.

2.3 Population coverage and timeliness in syndromic surveillance

A primary assumption of any syndromic surveillance system is that the behavior of the population changes when their health is affected, and that clusters (in space and/or time) of these behavioral changes can be detected if the population is continuously monitored [21]. Therefore, syndromic surveillance systems can be designed to minimize the main limitations of passive surveillance methods based on laboratory confirmation and disease reports by clinicians [20], namely: chronic under-reporting; a long time lag between outbreak onset and diagnosis; and a low sensitivity as a result of the high specificity of these methods. The low sensitivity of traditional surveillance relates to the focus on one disease or a list of reportable diseases, and the dependence on the ability of the clinician to recognize the clinical signs of specific diseases, a special limitation in cases of rare or emerging diseases [26, 27, 14].

In Figure 2.1 the timeline and population coverage associated with different surveillance strategies is schematically presented for three different target populations: humans, livestock and companion animals. Syndromic surveillance aims at reducing

the time lag associated with passive surveillance by monitoring populations before laboratory confirmation. Under-reporting is also minimized by the systematic, continuous screening of information at earlier stages in the disease process. As illustrated in Figure 2.1, population coverage is reduced as the timeline of the disease process continues from the general population to laboratory confirmation of diseases. Doherr and Audige (2001) [1] have noted that in this "pyramid of scrutiny" the animal owners and the veterinary practitioners act as a serial testing scheme, and the volume of laboratory submission reflects their judgement on the cost-benefit ratio associated with the laboratory tests.

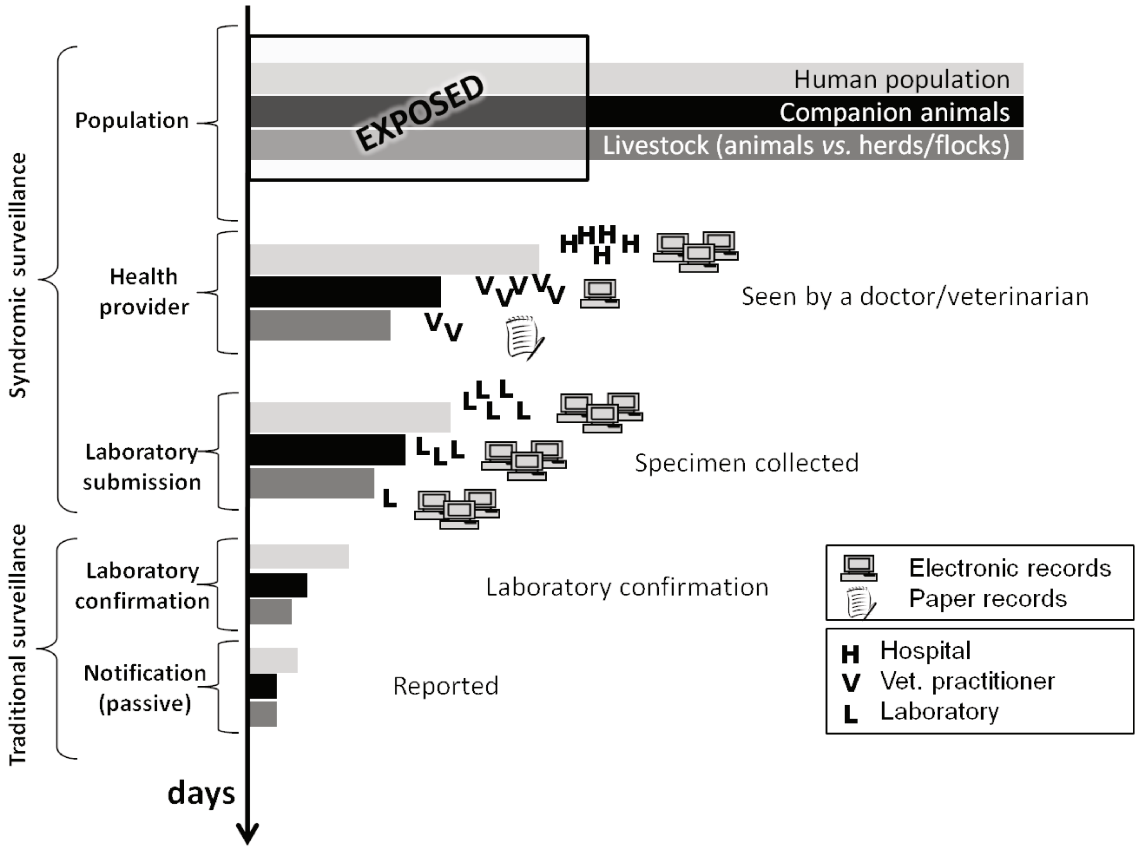


Figure 2.1: Schematic representation of the disease continuum in a population, and the surveillance opportunities according to population targeted, and type of data used. The scheme illustrates the proportions of subjects in each step of the disease process, for each of the three populations, in comparison to their initial population. The absolute number of livestock, companion animals and humans exposed to any given disease is not likely to be equal, and the top bars should be interpreted as the scaled total population. Proportions are illustrative only. Similarly, icons are not intended to represent a true count, but to illustrate comparative abundance.

The scheme in Figure 2.1 also indicates the loss in timeliness as surveillance is applied further along in the disease process. Timeliness refers to "the difference between the onset of an outbreak and the discovery of the outbreak" [22]. Buckeridge (2007) [28] reviewed the determinants of detection in automated surveillance systems in public health, pointing out characteristics of the system and of the outbreak that affect detection. The exact characteristics of the outbreak are unpredictable, but systems should be designed based on the expected characteristics of the disease(s) that it aims to detect [21]. The characteristics of the system listed by Buckeridge [28] were the choice of data source, the sampling strategy of the system, and the detection algorithm choice and settings.

The gain in timeliness as surveillance is applied closer to the top of the scheme shown in Figure 2.1, in comparison to the reporting of laboratory results, is usually based on the assumption that outbreak discovery closely follows the identification of positive cases [14]. In reality, this will only be true for the introduction of diseases in previously free zones/countries (any positive case is considered an alarm), and in this special case laboratory confirmation depends on the veterinarian having suspected the disease despite its absence in the region, and the laboratory having the specific test for it. Where the correct tests are not ordered/performed, or the outbreak event represents a sudden increase in the incidence of an endemic disease, its detection would likely occur much later in the disease process, if at all, in a situation where continuous statistical monitoring is not in place.

2.4 Syndromic surveillance initiatives in veterinary medicine

Scientific literature was reviewed using the following Medical Subject Head-ings (MeSH): cluster analysis, disease outbreak/veterinary, biosurveillance, medical informatics applications, and public health informatics. Keyword searches were pri-

marily applied on PubMed and CAB Abstracts. The search was last updated in January 2011. Electronic grey literature was searched using these terms and also "syndromic surveillance" and "early disease warning". Proceedings of the annual conferences of the International Society for Disease Surveillance (ISDS), symposiums of the International Society for Veterinary Epidemiology and Economics (ISVEE) and Conference of Research Workers in Animal Diseases (CRWAD) dated back to 2000 were screened individually. References within the papers found were also scrutinized.

A cursory look at syndromic surveillance initiatives in veterinary medicine reveals that this is an incipient field, and that a clear definition as to which systems should be classified as "syndromic" is hard to achieve. We focused our review on any surveillance systems based on the systematic monitoring of animal populations, using data sources that are timelier than traditional passive surveillance (as indicated in the left-most brackets of Figure 2.1). For the sake of structuring this review, systems that focus primarily on detection of emerging diseases, registering only atypical cases, are listed separately from systems that target animal health surveillance as a whole. The latter are based on monitoring all clinical cases, aiming at detecting not only disease introduction, but also changes in trends of endemic disease. Systems that monitor animal health with the primary purpose of detecting zoonotic threats for public health protection are also grouped separately. For reference, all the systems are listed in Table 2.1(in chronological order by publication date). Peer reviewed papers evaluating the potential of specific datasets for syndromic surveillance, but not reporting the implementation of any system, are listed in Table 2.2.

Table 2.1: Published initiatives in veterinary syndromic surveillance

System/Ref	Location	Data	Focus	Animal type	Syndromes	Additional Notes
VetPAD [29]	New Zealand	Clinical data from practitioners	Surveillance of animal diseases	Livestock	Not aggregated (all clinical cases)	Use of hand-held computers. Engages participation by providing software that contributes to practice management.
Émergences [30]	France	Clinical data from practitioners	Early detection of emerging diseases	Any species, any country, any disease		Access through website, information includes follow-ups.
Rapid Syndrome Validation Project for Animals [31]	United States	Clinical data from practitioners	Early detection of emerging diseases	Livestock	Focus on 6 groups of non-routine clinical syndromes	Various options for electronic transfer of data.
National cattle health surveillance system [32]	The Netherlands	Unsolved cases by farmers or veterinarians	Early detection of emerging diseases	Cattle	Focus on individual diseases	Data compilation and analyses is done weekly by a surveillance team, not automated.
BOSS [33]	Australia	Observations from producers and stock workers	Surveillance of animal diseases	Livestock	Software (BOVID) receives input concerning disease signs, and groups episodes into organ systems	Takes advantage of audience in daily contact with animals; Software to help producer with diagnosing the problem engages participation.
Purdue University-Banfield National Companion Animal Surveillance [34]	United States	Clinical and laboratory data, direct transfer	Sentinels for zoonotic diseases; portal for evidence-based medicine	Companion Animals	Retrospective pilot: tick and flea vector activity; leptospitosis and ILI. Plan to focus on other syndromes	Makes use of already computerized and centralized database, allowing for daily automated analysis and great geographical coverage.
Using pre-dx data from vet. lab. to detect disease outbreaks in companion animals [35]	United States	Laboratory microbiology tests submissions	Sentinels for zoonotic diseases	Companion Animals	Direct map of test orders into 11 syndromic groups	Makes use of already computerized database, allowing for daily automated analysis. Use of test orders is timelier than results.
LAHVA: Linked Animal-Human Health Visual Analytics [36]	United States	Clinical data from human and pet hospitals	Sentinels for zoonotic diseases	Companion animals	Pilot: seasonal flu and wastewater contamination	Links in one tool the surveillance in public and animal health
FarmFile [37]	United Kingdom	Laboratory results	Surveillance of animal diseases	Livestock	Focus on "Diagnostic Not Reached" events to assess the risk of new diseases emergence	Not real-time, post-result based, but the focus on non-diagnosed is innovative and adds values to the current surveillance.
SAVSNET [38]	United Kingdom	2 steps: 1) laboratory results; 2) real-time practice-based	Surveillance of animal diseases	Companion animals	Piloted using gastro-intestinal syndrome	Focus on information sharing to benefit not only population medicine, but also individual, evidence-based medicine.
Syndromic surveillance among livestock entering an auction market [39]	United States	Animal observations by veterinarian during auction market days	Surveillance of animal diseases	Livestock	Syndromic groups	Conceptually, can be implemented in handheld computers and give immediate feedback.
Alberta Veterinary Surveillance Network [40]	Canada	Disease and non-disease events from practitioners	Surveillance of animal diseases	Livestock	Syndromic groups	Part of a network supported also by pathologists and an investigation network.
Ontario Swine Veterinary-based Surveillance System (OSVS) [41]	Canada	Clinical data from practitioners	Surveillance of animal diseases	Livestock	Summarized by body system and production effects	Formally evaluated to assess compliance, completeness, coverage and timeliness. Results show good acceptance.

Table 2.2: Peer reviewed publications investigating the potential of different datasets in implementing veterinary syndromic surveillance systems.

Study	Location	Type of data	Goal	Animal type	Syndromes	Evaluation
Salmonella outbreaks detection [42]	United Kingdom	Laboratory results	Public and animal health surveillance	Livestock	Salmonella Typhimurium cases	Assess the improvements needed in the data collection process to allow for the implementation of early detection systems.
Laboratory data use for syndromic surveillance [13]	New Zealand	Laboratory submissions	Animal disease surveillance	Livestock	Test orders directly mapped into syndromic groups	Discusses the potential of the data for use in syndromic surveillance, and the inherent biases.
West Nile virus outbreak detection [43]	France	Clinical data from practitioners	Sentinels for humans	Horses	Neurological clinical cases	Retrospective analysis of an outbreak: alarm could have been 4 weeks earlier.
Early-warning system to reduce abortions in dairy cattle [44]	Denmark	Clinical data from practitioners	Animal disease surveillance	Livestock	Abortion	Evaluation of the system included costs of false alarms versus the cost of operating the system.
Detection of abortion in mares [45]	United States	Laboratory submission	Animal disease surveillance	Horses	Abortion	Retrospective analysis of an outbreak: Would have detected 1 week earlier.

2.4.1 Syndromic surveillance based on notification of atypical cases

Vourc'h et al. (2006) [46] presented a list of 14 emergence events associated with animal disease, and claimed that in most of these the key to detection was the observation of unusual signs or an unusual combination of signs. The same authors argued that focusing on solely reporting atypical cases (as opposed to monitoring trends for several unspecific clinical signs) can reduce the reporting load and requirement for disciplined coverage associated with the general syndromic surveillance approach.

The Èmergence system [30] was developed in France based on two components: a farmer component via routine surveys on farms, and a veterinarian component [30]. The veterinarian participation is based on atypical clinical case notification on a website (INRA – National Institute for Agricultural Research), and follow-ups. Monthly confirmation of vigilance is requested from veterinarians not reporting any atypical cases. The system also tracks diseases with emergence potential and/or known public health importance. The system currently focuses on bovines but it is built to be generic allowing its application to "any species, any country, any disease".

Passive reporting of atypical cases is one of the components of the National Cattle Health Surveillance System implemented in The Netherlands in 2003 [32]. Farmers or veterinarians report incidents not fully understood, motivated by the availability of specialists who visit the farm free of charge, in order to collect detailed information and investigate the problem. The system is complemented by the continuous collection of census data, pathological diagnosis of carcasses, toxicology tests, and periodical prevalence studies. While this represents an innovative system for early disease detection and information collection, the data compilation and analysis is performed by a surveillance team meeting weekly, rather than automated. The team looks for signs of introduction of specific emerging diseases, or analyzed trends of particular diseases, rather than grouping information into syndromes. Quarterly

reports are made available to the public.

The Rapid Syndrome Validation Project (RSVP) was first developed for public health, and later applied to cattle populations (RSVP-A) [31]. Clinical presentations are grouped into six syndromic groups that purposely focus on less common endemic disease presentations, and exclude the most common diseases and production problems. Several forms of data capture are available for veterinarians to report observed cases, including hand-held computers, cell phone, phone and fax lines, and the Internet.

2.4.2 Syndromic surveillance based on analysis of all clinical cases

Because the clinical signs of diseases observed in animals can vary depending on a great number of factors [47], disease introduction events may initially present as a collection of unspecific signs. Practitioners may therefore fail to diagnose diseases outside their sphere of experience. Alternatively the signs may not be specific enough to allow recognition that a new disease has been introduced [48]. Recognizing this, more surveillance systems are designed to monitor general signs, rather than specific diseases or only atypical cases.

In 2003 McIntyre et al. [29] reported on VetPAD, an initiative in New Zealand which aims to take advantage of veterinary practitioner data to improve disease surveillance capability. Understanding that for the system to be sustainable and keep veterinarians engaged it needed to be simple, and offer some advantages for participation, the initiative was based on providing software that would help the practitioner manage her/his practice using a handheld computer, which would electronically transfer data to the surveillance program. The data recorded includes all clinical cases attended by the veterinarian, and goes beyond diagnosis, recording also procedures, treatment, laboratory samples, medications, etc.

Initiatives to collect practitioner data are also being developed in Canada, through the Alberta Veterinary Surveillance Network [49, 50, 40] and the Ontario Swine Veterinary-based Surveillance System (OSVS) [41]. In the former veterinarians are encouraged to report all their daily animal health consultations. The veterinary surveillance system is also supported by pathologists and an investigation network, through which producers and other people in contact with livestock can report atypical observations. The OSVS is focused on swine veterinarians, using a variety of recording systems including paper forms and handheld computers, adapted to each clinic's management.

In Australia a system for electronic capture of syndrome data from livestock has been piloted [33, 27, 51]. The Bovine Syndromic Surveillance System (BOSS), a voluntary, producer-driven surveillance system, extends the target audience beyond the veterinarians, including lay observers who are in daily contact with cattle, such as stock inspectors, farmers and stock workers. The method used to engage participation is to provide a generic cattle disease diagnostic program — based on the BOVID system [52] — through which the producers can get a ranked list of differential diagnoses based on the signs observed in their cattle, and be advised of the precautions to take. The information that the producer feeds to the software for decision are exactly those that the surveillance program can take advantage of: animal characteristics, numbers affected, time and place of occurrence, duration of the disease event, and management information regarding the herd.

Recognizing that new or emerging diseases can go undiagnosed due to the lack of specific tests, in Great Britain a system of syndromic surveillance has been developed based on laboratory submissions for which a diagnosis was not reached. Building on the Veterinary Investigation Diagnosis Analysis (VIDA) system, the FarmFile system [37] among other improvements, included statistical monitoring of the ratio of "Diagnosis Not-Reached" (DNR) samples to the total samples processed. Even

though this system is not designed to operate in real-time and uses information from the test results phase, a number of syndromic surveillance techniques are adopted by FarmFile, including grouping test requests (including DNR) according to the body system affected, and an on-going monitoring of trends. Moreover, the focus on DNR samples represents an innovative initiative, potentially increasing the ability of the current surveillance to account for emerging diseases based largely on what was previously discarded data.

All the systems previously mentioned are focused on livestock. Syndromic surveillance systems targeting companion animals are usually designed with focus on public health, as discussed in the following topic. An exception is Small Animal Veterinary Surveillance Network (SAVSNET) [38], in development at the University of Liverpool. Besides monitoring disease trends, the project also aims at making the collected information via reports on a website.

Also in the United Kingdom, the National Animal Disease Information Service (NADIS) [53] deserves attention for its support to animal disease monitoring and evidence-based medicine. Even though the system is not syndromic or prospective, it does include a unique network of 60 veterinary practices and 6 veterinary colleges, monitoring diseases in cattle, sheep and pigs, and publishes publicly available reports of disease trends and parasite forecasts.

2.4.3 Syndromic surveillance focusing on public health (animals as sentinels for human diseases)

Ashford et al. (2000) [54] and Davis (2004) [55] reviewed the role of veterinarians in the preparedness against bioterrorism, based on the fact that almost all the biological bioterrorism agents listed by a group of experts gathered by the CDC in the United States in 1999 are zoonotic. Rabinowitz (2006) [56] reviewed several diseases with

bioterrorism potential and the role of animal populations in their detection. This is based on the assumption of one or more of the following factors being true: the 'sentinels' have increased susceptibility, would present with a shorter incubation period, are likely to be exposed sooner or more intensively and continuously through the environment; or simply because the concomitant observation in humans and animals would add confidence to the detection of a natural or introduced disease threat. As Figure 2.1 illustrates, depending on the disease, the number of sick animals can indeed exceed the number of sick humans. Moreover, domestic animal populations may be easier to observe or test.

Syndromic surveillance systems collecting animal health data for public health surveillance focus mainly on zoonotic diseases. These systems have thus far been largely based on companion animals, due to their proximity to humans, but their choice of targeted animals can also be based on the susceptibility of the different species, and their potential to signal disease before humans. Examples of the latter are the systems based on the higher susceptibility of crows and horses to West Nile virus [57, 47, 58, 59, 43]. One unique initiative highlighted the potential of zoo animals as sentinels, focusing also on West Nile virus detection [60].

Animal data have been incorporated into a few implemented syndromic surveillance systems for human populations. Those reported in the literature include: the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) [61], the North Dakota Electronic Animal Health Surveillance System [62] and the Multi-Hazard Threat Database (MHTD), a disaster preparedness project of the North Carolina Department of Agriculture and Consumer Services (2007) [63]. Brianti et al. (2007) [64] investigated the potential for improving public health surveillance of leishmaniosis by a retrospective survey which included data from veterinary practitioners and from hospitals.

Glickman et al. (2006) [34] highlighted the gap in our understanding of the dynam-

ics and disease burden in companion animals even though they are in daily contact with humans. The authors were responsible for implementing a National Companion Animal Surveillance Program (NCASP) in the United States in 2004 which took advantage of large amounts of computerized data from a major chain of pet hospitals in that country (450 hospitals), complemented by access to the computerized database from a network of diagnostic laboratories serving 18,000 pet hospitals. The system allows daily data analysis of all clinical visits to the hospital network. Results based on monitoring tick infestation, leptospirosis in dogs and the occurrence of influenza-like illness (ILI) in cats, have demonstrated the feasibility of conducting parallel syndromic surveillance in animals and humans.

Shaffer et al. (2007) [14] evaluated the use of companion animals as sentinels of infectious diseases in humans by the implementation of syndromic detection of diseases using laboratory submission requests, also taking advantage of the already available, electronic database of a laboratory network. Microbiology test orders were transferred daily, and directly mapped into 11 syndromic groups monitored independently. The authors report the positive results in using the system for population surveillance in a timely manner, and highlight the wide geographic coverage given by one single source of data.

Maciejewski et al. (2007) [36] reported the construction of a framework for joint analysis of human emergency room data and veterinary hospital data (mostly pets), called Linked Animal-Human Health Visual Analytics (LAHVA). Human data are processed daily, while animal data are received in batches every 13 weeks. The inclusion of animal data is considered to add sensitivity and specificity to the surveillance program, and takes advantage of the lower privacy concerns regarding animal data. Besides temporal analyses, the system's advantages include the integration of different data sources, and the visual analytic tools that integrate human and animal data. Testing of the system was performed by retrospective analysis using seasonal

influenza and wastewater contamination events.

2.5 Data sources for syndromic surveillance in veterinary medicine

In public health it has been noted that the ultimate choice of target for syndromic surveillance (according to the scheme in Figure 2.1) depends on the balance between quality and timeliness, and the weight of the costs of false alarms and missed alarms. In animal health the decision is further complicated by the scarce availability of suitable data [51, 65], which for the purposes of syndromic surveillance should be acquired continuously, in an automated routine, be electronically stored and timely available [21]. Moreover, animal data are subject to more non-disease variation than human disease data [42]. The rate of seeking care is not only related to the awareness and severity of diseases, as in humans, but also, especially in livestock, by cost. In turn, the rate of laboratory test submission is not only a result of diagnostic concerns; specimen collection can also take place for a variety of other reasons such as trade certification, food safety monitoring, etc.

Shephard (2006) [51] listed barriers to the development of syndromic surveillance systems in animal health as including the great diversity in species, production and purpose, and the hierarchical structure of animal populations (in food production). Additional barriers relate to the poor availability of data sources in comparison to human medicine, due to less frequent capture, often in a non-computerized format, as well as less well developed data standards. This section will review how some of the initiatives listed in the previous section have dealt with the problem of finding adequate datasets, and their strategies to increase population coverage compared to voluntary notification of confirmed cases.

2.5.1 Voluntary notification

Coverage of systems based on passive notification can be increased by understanding the behaviour of the reporting entities - veterinarians, animal owners, etc - and designing ways to positively influence them. The challenge is to find a strategy that is not only successful, but also sustainable [66]. As early as 1998 Gobar et al. [67] reported a program for surveillance of causes of death in dogs, using the Internet to survey small animal veterinarians. The novelty resulted in 25 veterinarians actively submitting case materials and promoting discussion, but no report of the sustainability of the system and rate of participation over time was found. Shephard (2006) [51] reported a study to investigate the sustainability of implementing a system based on veterinary voluntary reporting of clinical livestock cases. The results indicated that the system would likely not be sustainable, especially due to veterinarians' perceptions of limited personal value associated with participation, and a view of increased risk of penalty in case of reporting.

Systems that focus only on the reporting of atypical cases, such as RSVP-A [31], the national cattle surveillance system in The Netherlands [32], and Èmergences [30], aim at keeping veterinarians involved by reducing the time demanded of them - the systems provide easy and quick reporting, through handheld computers or websites. The RSVP-A and the national cattle surveillance system in The Netherlands also promote participation by giving information feedback to the public (in the form of publicly available quarterly reports on the latter, but restricted to participating veterinarians in the RSVP-A). However, maintaining compliance over time remains a great challenge [21]. Reports on evaluations of the sustainability of these systems could not be found in the literature.

2.5.2 Clinical data

In contrast to human medicine, in veterinary clinics the payment is due, in most cases, at the time of service, with no requirement to transfer data to third-party payers, such as insurance companies. This has caused veterinary clinic data recording to be primarily focused on client and invoice management, and there has been little incentive to develop and implement standards for disease coding [65].

Despite these bottlenecks, the use of computerized records is becoming standard practice in companion animal medicine, offering opportunities for the collection of syndromic data. The SAVSNET for instance [38], which plans to use practice-based, real-time collected data in its next implementation step, will take advantage of the fact that around 20% of pet clinics in the UK use the same software for practice management. The lack of data standards in veterinary medicine, however, means that data integration among clinics using different software remains problematic.

The opportunities for data integration increase with the growth of corporate veterinary practices [68]). The Purdue University-Banfield National Companion Animal Surveillance [34] reported a coverage of 2% of the total pet dog and cat population in the United States, by using the centralized database of Banfield, a pet hospital chain widely spread across the country [36], and whose demographic and medical information is completely computerized. Data from the same hospital network are also used by the LAHVA initiative [36].

Automated collection of clinical data is harder for systems targeting livestock due to the lower level of computerization in large animal practices, compared to companion animal practices. These systems depend on the willingness of the veterinarian to comply and take the extra effort of submitting their routine data to a surveillance system. Engagement is sought by adapting the recording system to the routine recording process of the practice or by offering feedback to the veterinarians and

farmers by means of a complete investigation network, as in the Alberta Veterinary Surveillance Network [49, 50, 40]. Assessments of system sustainability have not been reported.

Robotham and Green (2004) [69] stated that systems that depend uniquely on voluntary transference of routine clinical data by veterinarians are not sustainable without any return to the veterinarian. Proposed methods to increase veterinarian engagement include continuous training, return of the information collected with added value to the practitioners [70, 71], and financial incentives to reporting [40].

2.5.3 Herd management data

Automated monitoring of herd management data and indicators of production quality have been reported and reviewed [71, 72]. However no reports on implementations of syndromic surveillance systems based on these data were found. Mork et al. (2009) [73] compared data kept on farmers' records to the data reported by veterinarians to a dairy industry cattle database in Sweden, and showed that only 54% of the disease events registered by farmers were treated by a veterinarian. Even for those events that were reported by both groups, the farmers kept information that was more detailed and specific than that reported by the veterinarians.

The BOSS system [27], even though based on disease events, can be considered a system based on direct herd information, as it represents an effort to involve farmers directly. Rate of underreporting should theoretically be low, as it targets the population of animals becoming sick, not the population of animals for which veterinary care was sought. However, population coverage will be limited by access to (and willingness to use) a computer. This is becoming less and less of a problem, as an increasing number of herds are already managed with the help of computerized systems.

The increase in the use of computerized herd management tools could offer another opportunity for surveillance. The lack of uniform standards among systems may however complicate integration, and it would suffer from the same problems discussed for capture of computerized clinical data.

2.5.4 Laboratory data

Laboratory test requests are a type of syndromic data. They are timelier than results, and can be grouped in syndromes according to the nature of the disease and/or symptoms observed by the veterinarian [11, 74, 10]. Stone (2007) [13] investigated the potential of using laboratory test requests for syndromic surveillance in veterinary medicine and reviewed the potential biases associated with this type of data. The author also pointed out the variability in the submission rates year to year, and misclassification biases (veterinarian not submitting the right sample or requesting the correct test), but concluded that the data are suitable for syndromic surveillance.

Laboratory test requests are more often automated and electronically recorded than clinical data [75] and therefore these data allow for the construction of a sustainable surveillance system. Laboratories also represent a more centralized source of data, especially in livestock medicine. However, their use depends on the willingness of data owners to share these data [34].

It has been reported that laboratory test orders suffer from the low submission of specimens as part of the diagnostic process in veterinary medicine [76]. However, Shaffer (2007) [14] assessed the potential of microbiology test submissions for syndromic surveillance in companion animals assuming that the consistency of test orders over time allows for the use of these data in prospective monitoring, and that increases in the number of test orders can be used as indicators of an increase in disease burden. The availability of historic data is another advantage of laboratory

data in veterinary medicine over other types of data, since some estimation of a baseline of disease burden is needed in syndromic surveillance to compensate for the lack of denominator data.

Laboratory test requests screen a larger proportion of the animal population than sick animals, as animals can be tested for different purposes. Zhang et al. (2005) [15] reported that four different purposes were recorded as reason for test requests in their laboratory data: diagnostic, export testing, government monitoring, and industry monitoring.

The use of laboratory data in veterinary syndrome surveillance appears to be a growing field. The Canadian Animal Health Surveillance Network (CAHSN), part of the Canadian Food Inspection Agency, is establishing a network of federal, provincial and university animal health diagnostic laboratories to implement an early warning system for animal diseases in real-time, especially diseases with zoonotic potential [77]. The website of the Gluck Equine Research Center [78] reported that the Veterinary Diagnostic Laboratory in the United Kingdom is developing a syndromic surveillance system in near real-time, also based on monitoring sample submissions. Table 2.2 provides additional examples of investigations of the potential of laboratory data on early disease detection [42, 45].

2.5.5 Others

The limited number of implemented syndromic surveillance systems in veterinary medicine use the sources of data noted above. However, a variety of alternate data sources are being explored for their syndromic surveillance potential.

Egenvall et al. (1998) [79] and Penell et al. (2007) [80] have assessed the quality and completeness of computerized insurance data from dogs and cats, and horses respectively. If the use of health insurance grows in veterinary medicine, these data

may provide a source of centralized information, and the use of coding standards may become more widespread.

The work of Van Metre et al. (2009) [39] investigated the use of direct observation in auction markets. The advantage of this method is associated with the opportunity to screen a large number of animals at once, and especially of reaching smaller operations which may be systematically excluded of other surveillance methods due to a lower frequency of veterinary care [39]. Even for the population under veterinary care, observations in auction markets may be timelier than the observation of clinical cases.

Abattoirs represent a unique source of data for veterinary surveillance, compared to public health. Engle (2006) [81] used the condemnation data available through the electronic Animal Disposition Reporting System (eADRS) from the Food Safety and Inspection Services (FSIS) in the USA, and concluded that a swine erysipelas outbreak in Iowa and Minnesota during July 2001 could have been identified up to 10 months earlier if automated analysis of the data had been in place. Weber (2009) [82] also evaluated the potential for using condemnation data to set up an animal health monitoring system. Benschop et al. (2008) [83] provided a thorough temporal and spatial analysis of abattoir data collected by the Danish Swine Salmonellosis Control Programme, and its potential for temporal monitoring and to improve surveillance design.

McNamara (2007) [60] drew attention to the fact that zoos are an often overlooked source of surveillance data. The author highlighted their role as epidemiological monitoring stations, as they "contain a population of known individuals at a point-source location that are followed over time". Zoos have historical data on animal tests that are performed as animals are received and regularly throughout their life. The author reported the success of a "Surveillance for West Nile Virus in Zoological Institutions" that ran successfully for 4 years, and is now being used as a model for

expanding H5N1 surveillance in the United States.

Smith et al. (2006) [84] presented even more innovative ideas to collect livestock health information that goes beyond clinical, sporadic information. The authors are developing a telemonitoring system that continuously transfers animal health data from devices permanently worn by the animals. Data would be collected and monitored continuously through devices placed in points of animal agglomeration within the farm. Evaluations of the system, especially of its cost-effectiveness, are not yet available.

2.6 Implementation of disease aberration detection from animal health data

Figure 2.2 summarizes the process of using animal health data sources to monitor disease trends and detect temporal or spatial aberrations in the number of cases. Comprehensive reviews of each of the components of a syndromic surveillance system are available elsewhere [85, 21, 51, 35]. The focus of this review is on the particular characteristics of veterinary syndromic surveillance, in livestock and companion animals.

2.6.1 Definition of events and syndromes

Automated disease monitoring systems must make a clear definition of what constitutes one event in the data available, as the statistical analyses are typically based on observed counts. For companion animals each patient entry is usually considered to represent an event, as long as there is no evidence that repeated encounters are associated with the same health event. In the case of livestock, health events are usually enumerated at the herd level.

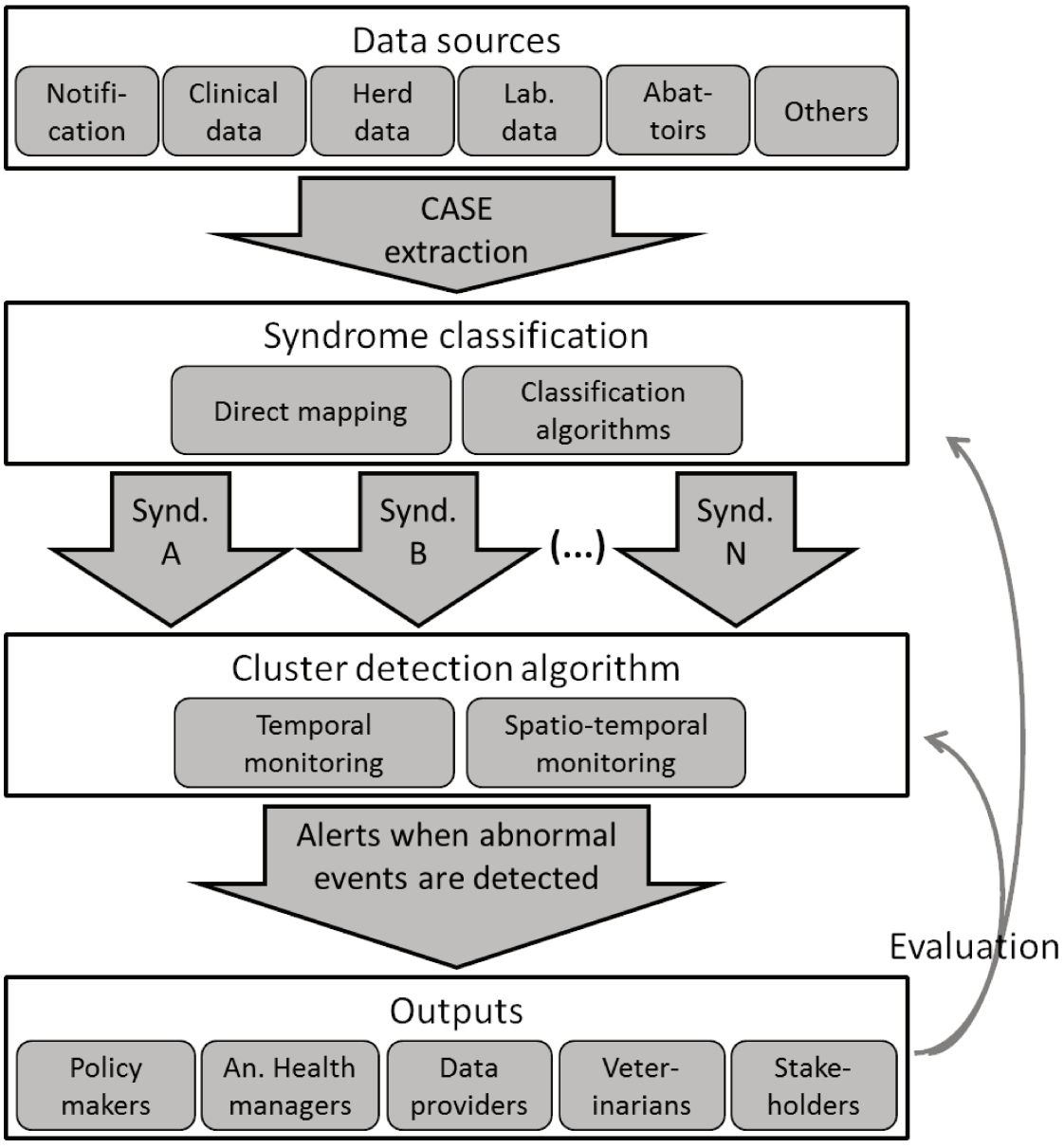


Figure 2.2: Process of monitoring disease trends and detecting clusters using animal health data sources. Synd = Syndrome.

Once the events have been identified, identifying the criteria to be used to group these into specific syndrome(s) and devising reliable/automated data classification protocols are essential components of an early epidemic detection system [86, 21]. The classification protocol must be based on the system's goals, but must also relate to the specific data in hand, as the data grouping will likely influence the performance of the alert detection algorithm [14].

In public health, clinical data are usually coded for billing purposes using standard nomenclatures, and coding standards for laboratory data are also available, allowing for the integration of multiple sources of data. When clinical or laboratory data are coded, classification can often be performed by directly mapping codes into syndromes. When data are not coded, automated classification algorithms must be trained to recognize relevant medical information in the data, and determine the syndrome associated with each event/unit. A common example is the use of text mining algorithms to extract information from text entered by nurses during triage in emergency rooms (chief complaint data) [86].

Vocabularies and standards for data classification are not as uniform in animal health [65] as is the case for human health. Wurtz and Popovich (2002) [87] reported on the range of codes that do exist, but noted that these are not widely used in clinics. Numerical codes are available through the Standardized Nomenclature for Veterinary Diseases and Operations (SNVDO), and the National Animal Health Reporting System (NAHRS). In addition veterinary input has been incorporated into more general health ontologies such as HL7, LOINC, and SNOMED (which has been renamed the "Systematized Nomenclature of Human and Veterinary Medicine"). Bartlett et al. (2010) [71] reported that the Veterinary Medical Data Base (VMDB), created in 1964 to store all clinical cases seen in veterinary teaching hospitals across North America, is not up to date because several schools are behind in coding their cases for upload to the database; a problem that would not exist if hospitals already

coded their cases routinely under a standard system. None of the surveillance systems presented in this paper reported using a standard classification system.

In the absence of standard nomenclature, a key element in the implementation of any syndromic surveillance system is the definition of syndrome groups and the rules to assign events membership. Shaffer (2007) [14] reported a consultation with a group of seven veterinarians, together with staff from the diagnostic laboratory who handle data regularly, to determine which laboratory test orders should be mapped into which syndromes. The final syndromic groups identified during this consultation were: respiratory, GIT, neurologic, dermal, reproductive, endocrine, hepatic, infectious, febrile, renal, and non-specific. Stone (2007) [13] also mapped laboratory data into groups based mainly on organ systems. After some standardization of the data these records were grouped into the following categories: reproductive system, abortion, alimentary system/oral, anorexia/depression/malaise, circulatory/oedema/anaemia; diarrhoea/dysentery, lymphoreticular, mastitis, musculoeskeletal, nervous system, perinatal losses, respiratory system, skin/photosensitivity, sudden death, and urinary/renal. Samples for which a diagnosis was not reached in the FarmFile system [37] were mapped into syndromes based on body system together with the information given by the veterinarian at submission concerning observed clinical signs. The final syndromic groups in that system were: systemic, digestive, respiratory, urinary, musculoskeletal, nervous, skin, circulatory, reproductive, other, disease type unknown, mastitis and fetopathy. The implementation of this syndrome mapping within FarmFile generated feedback which improved the data collection forms. The list of clinical signs to be used by veterinarians when submitting samples was revised, in order to improve syndromic classification of the data.

For most of the systems based on clinical data listed in this review the protocol for classifying data into syndromic groups could not be found in the literature, or none

had yet been implemented. A number of the systems are still being piloted using one or a few specific syndromes, and these have so far been identified retrospectively. The RSVP-A system uses six syndromic groups, but classification is decided and entered by the veterinarian reporting the atypical cases observed; this is also the case of the Alberta Veterinary Surveillance Network. In the BOSS system [51] the BOVID software, a rule-based diagnostic program designed to identify the most probable diagnosis based on clinical signs reported, classifies the reported cases into syndromic groups (based on organ system) to deliver counts by syndrome. The syndrome groups used within BOSS/BOVID are: body; ears/eyes; airways; GIT; genital and urinary system; nervous; skin; cardio-vascular; death or reduced production; and muscle, bone or gait abnormal.

2.6.2 Aberration detection algorithms

Monitoring of time series data in surveillance can be retrospective or prospective. Retrospective surveillance is used to explain temporal and spatio-temporal patterns in data, and is therefore used in the generation of hypotheses. In syndromic surveillance, where the focus is outbreak detection, statistical analysis is prospective, aiming at detecting meaningful changes from the expected range of data values, which are referred to as "aberrations" [88, 89]. Mandl et al. (2004) [21] summarized the methodological stages to process data for outbreak detection, once events have been classified into syndromic groups, as: evaluation of historical data to establish a baseline model for the expected number of cases; comparison of observed values to baselines, to detect abnormal activities if occurring; culminating in an evaluation of the alert and a decision as to whether notification and investigation should take place.

The choice of algorithm to detect abnormal activities is based on the type of data (number of time series to monitor, whether rates or counts are monitored, rare versus

frequent counts, temporal or spatio-temporal data); the availability of historical data to construct baselines; the nature of the disease being monitored (whether outbreaks are expected to occur as 'spikes', a sudden or slow increase); and an assessment of the desired balance between sensitivity (ability to detect true alarms) and specificity (ability to avoid false alarms). Algorithms for outbreak detection have been thoroughly reviewed elsewhere [90, 91, 92, 88, 24, 51, 68]. The goal here is to list the methods that have been cited in the veterinary syndromic surveillance systems covered in this review. However, as many of the systems are still in their initial implementation phase, the types of algorithms being used were often not identified; and a column detailing this information could not added in Table 2.1.

For temporal analysis, control charts (such as cumulative sums and exponentially weighted moving averages) are the most commonly employed algorithms [51, 62, 14, 40, 82]. This is not surprising, as for most of the data sources used there is limited availability of historical data. Control charts require limited baseline data, using a small number of previous observations to establish thresholds of expected values, based on the assumption that those observations came from a pre-specified parametric distribution. New observations are compared to the thresholds, and the system is determined to be "out-of-control" if the observations fall beyond the calculated expected limits [93]. Performance is not optimal, since these methods do not exploit the full information content of the data, and because health data often violates the basic assumptions of control charts - that events are independent, stationary and normally distributed [94]. However, the popularity of these methods in public and animal health surveillance attest for their usefulness, especially when historical data are limited.

When historical information is available regression methods can be used. Published work on regression methods applied to veterinary data have thus far focused on retrospective analyses, as a means of assessing their potential for prospective mod-

eling. For instance the work of the Purdue University-Banfield National Companion Animal Surveillance [34] with clinical data, the analysis of the Danish Salmonella Control Programme data [95]), and the work of Kosmider et al.(2006) [42] based on laboratory detection of Salmonella in British livestock, are all examples which adopt this approach.

Geographical information is often used to aggregate data into demographic areas, after which temporal analysis is applied to these areas independently (Shaffer, 2007) [14]. Public health systems are usually restricted by privacy concerns regarding address information from patients [36], while in animal health systems the problem is the lack of geo-location data relating to health events. Often the only geographical information in the system refers to the practitioner location or postal code [14]. This represents a challenge for use in spatial analysis of animal surveillance, since the geographical radius of clients attended by each practitioner is not usually determined and may vary greatly, particularly in regions of low farm and/or practitioner density.

Where spatial cluster analyses were performed in the systems reviewed here, the most commonly reported method was the scan-statistic [45, 96], which can be performed with the freely-available software SaTScan [97, 98]. Spatial cluster detection using algorithms available within the R statistical package was reported in the case of the Alberta Veterinary Surveillance Network [40].

2.6.3 Evaluation

The "framework for evaluating public health surveillance systems for early detection of outbreaks" was reviewed in 2004 by a working group promoted by the CDC [74]. The document contains an operations checklist to review system-wide issues, data sources, data processing, statistical analysis, and epidemiological analysis, interpretation and investigation. It sets out a framework for description and evaluation of any

system as a whole, including: usefulness, flexibility, acceptability, portability, stability and costs. In a similar way, Stone (2007) [13] stated that a veterinary syndromic surveillance system should be evaluated for: population coverage, automation of data capture and transfer, value to users, detection efficiency of programmed algorithms, and contribution to claims of disease freedom.

More quantitative evaluation methods have been proposed to specifically evaluate the performance of various detection algorithms, using real or simulated data; and thus to evaluate the system's performance at the population level in the similar way to which test diagnostic performance is evaluated for individual testing. This includes the measurement of sensitivity, and specificity [99]. Kleinman and Abrams propose methodologies which also include an evaluation of the timeliness of a system [100] and the number of lives saved [101], based on the traditional Receiver Operating Characteristic (ROC) curves used for diagnostic tests evaluation.

Ultimately, the factors that affect the ability of any system to detect outbreaks also depend on the nature of the outbreak [28]. Evaluation of outbreak detection algorithms based on simulated data has been suggested in the literature. These evaluations may use wholly simulated data sets or may superimpose various patterns of simulated outbreaks onto authentic data. An overview of these approaches can be found in Buckeridge et al. 2005 [88]. A holistic evaluation of how all system components operate in real time is only possible once the system has been implemented. None of the systems listed in this review have been formally evaluated using the metrics described above; this is not surprising as most of them are still in development and few are fully operational. However, various authors have attempted to assess the quality of different system components. In those cases where any, even limited, evaluation was reported, notes have been added to Table 2.1 and Table 2.2.

The Ontario Swine Veterinary-based Surveillance System (OSVS) [41] was the only system for which an evaluation of the characteristics of data acquired through

practitioners' reports was performed. The authors estimated the level of compliance by comparing the data provided by practitioners against the submissions made by the same veterinarians to Ontario's Animal Health Laboratory. Completeness of data (measured in terms of the completion of each form field), coverage of the program, and timeliness for reporting were evaluated. Completeness actually increased from the first to the second year of the study, as well as coverage of farms in the province.

The Emergences system was not formally evaluated, but in 2003 Vourc'h and Barnouin [30] reported that over a period of six months the system received 33 notifications, two of which were considered atypical. Shaffer (2007) [14] reported that during the pilot study of the described syndromic surveillance system based on microbiological test requests nine clusters were detected. Follow-up investigations were able to link two of these to a true increase in the incidence of disease. Assessing sensitivity and specificity was not considered viable due to the lack of a gold standard for determining when outbreaks were really happening. The BOSS system was also not evaluated due to the lack of a standard against which the completeness of the data received from producers could be assessed [51]. Retrospective analysis of the data on LAHVA indicated that respiratory symptoms in dogs occur approximately 10 days earlier than is the case for humans, and that detection of eye-inflammation in dogs would also have served as a sentinel for humans in a case of wastewater contamination [36] Also retrospectively, Odoi (2009) [45] showed that an outbreak of abortion in mares could have been detected 6 days earlier.

2.7 Discussion

A pre-conference workshop at the ISVEE meeting in 2009 discussed the development and application of methods for effective surveillance in livestock populations [66]. Syndromic surveillance systems can meet several of the surveillance goals proposed

during that meeting, including: comprehensive coverage of many diseases within a single monitoring system, detection of emerging diseases, maximizing the value of existing data sources, integration of public health with veterinary data, development of new analytical methods, technological innovation, flexibility in the type of data available and desired system outcome, encouraging stakeholder participation, and an increase in negative reporting. This paper has discussed how syndromic surveillance in animal populations can help meet many of these goals.

Only systems that explicitly address outbreak detection based on systematic monitoring of animal population data have been included in this review. However, there is little doubt that disease control capabilities have also been enhanced by systems for disease monitoring which adopt novel approaches to data sharing, integration and visualization. The authors recommend the following examples to those readers interested in exploring the broader application of information systems to veterinary surveillance: the Michigan equine monitoring system [102]; the Pathman project [103]; the Rapid Analysis & Detection of Animal-related Risks (RADAR) [104, 105]; the FMD BioPortal System [96]; geographical information systems for the surveillance of bluetongue in Australia [106] and Italy [107]; the swine industry initiative for disease data sharing in Minnesota [108]; GLiPHA [109]; and the papers of Egbert (2004) [110] and Durr & Estland (2004) [103].

The initiatives reviewed have made use of several sources of clinical and diagnostic data in order to implement syndromic surveillance system in veterinary medicine.

Due to the lack of commonly adopted data standards, each syndromic surveillance system implemented in veterinary medicine to date has tended to develop and validate their own classification system. As long as common standards are not adopted, new systems will have limited capability to take advantage of the progress made by existing systems. While each method may be valid within its own architecture, the use of standards would enable data integration across heterogeneous datasets,

and allow comparisons among geographical locations and veterinary practices over time [29]).

As a result of the current limitations, most efforts to date have been directed towards developing animal health data collection strategies, analyzing historical data already available for their potential to support syndromic surveillance, or solving problems of data classification and integration; rather than focusing on the development of automated syndromic analysis. The concentration of effort on these early stages of development is evident when one considers the relatively plentiful supply of papers dealing with potential data sources, in contrast to those reporting the use of various aberration detection algorithms, illustrating systems outputs or evaluating operational syndromic surveillance systems. In fact, none of the listed initiatives contain all the components which characterize the more mature systems for early disease detection in public health. In consequence, the term "syndromic surveillance" has been applied throughout this review in a rather loose manner, since the term has been coined in reference to early disease detection systems based on the systematic monitoring or large amount of pre-diagnosis data. Not all of the systems reviewed here are strictly based on the classification of data into syndromes.

Despite differences in structure, all of the initiatives reviewed are making efforts to improve the quantity, quality and speed of information extraction from animal health data, and the lessons learned will support further advances in the development of the field of syndromic animal surveillance.

Many of the systems developed in veterinary medicine have attempted to solve data limitations by encouraging passive notification of cases or transfer of clinical data, directly from farmers, or by enrolling private veterinarians in the system. All these systems are dealing with sustainability issues. Information feedback or financial incentives to participating veterinarians have been used as strategies to sustain participation, but in general the lesson learned is that if data transfer demands extra

effort from participants, long term sustainability may not be possible.

Given the current barriers, diagnostic laboratories appear to provide a readily available source of data for syndromic surveillance in animal health. The less timely nature of laboratory data is compensated, in veterinary medicine, by its greater specificity when compared to clinical data. In addition there is reasonable availability of current and historical laboratory data in digital format, both for companion and livestock animals. In companion animal medicine, where computerization of records is already common, investments in the use of data standards will increase the value of clinical data for syndromic surveillance use. In livestock health, however, the use of laboratory data remains the most readily available and reliable source of electronic, continuously recorded data. Laboratories are typically centralized and can cover large geographical areas. However, it is also important that investment be made in data standardization within the livestock laboratory sector as this would allow for the integration of databases across broader geographical areas.

The expansion of syndromic surveillance in public health has fomented great improvements in the development and adaptation of aberration detection algorithms for use in health data, as demonstrated for instance by the work of several teams within the BioALIRT project, which has been sponsoring research on improving the timeliness of outbreak detection since 2001 [111]. Implementation of syndromic surveillance in public health has also resulted in the expansion of the field of infectious disease informatics. Several teams have documented their experiences in creating information systems and provided guidelines on the architecture necessary to conduct prospective, real-time surveillance [112, 113, 114]. Therefore, as quality animal health data become more readily available, the development of veterinary syndromic surveillance will be able to take advantage of the statistical and computational advances made in the public health field.

In all syndromic surveillance systems the primary output is some form of alarm

in the event of aberration detection. However, syndromic surveillance is not a replacement for traditional surveillance [11], and therefore once an alarm is triggered by the detection algorithm it must be reviewed by epidemiologists [74]. The design of the system should take into account the information that will be needed when making subsequent decisions and the outputs of the system, in case of any alarm, should contain all the information available from the syndromic dataset that may be of value [115]. The investment in syndromic surveillance may be wasted if, once a decision is made, the epidemiologist cannot count on an investigation team ready to respond to an alarm; the process for aberration follow-up should therefore be described as part of the syndromic surveillance system design [13].

Syndromic surveillance systems can confer benefits which go beyond the detection of true alarms [91]. They can support additional goals associated with animal health surveillance, such as: monitoring disease trends; facilitating the control of disease or infection; supporting claims for freedom from disease or infection; providing data for use in risk analysis, for animal and/or public health purposes; and substantiating the rationale for sanitary measures [116]. In practice most systems designed for early detection of disease, due to the longitudinal nature of their data collection, will also contribute to situational awareness, building a foundation for epidemiological research and hypotheses generation and testing, and thus provide support for evidence-based medicine. A number of the systems reviewed here intend to deliver the information extracted from the syndromic surveillance process to the public [38] or to participating veterinarians [29, 31, 34].

The development of the field of syndromic animal surveillance progressively enhances the animal health community's ability to detect and to respond to outbreaks. The automated and continuous collection of animal health data also facilitates the integration with public health systems, and represents a further step towards One Medicine. A recent review [117] noted that there have been on-going efforts to in-

tegrate human and animal data in surveillance initiatives since 2000. However, as pointed out in that review, none of the integrated systems have yet been evaluated and there are several barriers to data sharing between the two domains. Ethical and privacy concerns are not as restrictive in animal health data, as they can be in public health. Nevertheless, barriers to data sharing, mainly related to data ownership and proprietary information, and barriers to data integration due to the lack of commonly adopted standards continue to impair the communication within and between animal and public health data sources.

The longer experience of public health systems with syndromic surveillance has indicated that the cost of system maintenance and response to false alarms can only be justified by the system's contribution to more than event detection [118]. In veterinary medicine, the progress in the development of early warning systems has stimulated the review, improvement and expansion of data collection methods in animal health, though sustainability issues are now evident for systems based on voluntary notification or passive data transfer from veterinarians.

While in public health, syndromic surveillance can be based on sales of over-the-counter medicine, or emergency visits, in animal health the earliest type of syndromic data available is clinical. Several initiatives have shown that there is potential for clinical data to be used in the continuous monitoring of animal populations, but implementations in real-time still depends primarily on finding sustainable ways to collect and process clinical data from practitioners. To achieve this, investment is needed in systems which enable information flow from livestock practitioners to surveillance teams, and financial incentives are often necessary to guarantee practitioner engagement. Clinical data from companion animal medicine is more often computerized, and offers greater potential, but authors have indicated problems concerning data sharing, and unreliable flow of data from providers to the surveillance teams.

Until investments have been made to solve these issues, laboratory data continues to offer the greatest potential for syndromic surveillance in veterinary medicine. Similar problems to those associated with clinical computerized data such as the need for data sharing agreements and investments in data classification to allow integrations across different plataforms, apply to laboratory data. However, until data integration problems are solved, monitoring each single source of laboratory data may still offer geographical coverage greater than any single source of clinical data. Systems implemented directly with the data provider will minimize data flow issues.

This review has illustrated that the field of syndromic surveillance in veterinary medicine is incipient, but fast growing. As syndromic animal surveillance systems have developed over the past decade, limitations in the data available on animal health have become apparent. The lack of automated data collection limited opportunities for implementation of systematic monitoring systems; lack of computerized records limited automated analysis; and the lack of standards limited the integration across multiple databases. The costs of overcoming these barriers and implementing real-time monitoring systems are justified by their utility. Syndromic surveillance systems offer opportunities that go beyond early detection of diseases, providing information to aid planning and policy development.

Chapter 3

Exploratory analysis of methods for automated classification of laboratory test orders into syndromic groups in veterinary medicine

[0]Fernanda C. Dórea, C. Anne Muckle, David Kelton, J.T. McClure, Beverly J. McEwen, W. Bruce McNab, Javier Sanchez and Crawford W. Revie. PloSONE 2013, In Press

3.1 Abstract

Background: Recent focus on earlier detection of pathogen introduction in human and animal populations has led to the development of surveillance systems based on automated monitoring of health data. Real- or near real-time monitoring of pre-diagnostic data requires automated classification of records into syndromes — syndromic surveillance — using algorithms that incorporate medical knowledge in a reliable and efficient way, while remaining comprehensible to end users.

Methods: This paper describes the application of two machine learning (Naïve Bayes and Decision Trees) and rule-based methods to extract syndromic information from laboratory test requests submitted to a veterinary diagnostic laboratory.

Results: High performance (F1-macro = 0.9995) was achieved through the use of a rule-based syndrome classifier, based on rule induction followed by manual modification during the construction phase, which also resulted in clear interpretability of the resulting classification process. An unmodified rule induction algorithm achieved an F1-micro score of 0.979 though this fell to 0.677 when performance for individual classes was averaged in an unweighted manner (F1-macro), due to the fact that the algorithm failed to learn 3 of the 16 classes from the training set. Decision Trees showed equal interpretability to the rule-based approaches, but achieved an F1-micro score of 0.923 (falling to 0.311 when classes are given equal weight). A Naïve Bayes classifier learned all classes and achieved high performance (F1-micro = 0.994 and F1-macro = 0.955), however the classification process was not transparent to the domain experts.

Conclusion: The use of a manually customised rule set allowed for the development of a system for classification of laboratory tests into syndromic groups with very high performance, and high interpretability by the domain experts. Further research is required to develop internal validation rules in order to establish automated

methods to update model rules without user input.

Keywords laboratory order data, laboratory records system, classification, medical records, veterinary, animal.

3.2 Background

Disease emergence and bioterrorism events, especially since 2001, have highlighted some of the short-comings of traditional surveillance, generally based on laboratory test results and direct reporting [20]. Focus has shifted to earlier detection of pathogen introduction in human or animal populations, leading to the implementation of new techniques using data sources upstream to those typically used in traditional surveillance [114]; especially pre-diagnosis data that are already available and automatically collected [119], such as sales of over-the-counter medicine, absences from work or school, and patients' chief complaint upon visits to an emergency center [22].

Due to the lack of sensitivity of pre-diagnostic data, surveillance systems using this information target general groups of diseases, or syndromes, and are therefore often referred to as "syndromic surveillance" [3]. Grouping pre-diagnostic data into syndromes is the first step of implementing a syndromic surveillance system [119]. Valid, reliable, and automatic classification of syndromes was an essential component of early computerized epidemic detection systems [86]. When data are structured using standardised codes, such as the Logical Observation Identifiers Names and Codes (LOINC®) used in laboratories, the International Classification of Diseases (now on its 10th revision, ICD-10), or the Systematized Nomenclature of Medicine (SNOMED®) [85], syndrome classification can be performed by mapping those codes into syndromes. However, text mining or other machine learning tools can

be invaluable when free-text or semi-structured data are being used [86]. Naïve Bayes classifiers have frequently been used in syndromic surveillance when the input data are chief complaints (free-text typed in by nurses) at emergency facilities [86, 120, 121, 122, 123].

Rule-based methods were widely used before the computational capacity of common computers made it possible for machine learning methods to be widely adopted [123]. Nevertheless, they have remained a popular choice in the health field due to their transparency and interpretability. In the 2008 challenge organized by i2b2 (Informatics for Integrating Biology to the Bedside), which consisted of automatic classification of obesity and comorbidities from discharge summaries [124], the top ten solutions were dominated by rule-based approaches, demonstrating their efficacy.

Decision trees are a third type of classification algorithm recommended when results must be delivered to a broader audience, such as health workers, as it is also a relatively simple method to interpret [125]. Other machine learning algorithms used in the medical field include: Artificial Neural Networks (ANN) [126]; and Support Vector Machines (SVM) [127]. These methods are powerful, but both adopt a "black-box" approach; so that the way in which decisions are made by the classifier is not transparent. They have been used in more complex medical tasks, such as the interpretation of radiographs and studies of drug performance [128, 129, 130]. However, to the authors' knowledge, the use of these algorithms to classify health data for the purposes of syndromic surveillance has not been documented in the peer-reviewed literature.

In contrast to laboratory test results, on which traditional surveillance is based, laboratory test orders can be a valuable data source for syndromic surveillance, since they are collected and stored electronically in an automated manner, but are more timely for surveillance purposes than laboratory test results. Laboratory submission data have, for example, been incorporated into CDC's BioSense Early Event

Detection and Situation Awareness System [10]. Moreover, because there are fewer laboratories than sites of clinical care, the use of laboratory databases can provide more complete records and cover larger areas [114]. Besides changing focus to early diagnosis, modern surveillance systems are evolving to complete biosurveillance systems. This term is intended to imply a broadening focus, addressing not only human health but all conditions that may threaten public health, such as a disruption in the food supply, or large social and economic disruptions resulting from outbreaks of diseases in animals [114, 131]. Besides their role in the food supply and agricultural economy, animals could serve as sentinels for the detection of certain zoonotic diseases that may be recognized earlier in animals than in humans [55].

Animal data have been incorporated into a few surveillance systems for human populations, including: the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) [132], the North Dakota Electronic Animal health Surveillance System [62] and the Multi-Hazard Threat Database (MHTD) [63]. Glickman et al (2006) [34] and Shaffer et al (2008) [133] have investigated the value of animal health data as sentinels for public health. Despite the less frequent requests for laboratory analyses made by veterinarians compared to human clinicians, the authors hypothesized that, "the consistency of test orders over time is such that increases in cases of disease will result in detectable increases in the number of test orders submitted by veterinarians that can be identified using prospective analysis" (Shaffer, 2008 [133], page2).

An overview of the development of syndromic surveillance systems in the veterinary context has been provided in a recent review of the literature [134]. This review indicated that initiatives using laboratory data had been based on establishing direct relationships between test codes and syndromic groups. The use of clinical data has typically relied on syndrome definition being provided by the veterinarian. Machine learning or rule-based methods applied to the identification of syndromes in

animal health data had not been documented. This paper describes the exploratory analysis of such methods to extract syndromic information from laboratory test requests submitted to a veterinary diagnostic laboratory. These steps are part of the development of a syndromic surveillance system taking advantage of the centralized, computerized, and routinely updated sources of data provided by the Animal Health Laboratory in the province of Ontario, Canada. The initial phase of implementation, described here, focused on cattle sample submissions.

3.3 Methods

3.3.1 Data source

The Animal Health Laboratory (AHL) at the University of Guelph is the primary laboratory of choice for veterinary practitioners submitting samples for diagnosis in food animals in the province of Ontario, Canada. The number of unique veterinary clients currently in the laboratory's database (2008 to 2012) is 326. The AHL has a laboratory information management system (LIMS) that is primarily used for reporting the results of diagnostic tests.

Three years of historical data from the AHL were available, from January 2008 to December 2010. Cattle were chosen as the pilot species due to high volume of submissions from dairy and beef herds in Ontario. All laboratory test orders for diagnoses in cattle were extracted from the database; all farm identification elements had been removed from these data.

3.3.2 Data structure

Test requests are entered into the AHL database on a daily basis. Individual test requests are recorded as unique data entries. A common *case code* (submission

number) is given to all samples from the same herd on any given day, allowing identification of samples related to the same health event. In human health, a case usually refers to one person at a time. Such that two people, with the same medical complaint, living in the same household, submitting samples on the same day would be counted as two cases. In veterinary medicine which often works in herds or flocks, samples submitted from one, two or more animals, of the same type, from the same herd ("household") with the same medical complaint on the same day, would be counted as one case.

The nature of the diagnostic sample is identified in the database by two fields: the *sample type* field, in which the laboratory staff chose from a pre-set list (blood, feces, brain tissue, etc); and the *client sample ID*, a free-text field used to enter the source animal identifier given by the client. The diagnostic tests are identified by codes pre-set in the system. All codes are textual.

Table 3.1 shows a sample of the data. Only the fields relevant for medical information extraction are shown. Submission numbers have been removed, but samples from the same submission are represented in the table with consecutive rows in the same shading.

3.3.3 Syndrome definition

All of the historical data available were reviewed manually to identify the potential for syndromic classification at the time of sample submission. Veterinarians do not often provide detailed case history information. Therefore the identification of syndromes was based only on the type of diagnostic test requested, and the type of sample submitted, which allowed identification of the organ system targeted for diagnosis.

A syndromic group was defined as a group of test requests that: (i) are related to diseases from the same organ system; (ii) are all diagnostic tests for the same spe-

Table 3.1: Sample of the data available, restricted to the fields relevant for syndrome classification. Keywords and test names relevant for classification are shown in bold.

Date	Sample ID*	Client Sample ID	Sample Typle	Diagnostic test code	Diagnostic test description
2010-01-04	10-AAAA-0001	Tulip	Milk	Beta-Lactamase Test	Beta-lactamase test
2010-01-04	10-AAAA-0002	Plum	Milk	Culture Bact	Bacterial culture
2010-01-04	10-BBBB-0005	A517	SMALL Intestine	Culture Bact	Bacterial culture
2010-01-04	10-BBBB-0009	B516	Tissue Pooled	RLA	Rotavirus A - latex agglutination
2010-01-04	10-BBBB-0010	517, 516	Tissue - Fixed	Histopathology	Histopathology
2010-01-07	10-CCCC-0002	139 W-H-1 - Pericardial	Fluid	Culture Bact	Bacterial culture
2010-01-07	10-CCCC-0004	139 W-H-1 - Heart	Tissue	Culture Bact	Bacterial culture
2010-01-05	10-DDDD-0001	Webb/None Given	Tissue - Fixed	IHC - Bov Corona	IHC - Bovine coronavirus
2010-01-05	10-DDDD-0002	Webb/None Given	Ear - Notch	BVDV Antigen ELISA	Bovine viral diarrhea virus - antigen ELISA
2010-01-05	10-DDDD-0001	11675 BOOSTER 110004	Semen	Culture Bact	Bacterial culture
2010-01-27	10-DDDD-0031	Black face w white spot	Blood - Serum	N. caninum ELISA	Neospora caninum - ELISA
2010-01-27	10-EEEE-0002	Lung	Tissue	Culture Bact	Bacterial culture
2010-01-27	10-EEEE-0003	LuLiKiSpThTy	Tissue Pooled	Cell Cult Isolation	Virus isolation in cell culture
2010-01-27	10-EEEE-0005	Stom. content	Tissue	Culture Bact	Bacterial culture
2010-01-27	10-EEEE-0006	liv/spl/kid	Tissue	Culture Bact	Bacterial culture

* The field containing Submission ID was removed to ensure confidentiality, and omitted in the Sample ID shown.

cific disease, in cases of tests requested so frequently that their inclusion in another group would result in their being, alone, responsible for the majority of submissions; or (iii) tests that have little clinical relevance and should be filtered out (e.g., tests in environmental samples, general haematology profiles, as well as a range of "non-specific" submissions). Despite the absence of clinical information, the sample description allows identification of abortion cases through keywords such as "placenta" or "fetus". "Abortion" is therefore the only syndromic group defined based on a clinical syndrome, rather than using the three criteria listed above. Based on those criteria, an initial list of syndromic groups was compiled and then reviewed by a pathologist (BJM), a bacteriologist (CAM) and a clinician (DK). Following this review, all historical data were manually classified into syndromic groups to serve as training examples for the machine learning algorithms. Syndromic definition and manual classification were discussed until consensus was achieved among all experts.

Each submitted case (one or more test requests from a herd on a given day) could have multiple types of samples and/or multiple diagnostic tests requested. Syndromic

classification was performed for each individual database entry (test request), and later collapsed by case submission numbers, eliminating repeated syndromes within the same case. As a result, a given case could be associated with multiple syndromes by virtue of clues relating to multiple organ systems found in the same submission.

3.3.4 Mapping of test codes

Based on the aforementioned list of syndromic groups, a list of all diagnostic test codes that could be mapped into a syndromic group was established. Mapping is used here to describe the direct relationship: "if test requested is X, then syndromic group is Y", and mapping rules of this type were established for all test request codes that could be classified into only one syndromic group with certainty. This is typically the case for serological tests, where the veterinarian specifies the pathogen or disease to be confirmed, and the sample type is not informative of the organ system affected, as it is "serum" or "blood".

This mapping was built as a model in RapidMiner 5.0 (Copyright 2001-2010 by Rapid-I and contributors), an open source data mining package, which provides tools for data integration, analytical ETL (extract, transform, load), data analysis and reporting. RapidMiner includes an option to code any learned model in XML format, which can subsequently be directly manipulated.

Observations where test code was not associated with any mapping rule were assigned "Unknown" as the syndromic group at this stage in the processing. These were test requests such as "bacterial culture", which are not informative of the disease suspicion or organ system targeted by the veterinarian. These observations formed an *unmapped* subset of the data.

3.3.5 Algorithms for automated syndrome classification

For the *unmapped* subset, text mining was used to separate all words found in the fields describing the sample type (*client sample ID* and *sample type*, Table3.1) in the three years of available data. A tokenization process was applied using any non-letter character as a break point to separate words. The list of all mined words in the historical data was manually reviewed to construct a dictionary of medically relevant terms, as well as acronyms frequently used, and common misspellings. This is similar to the process described in [135] and [136].

Once the dictionary was built, all data tokenization was performed searching only for those specific tokens. For each observation being evaluated, the fields *sample type* and *client sample ID* were tokenized, and a vector was created to designate the binary occurrence of each word in the dictionary. These vectors were then used by the classifier algorithms to learn from the training dataset and to classify test data.

The rule induction algorithm in RapidMiner [Repeated Incremental Pruning to Produce Error Reduction (RIPPER)] was used. Information gain was used as the criterion used for selecting attributes and numerical splits. The sample ratio and pureness were set at 0.9 and the minimal prune benefit 0.25. Using the XML model of rules induced by the RIPPER algorithm as a template, a manually modified set of rules was also explored.

The Naïve Bayes learner available in RapidMiner was used to develop and apply a Naïve Bayes classifier. The learner requires no parameters settings other than an indication of whether a Laplace correction should be used to prevent high influence of zero probabilities. Laplace correction was not used. Decisions trees were constructed using gain ratio as the criterion for selecting attributes and numerical splits. The minimal size for split was set at 4, minimal leaf size 2, minimal gain 0.1, maximal depth 20, confidence 0.25, and up to 3 pre-pruning alternatives.

The XML code of the models used, as well as the set of customised rules for classification, are available upon request from the first author.

3.3.6 Assessing algorithms performance

Due to the large variability in the free-text entered by veterinarians to describe the samples submitted, it was deemed important to have a large test set, in order to assure that classification would be satisfactory once applied to new data. Manually classified historical data were split in half. After sorting sample submissions according to date and submission number, observations were alternately assigned to two different sets. Each classification algorithm was trained using one of the two sets, and then used to classify the alternative set. The process was then repeated switching training and test subsets.

Based on a comparison to the manual classification which had been carried out with the help of experts, the following performance measures were assessed for each classifier (using overall results from both test datasets): recall (the fraction of relevant instances correctly identified by the algorithm); precision (the fraction of the identified instances that were correct), and F1-score, the harmonic mean of recall and precision; i.e.$(2 * precision * recall) * (precision + recall) - 1$. After computing recall, precision and F1-score for each of the classes, these measures were averaged over all classes to give macro-averaged scores. An average weighted according to the number of records in each of the classes was also calculated; often referred to as micro-averaged scoring.

Stability was investigated by producing slightly different training subsets (for instance removing small random samples from the training set, or eliminating individual syndromic groups at a time), and assessing the resulting difference in the performance of the classifier.

3.4 Results

The three years of historical data contained 23,221 cases (samples from the same herd on a given day), consisting of a total of 218,795 individual test requests from cattle (i.e. bovine, dairy or beef animals of any age).

Based on an evaluation of these three years of historical data, and input from experts, the syndromic groups listed in Table 3.2 were defined. The table also lists the criteria for syndromic group creation and the number of test requests and cases assigned to each syndromic group following manual classification.

After classifying all sample submissions, and eliminating repeated syndromic instances within the same case, the final number of "syndromic cases" in the historical dataset was 30,760. Given that there were 23,221 initial herd investigations, this implies an average of 1.32 recorded syndromes per case. The distribution of syndromes per case is shown in Figure 3.1.

Of all the samples submitted, 75.7% (165,649) could be directly mapped into syndromic groups based on the test request information alone.

For the syndromic groups created based on clinical signs, non-specific signs or specific organ systems (see Table 3.2), Figure 3.2 illustrates the percentage of test requests which could be allocated to a syndromic group via direct mapping versus those that fell into the unmapped subset. Around 25% (53,146) of all instances in the database could not be directly mapped into a syndromic group and these provided the material for which automated classification was explored. Although these unmapped instances contain 16 of the original 22 defined syndromic groups, the syndromic group "Mastitis" alone is responsible for over 70% of these instances, and three groups ("Mastitis", "Nonspecific" and "GIT") account for over 90% of the data, as shown in Table 3.3. For the groups Mastitis and GIT, 94% and 77% of the

Table 3.2: Syndromic groups, defined based on an evaluation of three years of diagnostic test requests.

Syndromic group	Criteria for syndromic group creation	Number of test requests	Number of cases
Abortion	Clinical sign	559	225
Circulatory	Organ Systems	57	50
Eyes and ears		37	20
GIT		8,733	2,564
Haematopoietic		231	199
Hepatic		135	119
Mastitis		49,246	6,766
Musculoskeletal		233	149
Nervous		150	129
Reproductive		857	192
Respiratory		8,501	1,452
Skin and Tegument		14	7
Systemic		3,328	700
Urinary		501	146
BSE*	Individual diseases with high number of test requests	5,306	158
BLV		34,468	3,321
BVD		12,689	2,354
Johnes disease		11,123	2,040
Neosporosis		6,198	1,467
Clinical Pathology (hematology/biochemistry)	Other types of tests	61,059	4,282
Environmental samples		655	58
Antimicrobial susceptibility		140	33
Toxicology		6,866	955
Nonspecific samples	Samples whose syndromic group could not be determined	7,708	3,374
Total		218,795	30,760**

GIT = Gastro-intestinal tract; BSE = Bovine Spongiform Encephalopathy; BLV = Bovine Leukemia Virus; BVD = Bovine Viral Diarrhea

* BSE test requests are large compared to counts of other test submissions that can be classified as "Nervous".

** The number of cases after classification is higher than the initial number of cases because multiple syndromes can be identified within a single submission.

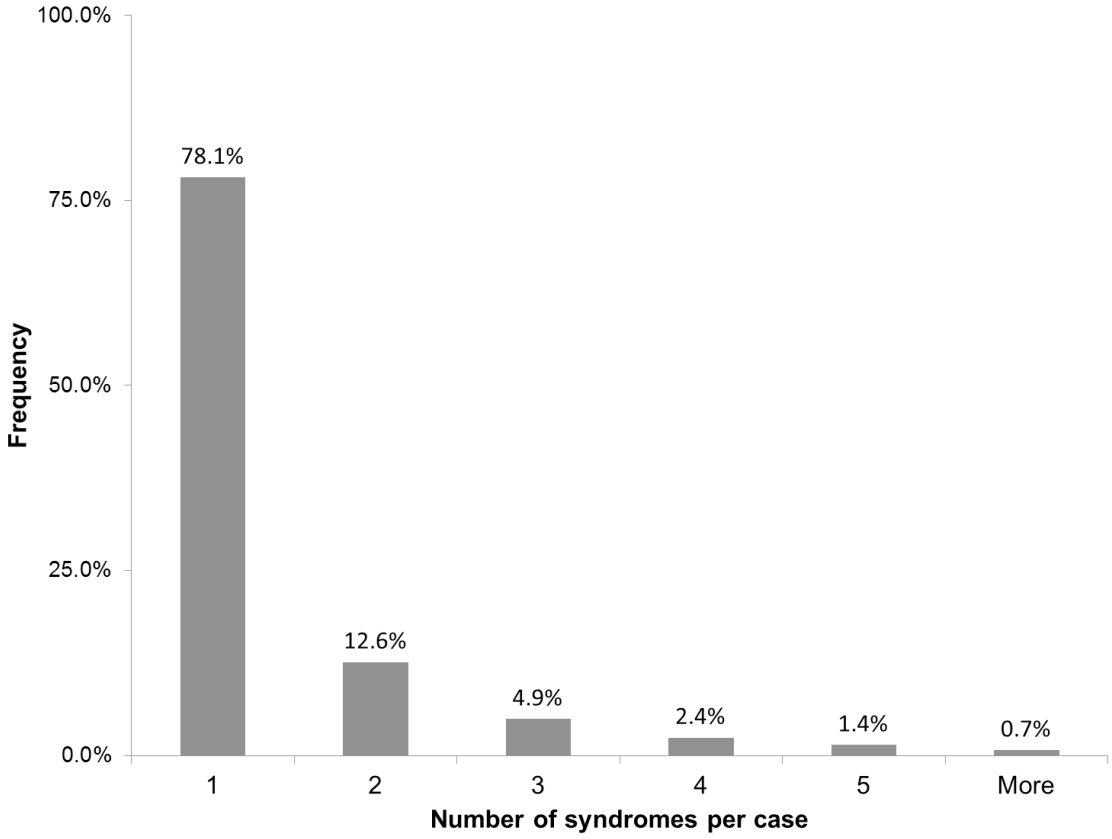


Figure 3.1: Number of syndromes identified in each case using information from individual test requests.

unmapped observations, respectively, refer to the test "Bacteria culture". Unmapped observations which are ultimately classified as "Nonspecific" contain a greater variety of test names, including the following which occur frequently: "Bacterial culture" (18%), "Histology" (27%) and "Necropsy" (18%).

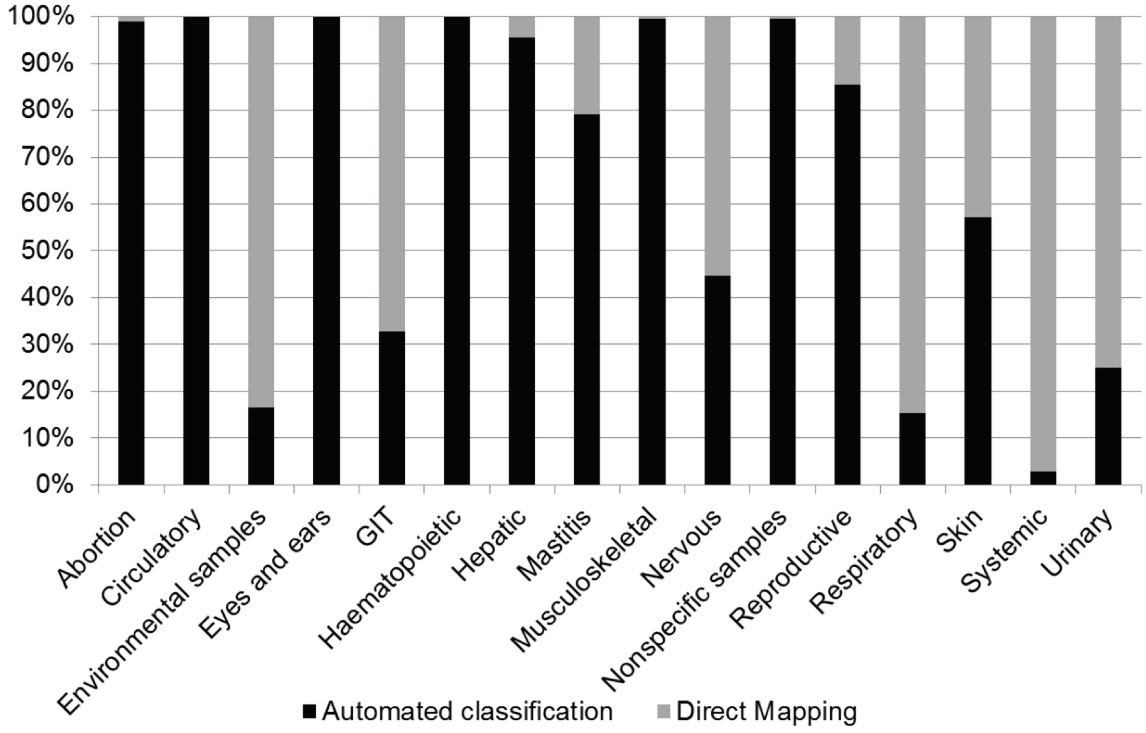


Figure 3.2: Percentage of test requests classified by direct mapping and automated classification.

The results of automated classification using different algorithms are shown in Table 3.4 and described in detail below.

The use of rule induction (RIPPER) achieved only moderate performance overall. Three groups with low frequency of test requests — "Environmental samples", "Skin", and "Eyes and Ears" — were not included in the rules, but as shown in Table 3.3 these groups represent only 0.3% of all instances subjected to automated classification. The F1-macro average was 0.677, but because the unlearned groups account for such a small proportion of the submissions, when the classes' performance is averaged accounting for the weight of each class, the F1-micro is 0.979 (Table 3.4). Upon

Table 3.3: Instances and syndromic groups in the 'unmapped' subset of the data

Syndromic group	Instances	Percentage of total	Cumulative percentage
Mastitis	38,934	73.26%	73.26%
Nonspecific	7,667	14.43%	87.68%
GIT	2,857	5.38%	93.06%
Respiratory	1,309	2.46%	95.52%
Reproductive	732	1.38%	96.90%
Abortion	553	1.04%	97.94%
Musculoskeletal	232	0.44%	98.38%
Haematopoietic	231	0.43%	98.81%
Hepatic	129	0.24%	99.06%
Urinary	125	0.24%	99.29%
Envir. samples	109	0.21%	99.50%
Systemic	98	0.18%	99.68%
Nervous	67	0.13%	99.81%
Circulatory	57	0.11%	99.91%
Eyes and ears	38	0.07%	99.98%
Skin and Tegument	8	0.02%	100.00%
Total	53,146		

Table 3.4: Performance measures for the algorithms implemented.

	Class average (Macro)*			Weighted average (Micro)		
Algorithm	recall	precision	F-score	recall	precision	F-score
Manually modified rules	.994	1.000	.997	1.000	1.000	1.000
Rule Induction**	.626	.793	.677	.991	.981	.979
Naïve Bayes	.983	.939	.955	.994	.996	.994
Decision Trees**	.290	.416	.311	.936	.937	.923

* The total number of groups in the training data was 16, and the total number of instances 53,146.

** The Rule Induction algorithms failed to learn 3 classes, and the Decision Tree 11 classes.

manual review of the rules created by the algorithm, it was found that the main source of error was failure of the algorithm to establish good decision rules when multiple medically relevant words were found in the same test request. This method was easy to implement and the rules generated are transparent and easily interpreted.

The rules produced by the RIPPER algorithm were manually modified to account for some of the relationships missed, producing a set of custom rules. Running the custom rule set against the entire *unmapped* subset resulted in an F1-macro score of 0.997, and F1-micro score of 0.9995 (Table 3.4). The remaining errors tended to be due to use of abbreviations not common enough to have been incorporated in the rules, misspellings or the absence of a space between two words, resulting in the tokenization process failing to identify these words.

The performance of the Naïve Bayes algorithm was high (F1-macro of 0.955 and F1-micro 0.994), as shown in Table 3.4. The main performance issue associated with this algorithm was its instability. Slightly different datasets resulted in very different performances (results not shown). With unbalanced training and test datasets, for instance, rather than assigning the label "Nonspecific" to samples that could not be classified, the Naïve Bayes algorithm would assign these samples, as well as misclassified samples from other groups, into one of the groups with a small number of submissions.

The classifier based on Decision Trees performed reasonably well in the micro score (F1-micro score of 0.923). However the classifier failed to learn 9 classes, which are biologically relevant, despite accounting for only 2% of the *unmapped* instances (which explains the high micro average). Moreover, the models appeared to be unstable: slight changes in the training data could result in a completely different 'shape' of decision tree, and a similar phenomenon was observed when the initial parameters for minimal gain and confidence where varied.

3.5 Discussion

This study evaluated the classification of structured data from animal laboratory test requests into syndromic groups for surveillance. This type of data lacks specificity not only because it precedes diagnostic results, but also due to the limited amount of clinical information provided by veterinarians. Previous work has focused on the direct mapping of specific test requests to syndromic groups [34, 133]. Here the use of text-mining was explored to extract information from fields containing a description of the sample collected by the veterinarian, in order to identify the organ system(s) affected in the clinical case being investigated.

Due to the structured format of the data, the text-mining task did not need to account for sentence semantics or other contextual information. Statistical methods were sufficient to capture the majority of medically relevant information from the fields mined. The binary occurrence of words from a manually constructed dictionary served as input to the classifier. The algorithms needed therefore to learn the relationship between these words, their co-occurrences and the target syndromic group.

Rule induction is suitable for uncovering these types of regular relations [135], and is recommended in cases when improvements in accuracy can be achieved by incorporating relationships among attributes [137]. However, upon manual review of the rules created by the algorithm, it was found that performance could be improved by including specific relationships in cases of multiple word occurrences. It was noted that the main relationships that the rule induction had failed to capture involved:

- (i) Sampling of multiple organs. For instance *heart* was associated with the "Circulatory" syndrome, and *liver* with "Gastro-intestinal", but the observation of samples from both organs in the same test request should be classified as "Systemic".

- (ii) Precedence being given to some words. "Abortion" is an actual clinical syndrome, in contrast to all other groups based on organ systems. Therefore the observation of any words related to abortion (*fetus, placenta, aborted*, etc) should result in classification of "Abortion", regardless of what fetal organ(s) was(were) collected.

- (iii) The co-occurrence of words which have a different meaning than when they occur on their own. For instance *ear* is a word included in the dictionary of relevant terms and would typically be associated with the "Eyes and ears" syndrome; however, this word should be ignored when it appears in the expression *ear tag*, which refers only to animal identification within a herd.

These relationships are still simpler than typical contextual challenges associated with full textual analysis, and the set of manually modified rules exhibited high performance. The remaining issues that prevented correct classification, such as misspellings and inconsistent abbreviations relate to the quality of the data, something which often complicates the interpretation of syndromic information [2].

The rule-based algorithm using manually modified rules was considered the most suitable algorithm for the classification of the animal laboratory dataset at hand, due to its high accuracy, ease of implementation, and high interpretability/transparency. Although simple, this rule-based solution is in line with research reporting from the i2b2 Obesity Challenge. Among the top 10 performing systems, rule-based approaches were the most successful in the textual task, which required classification based on documented information [124].

Rules also have the advantage that they are transparent and can typically be easily interpreted by the collaborating health experts [135]. Their main disadvantage is the knowledge acquisition bottleneck, in the case where rules are manually created, limiting portability and flexibility [135, 138]. Updates in the future to accommodate

changes in the language may have to be implemented manually, rather than in an automated manner.

The Naïve Bayes classifier demonstrated high performance. The main limitation observed with the use of this algorithm was its instability when groups with low frequency were included in the dataset. This behavior has been documented elsewhere [136]. The algorithm assumes that parameters are independent [138]. In this context the parameters were the binary occurrences, within each record, of the keywords from the dictionary built. Instability was however not observed to be due to occurrence of multiple keywords; rather it was associated with groups having small numbers of training examples. Due to the fact that the Naïve Bayes approach exhibits low transparency, it was not possible to track the specific mechanisms causing the problems observed in these low frequency categories, or to instigate measures to improve the way the algorithm was recording and using relationships between samples and the classification groups.

If transparency is not a limiting issue, that is, if domain knowledge experts are not required to understand and review the way by which the classifier is making decisions and classifying each instance, the Naïve Bayes algorithm can be an alternative to manually modified rules. Besides the high performance — though not as high as the custom set of rules — its implementation was the easiest of all algorithms evaluated, and automated updates can be planned by retraining the algorithm at regular intervals.

Nonmetric methods, such as Decision Trees, provide a "natural way to incorporate prior knowledge from human experts" [137]. However, this algorithm performed very poorly when small frequency groups were present; completely missing up to nine syndromic groups. Decision Trees were also very unstable to small changes on the data. This type of behaviour, in terms of training set sensitivity, has been well documented for Decision Trees [137].

The high performance reported in this study for the rule-based classifier refers to the algorithm's ability to reproduce the manual classification of records provided by a human expert. This classification, however, is based on an active review of test orders and diagnostic specimens submitted. Clinical descriptions are not normally submitted by veterinarians, and were not available for use in the classification of records, which constitutes a limitation to the classification process. While the lack of clinical information is expected to reduce the precision and recall of the system in comparison to the actual syndromes observed by the veterinarians, the consistency of the classifier and its high accuracy in utilising the information that is available should allow the system to capture increases in the number of submissions across different syndromic groups. Figure 3.3 illustrates the time series of daily counts, constructed after data had been classified using the rule-based algorithm, for two syndromic groups with expected seasonal behaviour: Bovine Viral Diarrhea and Mastitis. The series reflect the expected seasonal patterns, which supports the conjecture that classified records successfully reflect real trends in the number of submissions for various syndromes.

The development of this system has been conducted at the request of the data providers and the Ontario Ministry of Agriculture Food and Rural Affairs, which is responsible for the animal surveillance activities in the province of Ontario. The system has benefited greatly from the automated extraction of surveillance information from this animal health database. As the information extraction was based on data already regularly submitted to the AHL without any requirement for passive or active collection of additional data, sustainability of the system is not expected to be an issue.

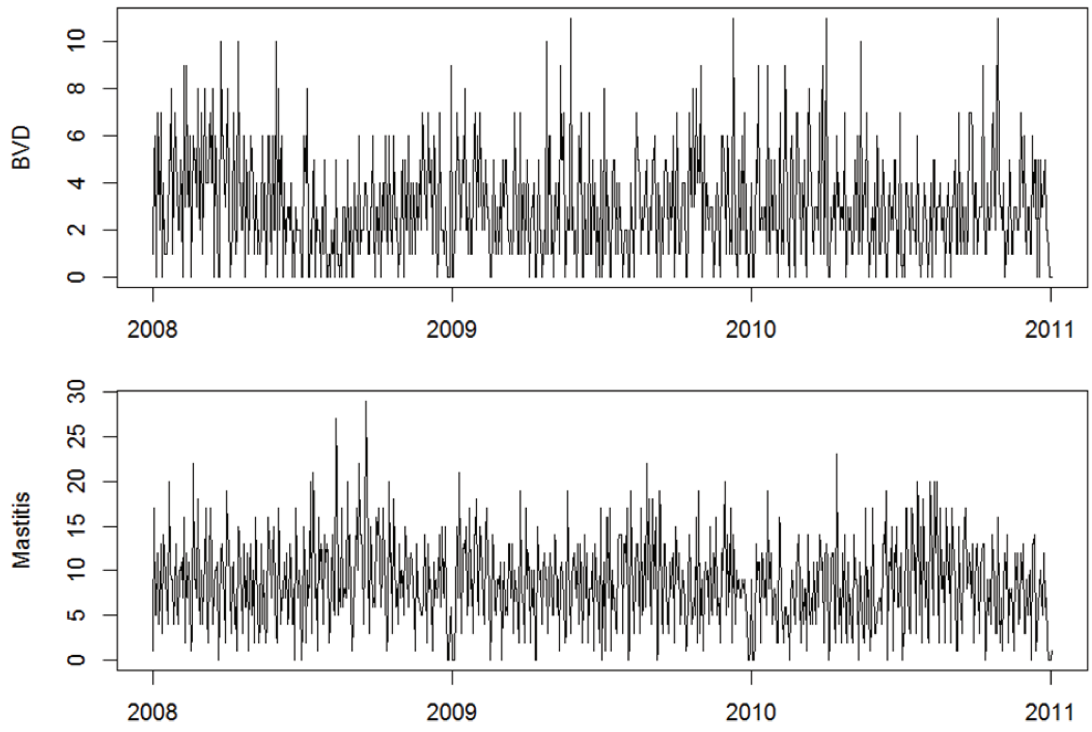


Figure 3.3: Daily counts of cases allocated to Bovine Viral Diarrhea (top) and Mastitis (bottom) syndromes.

3.6 Conclusion

Real-time monitoring of animal health data depends on establishing reliable models that reflect medical knowledge and that can be applied in an automated manner. Such models should be efficient, but also comprehensible to end users.

In this study the structured format of laboratory data, and the use of standard test codes, allowed for classification of approximately 75% of test requests into syndromic groups using direct mapping. For the remainder of the data, high accuracy (F1-macro = 0.997) was achieved through the use of a rule-based syndrome classifier. Induced rules were manually modified during the construction phase, but resulted in clear interpretability of decisions and resulting classification. While the use of rules was easy to implement and interpret, the construction of a dictionary of medically relevant terms and the manipulation of rules were time-consuming steps. Implementation of similar systems making use of other sources of laboratory data should be easier facilitated as standardized languages are more widely adopted in animal health laboratories, avoiding the repetition of this process for every new database.

The use of a custom rule set limits the potential for automatic revision of the classification model. Further research is required to establish internal validation rules, possibly based on the results available from historical data, in order to define automated ways to carry out model updates in the future.

Chapter 4

Retrospective time series analysis of veterinary laboratory data: preparing a historical baseline for cluster detection in syndromic surveillance

[0]Fernanda C. Dórea, Crawford W. Revie, Beverly J. McEwen, W. Bruce McNab, David Kelton and Javier Sanchez. Preventive Veterinary Medicine 2012, http://dx.doi.org/10.1016/j.prevetmed.2012.10.010

4.1 Abstract

The practice of disease surveillance has shifted in the last two decades towards the introduction of systems capable of early detection of disease. Modern biosurveillance systems explore different sources of pre-diagnostic data, such as patient's chief complaint upon emergency visit or laboratory test orders. These sources of data can provide more rapid detection than traditional surveillance based on case confirmation, but are less specific, and therefore their use poses challenges related to the presence of background noise and unlabeled temporal aberrations in historical data. The overall goal of this study was to carry out retrospective analysis using three years of laboratory test submissions to the Animal Health Laboratory in the province of Ontario, Canada, in order to prepare the data for use in syndromic surveillance. Daily cases were grouped into syndromes and counts for each syndrome were monitored on a daily basis when medians were higher than one case per day, and weekly otherwise. Poisson regression accounting for day-of-week and month was able to capture the day-of-week effect with minimal influence from temporal aberrations. Applying Poisson regression in an iterative manner, that removed data points above the predicted 95^{th} percentile of daily counts, allowed for the removal of these aberrations in the absence of labeled outbreaks, while maintaining the day-of-week effect that was present in the original data. This resulted in the construction of time series that represent the baseline patterns over the past three years, free of temporal aberrations. The final method was thus able to remove temporal aberrations while keeping the original explainable effects in the data, did not need a training period free of aberrations, had minimal adjustment to the aberrations present in the raw data, and did not require labeled outbreaks. Moreover, it was readily applicable to the weekly data by substituting Poisson regression with moving 95^{th} percentiles.

Keywords time series analysis, retrospective analysis, laboratory, syndromic surveillance, disease trends, animal health surveillance.

4.2 Introduction

Surveillance has shifted in the last two decades towards systems capable of early detection of disease [2]. Modern biosurveillance systems are designed to take advantage of data assumed to contain signatures of healthcare-seeking behaviors, which are not as specific as diagnosis, but allow for more rapid detection, and can be aggregated as syndromes. Surveillance based on these types of data is therefore referred to as *syndromic surveillance* [3]. A recent review of syndromic surveillance initiatives in veterinary medicine [134] indicated that opportunistic data sources are difficult to find in animal surveillance due to the scarcity of computerized, automatically collected data.

The secondary use of clinical animal data, whether computerized or not, also relies on the voluntary participation of veterinarians and/or producers. One alternative to relying on data shared voluntarily is the exploitation of automatically collected laboratory submission data [13]. Laboratory test results have been analysed retrospectively to detect temporal clustering of bacterial pathogens in public health [139, 140, 9] and veterinary medicine [141, 15]. The use of submission data, however, more properly fits the purposes of syndromic surveillance, as test requests are available earlier, though provide less specificity, than test results. Despite having lower population coverage than clinical data, laboratory data are generally stored in computerized systems, and have been available over relatively lengthy periods of time, meaning that historical analyses are usually possible.

When historical computerized data are available, a key challenge involves the construction of outbreak-free baselines, as any outbreaks will typically not be labeled,

nor will their shape and magnitude be known [2]. Detection of abnormal behavior in prospective analysis is based on either modeling and removing expected background (model-driven methods) or comparing profiles to similar data from unaffected populations (data-driven methods) [25, 2]. In both cases, a baseline free of outbreaks is necessary: in the former case to create models of expected behavior, and in the latter to serve as a comparison to the data being tested. Historical data can provide a baseline for temporal aberration detection algorithms, but data quality and influence of past outbreaks are challenges to overcome when determining 'typical' background behavior against which the presence of abnormalities can be investigated [2].

The overall goal of this study was to carry out retrospective analysis using three years of laboratory test submissions, related to health events in cattle, made to the Animal Health Laboratory in the province of Ontario, Canada. These historical data were analyzed for their potential use in syndromic surveillance. The retrospective analysis had two specific objectives. The first was to conduct time series analysis in order to discover explainable patterns in the data, such as day-of-week or seasonal effects as well as global trends. The second objective was to identify a procedure that could adequately describe the "normal behavior" for each syndrome, separating the background behaviour from temporal aberrations present in the historical laboratory test request data.

4.3 Methods

4.3.1 Data source

The Animal Health Laboratory (AHL) is a full-service veterinary diagnostic laboratory that serves livestock, poultry and companion animal veterinarians in the province of Ontario, Canada. The AHL is part of the University of Guelph and is

an integral part of the Ontario Animal Health Surveillance Network (OAHSN).

The AHL has a Laboratory Information Management System (LIMS) that is primarily used for reporting the results of diagnostic tests and for administrative purposes, but can also be used as a data retrieval platform for surveillance. Test requests are entered into the AHL database daily (only in exceptional circumstances are tests not entered in the computerized system on the same day that they are received). Individual tests are recorded as unique data entries. A common *case code* (submission number) is given to all samples from the same herd submitted on the same day. Retrospective analysis was performed on a dataset created by extracting three years (2008 to 2010) of data from all cattle sample submissions.

4.3.2 Case definition and syndromic groups

Individual health events were defined as one syndrome occurrence per herd. Individual herds can be identified in the database by the *case code* (a unique submission number), however it is not possible to consistently identify repeating submissions from the same herd if received on different days, and so recurring instances related to the same health event are recognized as multiple events.

Syndrome classification was performed based on the type of sample submitted and the diagnostic test requested by the veterinarian, which are the only pieces of information available at the time of submission. The full list of syndromes defined by the diagnosticians involved in this work is shown in Table 4.1.

Classification is first performed for each requested test. For pathogen specific tests, a direct correspondence was established between tests and syndromes. For instance: rabies tests are mapped to the nervous syndrome; brucellosis tests are mapped to the reproductive syndrome; etc. For non-specific tests, such as "bacteriological investigation", or "histology", text mining methods were used to search the

Table 4.1: Syndromic groups identified after analysis of three years (2008-2010) of diagnostic test requests to the Animal Health Laboratory, at the University of Guelph, Ontario, Canada.

DAILY monitoring	Days with 0 counts	Percentiles			
		25%	50%	75%	100%
Bovine Leukaemia Virus (BLV) tests	10.1%	2	4	7	33
Bovine Viral Diarrhoea tests	13.5%	1	3	4	11
Biochemical profile	12.4%	1	2	4	10
Clinical pathology tests (others)	12.8%	1	2	4	17
Gastro-intestinal	10.0%	2	3	5	12
Johnes disease tests	16.4%	1	2	4	12
Neospora caninum tests	25%	0.75	2	3	11
Mastitis	3.6%	5	9	12	29
Respiratory	20.9%	1	2	3	9
Nonspecific test requests*	6.2%	2	4	5	14
WEEKLY monitoring	Weeks with 0 counts	Percentiles			
		25%	50%	75%	100%
Circulatory, hepatic and hematopoietic[1]	39.2%	0	1	2	6
Nervous	37.8%	0	1	2	5
Reproductive and Abortion[2]	32.5%	0	1	2	8
Systemic	26.3%	0	2	4	10
Toxicology tests	29.2%	0	2	5	30
Urinary	70.0%	0	0	1	4
Others[3]	64.1%	0	0	1	5

[1]This group merges syndromic groups which initially contained very small numbers of submissions: “circulatory”, “hepatic” and “hematopoietic”;
[2] merges “reproductive” and “abortion”;
[3] merges “skin”, “eyes” and “ears”.
*Test requests that could not be classified into any of the other groups.

text entered freely by veterinarians describing the sample submitted, as well as the information from the field "sample type" used by laboratory staff. A dictionary of medically relevant words was constructed, and their relationship to different organ systems was established. For instance samples in which the word "lungs" is found are classified as respiratory syndrome, but if multiple organs from different systems are found, the syndrome type is "systemic". Abortion keywords have precedence, so that for instance "fetus lungs" are classified as the abortion syndrome, rather than respiratory. These correspondences compose a set of classification rules. The process was automated using rule-based classification algorithms, and is described in detail in [142].

Once each test request is classified into a syndrome, the data are collapsed by syndrome and case code for each day. This assures that multiple tests referring to the same syndromic type are not counted multiple times when related to the same case. However if clues to more than one syndrome are found within the same case all possible syndromes are counted.

4.3.3 Data characterization

All statistical analyses were performed in the R environment [143]. Complete data series, with counts for every calendar day from January 1^{st} 2008 to December 31^{st} 2010, were generated for each syndromic group by inserting missing days and assigning to them a count of zero (R packages {timeDate} [144] and {chron} [145]). When median daily counts for a given syndromic group were equal to or less than one count per day, the merging of two or more groups was considered, based on clinical similarities according to the opinion of the experts involved in syndrome definition. For instance abortion cases are classified into an individual category, which may be merged with other reproductive cases if their median count is not higher than one per day.

All syndromic series were further aggregated into weekly counts. Both daily series and weekly series were evaluated when medians were greater than one case per day, and only weekly aggregated data were evaluated otherwise. Further aggregation (for instance into monthly counts) was not considered as a key goal of the system being developed was early detection.

Initial characterization of the individual time series were performed using summary statistics by day-of-week, month and year, time plots, moving average and moving standard deviation charts [94].

Regression models were used to model any temporal effects observed in the data upon analysis of summary statistics, such as day-of-week, seasonal effects and global linear trends. Regression models appropriate for count data, such as Poisson regression [146], negative binomial regression, and zero-inflated versions of these methods (R package pscl [147]) were explored. Fit was assessed individually for each model (analysis of residuals and goodness-of-fit), and compared among models using the Akaike Information Criterion (AIC).

4.3.4 Aberration removal

To address the second objective, which was to define an outbreak-free historical baseline for each syndrome by separating the background behavior from temporal aberrations present in the historical data, two methods were investigated.

Smoothing was attempted using Holt-Winters exponential smoothing [148, 149], a method chosen due to its ability to model the temporal effects present in the data. Initializing smoothing coefficients (alpha for level, beta for trend and gamma for seasonality) can be provided when implementing this method. Lotze et al. (2008) [94] suggest using $\alpha = 0.4$, $\beta = 0$ and $\gamma = 0.15$ for surveillance data with seasonal components, and $\alpha = 0.1$ when there is no season component. The seasonal compo-

nent can be modeled as additive to the baseline (for each season effects of different magnitudes are added to the time series); or multiplicative (effects are modeled as a ratio from the baseline time series). The latter is not appropriate when there are zero-count days in the time series, as was the case in most of the syndrome time series evaluated here [148].

The second method was based on the procedure used by Tsui et al. (2001) [150]. The procedure is based on the assumption that after fitting the entire data to a regression model, data points above the 95% confidence interval of model predictions represent data occurring within epidemic time points. Data points above a one-sided 95% confidence interval are removed. The regression model used by the authors was the Serfling method [151, 150], a linear regression model based on weekly counts that introduces sine and cosine terms in order to account for seasonal waves. To explore the method for the data available in this study, the substitution of the Serfling method with the Poisson regression used during data characterization was tested. Replacement of detected outliers was evaluated using the limit of the confidence interval, and alternatively by the expected value for that time point, based on model predictions. To identify outliers in a Poisson model, the 95^{th} percentile of the Poisson distribution with mean equal to the estimated value for each time point was used as the threshold limit of that point. That is, for each estimated value λ_i, the upper limit is the smallest integer x such that $P(\lambda_i \leq x) \geq 0.95$. Lastly, an assessment was carried out as to whether repetition of the steps of model-fitting and outlier removal, in an iterative process, would improve anomaly elimination.

4.4 Results

4.4.1 Case definition and syndromic groups

The complete list of syndromic groups is shown in Table 4.1. A choice to monitor daily only those syndromes with median counts greater than one submission per day was made; the remaining syndromes were grouped into weekly counts. Syndromic groups merged into larger groups are also shown in the table, with details provided in the numbered footnotes. The AHL primarily operates on weekdays, with selected emergency testing available outside of usual business hours. Test requests are entered in the database daily and the date registered is that on which the sample was received. Sample submissions assigned to Saturdays and Sundays in the database were allocated to the following Monday. Daily medians in Table 4.1 therefore refer to the weekday median. All the time series described are based on 5-day weeks, and 260-day years.

4.4.2 Data characterization

Time series for six of the syndromic groups listed in Table 4.1 are shown in Figure 4.1, three daily series — requests for serological tests of Bovine Leukemia Virus (BLV), counts of tests related to mastitis diagnostics, and counts of tests for respiratory diseases; and three series chosen to be monitored weekly — test requests related to systemic diseases, reproductive diseases, and toxicology tests. Mastitis is the group with the highest daily average, the BLV series was chosen due to the evident presence of temporal aberrations in the historical baseline, while Respiratory was selected based on the assumption that it was more likely to exhibit seasonal variation. The weekly series were chosen to illustrate different weekly averages and presence of aberrations.

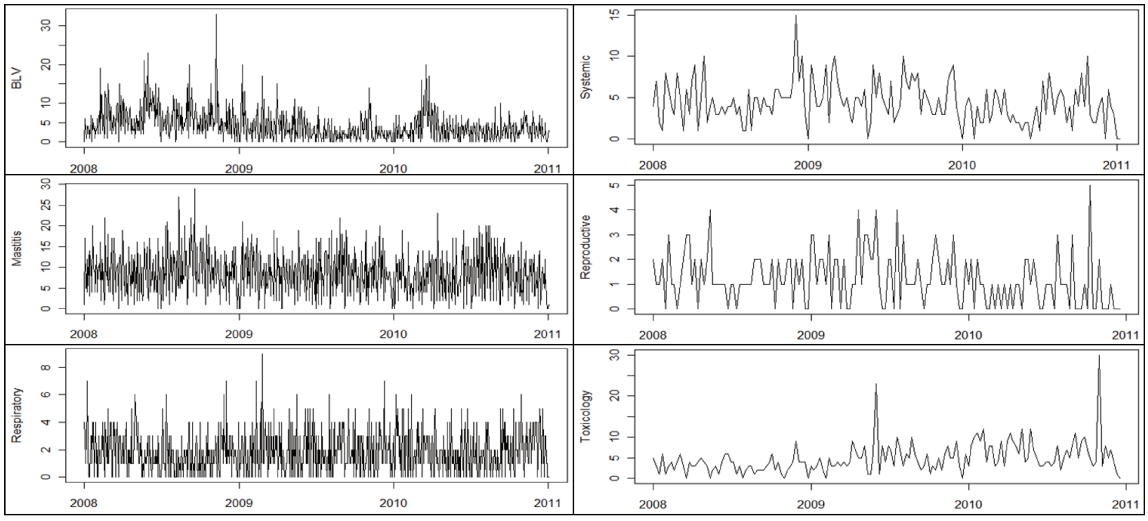


Figure 4.1: Examples of time series of daily (left) and weekly (right) counts of requests for tests associated with various syndromic groups

4.4.2.1 Time series for syndromes monitored daily

All daily series showed strong DOW effects. A zoomed view of 7 weeks at the beginning of 2010 for the Mastitis series is shown in Figure 4.2-A. Mondays are labeled in the graph. Box-plots of the quartiles of daily counts for the whole Mastitis series, per day-of-week, are shown in Figure 4.2-B. The peak of diagnostic sample submissions on Tuesdays is a result of the large number of sample submission through courier — because this laboratory serves the entire province of Ontario, many samples are mailed to the laboratory. Samples collected at the beginning of a week are therefore often received on Tuesday.

Month was, for most syndromic time series, a significant predictor in the Poisson regression model (at the 5% significance level). Monthly box-plots are shown for the Mastitis series in Figure 4.2-C. Year was not a significant predictor in the Poisson model for any syndromic series but BLV submissions. In that case, however, the effect was due to a high number of submissions in 2008 compared to 2009 and 2010, while the number of submissions in the latter two years was not significantly different from each other. No global linear trend was detected in any of the time series studied.

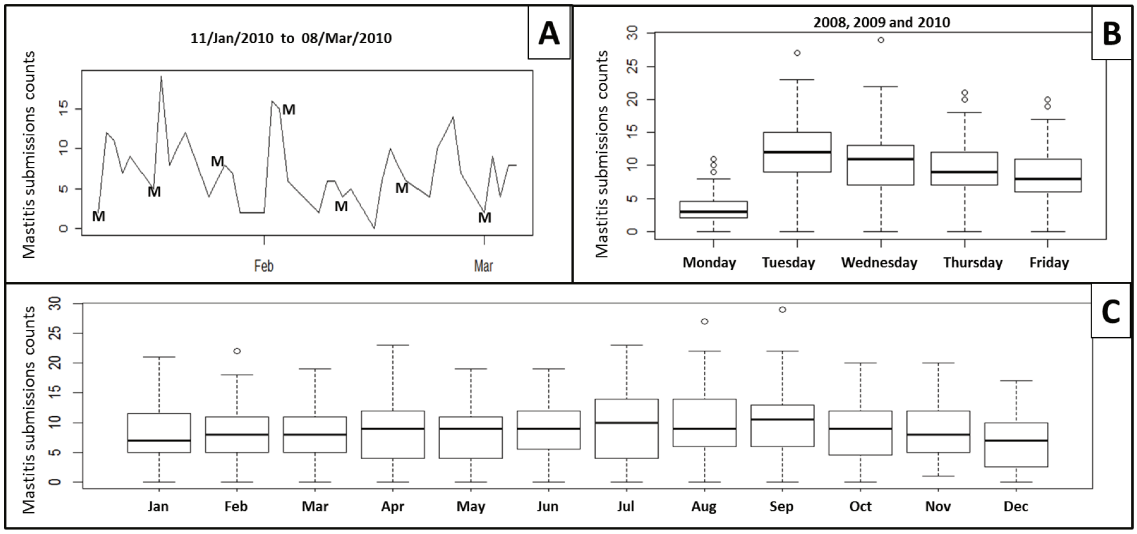


Figure 4.2: Day-of-week and month effects exemplified using the mastitis series. A- 7 week zoomed view of the series. Mondays are labeled with "M". B- Box-plots of all counts in each day of the week for the entire series (2008 through 2010). C- Box-plots of all counts in each month for the entire series (2008 through 2010).

Moving averages and standard deviation charts using several window sizes indicated that all series evaluated were non-stationary. The predicted values from the Poisson model are shown in Figure 4.3 for the BLV series, focusing on the year 2010 for visualization purposes (model fitting also included 2008 and 2009). No improvement (based on the reduction in the AIC) was obtained when using negative binomial or zero-inflated models to account for the substantial numbers of zero counts in the data. Analysis of residuals, deviance and goodness of fit (based on Pearson residuals) did not give reason to suspect of lack of fit to the Poisson regression model in any of the daily series evaluated. This result is restricted to the series chosen for daily monitoring, that is, those with a daily median greater than one submission per day.

4.4.2.2 Time series for syndromes monitored weekly

When counts are aggregated by week, the syndromic time series are reduced to 157 observations, rather than the 782 weekdays of the original daily data. Exploratory analysis using Poisson regression and the Serfling method [151, 150] indicated that

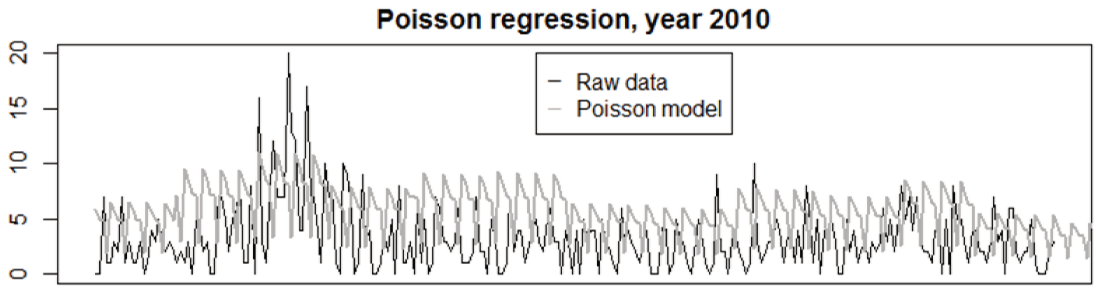


Figure 4.3: Poisson model predictions (grey) and the original BLV time series used to fit the model (black). Model variables were day-of-the-week and month.

the use of non-parametric methods, such as moving percentiles, were best suited to characterize these series.

4.4.3 Aberration removal

4.4.3.1 Time series monitored daily

The seasonal component of the Holt-Winters exponential smoothing reflected mainly the weekly effects, and no global trend was detected. It was therefore hypothesized that recognition of years was not relevant, and that modeling performance could be improved if the period was set to represent each week, rather than each year. A time series was created in which the cycles were set to 5 days, and the Holt-Winters smoothing was reapplied. Using shorter cycles allowed refitting of the parameters much more frequently (a great number of 5-days cycles within each year of data), resulting in the same final empirically calculated smoothing coefficients regardless of the choice of initializing coefficients.

The Holt-Winters exponential smoothing was able to reproduce closely the temporal effects and the random behavior of the data, but aberrations present in the raw data were incorporated in the model predictions. This is in contrast to the Poisson regression applied to all data (global model). Because day-of-week and month were

the only predictors incorporated, the Poisson model provided estimates that will be identical for each day of the week and month in different years, but are closer to what is expected in terms of baseline data.

Considering these results, Poisson regression was considered an appropriate method for modeling global behavior, when the main goal is to capture baseline activity with minimal influence of temporal aberrations present in the data, especially when these aberrations (potential outbreaks) have not been identified. The disadvantage of losing some of the original variation in the data through the application of a global model was addressed by applying a procedure similar to that suggested by Tsui et al. (2001) [150], in which most of the original data are kept, and a fitted model is used only to detect and replace outliers.

Applying the method of outlier removal as an iterative process confirmed that the subsequent steps of model fitting provided further aberration removal. Setting the process to repeat iterations for as long as outliers were detected typically required 3-4 iterations of model fitting and outlier removal for each syndromic time series, after which all observations fell within the 95^{th} percentile interval of the Poisson estimates.

When outliers detected using the percentile limit were substituted by the model prediction for that data point, rather than the limit of detection itself, an additional 1–2 iterations were necessary until no outliers were detected. The resulting time series after applying the iterative process of outlier removal based on Poisson regression, for the BLV series, are shown in Figure 4.4. Outlier substitution using the upper limit of the confidence interval (Figure 4.4-B) represented a better balance between removing temporal aberrations and keeping the original variation of the original data, without over-smoothing.

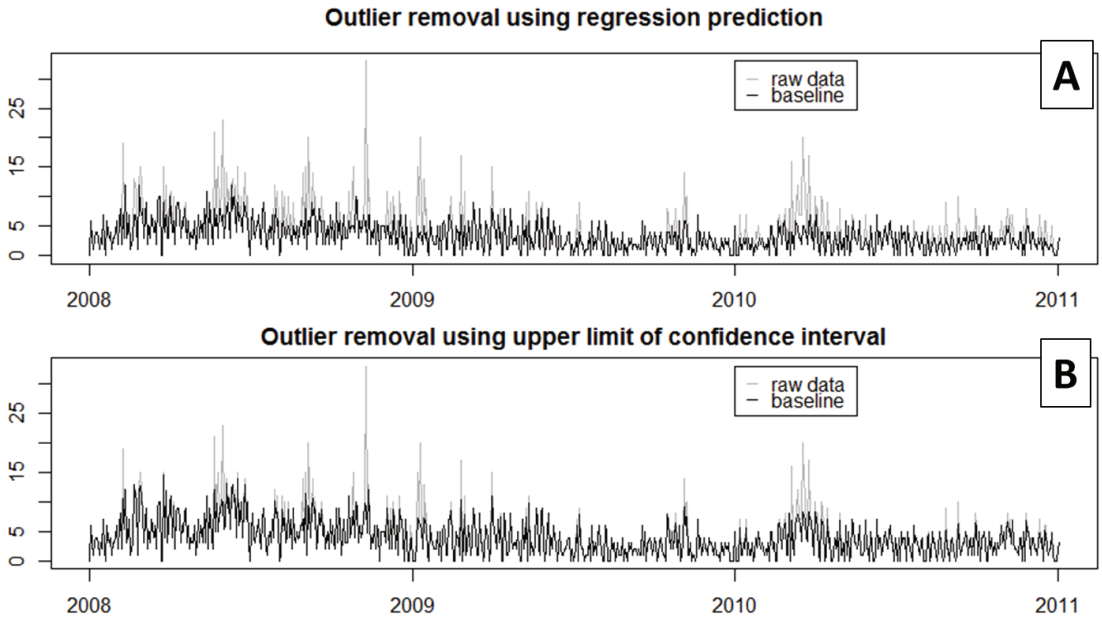


Figure 4.4: Raw data of daily counts for the BLV series (grey), superimposed by a baseline constructed after removal of temporal aberrations (black) using an iterative process based on Poisson regression. In the top graph, data points greater than the 95th percentile interval for the Poisson estimates were substituted by the model predicted value in each iteration. In the bottom graph the outliers were substituted by the upper limit of the 95th percentile interval in each iteration.

4.4.3.2 Time series monitored weekly

As the results of the exploratory analyses indicated that non-parametric methods were suitable for handling weekly data, the use of moving percentiles was investigated to remove temporal aberrations. In a manner similar to moving averages, a number of observations to the left and to the right of each value in a vector are used to calculate the statistic — in this case a percentile. Following the process previously used for daily data, the 95^{th} percentile was used to construct an upper limit for each value, and moving windows of 10 to 52 weeks were evaluated. These upper limits were used in the same iterative process described for daily data, to remove temporal aberrations.

This process applied to weekly series demonstrated better results using the 26 week window. Using shorter windows tended to result in inconsistent results, failing to eliminate temporal aberrations in some series or some specific periods within a

series, and over-smoothing others. Larger windows tended to over-smooth the series, eliminating most of the random variation. The result of the process based on 26 weeks moving windows is shown in Figure 4.5 for the time series of counts of test requests for systemic diseases.

The iterative procedure was performed consistently for all time series in Table 4.1, with results similar to those shown in Figures 4.4 and 4.5.

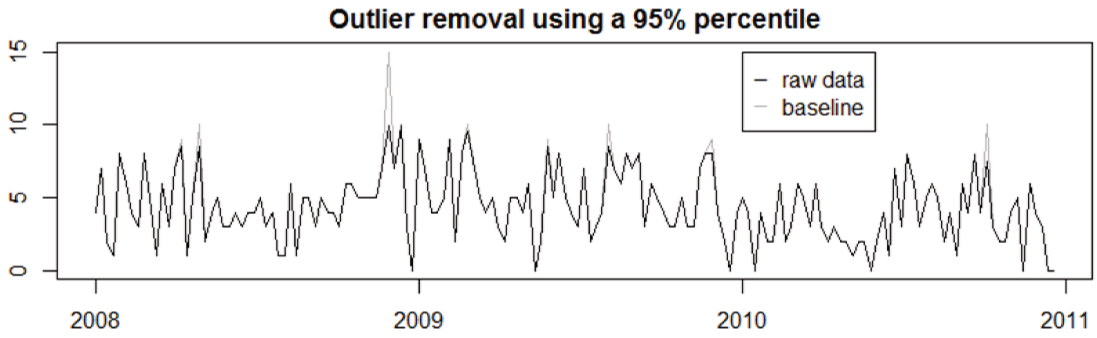


Figure 4.5: Raw data of weekly counts for the systemic syndrome series (grey) superimposed by a baseline constructed after removal of temporal aberrations (black) using an iterative process based on moving 95^{th} percentiles, in windows of 26 weeks.

4.5 Discussion

Syndromic surveillance operates under the assumption that anomalies indicative of disease outbreaks can be detected when information is monitored continuously [2]. Signatures of outbreaks can be obscured in the data by explainable factors, such as day of the week or seasonal effects, autocorrelation and global trends [94].

In this work three years of laboratory test requests from the Animal Health Laboratory at the University of Guelph, Ontario, were evaluated. The aim was to evaluate statistical approaches that would account for temporal effects in order to establish the baseline behaviour of the data for aberration detection in real time. Cases were counted daily, and repeating health events from the same herd were not discarded. This was considered to affect the specificity of the system, rather than sensitivity.

Once data were separated into syndromic groups, all time series of daily counts showed strong day-of-week effects. Even though the effect is not always consistent it was successfully reproduced by a Poisson regression, in which month was also a significant predictor. No global linear trends were found. In the series of counts aggregated weekly non-parametric methods such as moving percentiles were sufficient to model the data. Very low counts (medians are shown in Table 4.1) and weak effect of month explain why modeling and/or removing temporal effects was not an important condition in the weekly aggregated time series, and non-parametric methods could be used.

The Holt-Winters exponential smoothing was not able to separate the temporal aberrations from normal, background behavior. The attempt to change the settings of the Holt-Winters smoothing to recognize week as the data period, rather than year, proved to be valid in simulating the day-of-week effect. However, since local regression methods such as this adapt closely to the background variation in the data, the method will only be useful in modeling the data once an outbreak-free baseline is available.

On the other hand, the Poisson regression model fit to the whole data allowed all days of normal behavior in the data to contribute to the estimates, and therefore the resulting estimates were closer to the expected baseline of normal behavior. Regression has been used in several implemented biosurveillance systems, and it is a natural choice when this amount of historical data are available [152, 2]. The Poisson regression, specifically, can be more robust than other linear models since it does not require constant variance [88]. The Poisson regression, however, assumes variance equal to the mean of the distribution of observed counts. This proved to be a reasonable assumption for the syndromic time series evaluated in this work, since neither a negative binomial nor zero-inflated models indicated a better fit for the daily counts. If this assumption is not met, models which can account for zero-

inflated distributions and/or overdispersion should be explored [153].

By reducing the model variables to key explainable factors, such as day-of-week and month, it was possible to model the baseline behavior, while preventing adaption to temporal aberrations. Removing such aberrations from training data has been noted as a key challenge of implementing any system for early detection of outbreaks [2]. There are statistical methods to identify whether outbreaks are present or not [154], and the use of diagnostic information to label outbreaks has also been suggested [86]. However, even if outbreaks can be identified, the problem of how to remove the outbreak signature from the background data of normal behavior remains. The challenge addressed in this work was that of identifying an algorithm that could be used despite the absence of clean training data and the lack of knowledge about the shape and duration of any outbreaks.

When removing aberrations it is desirable to keep as much of the original data as possible, using model predictions only to replace days in which temporal aberrations are present. To achieve this, the method proposed by Tsui et al. (2001) [150] was adapted by substitution of the Serfling algorithm, more appropriate for time series with strong seasonal effects, with a Poisson regression and application of the steps iteratively. Fitting a global model such as Poisson regression assumed that the covariates chosen were sufficient to capture the systematic behavior of the data, and that their relationship to the counts is homogeneous across the entire period [148].

The substitution of outliers by modeled values resulted in over-smoothing of the data. If adopted, this would generate a baseline which would likely lead to the detection of excessive number of false alarms when used to train aberration detection algorithms. When our adjusted method substituted outliers with the upper limit of the 95^{th} percentile it proved to be efficient in removing temporal aberrations, while keeping most of the original data, and maintaining the day-of-week effect.

The BLV series was used to illustrate the method, since it had the most noticeable set of temporal aberrations in historical data, but the method also performed well when applied to the other daily time series. It is therefore an efficient procedure for the automated cleaning of historical data, producing baselines that can be used in prospective analysis.

This iterative smoothing method also proved useful for removing aberrations from syndromic data that was aggregated at a weekly level, when substituting the Poisson regression with moving percentiles. One setting (using a 26 week window and the 95^{th} percentile) worked well for all data series, again allowing for automated implementation. This assumes that no predictable effects, such as seasonal patterns, are present in the data. This assumption was met for the time series evaluated due to the decision to aggregate sparsely occurring syndromes into weekly counts, thus removing any day-of-week effect.

It was not the intention of this work to investigate the reasons for the aberrations documented in the historical data available. Some of these aberrations could constitute random variation in the data, rather than true alarming health events, in which case removal of all aberrations could cause over-smoothing of the data, reducing the specificity of a system based on these developed baselines ("over-sensitive" detection alarms). This was however chosen over the risk of developing a system with low sensitivity, due to noisy baseline data. Once a baseline is available, system implementation can be simulated retrospectively, that is, the system can be set to run daily in data starting 6 months to a year before the actual date of implementation, so that a buffer, or "purging" time is used to let the system re-adjust to real data, and in case of excessive false-alarms adjustments to system settings can be made to maximize sensitivity without decreasing the specificity of the system.

This work assumed that prospective monitoring in real-time, the next stage of a syndromic system development, will be based on monitoring count events, as opposed

to monitoring the time between occurrences, which are better suited for monitoring rare events. A decision was therefore made to monitor daily counts for only those syndromes with a median value greater than one, and to group the remaining time series into weekly counts. Further grouping into monthly counts was not considered appropriate for the early-detection warnings to be captured by this system. Without such further grouping, however, some series may be better monitored with methods specifically developed for the monitoring of rare events. Such methods are beyond the scope of this work, as most of the series with low counts were a result of an attempt to classify all laboratory tests into a syndromic group, rather than a true interest in rare events in these data.

4.6 Conclusion

Successful identification of outbreak signatures in population data, the primary goal of syndromic surveillance, depends on identifying and removing explainable variation from the noisy background of normal behavior. Three years of laboratory test request data from the Animal Health Laboratory in Ontario were analyzed retrospectively in order to identify such explainable factors. Day-of-week and month effects were found to be the only relevant effects that required removal. Poisson regression accounting for day-of-week and month was able to capture these effects with minimal contamination by temporal aberrations.

The results of the exploratory analyses were used to identify temporal aberrations in the historical data. By applying Poisson regression in an iterative manner, that removed data points above the 95^{th} percentile, it was possible to remove these aberrations in the absence of labeled outbreaks, while keeping the temporal effects from the original data. This resulted in the construction of time series that represent the baseline pattern over a three year period, free of temporal aberrations. The final

method proposed did not require a training period free of aberrations, had minimal adjustment to these aberrations present in the raw data, and did not require labeled outbreaks. Moreover, it could be readily adapted for weekly data by substituting Poisson regression with moving 95^{th} percentiles.

Chapter 5

Syndromic surveillance using veterinary laboratory data: data pre-processing and algorithm performance evaluation

[0]Fernanda C. Dórea, Beverly J. McEwen, W. Bruce McNab, Crawford W. Revie and Javier Sanchez. Submitted for peer-review. Journal of the Royal Society Interface.

5.1 Abstract

Diagnostic test orders to an animal laboratory were explored as a data source for monitoring trends in the incidence of clinical syndromes in cattle. Four years of real data and over 200 simulated outbreak signals were used to compare pre-processing methods that could effectively remove temporal effects in the data, as well as temporal aberration detection algorithms that provided high sensitivity and specificity in such monitoring systems. Weekly differencing demonstrated solid performance in removing day-of-week effects, even in series with low daily counts. For aberration detection, the results indicated that no single algorithm consistently performed better than all others across the range of outbreak scenarios simulated. Exponentially Weighted Moving Average charts and Holt-Winters exponential smoothing demonstrated complementary performance, with the latter offering an automated method to adjust to changes in the time series that will likely occur in the future. Shewhart charts provided lower sensitivity but earlier detection in some scenarios. Cumulative Sum charts did not appear to add value to the system. These findings indicate that automated monitoring, aimed at early detection of temporal aberrations in syndromic data sets of this type, will likely be most effective when a number of different algorithms are implemented in parallel.

Keywords: laboratory, syndromic surveillance, temporal aberration detection, outbreak detection, control charts.

5.2 Introduction

During the last decade, increased awareness of the need to recognize the introduction of pathogens in a monitored population as early as possible has caused a shift in disease surveillance towards systems that can provide timely detection [20, 2].

Some monitoring has shifted to pre-diagnostic data, which are available early, but lack specificity for detection of particular diseases. These data can, however, be aggregated into syndromes, a practice which has led to an increase in the use of the terms "syndromic data", and "syndromic surveillance" [3, 2].

Disease outbreak detection is a process similar to that of statistical quality control used in manufacturing, where one or more streams of data are inspected prospectively for abnormalities [2]. For this reason, the use of classical quality control methods has been used extensively in public health monitoring [93, 155]. However, these types of control charts are based on the assumption that observations are independently drawn from pre-specified parametric distributions, and therefore their performance is not optimal when applied to raw, unprocessed health data [94], which are typically subjected to the effect of factors other than disease burden. Some of these factors are predictable, such as day-of-week effects, seasonal patterns or global trends in the data [2]. These predictable effects can be modelled and their influence removed from the data [146, 94, 25]. An alternative is to make use of data-driven statistical methods, such as the Holt-Winters exponential smoothing, which can efficiently account for temporal effects [149].

The use of real data is an essential step in the selection of algorithms and detection parameters because the characteristics of the baseline (such as temporal effects and noise) are likely to have a significant impact on the performance of the algorithms [99]. However, the limited amount of real data and lack of certainty concerning the consistent labelling of outbreaks in the data prevent a quantitative assessment of algorithm performance using standard measures such as sensitivity and specificity. These issues can be partially overcome through the use of simulated data which can include the controlled injection of outbreak signals. Furthermore this approach has the advantage of allowing for the evaluation of algorithm performance over a wide range of outbreak scenarios [156].

A recent review [134] indicated that few systems have been developed for real- or near-real time monitoring of animal health data. Previous work by the authors [157] has addressed the possibility of using laboratory test requests as a data source for syndromic surveillance in aiming to monitor patterns of disease occurrence in cattle. In this paper these same data streams are evaluated with the aim of constructing a monitoring system that can operate in near-real time (i.e. on a daily and weekly basis).

The points outlined above were addressed in an exploratory analysis designed to:

- (i) identify pre-processing methods that are effective in removing or dealing with temporal effects in the data;
- (ii) explore methods that combine these pre-processing steps with detection algorithms, with the data streams available and being aware of the importance of having a detection process interpretable by the analysts;
- (iii) identify the temporal aberration detection algorithms that can provide high sensitivity and specificity for this specific monitoring system.

A variety of algorithms and pre-processing methods were combined and their performance for near-real time outbreak detection assessed. Real data were used to select algorithms, while sensitivity and specificity were calculated based on simulated data which included the controlled injection of outbreaks.

5.3 Methods

All methods were implemented using the R environment [143].

5.3.1 Data source

Four years of historical data from the Animal Health Laboratory (AHL) at the University of Guelph in the province of Ontario, Canada, were available — from January 2008 to December 2011. The AHL is the primary laboratory of choice for veterinary practitioners submitting samples for diagnosis in food animals in the province of Ontario, Canada. The number of unique veterinary clients currently in the laboratory's database (2008 to 2012) is 326. The laboratory receives approximately 65,000 case submissions per year, summing to over 800,000 individual laboratory tests performed, of which 10% refer to cattle submissions, the species chosen as the target for syndromic surveillance implementation.

A common standard for the classification of syndromes has not been developed in veterinary medicine. Classification was therefore based on manual review of three years of available data, and then creating rules of classification reviewed by a group of experts (a pathologist, a microbiologist and two clinical veterinarians) until consensus was reached by the group. These rules were implemented in an automated system classification as documented in Dorea et al. (2013) [142].

An effort was made to classify every laboratory submission record into at least one syndromic group. Therefore, the final syndromic classification was not only based on a direct relation to clinical syndromes. A "syndromic group" was defined in this system as laboratory submissions: (i) related to diseases from the same organ system; (ii) comprising diagnostic tests for the same specific disease, in cases of tests requested so frequently that their inclusion in another group would result in their being, alone, responsible for the majority of submissions; or (iii) that have little clinical relevance and should be filtered out. Seventeen syndromic groups were created. Nine groups referred to clinical syndromes: gastro-intestinal; mastitis; respiratory; circulatory, hepatic and haematopoietic; nervous; reproductive and abortion; systemic; urinary;

and "other". Diagnostics for specific diseases assigned to one of four groups based on a higher volume of submissions (ii above) were: bovine leukosis; bovine viral diarrhoea; Johnes disease and *Neospora caninum*. Lastly, the four groups created to classify general tests were: biochemical profile; other clinical pathology tests; toxicology tests; and nonspecific tests (those which could not be classified into any previous group).

Individual health events were defined as one syndromic occurrence per herd. Classification is first performed for each requested test. Once each test request is classified into a syndromic group, the data are collapsed by the unique herd identifier for each day. Due to a very low number of submissions on weekends, any cases in the database assigned to weekends were summed to the following Monday, and weekends were removed from the data. Only syndromic groups with a median greater than one case per day were monitored daily [157]. It was proposed that the remaining syndromes (7 of 17 in total) would be monitored on a weekly basis; these series are not discussed further in this paper. All the methods described in this paper were carried out for all the syndromic groups monitored daily. As documented in [157], the time series of daily cases for each of these groups showed very similar statistical properties: daily medians between 2 and 4, except for tests for diagnosis of mastitis and respiratory syndromes, for which daily medians were 9 and 1, respectively; strong day of week effect; no global monotonic trends; and weak seasonal effects, especially for the syndromes with lower daily medians.

Methods and results are illustrated using the daily counts of laboratory test requests for identification of Bovine Leukaemia Virus (BLV). No more than 5% of cattle infected by the virus will develop clinical signs of bovine leukosis, which is characterized by a reduction in condition, diarrhoea, and tumours in several organs, which can sometimes be palpated through the skin. Tests for BLV are often requested in animals with a decrease in body condition as well as milk production. This series was

chosen due to the statistical similarities to the time series of other syndromic groups, while being the only times series showing evident presence of temporal aberrations (outbreak signals) documented in the historical data. Additionally, the counts of test requests for diagnosis of mastitis are used to illustrate the particular effect of working with time series with stronger seasonal effects; while the daily counts of laboratory submissions for diagnostic of respiratory syndromes is used to illustrate the particular challenges of working with time series with lower daily median. The three time series are shown in Figure 5.1.

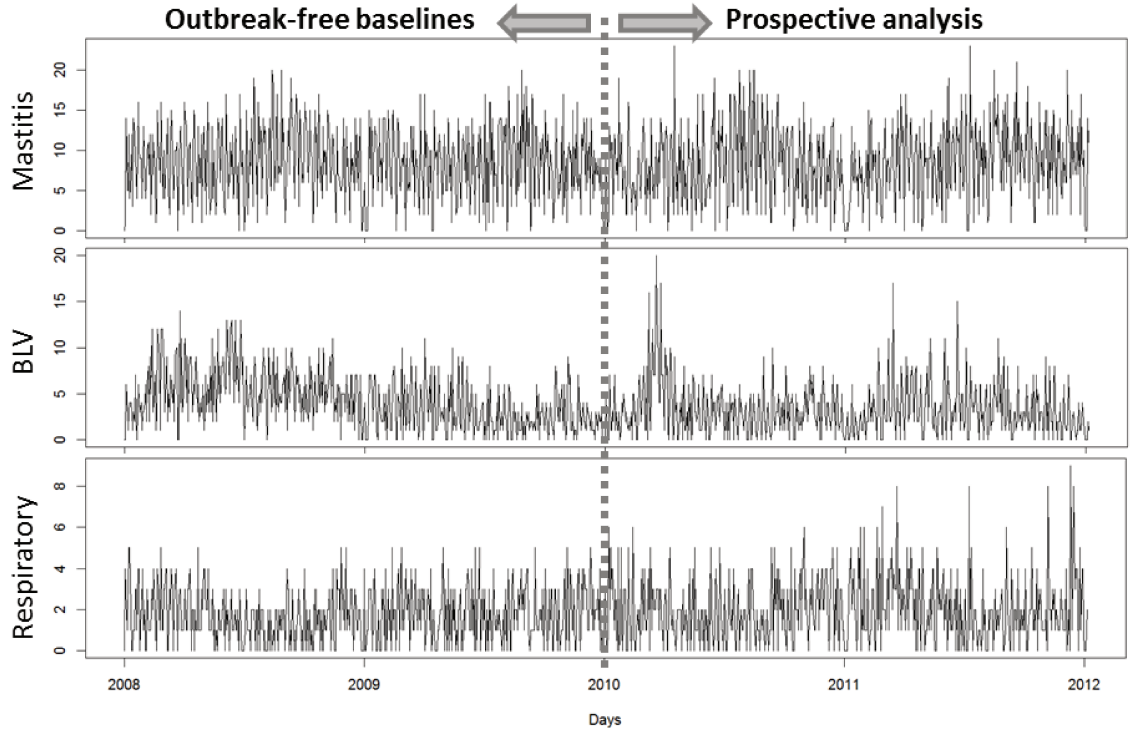


Figure 5.1: Syndromic groups used to exemplify the times-series used in this work. Data from 2008 and 2009 have been analysed in order to remove temporal aberrations, constructing an outbreak-free baseline.

Data from 2008 and 2009 were used as training data. These data had been previously processed to remove temporal aberrations, creating *outbreak-free baselines* for each syndromic group [157]. Data from 2010 and 2011 were used to evaluate the performance of detection algorithms trained using those baselines.

5.3.1.1 Simulated data

In order to simulate the baseline (background behaviour) for each syndromic group the four years of data were fit to a Poisson regression model with variables to account for day-of-week and month, as previously documented [157]. The model predicted value for each day of the year was set to be the mean of a Poisson distribution, and this distribution was sampled randomly to determine the value for that day of a given year, for each of 100 simulated years.

To simulate outbreak signals that also preserved the temporal effects from the original data, different outbreak signal magnitudes were simulated by multiplying the baseline data by selected values. Magnitudes of 1, 2, 3 and 4 were used. These values were selected arbitrarily as the daily counts were too low to apply an increase based on standard deviations, as previously suggested [21, 156].

Outbreak signal shape (temporal progression), duration and spacing were then determined by overlaying a filter to these outbreak series, representing the fraction of the original magnified count which should be kept. For instance, a filter increasing linearly from 0 to 1 in 5 days (explicitly: 0.2, 0.4, 0.6, 0.8 and 1), when superimposed to an outbreak signal series, would result in 20% of the counts in that series being input (added to the baseline) on the first day, 40% in the second, and so on, until the maximum outbreak signal magnitude would be reached in the last outbreak day. The process and resulting series are summarized in Figure 5.2.

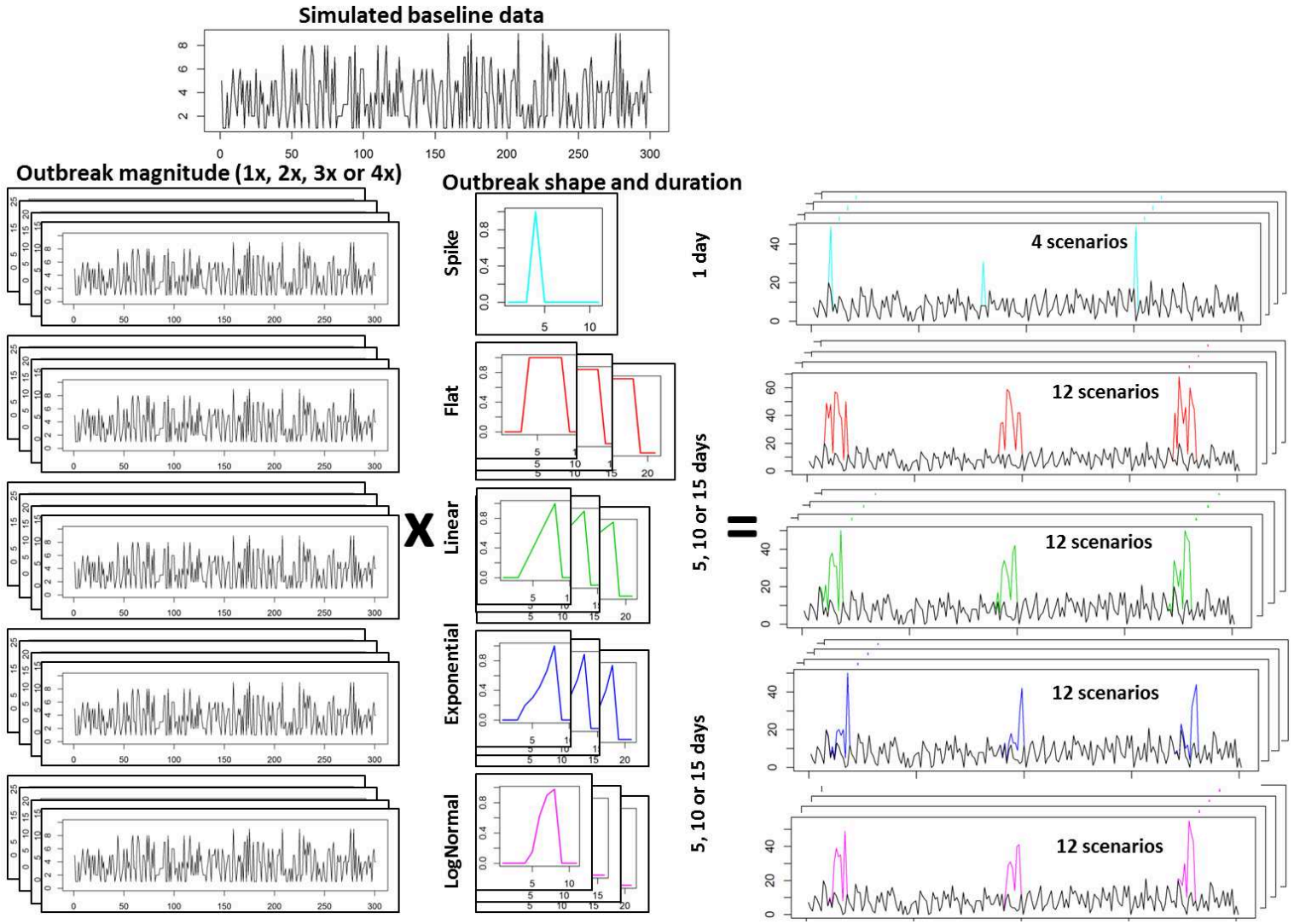


Figure 5.2: Synthetic outbreak simulation process. Data with no outbreaks were simulated reproducing the temporal effects in the baseline data. The same process was used to construct series that were for outbreak simulation, but counts were amplified up to 4 times. Filters of different shape and duration were then multiplied to these outbreak series. The resulting outbreaks were added to the baseline data.

The temporal progression of an outbreak is difficult to predict in domestic animal populations, where the epidemiological unit is the herd rather than the individual animal, because a large proportion of transmission is due to indirect contact between farms locally and also over large distances [158]. The same pathogen introduction can result in different temporal progressions in different areas as a result of spatial heterogeneity, as seen in the foot-and-mouth disease outbreak in the UK in 2001 [159] and the bluetongue outbreak in Europe in 2006 [160]. For this reason, several outbreak signal shapes previously proposed in the literature ([119, 161]) were simulated. These shapes were combined to generate the following filters:

- Single spike outbreaks: A value of 1 is assigned to outbreak days, while all other days are assigned a value of zero.
- Moving average (flat) outbreaks: Each outbreak signal is represented by a sequence of 5, 10 or 15 days (one to three weeks) with a filter value of 1 (outbreak days), separated by days of non-outbreak in which the filter value is zero.
- Linear increase: The filter value increases linearly from 0 in the first day, to 1 in the last day. This linear increase was simulated over 5, 10 and 15 days.
- Exponential increase: The filter value increases exponentially from 0 in the first day, to 1 in the last day. For the duration of 5 days this was achieved by assigning 1 to the last day, and dividing each day by 1.5 to obtain the value for the preceding day. For the durations of 10 and 15 days a value of 1.3 was used.
- Log-normal (or sigmoidal) increase: The filter value increases following a log-normal curve from 0 in the first day, to 1 in the last day. The same values for the distribution are used for any outbreak signal length [$lognormal(4, 0.3)$], but the value corresponding to 5, 10 and 15 equally distributed percentiles from this distribution are used to assign the filter value for outbreaks with these respective durations.

Each filter was composed using one setting of outbreak signal shape and duration, repeated at least 200 times over the 100 simulated years, with a fixed number of non-outbreak days between them. The space between outbreak signals was determined after real data were used to choose the initial settings for the aberration detection algorithms, in order to ensure that outbreak signals were spaced far enough apart to prevent one outbreak from being included in the training data of the next. Each of these filters was then superimposed on the 4 different outbreak signal magnitude series, generating a total of 52 outbreak signal scenarios to be evaluated independently by each detection algorithm.

5.3.2 Detection based on removal of temporal effects and use of control charts

5.3.2.1 Exploratory analysis of pre-processing methods

The retrospective analysis [157] showed that day-of-week (DOW) effects were the most important explainable effects in the data streams, and could be modelled using Poisson regression. Weekly cyclical effects can also be removed by differencing [94]. Both of the following alternatives were evaluated to pre-process data in order to remove the DOW effect:

- Poisson regression modeling with day-of-week and month as predictors. The residuals of the model were saved into a new time series. This time series evolves daily by refitting the model to the baseline plus the current day, and calculating today's residual.
- Five-day differencing. The differenced residuals (the residual at each time point t being the difference between the observed value at t and $t-5$) were saved as a new time series.

Autocorrelation and normality in the series of residuals were assessed in order to evaluate whether pre-processing was able to transform the weekly- and daily-autocorrelated series into independent and identically distributed (i.i.d.) observations.

5.3.2.2 Control charts

The three most commonly used control charts in biosurveillance were compared: (1) Shewhart charts, appropriate for detecting single spikes in the data; (2) cumulative sums (CUSUM), appropriate for use in detecting shifts in the process mean; and (3) the exponentially weighted moving average (EWMA), appropriate for use in detecting gradual increases in the mean [155, 94].

The Shewhart chart evaluates a single observation. It is based on a simple calculation of the standardized difference between the current observation and the mean (Z-statistic); the mean and standard deviation being calculated based on a temporal window provided by the analyst (*baseline*).

The CUSUM chart is obtained by:

$$CUSUM : C_t = max(0, (D_t + C_{t-1})) \tag{5.1}$$

where t is the current time point,D_t is the standardized difference between the current observed value and the expected value. The differences are accumulated daily (since at each time point t the statistic incorporates the value at $t-1$) over the *baseline*, but reset to zero when the standardized value of the current difference, summed to the previous cumulative value, is negative.

The EWMA calculation includes all previous time points, with each observation's

weight reduced exponentially according to its age:

$$EWMA : E_t = (1-\lambda)^t E_o + \sum_{i=1}^{t} (1-\lambda)^t \lambda I_t \tag{5.2}$$

where λ is the smoothing parameter (> 0) that determines the relative weight of current data to past data, I_t is the individual observation at time t and E_o is the starting value [162, 155].

The mean values from the *baseline* are used as the expected value at each time point. Baseline windows of 10 to 260 days were evaluated for all control charts.

In order to avoid contamination of the baseline with gradually increasing outbreaks it is advised to leave a buffer, or *guard-band gap*, between the baseline and the current values being evaluated [163, 88, 164]. Guard-band lengths of one and two weeks were considered for all algorithms investigated. Weekly (rather than daily) guard-bands were used in order to align the day of the week being evaluated with the days of the week in the baseline.

One-sided standardized *detection limits* (magnitude above the expected value) between 1.5 and 3.5 standard deviations were evaluated. Due to the large variation in the detection limits documented in the literature, this large range of values was used for initial evaluation [140, 161, 141, 165].

For the EWMA chart, smoothing coefficients from 0.1 to 0.4 were evaluated based on values reported in the literature [166, 167, 165].

The three algorithms were applied to the residuals of the pre-processing steps.

5.3.3 Detection using Holt-Winters exponential smoothing

As an alternative to the removal of DOW effects and sequential application of control charts for detection, a detection model that can handle temporal effects directly was explored [168]. While regression models are based on the global behaviour of the time series, the Holt-Winters generalized exponential smoothing is a recursive forecasting method, capable of modifying forecasts in response to recent behaviour of the time series [148, 149]. As the name suggests, the method is a generalization of the exponentially weighted moving averages calculation. Besides a smoothing constant to attribute weight to mean calculated values over time (level), additional smoothing constants are introduced to account for *trends* and *cyclic* features in the data [149]. The times-series cycles are usually set to one year, so that the cyclical component reflects seasonal behaviour. However retrospective analysis of the time series presented in this paper [157] showed that Holt-Winters smoothing [148, 149] was able to reproduce DOW effects when the cycles were set to one week. The method suggested by Elbert and Burkom (2009) [149] was reproduced using 3 and 5-day-ahead predictions ($n = 3$ or $n = 5$), and establishing alarms based on confidence intervals for these predictions. Confidence intervals from 85% to 99% were evaluated. Retrospective analysis showed that a long baseline yielded stabilization of the smoothing parameters in all time series tested when 2 years of data were used as training. Various baseline lengths were compared relative to detection performance. All time points in the chosen baseline length, up to n days before the current point were used to fit the model daily. Then the observed count of the current time point was compared to the confidence interval upper limit (detection limit) in order to decide whether a temporal aberration should be flagged [157].

5.3.4 Performance assessment

Two years of data (2010 and 2011) were used to qualitatively assess the performance of the detection algorithms (control charts and Holt-Winters). Detected alarms were plotted against the data in order to visually compare the results. This preliminary assessment aimed at reducing the range of settings to be evaluated quantitatively for each algorithm using simulated data. It also served to assess interpretability of the outputs by the end users of the system.

The choice of values for *baseline, guard-band* and *smoothing coefficient* (EWMA) were adjusted based on the visual assessments of real data, to ensure that the choices were based on the actual characteristics of the observed data, rather than impacted by artefacts generated by the simulated data. These visual assessments were performed using historical data where aberrations were clearly present, as in the BLV time series,in order to determine how different parameter values impacted the first day of detection, subsequent detection after the first day, and any change in the behaviour of the algorithm at time points after the aberration. In particular, an evaluation of how the threshold of aberration detection was impacted during and after the aberration days was carried out. Additionally, all data previously treated in order to remove excessive noise and temporal aberrations [157] were also used in these visual assessments, in order to evaluate the effect of parameter choices on the generation of false alarms. The effect of specific data characteristics, such as small seasonal effects or low counts, could be more directly assessed using these visual assessments rather than the quantitative assessments described later.

To optimize the detection thresholds quantitative measures of sensitivity and specificity were calculated using simulated data. *Sensitivity* of outbreak detection was calculated as the percentage of outbreaks detected from all outbreaks injected into the data. An outbreak was considered detected when at least one outbreak day

generated an alarm. Algorithms were also applied to the simulated baselines directly, without the injection of any outbreaks, and all the days in which an alarm was generated in those time series were counted as *false-positive* alarms. Specificity (the number of non-outbreak days correctly identified as such) was calculated as $(100\% - percentageoffalsepositivedays)$. *Time to detection* was recorded as the first outbreak day in which an alarm was generated, and therefore can only be evaluated when comparing the performance of algorithms in scenarios of the same outbreak duration. Sensitivity of outbreak detection were plotted against false positives in order to calculate the Area Under the Curve (AUC) for the resulting Receiver Operating Characteristic (ROC) curves.

5.4 Results

5.4.1 Preprocessing to remove the DOW effect

Autocorrelation function plots and normality Q-Q plots are shown in Figure 5.3 for the BLV series, for 2010 and 2011, to allow the two pre-processing methods to be evaluated. Neither method was able to remove the autocorrelations completely, but differencing resulted in smaller autocorrelations and smaller deviation from normality in all time series evaluated. Moreover, differencing retains the count data as discrete values. The Poisson regression had very limited applicability to series with low daily counts, cases in which model fitting was not satisfactory.

Due to its ready applicability to time series with low as well as high daily medians, and the fact that it retains the discrete characteristic of the data, differencing was chosen as the pre-processing method to be implemented in the system and evaluated using simulated data.

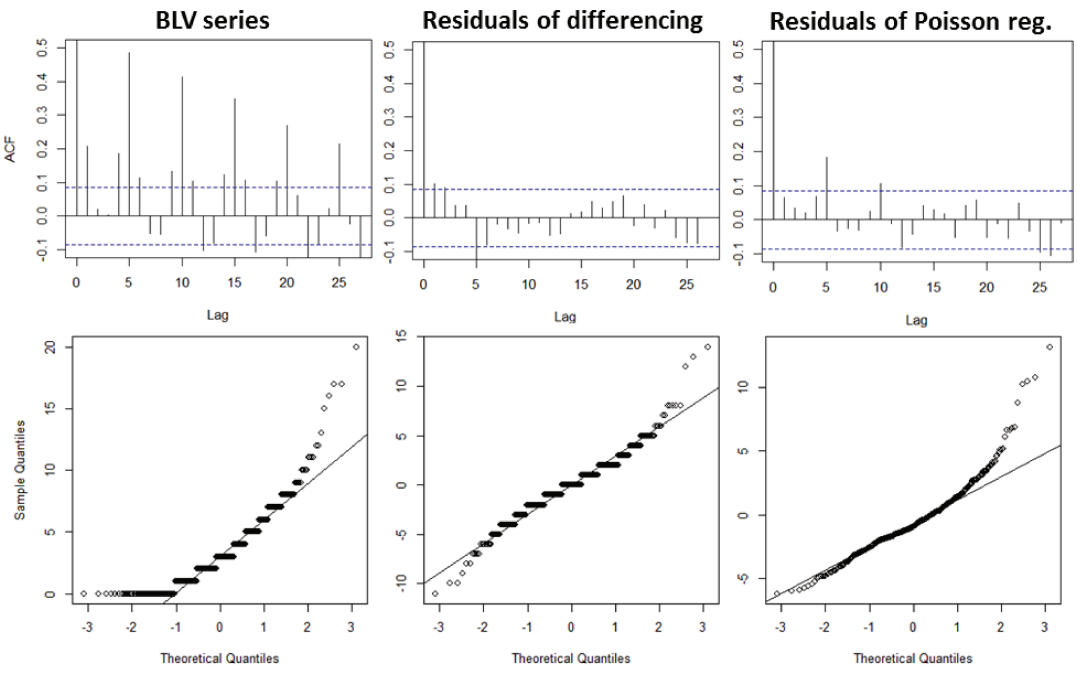


Figure 5.3: Comparative analysis of the autocorrelation function and normality plots for the BLV series (years 2010 and 2011) before and after pre-processing.

5.4.2 Qualitative evaluation of detection algorithms

Based on graphical analysis of the aberration detection results using real data, a baseline of 50 days (10 weeks) seemed to provide the best balance between capturing the behaviour of the data from the training time points and not allowing excessive influence of recent values. Longer baselines tended to reduce the influence of local temporal effects, resulting in excessive number of false alarms generated by random effects. Shorter baselines gave local effects too much weight, allowing aberrations to contaminate the baseline, thereby increasing the mean and standard deviation of the baseline, resulting in a reduction of sensitivity.

For the guard-band the use of one week did not prevent contamination of the baseline with aberrations when these were clearly present. The guard-band was therefore set to 10 days.

For the EWMA control charts, the number of alarms generated was higher when the smoothing parameter was greater. When evaluating graphically whether these alarms seemed to correspond to true aberrations, a smoothing parameter of 0.2 produced more consistent results across the different series evaluated and so this parameter value was adopted for the simulated data.

EWMA was more efficient than CUSUM in generating alarms when the series median was shifted from the mean for consecutive days, but no strong peak was observed. EWMA and Shewhart control charts appeared to exhibit complementary performance — aberration shapes missed by one algorithm were generally picked up by the other. CUSUM charts seldom improved overall system performance if the other two types of control chart had been implemented.

The performance of the Holt-Winters method was very similar with 3- and 5-day ahead predictions. Five-days ahead prediction was chosen because it provides a longer guard-band between the baseline and the observed data. Since this method is data-driven, using long baselines (2 years) did not cause the model to ignore local effects, but it did allow convergence of the smoothing parameters, eliminating the need to set an initial value. The method was set to read two years of data prior to the current time point. The use of longer baselines (up to 3 years) did not improve performance, but it would require longer computational time. The method did not appear to perform well in series characterised by low daily medians. In the case of the respiratory series, for instance, the Holt-Winters method generated 19 alarms over a period of 2 years, most of which seemed to be false alarms based on visual assessment (the control charts generated only 5-8 alarms for the same period).

Based on qualitative assessment alone, the range of detection limits to be evaluated using the simulated data could not be narrowed by more than half a unit for the control charts. It was therefore decided to evaluate 8 detection limits (in increments of 0.25) when carrying out the quantitative investigation: 2 to 3.75 for the Shewhart

charts, 1.75 to 3.5 for CUSUM charts and for EWMA. For the Holt-Winters method confidence intervals greater or equal to 95% were investigated using simulated data.

5.4.3 Evaluation using simulated data

Based on the results of the qualitative analysis (baselines of 50 days and a range or guard-band of 10 days) outbreaks were separated by a window of 70 non-outbreak days. In case of single-day spikes the separation was 71 days, to ensure that spikes always fell on a different weekday.

As expected, the effect of increased outbreak magnitude was to increase sensitivity (both per outbreak and per day) and reduce time to detection. Longer outbreak lengths increased the sensitivity per outbreak, but reduced the sensitivity per day in shapes with longer initial tails, as linear, exponential and lognormal. For these shapes a longer outbreak length also resulted in longer time to detection.

Receiver operating characteristics (ROC) curves for system sensitivities plotted against the number of false alarms are shown in Figure 5.4 for each of the four algorithms evaluated and the three syndromes. Lines in each panel show the median sensitivity for the five different outbreak shapes, along the eight detection limits tested. Error bars represent the 25% to 75% percentile of 12 scenarios, combining the four scenarios of outbreak magnitude (one to four times the baseline) and the three scenarios of outbreak duration (one to three weeks) simulated. Area under the curve (AUC) for the plots are shown in Table 5.1, as well as median time to detection for the specific scenario of an outbreak of 10 days. A limited number of detection limits are shown in Table 5.1.

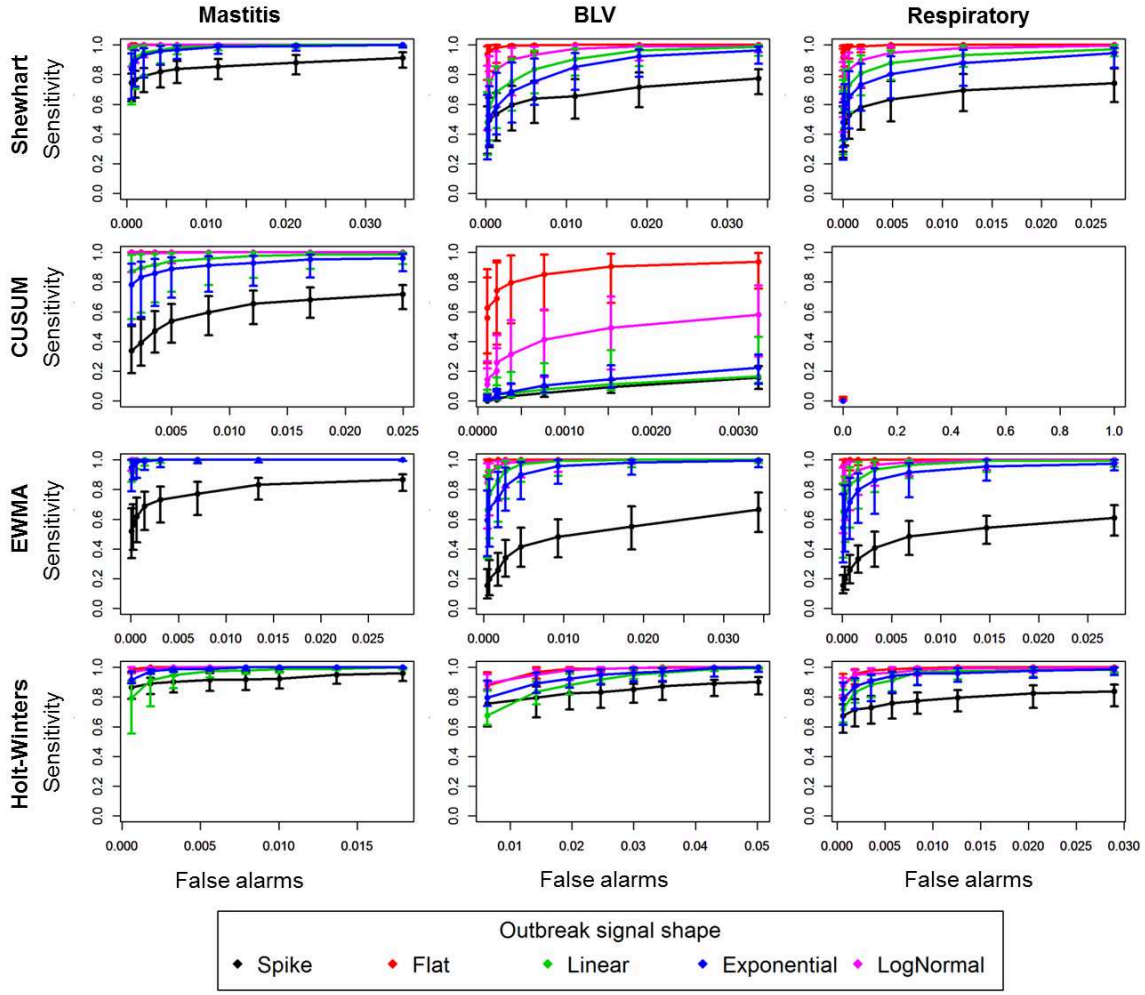


Figure 5.4: ROC curves representing median sensitivity of outbreak detection, plotted against number of daily false alarms, for four different algorithms evaluated (rows), applied to data simulating three different syndromes (columns), and using five different outbreak shapes. Detection limits for each plotted point are shown in Table5.1. Error bars show the 25% to 75% percentile of the point value over four different scenarios of outbreak magnitude (one to four times the baseline) and three different scenarios of outbreak duration (one to three weeks).

Table 5.1: Performance evaluation of different detection algorithms. Area under the curve (for sensitivity of outbreak detection) was calculated using the median sensitivity for all scenarios of each outbreak shape (four outbreak magnitudes and three durations), plotted against false positive alarms, for the different detection limits shown. These curves are shown in Figure 5.4. The median detection days for the four outbreak magnitudes simulated for each outbreak shape, in the scenario of a 10 days outbreak length, are also shown.

			Mastitis					BLV					Respiratory				
		Detection limits	Spike	Flat	Linear	Expon.	LogN.	Spike	Flat	Linear	Expon.	LogN.	Spike	Flat	Linear	Expon.	LogN.
	AUC-sens.outb.		0.843	0.965	0.899	0.884	0.953	0.694	0.934	0.709	0.686	0.806	0.676	0.930	0.715	0.673	0.791
Shewhart	Mean	3.75	–	1.11	3.39	4.93	5.07	–	1.33	4.48	5.69	5.64	–	1.37	4.61	5.92	5.90
	detection	3.00	–	1.20	4.47	6.63	5.83	–	1.61	5.84	7.47	6.74	–	1.71	5.90	7.74	6.86
	day	2.75	–	1.22	4.85	6.97	5.97	–	1.72	6.27	7.94	6.91	–	1.83	6.44	8.40	7.09
		2.00	–	1.30	5.87	8.11	6.52	–	2.12	6.99	8.83	7.49	–	2.23	7.27	8.88	7.52
	AUC-sens.outb.		0.654	0.975	0.912	0.868	0.972	0.501	0.777	0.504	0.505	0.554	–	–	–	–	–
CUSUM	Mean	3.50	–	1.35	5.31	8.05	6.43	–	2.90	8.27	9.76	8.26	–	–	–	–	–
	detection	2.75	–	1.56	6.15	8.79	6.80	–	3.57	9.03	10.00	8.60	–	–	–	–	–
	day	2.50	–	1.68	6.39	8.97	6.91	–	3.72	9.10	9.83	8.73	–	–	–	–	–
		1.75	–	2.01	7.05	9.40	7.28	–	4.07	9.00	5.00	9.02	–	–	–	–	–
	AUC-sens.outb.		0.737	0.971	0.965	0.946	0.971	0.559	0.961	0.797	0.764	0.889	0.563	0.952	0.800	0.747	0.859
EWMA	Mean	3.50	–	1.09	2.85	3.96	4.70	–	1.27	3.81	5.10	5.15	–	1.44	3.93	5.60	5.50
	detection	2.75	–	1.27	4.00	6.22	5.91	–	1.76	5.56	7.38	6.67	–	1.94	5.53	7.32	6.80
	day	2.50	–	1.37	4.38	6.79	6.14	–	1.98	5.96	7.86	6.93	–	2.14	5.98	7.76	7.10
		1.75	–	1.66	5.34	7.94	6.68	–	2.56	7.05	8.75	7.51	–	2.68	7.03	9.07	7.64
	AUC-sens.outb.		0.916	0.976	0.879	0.940	0.966	0.835	0.890	0.793	0.851	0.897	0.814	0.912	0.832	0.865	0.910
Holt-Winters	Mean	0.995	–	1.23	4.27	5.44	5.37	–	1.45	4.81	5.74	5.71	–	1.48	4.65	5.90	5.93
	detection	0.980	–	1.35	5.37	6.56	5.85	–	1.74	5.74	6.69	6.24	–	1.83	5.60	6.88	6.42
	day	0.975	–	1.42	5.72	6.94	6.00	–	1.81	6.07	6.86	6.41	–	1.96	5.79	7.14	6.55
		0.960	–	2.11	7.32	8.39	7.03	–	2.36	7.14	8.22	7.37	–	2.42	7.11	8.31	7.29

Starting at the first column of Figure 5.4 and Table 5.1, the results for the Mastitis simulated series, the sensitivity of detection of spikes and flat outbreaks was highest for the Holt-Winters method. EWMA charts showed low sensitivity for those, but the highest performance for all slow raising outbreak shapes (linear, exponential and lognormal). The lowest sensitivity within each algorithm was for the detection of spikes, which is an artefact of the short duration of these outbreaks, compared to all other shapes. Similarly, the relatively high sensitivity for flat outbreaks can be interpreted as a result of the higher number of days with high counts in this scenario. Similarly, the performance for detection in lognormal shapes closely related to the flat outbreaks, being superior to linear and exponential increases. The CUSUM algorithm showed good performance in the Mastitis series, but its performance very quickly deteriorated for other series with smaller daily medians, as discussed below.

Median day of first signal for each outbreak, in the scenario of a 10 days to peak outbreak, are shown in Table5.1 for a few key detection limits. Based on the median day of detection for the flat and exponential outbreaks in the Mastitis series, it is possible to see, for instance, that even though the AUC is higher for the Holt-Winters (more outbreaks detected) when compared to the Shewhart chart, in case of detection the latter algorithm detects outbreaks earlier than the first.

Syndromes with lower daily counts, Figure 5.4 shows that the performance of all algorithms decreases as daily counts decrease. The problem is critical with the CUSUM algorithm. Because this algorithm resets to zero if the difference in observed counts is *lower* than the expected counts, its application to a series with a large number of zero counts (Respiratory) resulted in no alarm being detected, true or false.

The results show that algorithm performance is not only a function of the syndrome median counts, but also impacted by the baseline behaviour of the syndromic series. EWMA charts, which performed better than Holt-Winter for slow rising outbreaks

in the Mastitis series, also performed better for flat shapes in the BLV series, but Holt-Winters performed better for exponentially increasing outbreaks. In syndromes with even lower daily counts, as the Respiratory series, the Holt-Winters method outperformed EWMA charts in all outbreak shapes but flat, the case for which both the EWMA charts and the Shewhart charts showed better performance than Holt-Winters.

The impact of the underlying baseline in the absence of outbreaks is also seen in the range of false positive values. The same detection limits generated a greater number of false alarms in the BLV series for all algorithms. Except for the BLV series, the number of false alarms generated in every scenario was smaller than 3% (1 false alarm in each 30 days of system operation). For the Holt-Winters method, a detection limit of 97.5% would always result in specificity greater than 97%, without loss of sensitivity compared to the lowest detection limits evaluated. For the EWMA charts a detection limit of 2 standard deviations represents the maximum attained specificity without starting to rapidly decrease sensitivity, but the behaviour should be evaluated individually for different syndromes. For the Shewhart chart such a cut-off seemed to rest on a detection limit of 2.25 standard deviations for the lower count series, but for the Mastitis series a limit of 2.5 would reduce false alarms with very little reduction in sensitivity.

5.5 Discussion

A recent review of veterinary syndromic surveillance initiatives [134] concluded that, due to the current lack of computerized clinical records, laboratory test requests represent the opportunistic data with the greatest potential for implementation of syndromic surveillance systems in livestock medicine. In this paper we have evaluated two years of laboratory test request data, using the two preceding years as training

data, and illustrated the potential of different combinations of pre-processing methods and detection algorithms for the prospective analysis of these data where the primary aim is aberration detection.

A large number of studies have documented the use of public health data sources in syndromic surveillance, such as data from hospital emergency departments, physician office visits, over-the-counter medicine sales, etc [169]. In livestock medicine, however, the epidemiological unit for clinical data is usually the herd, rather than individual animals [134]. The number of epidemiological units in a catchment area for individual data sources is therefore generally smaller than in public health monitoring, resulting in challenges around handling data with low daily counts, such as those described in this paper. It is hoped that the description of the steps taken to prepare these data and to select appropriate detection algorithms together with the results of this evaluation can guide the work of other analysts investigating the potential of syndromic data sources in animal health.

The data used for algorithm training had been previously evaluated retrospectively [157] and were found to have a strong day-of-week (DOW) effect. This effect prevented the direct use of control-charts without data pre-processing. Regression (using a Poisson model) was not an efficient method to remove daily autocorrelation; in line with a finding previously reported by Lotze et al (2008) [94]. Differencing has been recommended not only to remove DOW effects, but any cyclical patterns in addition to linear trends [94]. Five-day (weekly) differencing demonstrated solid performance in removing the DOW effect, even in series with low daily counts, and preserved the data as count data (integers). Preserving the data as integers is important when using control-charts based on count data, and also in order to facilitate the analysts' comprehension of both the observed and the pre-processed data series.

When pre-processed data were subjected to temporal aberration detection using control charts, EWMA performed better than CUSUM. EWMA's superiority in de-

tecting slow shifts in the process mean is expected from its documented use [94]. In the particular time series explored in this paper the general poor performance of the CUSUM was attributed to the low median values, when compared to traditional data streams used in public health. The injected outbreak signals were simulated to capture the random behaviour of the data, as opposed to being simulated as monotonic increases of a specific shape. Therefore, as seen in Figure5.2, often the daily counts were close to zero even during outbreak days, as is common for these time series. As a result, the CUSUM algorithm was often reset to zero, decreasing performance. Shewhart charts showed complementary performance to EWMA charts, detecting single spikes that were missed by the first algorithm.

The use of control-charts in pre-processed data was compared to the direct application of the Holt-Winters exponential smoothing. Lotze et al. (2008) [94] have pointed out the effectiveness of the Holt-Winters method in capturing seasonality and weekly patterns, but highlighted the potential difficulties in setting the smoothing parameters as well as the problems of one-day-ahead predictions. In this work the temporal cycles were set to weeks, and the availability of two years of training data allowed convergence of the smoothing parameters without the need to estimate initialization values. Moreover, the method worked well with predictions of up to 5 days ahead, which allows a guard-band to be kept between the training data and the actual observations, avoiding contamination of the training data with undetected outbreaks [163, 88, 164]. Our findings confirm the conclusions of Burkom, et al., 2007 [148] who found, working in the context of human medicine, that the method outperformed ordinary regression, while remaining straight-forward to automate.

Analyses using real data were important in tuning algorithm settings to specific characteristics of the background data, such as baselines, smoothing constants and guard-bands. However, analysis on real data can only be qualitative due to the limited amount of data available [28]. The scarcity of data, especially those for which

outbreaks days are clearly identified, has been noted as a limitation in the evaluation of biosurveillance systems [170]. Data simulation has been commonly employed to solve the data scarcity problem, the main challenge being that of capturing and reproducing the complexity of both baseline and outbreak data [171, 170]. The temporal effects from the background data were captured in this work using a Poisson regression model, and random effects were added by sampling from a Poisson distribution daily, rather than using model estimated values directly. Amplifying background data using multiplicative factors allowed the creation of outbreaks that also preserved the temporal effects observed in the background data.

Murphy and Burkom (2008) [164] pointed out the complexity of finding the best performance settings, when developing syndromic surveillance systems, if the shapes of outbreak signals to be detected are unknown. In this work the use of simulated data allowed evaluation of the algorithms under several outbreak scenarios. Special care was given to outbreak spacing, in order to ensure that the baseline used by each algorithm to estimate detection limits was not contaminated with previous outbreaks [172, 119, 173, 156].

As the epidemiological unit in animal health is a herd, transmission by direct contact is not usually the main source of disease spread. Although the movement of cattle with poor biosecurity measures plays an important role in pathogen transmission, indirect contact between farms through the movement of people and vehicles is often a large component of disease spread [174]. The shape of the outbreak signal that will be registered in different health sources is hard to predict, and depends on whether the contacts, which often cover a large geographical area [158], will also be included in the catchment area of the data provider. The temporal progression of outbreaks in animal herds is often modelled as an exponential progression [175, 176], but data from documented outbreaks [160], and the result of models which explicitly take into account the changes in spread patterns due to spatial heterogeneity [177]

more closely resemble linear increases. Linear increases may also be observed when an increase in the incidence of endemic diseases is registered, as opposed to the introduction of new diseases. Due to these uncertainties, all the outbreak signal shapes previously documented in simulation studies for development of syndromic monitoring were reproduced in this paper [172, 119, 173, 156].

Evaluation of outbreak detection performance was based on sensitivity and specificity, metrics traditionally used in epidemiology, combined using the area under the curve (AUC) for a traditional ROC curve [111]. The training data used in this work to simulate background behaviour was previously analysed in order to remove aberrations and excess noise [157]. The number of false alarms when algorithms are implemented using real data is expected to be higher than that observed for simulated data. However, all the detection limits explored, generated less than 3% false alarm days (97% specificity) in the simulated data, which is the general fixed false-alarm rate suggested for biosurveillance system implementations [172]. Because the right tail of the ROC curves was flat in most graphs, it was possible to choose detection limits that provide even lower rates of false alarms, with little loss of sensitivity.

Metrics used in the industrial literature to evaluate control charts, such as average run length, are specifically designed for detection of a sustained shift in a parameter [178], which corresponds to the flat outbreak shape simulated in this work, but would be misleading when used to interpret the algorithms' performance for other outbreak scenarios. Therefore, although at times recommended for the evaluation of prospective statistical surveillance [179], performance measures from the industrial literature were not used [178].

The results showed that no single algorithm should be expected to perform optimally across all scenarios. EWMA charts and Holt-Winters exponential smoothing complemented each other's performance, the latter serving as a highly automated

method to adjust to changes in the time series that can happen in the future, particularly in the context of an increase in the number of daily counts or seasonal effects. However, Shewhart charts showed earlier detection of signals in some scenarios, and therefore its role in the system cannot be overlooked. The CUSUM charts, however, would not add sensitivity value to the system.

Besides the difference in performance when encountering different outbreak signal shapes, the "no method fits all" problem also applied to the different time series evaluated. The performance of the same algorithm was different between two series with similar daily medians (results not shown). This was likely due to non-explainable effects in the background time series, such as noise and random temporal effects. Therefore, the choice of a detection limit which can provide a desired balance between sensitivity and false alarms would have to be made individually for each syndrome.

The use of these three methods in parallel — differecing + EWMA; differencing + Shewhart; and Holt-Winters exponential smoothing — assures that algorithms with efficient performance in different outbreak scenarios are utilised. Methods to implement automated monitoring aimed at early detection of temporal aberrations occurrence using multiple algorithms in parallel will be evaluated in future steps of this work.

Chapter 6

Syndromic surveillance using veterinary laboratory data: algorithm combination and customization of alerts

6.1 Abstract

Syndromic surveillance research has evolved from the search for efficient temporal aberration detection algorithms, to focus on methods to combine multiple approaches. This work combines three algorithms that have demonstrated solid performance in detecting outbreaks signals of varying shapes when clinical syndromes in cattle had been subjected to monitoring. These are: two control-charts which detect different shapes of outbreak – Shewhart control charts designed to detect sudden spikes and EWMA control charts developed to detect slow increases in counts – together with a method which can explicitly account for temporal effects, Holt-Winters exponential smoothing. A scoring system to detect and report alarms using these algorithms in a complementary way is proposed. This system also provides robustness and flexibility in the establishment of what signals constitute an alarm. This flexibility also allows an analyst to customize the system for different syndromes. The use of multiple algorithms in parallel resulted in increased system sensitivity. Specificity measured using simulated data was slightly decreased, but the number of false alarms when the approach was applied to real data was limited (between 1 and 3 per year for each of ten syndromic groups monitored), so the credibility of the system should not be affected. The automated implementation of this approach, including a method for on-line removal of temporal aberrations after they have been detected, is described. Regular reports on the data streams being monitoring in the laboratory are emailed to analysts regardless of the presence of alarms, a feature that contributes to situational awareness in animal health.

Keywords: laboratory, syndromic surveillance, temporal aberration detection, outbreak detection, simulation, sensitivity,specificity, evaluation

6.2 Introduction

The emergence of new diseases and the increasing threat of bioterrorism have motivated the development, especially since the turn of the century, of surveillance systems focused on the early detection of disease. Early work in the field focused on identifying data that could contain signatures of disease introduction, resulting in the exploration of various data sources registering healthcare-seeking behaviours, such as sales of over-the-counter medicine, emergency hospital visits and laboratory test requests [20]. While these data precede diagnostic confirmation, observations can be aggregated and monitored based on syndrome characteristics; an approach which led to the term "syndromic surveillance" entering the scientific literature [3].

The next steps in syndromic surveillance research focused on the development of detection algorithms [88]. Algorithm development and evaluation took into consideration the specific temporal characteristics of surveillance data, such as daily autocorrelations, seasonal trends and day-of-week effects [94]. This also had to consider the context of any detection, such as the availability of temporal and/or spatial data, the influence of external factors in any particular source of data, or even the availability of multiple, and sometimes conflicting, data streams [171]. This research indicated that different algorithms demonstrate better performance in different scenarios (e.g. different 'shapes' of temporal aberration) [88], and efforts are being made to combine approaches, rather than settle on one 'best' algorithm [25, 2].

Attention has also been given to the issue of preventing aberrations that do occur from reducing the performance of detection algorithms that operate prospectively and in any automated manner. Researchers using data-driven methods have demonstrated that sensitivity of detection can be increased by the use of a "guard-band" between the baseline data and the time point being evaluated, in order to avoid contamination of the baseline with an outbreak signal [94, 171, 180]. Methods

for preventing parameters from being automatically updated in case of an alarm, for model-based systems, have also been discussed [149, 181]. However, the use of detection algorithms to remove detected aberrations from the time series, during automated monitoring, has not to the knowledge of the present authors been discussed.

In previous work we addressed the use of diagnostic test requests made to an animal health laboratory as a syndromic data source, first preparing the data for use [142, 157] and then evaluating the performance of different aberration detection algorithms [182]. The results indicated that Shewhart and Exponentially Weighted Moving Averages (EWMA) control charts, as well as Holt-Winters exponential smoothing, could detect temporal aberrations in the data with high sensitivity. However, none of these approaches was superior to the others in all scenarios of outbreak signal shape and duration. The previous results also highlighted the need to customize the system for the different time series (i.e. syndromic groups) being evaluated, given the effect of the baseline data on algorithm performance, something also discussed in Buckeridge et al. 2007 [28].

In this paper the use of all three algorithms, in combination, is explored. The automated implementation of this approach, including a method for the removal of temporal aberrations after they have been detected, is described. A scoring system to detect and report alarms using these algorithms in a complementary way is proposed. This system also provides robustness and flexibility in the establishment of what signals constitute an alarm. This flexibility also allows an analyst to customize the system for different syndromes.

6.3 Methods

6.3.1 Data source

The Animal Health Laboratory (AHL) is a full-service veterinary diagnostic laboratory that serves livestock, poultry and companion animal veterinarians in the province of Ontario, Canada. The laboratory receives around 65,000 case submissions per year, resulting in the execution of over 800,000 individual laboratory tests, of which 10% relate to cattle submissions. Test requests for diagnoses of disease in cattle were monitored at the day of submission – pre-diagnostic. Syndromic groups were created based on the type of sample submitted and the diagnostic test requested by the veterinarian. The automated rule-based classification process to determine the appropriate syndrome is described in Dorea et al. (2013) [142]. Individual health events were defined as any single syndromic occurrence per herd on a given day. Time series composed of daily counts of events for each specific syndromic group will be referred as "syndromic series".

Seventeen syndromic groups were defined. Only syndromic series with a median of greater than one case per day (10 from the total 17) were monitored daily [157]. Two of these are presented here as they help illustrate the methods developed: daily counts of laboratory test requests related to mastitis diagnostics in cattle (*mastitis series*) and for identification of Bovine Leukemia Virus (*BLV series*).

Data from 2008 and 2009 were used as training data. These data had been previously analysed in order to remove temporal aberrations, creating *outbreak-free baselines* for each syndrome [157]. Data from 2010 and 2011 were used to evaluate the methods described.

6.3.2 Simulated data

The data simulated in a previous study, which evaluated the performance of each detection algorithm individually [182], was also used to evaluate the performance achieved by combining algorithms in this study. These data were simulated using a Poisson regression model with variables to account for day-of-week and month to reproduce the normal behaviour of the baseline series. The predicted value for each day of the year was set to be the mean of a Poisson distribution, and this distribution was sampled randomly to determine the value for that day in a given year, for each of 50 simulated years. Outbreak signals were then injected simulating five different shapes (single spike, moving average, linear increase, exponential increase and lognormal increase), four magnitudes (one to four times the baseline counts) and three lengths to peak (one, two and three weeks). Several simulated time series were generated, each containing only one specific outbreak type, repeated over 200 times, and separated by at least 70 days of non-outbreak data. Details are described in Chapter 5.

6.3.3 Algorithms for aberration detection

Based on previous quantitative evaluations [182], using the actual syndromic series as well as simulated data with controlled injection of outbreaks, three detection algorithms were selected with the following detection settings:

- Exponentially Weighted Moving Averages control charts (EWMA) with a smoothing parameter of 0.2, baseline of 50 days, guard-band of 10 days (time between the point being evaluated and the baseline), and a detection limit of 2 standard deviations.
- Shewhart control charts with a guard-band of 10 days, baseline of 50 days, and detection limit of 2.25 standard deviations.

- Holt-Winters exponential smoothing (HW) with a baseline of 2 years, and detection limit based on the upper limit of the 97.5% confidence interval for model prediction; using 5-day-ahead predictions (guard-band).

The control charts were applied to data pre-processed by weekly differencing, while the HW method was applied to data directly.

6.3.4 Correcting temporal aberrations

In order to develop a robust method for correcting the baseline series in cases of temporal aberrations, applicable to all algorithms being tested, a correction value based on the detection limits for each algorithm was specified. Algorithms vary in the way they calculate a detection limit (standard deviations above the mean for control charts, and construction of a confidence interval for the Holt-Winters method), but the existence of such a threshold for the generation of an alarm is a common feature among all temporal aberration detection methods.

Each algorithm was trained using *outbreak-free baseline* data constructed using data from 2008 and 2009. During prospective monitoring, on-line automated temporal aberration correction was implemented by specifying that, in case of alarm, the detection limit should be stored as part of the *outbreak-free baseline*, rather than the observed value for that time point. This process is outlined schematically in Figure 6.1.

This process was implemented individually – in parallel – for each of the three algorithms. Real data available for the years 2009 and 2010 were used to assess, upon evaluation of plotted results, whether the correction method was successful in avoiding contamination of the baseline with temporal aberrations. Plots were also used for qualitative evaluation of the ability of each algorithm to remain sensitive to outbreak signals after the first aberration days had been incorporated into the

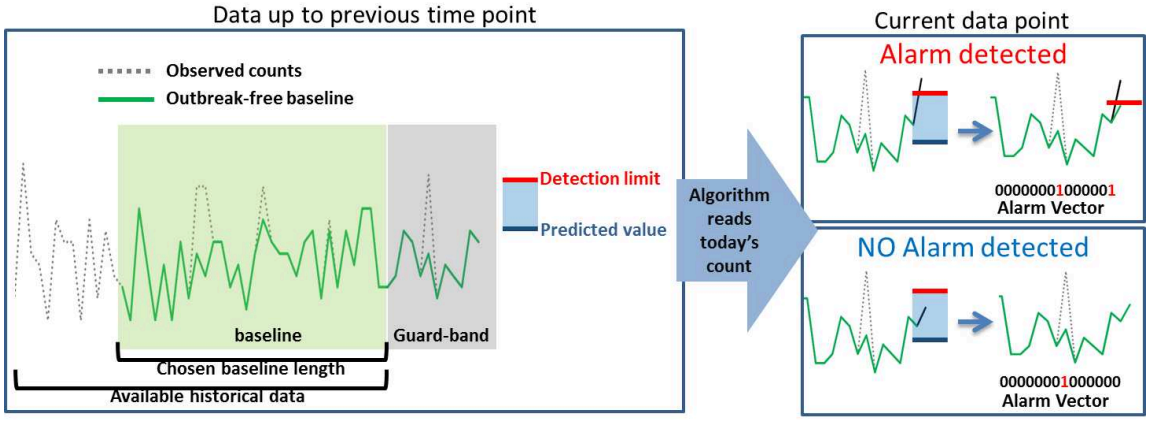


Figure 6.1: Schematic representation of the on-line process of aberration detection and correction of the *observed counts* series in case of alarm, in order to continually store an *outbreak-free baseline*.

baseline. Quantitative evaluation was then performed using simulated data. Using the *BLV series*, 100 years of baseline activity were simulated as described above. Two-hundred flat outbreak signals of two weeks duration and magnitude equal to three times the baseline data were injected in pairs, one outbreak pair per year. Each pair was composed of two outbreaks separated from each other by only 10 days. The percentage of outbreak signal days detected (sensitivity per day) was compared for outbreak signals which were from the first or the second in a pair.

6.3.5 Combining algorithms: Scoring system

During the process of evaluating the detection limits which would provide the best balance between sensitivity and specificity of aberration detection, for each algorithm, it was observed that no single detection limit would provide this optimum balance for all 10 syndromic series evaluated [182]. It became evident that the system should be able to operate under multiple detection limits. This was explored by maintaining several detection limits for each algorithm in all syndromic series, and using these to generate an overall score representing the "severity" of any alarm.

For each algorithm, five detection limits were implemented: the detection limit

that should result in the preferred balance between sensitivity and specificity for most of the ten syndromic series evaluated, as noted above for each algorithm; and two additional detection limits above and below this initial value. The lower detection limits are more sensitive, and the higher limits more specific. The lowest detection limit for each algorithm was determined as one which would result in specificity equal to 97% in at least 6 of the 10 syndromic series evaluated [182]. The five detection limits for each algorithm are shown in Table 6.1.

Table 6.1: Detection limits for each of the three algorithms implemented, and corresponding alarm scores. EWMA=Exponentially Weighted Moving Averages; Shewhart=Shewhart Control Charts; HW=Holt-Winters Exponential Smoothing.

	EWMA*	Shewhart*	HW**
Score=1	1.50	1.75	95.5%
Score=2	1.75	2.00	96.5%
Score=3	2.00	2.25	97.5%
Score=4	2.25	2.50	98.5%
Score=5	2.50	2.75	99.5%

*sd = standard deviation. **confidence interval.

Each detection algorithm evaluates the current count for the syndromic series being monitored using all five detection limits, and a *detection score* is generated corresponding to how many of these thresholds the current value reached, that is, a *detection score* between 0 and 5.

Combining the decision of the different algorithms in this context became straightforward, as the detection scores for each algorithm could be added, in order to generate a *final alarm score* between 0 and 15. Customization of detection for individual syndromic series was implemented by allowing the analyst to set a *reporting threshold* independently for each syndrome. That is, the analyst can manipulate the minimum *final alarm score*, which causes the system to report an alarm, by syndrome. This threshold can be changed at any time in order to increase sensitivity (using a lower threshold) or specificity (setting a higher threshold).

Data from 2010 and 2011 were used to test the scoring system, in order to visualize

the alarms generated by the system against real data streams. Then, using simulated data, system sensitivity and specificity were estimated. The scoring system was applied to over 100 simulated outbreaks of each shape, magnitude and duration, in order to calculate the sensitivity of the system using different *reporting thresholds* (1 to 15). *Sensitivity per outbreak* was calculated as the percentage of outbreak signals detected from all outbreak signals injected in the data. An outbreak was considered to have been detected when at least one outbreak day generated an alarm. *Sensitivity per day* was also calculated as the percentage of days that generated an alarm from all outbreak signal days.

The percentage of days with *false alarms* was calculated after applying the same thresholds to 35 years of simulated data which had not been injected with outbreaks. The modeled variability in syndromic counts according to month and day of the week, and the stochastic elements added by sampling values from a Poisson distribution, were assumed to mimic the natural variability in real data that would tend to generate false alarms.

6.3.6 System reports

Once a set *reporting threshold* is reached for a given syndromic series, an alarm is generated, that is, a report is triggered. Syndromic surveillance development based on this data source has been an initiative of the data provider (the AHL) and the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA), responsible for the programs of animal disease surveillance in the province. A designated pathologist from AHL and a designated epidemiologist from OMAFRA are the end users of the system developed, and are referred as the "analysts". These analysts will be responsible for receiving system outputs, interpreting them, and if necessary following up on alarms. The contents of the reports generated in case of an alarm were discussed with analysts, and the final format adopted is presented in the results. These reports

were generated as PDF files, which were then automatically emailed to analysts in case of alarms. Analysts also receive reports for every syndromic series in a regular weekly email.

All methods were implemented using modules from the R environment (http://www.r-project.org/) [143].

6.4 Results

6.4.1 Correcting temporal aberrations

The results of time series correction using the detection algorithms are shown in Figure 6.2 for the *BLV series* in 2010. For the control charts the series subjected to monitoring is composed of the residuals of weekly differencing (applied to remove temporal effects), rather than the observed time series, which are shown in green. The results indicated that automated aberration removal using the detection algorithm was effective. Besides a visual comparison between the original and the corrected time series, this conclusion is based on the fact that all algorithms were able to flag aberrations in multiple days, separated by a time period of over 10 days, which is the maximum guard-band used. This implies that the aberrations observed did not contaminate the baseline, and that the algorithms remained sensitive to consecutive aberration days subsequent to the initial day of outbreak.

Quantitative evaluation showed that there was a reduction in sensitivity when outbreaks signals were present in the baseline period used to train the aberration detection algorithms, as was expected. If the on-line correction of aberration is not implemented, the difference in sensitivity per day between the two outbreak signals injected in a pair is 24.3% for the Shewhart control chart, 14.9% for the EWMA control chart, and 8.8% for the Holt-Winters exponential smoothing. When

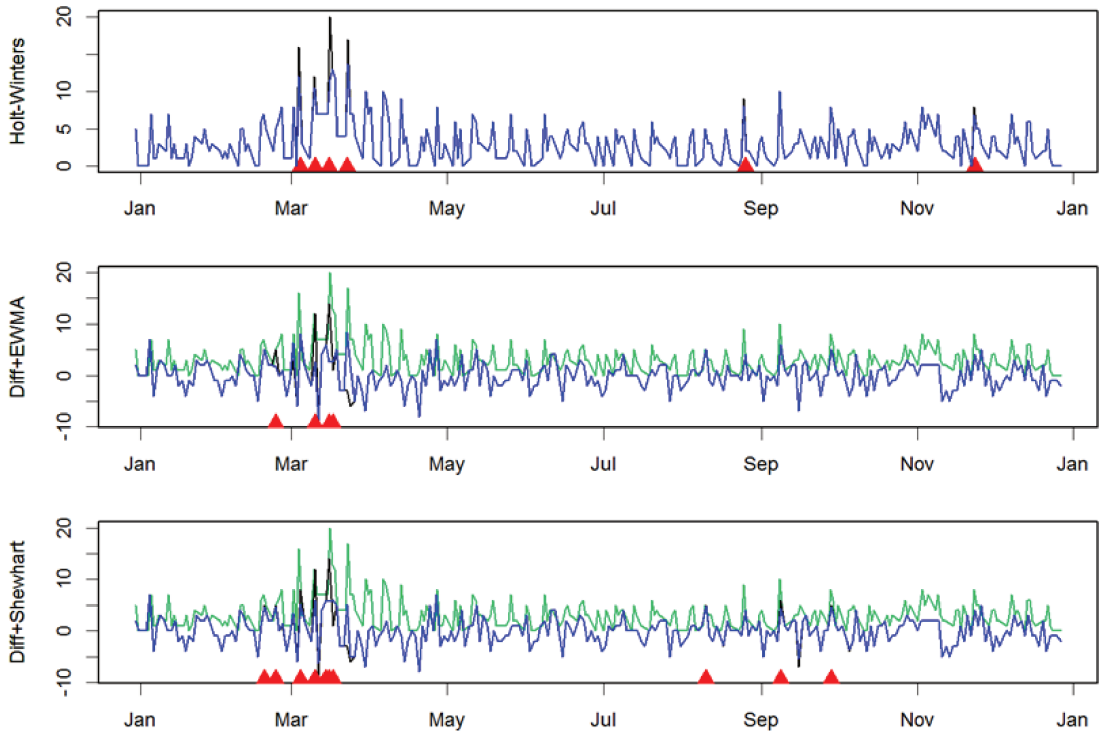


Figure 6.2: Alarm detection and automated correction in the *BLV series* for 2010. Before attempting to use control charts to detect aberrations the data have been pre-processed to remove temporal effects using weekly differencing. The original data is represented in green lines. The series subjected to monitoring are shown in black, superimposed on by blue lines showing the adjusted series after detection and removal of temporal aberrations. Alarms are shown as red triangles along the bottom of each graph.

automated correction was implemented, these differences were reduced to 12.6%, 6.2% and 3% respectively. In addition to the smaller reduction in sensitivity, the HW correction approach presents two qualitative advantages over the other two methods. First, the results are simpler to interpret. Since the time series is not altered by differencing, analysts can readily compare the corrected series with the original observed counts, in order to qualitatively assess the performance of the detection algorithm in correcting aberrations. Second, because this algorithm can deal with the temporal effects present in the data, its predictions for each time point reproduce these effects.

The implementation of a combined scoring system allowed the three detection algorithms to be implemented in parallel. However, for their results to be combined sensibly it was considered essential that the algorithms were operating under the same conditions, that is, that they were using the same baseline. If aberration correction was implemented in parallel using all three algorithms, as time passed, and most especially in case of repeated temporal aberrations, each of them would effectively utilise different baselines. Based on the conclusion presented above, HW exponential smoothing was selected as the sole method to remove aberrations from the data.

6.4.2 Scoring system

6.4.2.1 Detection using real data: qualitative analysis

Figure 6.3 shows the results of applying the scoring system to 2010 data, for the *mastitis* and *BLV* syndromic series. In this figure a reporting threshold of 7 for both syndromes was used to illustrate the method. That is, the analyst will only receive a report when the vertical bars representing the summed detection scores for all algorithms *is equal or greater than 7*(and therefore is higher than the grey shaded

area, which limit was set to 6.5). This would have happened only once, in April, for the *mastitis series*, and on three occasions, likely related to the same ongoing aberration in March, for the *BLV series*. Visual evaluation of aberration detection performance is difficult due to the day-of-week effects in the data, which can be misleading when judging the presence of aberrations. True quantitative analyses are reported below.

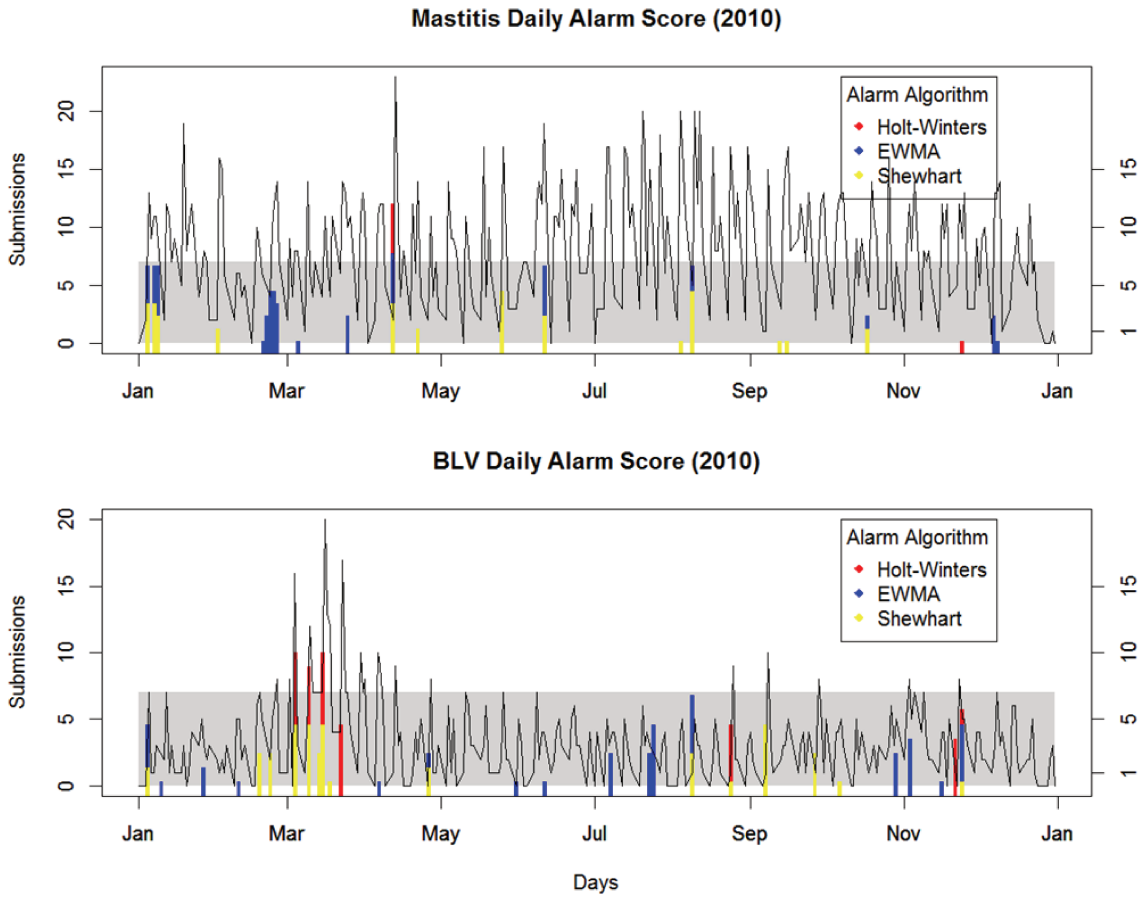


Figure 6.3: Aberration detection using three algorithms (Shewhart control charts, EWMA control charts and Holt-Winters exponential smoothing) combined using the scoring system and applied to real data. The top panel plots the *mastitis series* for the year 2010. Detection scores for each algorithm are shown as vertical bars, stacked to give a final alarm score which scale is shown in the secondary axis. The gray rectangle is used to mark the limit in the secondary axis which corresponds to the reporting threshold – here 7. The bottom panel shows a similar graph for the *BLV series*.

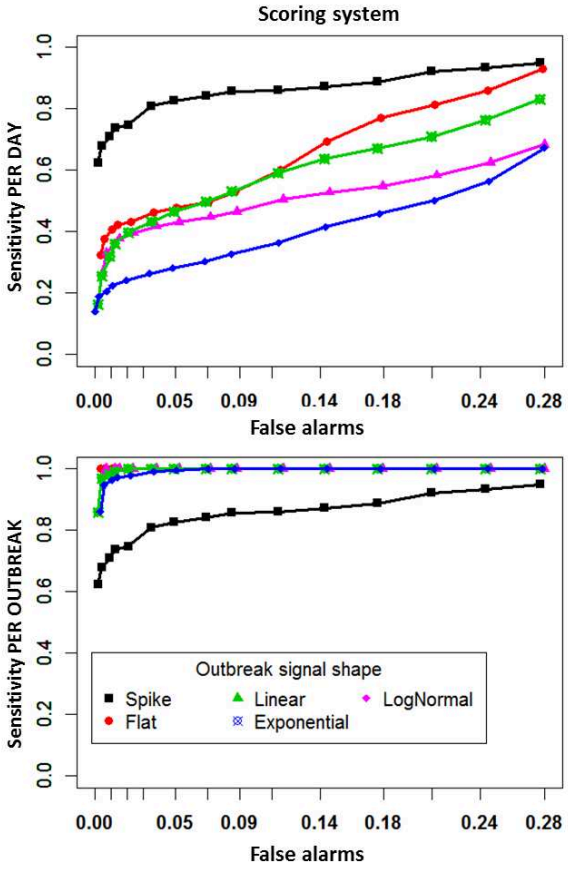
Application of the scoring system to real data was an important step in evaluating

how the system might add value to the analysis performed. The analyst can, by looking at the graphs illustrated in the two top panels of Figure 6.3, compare the final alarm score to the information regarding the behaviour of the data. The analyst will also know which algorithms were responsible for the alarm signal, and the individual scores generated. For the *BLV series*, for instance, small absolute signals between July and September indicate that some days with seemingly normal activity resulted in the generation of detection signals by the EWMA algorithm; likely indicating that these observations were somewhat unusual for that day-of-week. For the *mastitis series* a generally higher concentration of test requests can be observed between July and September. However, looking at the detection scores, indicates that these were not temporal aberrations, with only a few, low level, detection signals being generated by the control-chart algorithms. The HW algorithm, which can account for temporal effects more explicitly, did not generate any detection signals during this period, leading to the hypothesis that the generally increased numbers may be a temporal effect, such as a seasonal trend. It is also evident that lowering the *reporting threshold* of the *mastitis series* to 6, for instance, would have generated a much larger number of alarms (likely false ones) in the 2010 syndrome data set. Eight other syndromes monitored as part of this research were evaluated (graphs not shown), and three of these required adjustment of their combined *reporting threshold* to a value of 9 or 10 in order to prevent excessive numbers of false alarms.

6.4.2.2 Detection using simulated data: quantitative analysis

Figure 6.4 shows the results of applying temporal aberration detection using the scoring system against the simulated *mastitis series*. As many scenarios were evaluated, only the median performances are shown, and the graphs aim at highlighting the comparative performance of system settings across a range of outbreak shapes. The results indicate that decreasing the reporting threshold of the scoring system

can result in great sensitivity, but at the cost of higher levels of false detection.



Reporting Threshold	False Alarms	Sensitivity per Outbreak Day				
		Spike	Flat	Linear	Exponential	LogNormal
1	27.7%	95.0%	93.2%	83.2%	67.2%	68.2%
2	24.3%	93.4%	86.1%	76.3%	56.3%	62.4%
3	21.0%	92.2%	81.3%	70.9%	50.1%	58.3%
4	17.6%	88.8%	76.9%	67.1%	45.8%	54.7%
5	14.3%	87.2%	69.3%	63.7%	41.4%	52.6%
6	11.4%	86.0%	60.1%	58.9%	36.2%	50.4%
7	8.5%	85.7%	52.8%	53.0%	32.5%	46.4%
8	6.9%	84.1%	49.6%	49.6%	30.1%	44.6%
9	4.9%	82.6%	47.6%	46.4%	28.0%	43.1%
10	3.5%	81.0%	46.2%	43.1%	26.1%	41.5%
11	2.1%	74.8%	43.1%	39.6%	24.0%	39.4%
12	1.2%	73.6%	42.2%	35.9%	22.3%	37.6%
13	0.9%	70.9%	40.6%	31.7%	20.5%	35.8%
14	0.4%	67.8%	37.7%	25.3%	18.6%	32.9%
15	0.2%	62.4%	32.3%	16.1%	13.9%	26.3%

Reporting Threshold	False Alarms	Sensitivity per Outbreak				
		Spike	Flat	Linear	Exponential	LogNormal
1	27.7%	95.0%	100.0%	100.0%	100.0%	100.0%
2	24.3%	93.4%	100.0%	100.0%	100.0%	100.0%
3	21.0%	92.2%	100.0%	100.0%	100.0%	100.0%
4	17.6%	88.8%	100.0%	100.0%	100.0%	100.0%
5	14.3%	87.2%	100.0%	100.0%	100.0%	100.0%
6	11.4%	86.0%	100.0%	100.0%	100.0%	100.0%
7	8.5%	85.7%	100.0%	100.0%	100.0%	100.0%
8	6.9%	84.1%	100.0%	100.0%	100.0%	100.0%
9	4.9%	82.6%	100.0%	100.0%	99.6%	100.0%
10	3.5%	81.0%	100.0%	100.0%	99.1%	100.0%
11	2.1%	74.8%	100.0%	100.0%	97.8%	100.0%
12	1.2%	73.6%	100.0%	99.1%	97.3%	100.0%
13	0.9%	70.9%	100.0%	98.1%	96.3%	100.0%
14	0.4%	67.8%	100.0%	97.0%	94.9%	100.0%
15	0.2%	62.4%	100.0%	85.9%	86.0%	97.8%

Figure 6.4: Sensitivity of detection and false alarm rates when the combined algorithms are applied to the simulated *mastitis series*, with five different shapes of simulated outbreaks. The table rows and graph nodes show different final alarm scores used as the reporting threshold. Values in the table correspond to the median among 3 different outbreak magnitudes (1 to 4 times the background activity of the series) and 3 different outbreak lengths (1, 2 and 3 weeks; except for the spike, which is always one single day).

Besides high sensitivity, an advantage of using multiple algorithms is shown when the detection is compared to individual algorithms for each shape of outbreak. Figure 6.5 compares the results associated with sensitivity and false alarms in the *BLV series* by contrasting the performance of the combined approach with the individual algorithms as documented in previous work [182]. The EWMA algorithm, for instance, would show performance superior to the combined approach for spike signals, but the sensitivity of the algorithm is lower for all other outbreak signal shapes. A similar result can be observed for the Shewhart control chart. For a set rate of false

alarms the scoring system can generally achieve higher sensitivity than the EWMA charts.

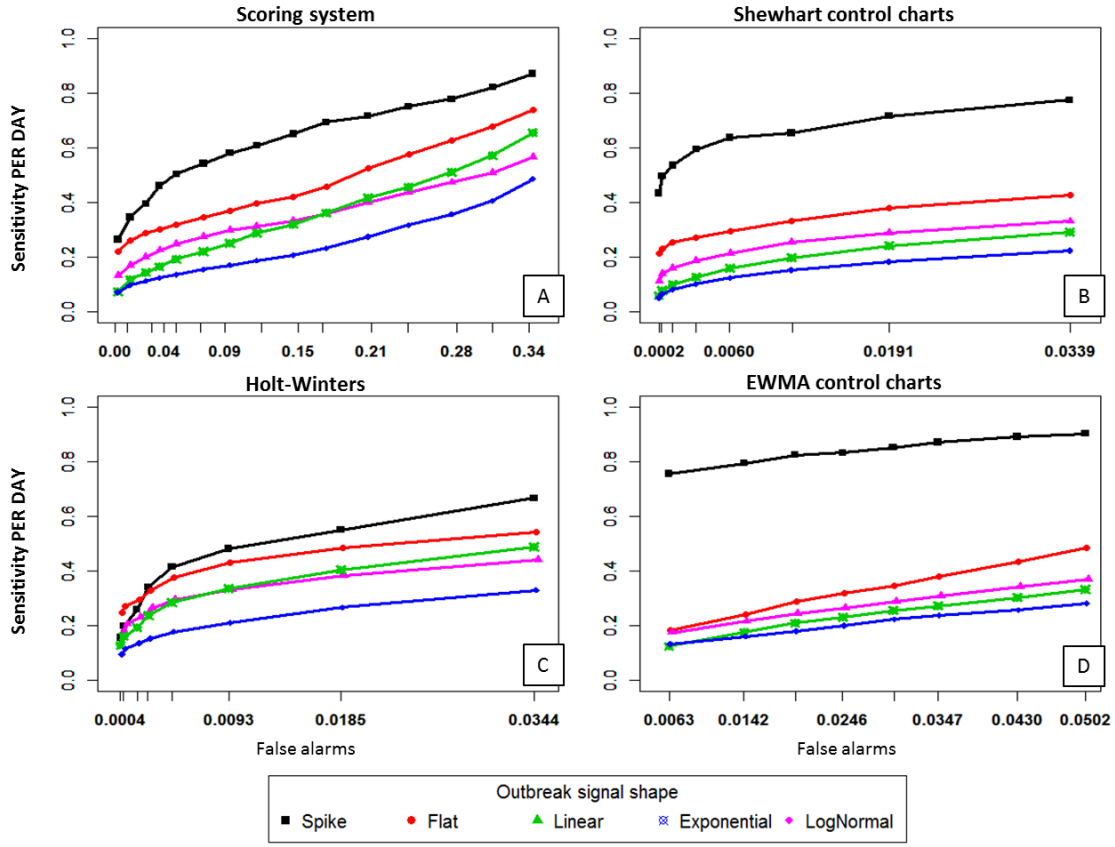


Figure 6.5: Sensitivity of detection and false alarms rate in the *BLV series*. Panel A shows the sensitivity of detection compared to false alarms rate when all three algorithms combined using the scoring system are applied to the simulated BLV series, with five different shapes of simulated outbreaks injected. The graph nodes show different final alarm scores used as the reporting threshold. Points represent the median among 3 different outbreak magnitudes (1 to 4 times the background activity of the series) and 3 different outbreak lengths (1, 2 and 3 weeks; except for the spike, which is always one single day). The remaining panels show sensitivity and false alarm when each detection algorithm is applied individually, as previously documented in [182].

Overall, Figure 6.5 illustrates that the use of multiple algorithms in combination allows the system to achieve higher sensitivity for a range of outbreak types, as outbreak shapes not detected by one algorithm tend to be able to be detected by another.

6.4.3 System reports

In the event of an alarm, the analyst receives an e-mail with an attached PDF file. The first page of the file contains the list of all the syndromes being monitored, with all those for which an alarm has been generated on the given day highlighted in red. Individual reports for the syndrome(s) which generated alarm(s) follow on individual pages. An example report page is shown in Figure 6.6. This report was generated because the *final alarm score* for the Respiratory series was 12, against a defined *reporting threshold* of 7. In the top panel, the analyst could see the *final alarm score* for the current day, which shows why the report was generated. The analyst could also quickly glance at the previous 4 days. In the next panels the analyst could have a broader view of the data behavior, as well as detection algorithm performance, over the last 6 months. In this middle panel detection scores are plotted as a secondary axis, and the reporting threshold is shown as a gray box in the background. The bottom panel allows the analyst to assess visually the performance of aberration correction.

6.5 Discussion

The role of laboratory data in the rapid detection of outbreaks has been recognized in public health, partly due to the extensive area coverage provided by these data in comparison to clinical data coming from individual practitioners or hospitals [9]. In a series of steps we have developed methods and a system to implement syndromic surveillance in animal health based on veterinary laboratory data. Having concluded that laboratory test requests represented an opportunistic data source with great potential for syndromic surveillance systems in livestock medicine [134], we explored diagnostic submissions for cattle made to the Animal Health Laboratory in the province of Ontario, Canada, in order to construct a monitoring and early

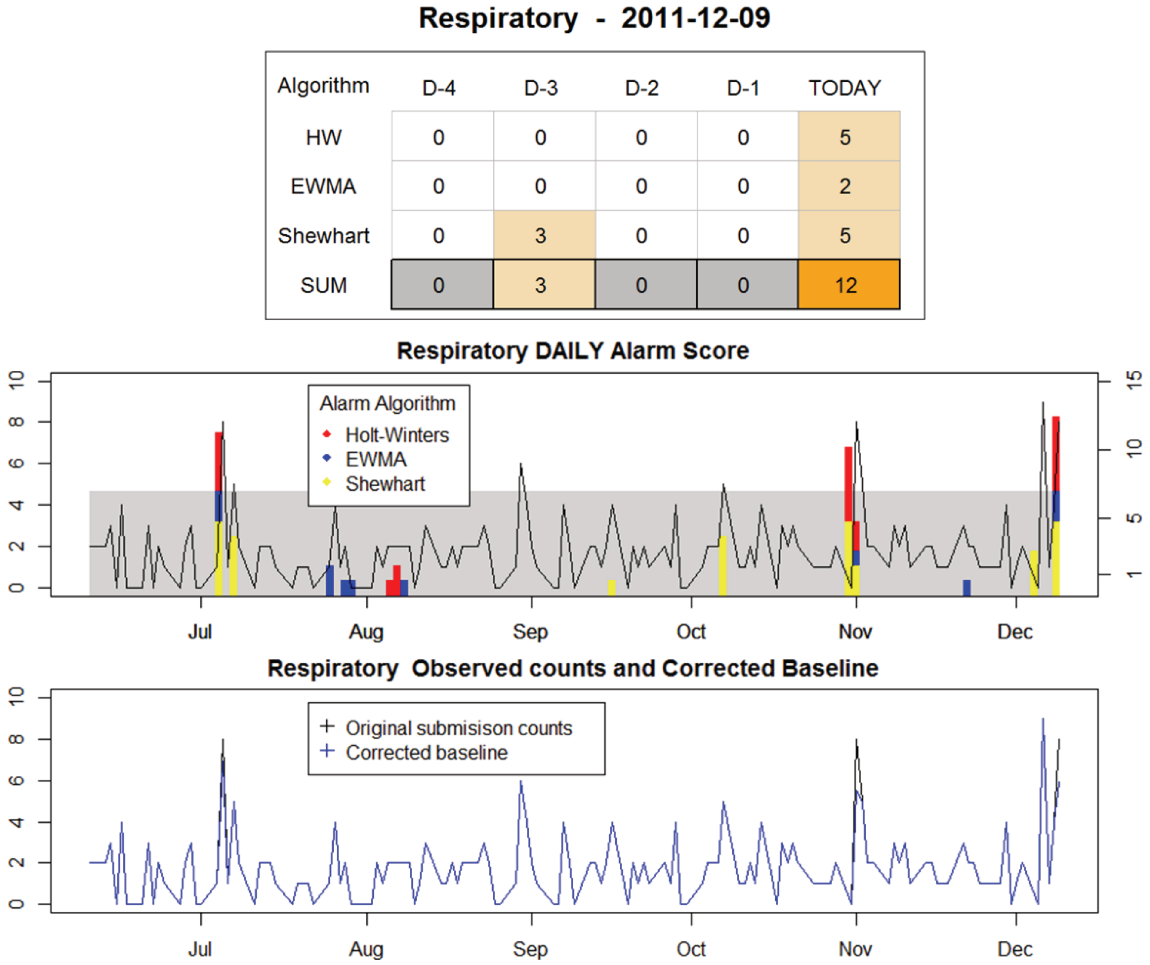


Figure 6.6: Example page of a daily report sent to analysts in case of alarm. The top table shows the detection score for the three algorithms used, in the last 5 days. Next the data of the last 26 weeks are plotted against the detection score for all three detection algorithms used, stacked to give a final alarm score. The main axis is the scale for the data, and the secondary axis gives the scale for the detection scores. The gray rectangle shows the range of final alarm score which will not generate an alarm. The bottom panel shows the observed data, superimposed by the data after aberration correction by the detection algorithm.

disease warning system for that province.

Steps to classify data into syndromes [142], and to evaluate these data retrospectively so they could be prepared for monitoring [157], were documented. Using the data available together with simulated data, the performance of different temporal aberration detection algorithms was evaluated [182]. This indicated that algorithm performance depended on the shape of the outbreak signals encountered, as well as the baseline characteristics of each individual syndromic series being monitored.

In the current paper, the implementation of multiple algorithms in parallel has been explored, together with the challenge of preventing aberrations from contaminating the training data set. The latter goal was addressed first. The correction of baseline series in case of alarms has mainly been discussed for regression methods [149, 181], and it is generally based on preventing model parameters from updating in cases where an alarm has been generated. The use of a guard-band between the baseline data and the time point being evaluated [94, 88, 180] can avoid contamination of the baseline with an outbreak signal before its first detection, but it does not prevent baseline contamination by the aberration after that point. In the present work the detection limit of each algorithm is used to correct the observed data, continuously storing an *outbreak-free baseline* which is used by the algorithms as training data. The method proved effective for all three algorithms explored, however, Holt-Winters exponential smoothing was chosen due to its advantages in terms of interpretability and explicit modeling of temporal effects in the data.

A detection system should be able to detect a variety of outbreaks with different signatures [172]. However, different temporal aberration detection algorithms typically demonstrate optimal performance for outbreaks with a specific temporal progression pattern; a challenge if one specific algorithm has to be selected [25]. The use of multiple algorithms in parallel has been explored through the use of decision rules which pool the binomial results from different algorithms [25], or by using

goodness-of-fit tests to decide when to switch between algorithms [183].

This work combined three algorithms that had demonstrated solid performance in detecting outbreaks of varying shapes across a range of syndromes which had been subjected to monitoring [182]. These algorithms were: two control-charts designed to detect different shapes of outbreak – Shewhart control charts designed to detect sudden spikes and EWMA control charts developed to detect slow increases in counts – together with a method which can explicitly account for temporal effects, Holt-Winters exponential smoothing. For each algorithm, multiple detection limits were used, in order to transform the outcome of each method into a magnitude score, rather than a binomial signal indicating whether an aberration was present or not. These detection scores were then combined to produce a final alarm score. All algorithms contribute to the measure of alarm magnitude, and this combined magnitude is used to decide whether analysts should receive an alarm report or not. In case of any alarm, in addition to regular weekly reports, analysts can review the output of all three detection algorithms across the range of monitored syndromes.

The use of magnitude scores, rather than a binary alarm decision, results in the analyst being responsible for the definition as to when an alarm will be triggered. This is seen as a positive feature. Considering the number of external factors that can influence fluctuations in the data being entered into any syndromic surveillance system it is expected that, once an alarm has been raised, a human analyst will review the output in the light of relevant factors and decide whether a true problem exists [148]. This is even more critical in animal health data than in the human case, since laboratory submission is not just a function of disease but also of animal value [1], and several economic factors have been associated with the rate of diagnostic submission to laboratories [184].

It becomes critical, therefore, to develop system outputs that provide as much information as possible, according to the capabilities of the data at hand and the

system. This was addressed by developing output charts that combine observed data with the detection scores for all three algorithms, plotted over time, and provided frequently to the analyst. Although the charts combine a lot of information, the consistency of the presentation results in rapid familiarization. Once the analyst gets used to their interpretation, the frequent inspection of reports will train the analyst in the behaviour of the data and the algorithms. For this reason the monitoring results for all syndromes are emailed to analysts weekly, regardless of the detection of any signal. Should an alarm be detected, the analyst will be able to judge, based on the past behaviour of the data, whether to challenge that alarm.

User autonomy, however, should not endanger the system performance. The fact that analysts can change the reporting threshold but not the actual settings of the algorithms – i.e., the detection limits applied to the data and the thresholds which trigger data correction to remove temporal aberrations and excessive noise – assures that users' choices will not deprecate the system in time. The performance of the algorithms is expected to remain unchanged regardless of the choices made by the analysts concerning reporting. If for instance an analyst reduces drastically the sensitivity of reporting by setting excessively high reporting thresholds, as soon as this is corrected the reporting sensitivity will be restored. Because data correction will have continued as the system runs, the outbreak-free baseline will not be contaminated with temporal aberrations, which are corrected, even if a reporting threshold was not reached, every day in which the middle detection threshold for the Holt-Winters exponential smoothing method is exceeded.

In evaluating the performance of a syndromic surveillance system when applied to historical data, it is generally difficult to estimate its performance due to lack of documentation as to the causes of the temporal aberrations registered in the data [9], such as those seen in the *BLV series* in Figure 6.3. However, the continuous inspection of system outputs should enable analysts to progressively tailor the system

for optimum performance. The system described here allows for a high degree of customization by the analyst, who can change the reporting threshold that triggers an alarm individually for each series, according to the observed behaviour of the algorithms, or to comply with institutional objectives. For instance, if too many false alarms are being observed for a specific syndrome, the reporting threshold for that individual syndrome can be raised, in order to increase specificity. Reporting thresholds can be set high in order to generate fewer reports; perhaps limiting analysis to the inspection of the regular weekly reports. If a specific syndrome required more intensive monitoring, the threshold could be lowered to increase sensitivity.

The choice to combine all three algorithms, however, came at the cost of a decreased specificity for the system as a whole (slightly higher rates of false alarms), which is expected behaviour when multiple diagnostic tests are applied in parallel [185]. Surveillance systems based on laboratory data in general should prioritize sensitivity and timeliness over specificity, since the coverage of laboratory data is small (that is, "small increases in laboratory data often indicate larger communitywide outbreaks" [9]). Widdowson et al. (2003), however, also highlighted that this increase in sensitivity should not result in an unmanageable number of signals. Despite the higher percentage of false alarms identified when using simulated data, the results of applying detection based on the scoring system to real data showed that the number of detected outbreaks was never greater than 4 per year for any of the 10 evaluated syndrome series, provided that the individual reporting thresholds were optimized for each syndrome. With continuous system optimization, the number of false alarms is expected to decrease [9], without any significant loss in system sensitivity.

The potentially high number of false alarms during the initial phases of system implementation, as well as the need for continuous inspection of system reports and parameters optimization, have been key points of contention in an ongoing debate regarding the value of syndromic surveillance [186]. After comparing syndromic

surveillance results to outbreaks detected locally by traditional surveillance van den Wijngaard et al. (2011) recommended, "the use of syndromic surveillance to reveal blind spots of traditional surveillance", as well as for, "monitoring disease burden and virulence shifts of common pathogens". The system discussed here, developed using laboratory submission data to the AHL, will serve as a backup to traditional animal health surveillance in the province of Ontario, detecting outbreaks that are widespread across the province. Moreover, the second recommendation made by van den Wijngaard et al. (2011), is a key feature of this system with regular compilations of observed data being delivered to analysts, which will contribute to situational awareness in animal health surveillance.

6.6 Conclusion

A system for automated monitoring of clinical syndromes in cattle has been developed using laboratory data. The system will be used for near real-time detection of temporal aberrations which can be indicative of disease introduction or increase in the incidence of endemic cattle diseases in the province of Ontario, Canada. In order to maintain automation the system has been optimized not only to detect but also to correct temporal aberrations in the data streams monitored, which is expected to ensure that the algorithms remain sensitive to multiple aberrations in the future. Holt-Winters exponential smoothing showed superior quantitative and qualitative performance as the algorithm of choice for on-line removal of aberrations.

A method was proposed to implement temporal aberration detection using this algorithm in parallel with two control charts: Shewhart and exponentially weighted moving averages. The use of a scoring system to combine these temporal aberration detection algorithms allowed great flexibility in the definition of what constitutes an alarm, and also flexibility to customize system parameters for each syndrome being

monitored, based on ongoing evidence. The choice to combine all three algorithms, however, came at the cost of a decreased specificity for the system. If the shape of the next outbreak to be detected was known, the best performing algorithm could be applied individually, achieving the same sensitivity with much higher specificity. As this is not possible, however, combining the three algorithms allows the system to take advantage of the strengths of each of them. The resulting system is effective in aberration detection, flexible to adjustments aimed at increasing performance for specific syndromes or complying with analysts' goals, and contributes to situational awareness in animal health.

Chapter 7

Conclusions and future directions

After a decade of development in public health, the development of modern biosurveillance systems takes speed in the area of veterinary medicine at a time when there is an increasing concern regarding the utility of syndromic surveillance [5]. The general conclusion is that the real utility can only be realized if attention is given to the potential of these systems for more than early disease detection. Already in 2005 Hurt-Mullen and Coberly [169] stated that syndromic surveillance, which was first developed with the goal of detecting disease outbreaks, "are becoming a basic tool for public health epidemiologist".

Current focus rests on developing holistic biosurveillance systems which work in parallel — rather than as a substitute — and help inform and complement all other types of surveillance [5, 187]. Authors, who in the last decade researched the potential of data sources and the use of different aberration detection algorithms, now highlight the need to continue research into system design and implementation. They also point out the need to compare methods developed under different conditions and to document their performance, fomenting the sharing of experience in order to enhance syndromic surveillance utility [5, 188, 4].

This dissertation has documented the process of developing and implementing syndromic surveillance methods in a source of data under explored for this purpose in animal surveillance, laboratory test requests for diagnosis of diseases in livestock.

Chapter 1 of this document presented a thorough review of the initiatives of syndromic surveillance in animal health. Struggles documented in the literature regarding acquisition of clinical data, and especially sustainability of systems based on voluntary participation of veterinarians or data providers in scattered locations, ratified the choice of using laboratory data in this research.

Chapters 2 through *6* of this dissertation described the steps of development of a system for automated extraction of surveillance information from the database of

the Animal Health Laboratory (AHL), in the province of Ontario. The collection of specimens for diagnostic purposes happens late in the disease continuum, and covers a small percentage of the population affected. For this reason, sensitivity and timely detection should be prioritized [9]. In the system developed, and described in this document, the data are monitored directly at the institution where they are collected and stored. A single, centralized, electronic source circumvents the need to gather data from scattered locations, as well as the need to develop standards that allow integration of these data from different sources. Issues of data transfer, sharing agreements and security are all also avoided if the data are monitored directly at the source institution.

Chapter 2 described the methods explored in order to implement automated classification of laboratory records into syndromes. The speed of extraction of surveillance information from health data does not depend solely on the rate of data collection, but also on the ability to translate data from the format they are collected to a meaningful surveillance output [8]. A challenge of working with laboratory data was that of determining how to transform diagnostic into epidemiological information, especially due to the limited clinical information received with each submission. On the other hand, due to the structured nature of laboratory data, when compared to clinical narratives, automated recognition of syndromes could be performed with simple methods based on keywords search. It was not necessary, for instance, to deal with contextual information, such as negation, expressions of uncertainty, etc [189]. The most time-consuming step of classification was the creation of a dictionary of keywords relevant to each classification task, and the definition of the relationship between these words, their co-occurrences and the target syndromic group. Once defined, however, these relationships were easily packed into a set of rules that achieved high classification performance.

Once classification was performed, the data were reduced to multiple time se-

ries registering daily (or weekly) submissions to the different syndromes monitored. *Chapter 3* described retrospective evaluation of these time series in order to identify temporal effects present, and define methods to model or remove them on-line, that is, when prospective, daily analyses were implemented. A method was presented for automated removal of excessive noise and historical outbreaks in historical data, in order to construct baselines of normal behaviour, which could be used as training data for the algorithms implemented in the next stages.

Chapters 4 and *5* described the prospective phases of system development, that is, the analyses which scan the time series in an on-line process, one day at a time, in order to detect temporal aberrations in comparison to a baseline of historical data. In *Chapter 4* several aberration detection algorithms were evaluated, using real and simulated data with injection of synthetic outbreaks. Upon the conclusion that no single algorithm was superior in all outbreak scenarios, in Chapter 5 a scoring system to combine algorithms was presented. This chapter also presented a method to continuously clean data from excessive noise and temporal aberration, which allows the system to keep an outbreak-free baseline for use by the detection algorithms.

The results from *Chapters 4* and *5* showed that a large scale simulation study was not able to prevent the need to tune system parameters individually for each syndrome monitored. The scoring system presented in Chapter 5 allowed fine tuning of detection limits per syndrome. On-line evaluation of the data was implemented beginning on January 1st, 2010, in order to simulate system implementation and use for two years (until the end of 2011), evaluate the outputs that would have been generated during those two years, and set alarm thresholds individually for each syndrome in order to improve sensitivity and specificity (reduce false alarms).

The delivery of an operational syndromic surveillance system to AHL and OMAFRA, as a result of this research, is expected to improve animal disease surveillance in the province, with very low cost of operation. However, the main outcome of this research

project is the scientific contribution to the field of veterinary syndromic surveillance, as attested by the five research manuscripts presented in Chapters 2 through 6. Statistical methods developed for public health data, and only extensively documented for use with those data, were through this research adapted and validated for use in data sources for which the epidemiological unit is the herd, rather than the individual, namely livestock health data. The issue of data classification was resolved through the use of relatively straight-forward algorithms which are also easily interpretable by medical experts. A novel method to produce outbreak-free baselines from historical data was presented. Algorithms for prospective detection of temporal aberrations were evaluated using large quantities of real and simulated health data in which herds were the epidemiological unit. Further, a method was proposed to use the detection algorithms themselves to clean temporal aberrations in time series in real-time, allowing the process of outbreak-free baselines to be reproduced in an automated manner when the system is operational. Finally, a novel method to combine multiple algorithms was proposed and evaluated, and its implementation using free-software was documented.

7.1 System delivery to AHL and OMAFRA

All the steps described in this dissertation have been implemented "off-line", using a batch of four years of data provided by the AHL. The research documented in *Chapters 3* through *6* resulted in the development of computational routines that could be automated and implemented "on-line", that is, evaluating data daily as they become available in the AHL database. The classification steps described in *Chapter 2* were implemented using the freely available data mining software RapidMiner 5.0 (Copyright 2001-2010 by Rapid-I and contributors), and the remaining statistical analyses were implemented in the open-source environment for statistical computing R [143]. The final result was a simple desktop application which could

be transferred to a computer inside the AHL. A RapidMiner process, containing R codes, was scheduled to run daily using Window's task scheduler. Security is assured because identification fields are never queried, and the data queried are never stored in the computer. The data pulled are converted into syndromic time series, and only matrices with daily counts for each syndrome are stored.

The process is described schematically in Figure 7.1

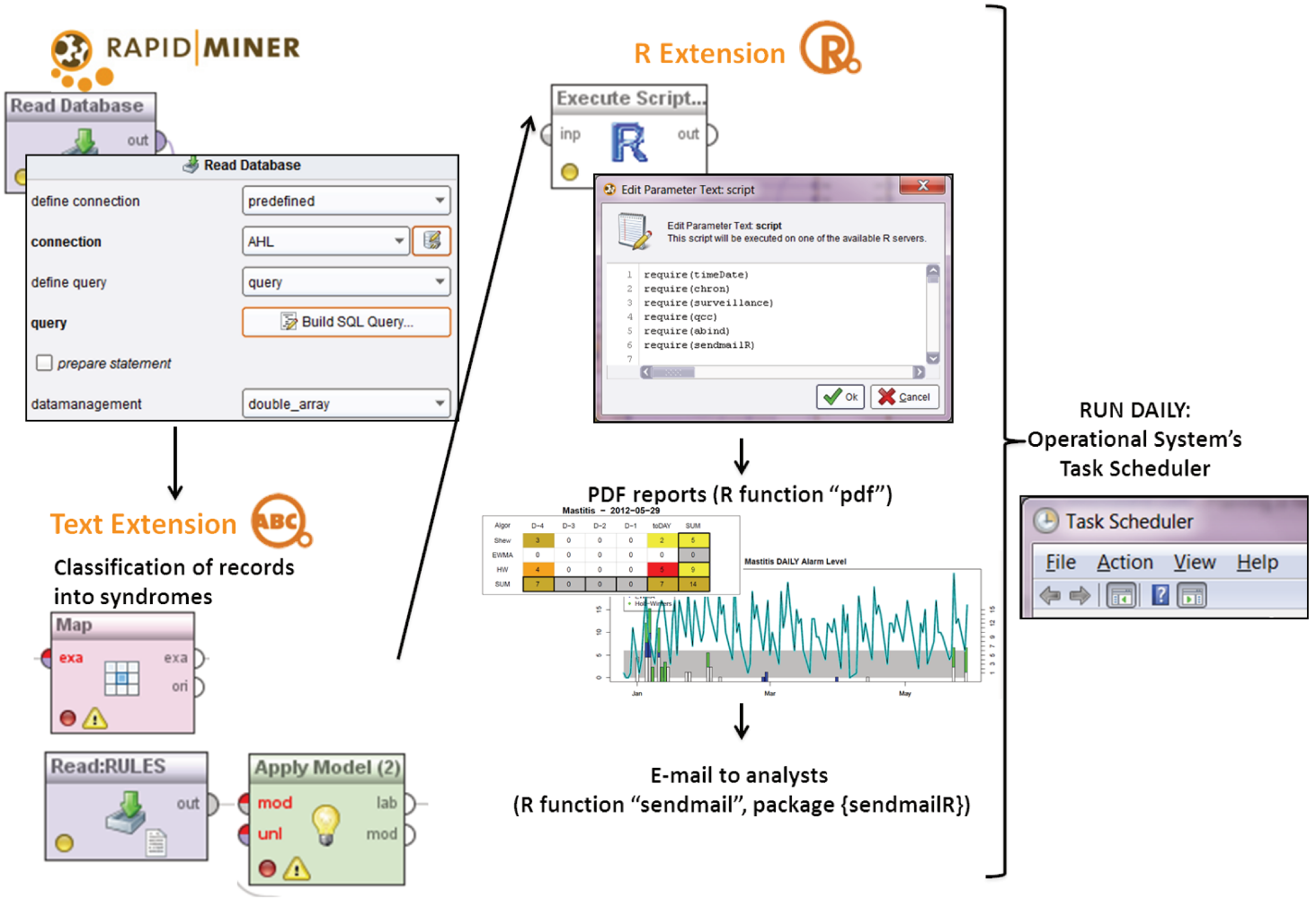


Figure 7.1: Overview of system implementation.

7.2 System limitations and potential for further development

The step of classification of laboratory data into syndromic groups allows the data to be converted into a format that can be useful for epidemiological purposes. In this initial development, focus was given to the monitoring of syndromic trends and early detection of temporal aberrations which could be indicative of a disease outbreak. However, once the initial framework is set, it can be built upon in order to increase system utility and tailor outputs to the needs of the AHL, the data provider, as well as the Ontario Ministry of Agriculture Food and Rural Affairs (OMAFRA), responsible for the animal surveillance in that province.

The final system delivered is not a computer application on its own, but a set routines for analysis using freely available statistical software. This resulted in fast achievements for the project, but outputs are also simple — pdf charts emailed to analysts — and little interaction and customization of outputs is allowed. Some previous attempts to develop syndromic surveillance in veterinary medicine have failed to sustain engagement and become established systems [134]. The simplicity of the system developed may facilitate its incorporation into AHL's and OMAFRA's routine activities, becoming a basic tool for diagnosticians and epidemiologists [169]. Once that happens, more resources may be deployed in order to involve technical informatics assistance, and develop more interactive outputs, allowing users to query surveillance information directly from the system.

The development of a system robust for several syndromes and with highly flexible definition of what constitutes an alarm represents a trade-off: fine customization is possible, but at the expense of analysts' time to review system outputs and tune parameters. As the analysts involved in the project become familiar with the system, the cycle of feedback between the analyst and the system will both increase performance of the system and feed the institution with surveillance information which

goes beyond early disease detection [9]. If efforts are not taken inside the institutions (AHL and/or OMAFRA) to allocate time from an analyst for this task, the utility of the system will not be employed to its fullest.

Animal health data are subjected to several sources of variation not related to diseases, such as economic factors [190]. The low counts for most syndromic groups monitored makes it difficult to model all these factors and take them into consideration when removing temporal effects. It may be necessary to deal with the false-alarms, which is easier when an analyst is constantly reviewing outputs and understands the data for each syndrome. Monitoring directly at the source institution has further advantages for reviewing each alarm. Analysts can seek more information in real time, or follow-up with the samples while they are still being processed at the institution, taking action for immediate response after diagnostic, or even before that depending on how strong was the evidence at the time of submission. The gain in early detection can only be realized if it triggers immediate reaction [8].

Analysis of data, currently, does not take into account the herd production type — beef versus dairy cattle —, as this information is not available within the AHL database. A main quality issue related to this limitation is that some syndromic groups may be largely represented by one production category (mastitis and dairy cattle being an obvious example), limiting the chances of detecting increases in the incidence of clinical cases when restricted to the the production type representing the minority of cases. If production type information can be collected for laboratory submissions in the future, a study could be performed to determine the impact of treating herds as homogeneous groups in the sensitivity of the system, in order to consider whether adding this source of variation could improve system performance.

A natural next step on the development of this syndromic surveillance system would be the inclusion of other animal species, besides cattle. The structure set up

during the course of this project would greatly facilitate this task. The mapping of any species-specific laboratorial tests would have to be defined, but the augmented list could be easily incorporated into the existing analysis flow. For the text mining steps, a few species-specific words would also have to be included in the dictionary of medically relevant words, but the majority of the words referring to organs and laboratorial specimens would be the same already included for cattle. Once data are classified into syndromes, the temporal effects of the newly constructed syndromic series would need to be evaluated individually. If the temporal effects observed are similar to those found in the cattle syndromic series, the additional time series could be directly incorporated into the analysis flow currently in place.

Detection of aberrations was only implemented using temporal information. Methods to detect spatial clusters were not explored due to limitations of the spatial data available in the database, mainly the recording of the postal code of the veterinarian attending the health event, rather than the postal code of the farm where the health event occurred. If the spatial location of the farm of origin of the samples can be recorded for all submissions in the future, the inclusion of spatial analysis in the system developed could increase the sensitivity of cluster detection.

An on-going challenge for the field of syndromic surveillance is the combination of multiple data sources when generating information that will guide surveillance activities and policy. Further research is needed in order to incorporate, for instance, clinical data into the existing system. Several challenges of combining multiple data sources have been documented in the literature [191]; and in veterinary medicine in particular this goal is hard to achieve due to the scarcity of centralized, routinely collected clinical data, in particular if those data need to be computerized. A more immediate step towards complementing the current system with extra information would be the inclusion of test results, which are available from the same data source (in this case the AHL), only at different days after submission. Currently, only in-

formation received for each sample at submission is taken into consideration. The system uses the type of samples and the tests performed in order to classify submissions into syndromic groups and count syndromic cases at the day of the submission. Further research could be conducted in order to use the results of the diagnostic tests, available days to weeks after depending on the test requested, in order to validate the syndromic classification performed by the system.

Besides classification validation, the incorporation of test results would allow the system to monitor the proportion of test requests that ultimately receive a diagnostic confirmation. A complementary approach to the work of diagnosticians would be the monitoring of the proportion of samples which do *not* get a final diagnosis. This can add to disease surveillance in two main ways. Firstly, an increase in the number of non-diagnosed samples may indicate the presence of an emerging pathogen. That is, veterinarians in the field are observing clinical signs which lead them to order tests for diseases they are familiar with, but those tests are consistently resulting negative because the agent responsible for the clinical manifestation is new in the area (emerging or foreign animal disease), or a new presentation of a disease is developing. O'Sullivan et al. (2012) [192] demonstrated that during a porcine circovirus associated disease outbreak in Ontario from 2004 to 2006, the probability of a positive polymerase chain reaction (PCR) for porcine reproductive and respiratory syndrome virus (PPRSV) at the AHL decreased. Monitoring the proportion of samples with "diagnostic not reached" for the purpose of detecting disease emergence has been documented in the UK [66]. The second important way in which the monitoring of samples with final diagnostic can aid surveillance is for documentation of disease freedom. A team responsible for investigating the use of laboratory data for surveillance purposes in Australia [193] has proposed that samples submitted for diagnostic of endemic diseases, which have a final diagnostic confirmed, should be used to document an existing passive surveillance system. The authors of the report point out that the confirmation of an endemic diseases rules out a number of foreign diseases

which may have similar clinical presentation, and represents evidence that certain diseases are not present in the country, as opposed to present but undetected. While the monitoring of results can only be performed with larger time delays than submissions, further research to implement monitoring of results in the system delivered to the AHL and OMAFRA could greatly benefit animal surveillance in the province of Ontario, and it is also a largely under studied subject in veterinary syndromic surveillance.

7.3 Conclusion

The development of modern biosurveillance systems is increasing in veterinary medicine after a decade of experience accumulated in public health. An automated computational routine has been developed for automated extraction of surveillance information from laboratory data, specifically laboratory test requests for diagnostic in cattle. All steps were set up using open source software, as a simple desktop application, which could be readily reproduced in other institutions. Fast development and simple maintenance is expected to lead to incorporation of this system into the routine of the data provider — the Animal Health Laboratory in the province of Ontario, becoming an indispensable tool for diagnosticians and epidemiologists, and fomenting further technical development. Suggested further developments are the inclusion of other animal species, and the collection of more accurate spatial information in order to allow the detection of spatial clusters.

Bibliography

[1] M. G. Doherr and L. Audige. Monitoring and surveillance for rare health-related events: a review from the veterinary perspective. *Philosophical transactions of the Royal Society of London.Series B, Biological sciences*, 356(1411):1097–1106, Jul 29 2001.

[2] Galit Shmueli and Howard Burkom. Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics*, 52:39–51, 2010.

[3] Centers for Disease Control and Prevention (CDC). Annotated bibliography for syndromic surveillance, 6 2006. Acessed on June 17th, 2010.

[4] Stephen S. Morse. Public health surveillance and infectious disease detection. *Biosecur Bioterror*, 10(1):6–16, Mar 2012.

[5] Ronald D. Fricker Jr. Some methodological issues in biosurveillance. *Statistics in Medicine*, 30:403–415, 2011.

[6] Kelly J. Henning. What is syndromic surveillance? *Morbidity and Mortality Weekly Report*, 53 Suppl:5–11, Sep 2004.

[7] Rebecca Katz, Larissa May, Julia Baker, and Elisa Test. Redefining syndromic surveillance. *Journal of Epidemiology & Global Health*, 1(1):21, 2011.

[8] S. E. Fienberg and G. Shmueli. Statistical issues and challenges associated with rapid detection of bio-terrorist attacks. *Statistics in Medicine*, 24:513 – 530, 2005.

[9] Marc-Alain Widdowson, Arnold Bosman, Edward van Straten, Mark Tinga, Sandra Chaves, Liesbeth van Eerden, and Wilfred van Pelt. Automated, laboratory-based system using the internet for disease outbreak detection, the Netherlands. *Emerg Infect Dis*, 9(9):1046–1052, Sep 2003.

[10] Haobo Ma, H. Rolka, K. Mandl, D. Buckeridge, A. Fleischauer, and J. Pavlin. Implementation of laboratory order data in biosense early event detection and situation awareness system. *Morbidity and Mortality Weekly Report*, 54 Suppl:27–30, Aug 2005.

[11] Julie A. Pavlin, Farzad Mostashari, Mark G. Kortepeter, Noreen A. Hynes, Rashid A. Chotani, Yves B. Mikol, Margaret A K. Ryan, James S. Neville, Donald T. Gantz, James V. Writer, Jared E. Florance, Randall C. Culpepper, Fred M. Henretig, and Patrick W. Kelley. Innovative surveillance methods for

rapid detection of disease outbreaks and bioterrorism: results of an interagency workshop on health indicator surveillance. *American Joumal of Public Heatth*, 93(8):1230–1235, Aug 2003.

[12] James W. Buehler, Ruth L. Berkelman, David M. Hartley, and Clarence J. Peters. Syndromic surveillance and bioterrorism-related epidemics. *Emerging Infectious Diseases*, 9(10):1197–1204, Oct 2003.

[13] Matthew Stone. The potential for exotic disease syndromic surveillance within veterinary laboratory submissions data. In *Proceedings of the Epidemiology and Animal health management Branch of the NVZA 2007*, 2007.

[14] Loren E. Shaffer, Julie A. Funk, Paivi Rajala-Schultz, Michael M. Wagner, Thomas E. Wittum, and William J.A. Saville. Evaluation of veterinary diagnostic laboratories as a possible data source for prospective outbreak surveillance. *Advances in Disease Surveillance*, 2:119, 2007.

[15] X. Zhang, B. McEwen, E. Mann, and W. Martin. Detection of clusters of salmonella in animals in Ontario from 1991 to 2001. *The CanadianVeterinary Journal*, 46(6):517–9, 522–3, Jun 2005.

[16] Calcin Schwabe. The current epidemiological revolution in veterinary medicine. part I. *Preventive Veterinary Medicine*, 1:5–15, 1982.

[17] H. Chen, S. S. Fuller, C. Friedman, and W. Hersh. *Medical Informatics: Knowledge Management and Data Mining in Biomedicine*. Springer Science+business Media, Inc., New york, USA, 2005.

[18] D. Zeng, H. Chen, C. Lynch, M. Eidson, and I. Gotham. *Infectious Disease Informatics and Outbreak Detection*. Knowledge Management and Data Mining in Biomedicine. Springer, 2005. ID: 1226.

[19] I.J. McKendrick, G. Gettinby, Y. Gu, A. Peregrine, and C. Revie. Hybrid information systems for agriculture: The case of cattle trypanosomiasis in Africa. *Outlook on Agriculture*, 23:262–267, 1995.

[20] Dena M. Bravata, Kathryn M. McDonald, Wendy M. Smith, Chara Rydzak, Herbert Szeto, David L. Buckeridge, Corinna Haberland, and Douglas K. Owens. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. *Annals of Internal Medicine*, 140(11):910–922, Jun 2004.

[21] Kenneth D. Mandl, J Marc Overhage, Michael M. Wagner, William B. Lober, Paola Sebastiani, Farzad Mostashari, Julie A. Pavlin, Per H. Gesteland, Tracee Treadwell, Eileen Koski, Lori Hutwagner, David L. Buckeridge, Raymond D. Aller, and Shaun Grannis. Implementing syndromic surveillance: a practical guide informed by the early experience. *Journal of the American Medical Informatics Association*, 11(2):141–150, 2004.

[22] M. M. Wagner, F. C. Tsui, J. U. Espino, V. M. Dato, D. F. Sittig, R. A. Caruana, L. F. McGinnis, D. W. Deerfield, M. J. Druzdzel, and D. B. Fridsma.

The emerging science of very early detection of disease outbreaks. *Journal of public health management and practice : JPHMP*, 7(6):51–59, Nov 2001.

[23] Daniel M. Sosin and J. DeThomasis. Evaluation challenges for syndromic surveillance–making incremental progress. *Morbidity and Mortality Weekly Report*, 53 Suppl:125–129, Sep 2004.

[24] R. D. Fricker Jr. Syndromic surveillance. an article for the encyclopedia for quantitative risk assessment. draft., 2006. Available at: http://faculty.nps.edu/rdfricke/docs/SyndromicSurveillance.pdf. Accessed on April 23rd, 2010.

[25] Inbal Yahav and Galit Shmueli. Algorithm combination for improved performance in biosurveillance systems. In *Proceedings of the 2nd NSF conference on Intelligence and security informatics: BioSurveillance*, BioSurveillance'07, pages 91–102, Berlin, Heidelberg, 2007. Springer-Verlag.

[26] M. D. Salman. *Animal disease suvrveillance and survey systems: methods and application*. Blackwell Publishing Professional, Iowa, USA, 1st edition, 2003.

[27] R. W. Shephard, J. A. Toribio, A. R. Cameron, P. C. Thomson, and F.C. Baldock. Development of the Bovine Syndromic Surveillance System (BOSSS). In *Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics, 2006*, 2006.

[28] David L. Buckeridge. Outbreak detection through automated surveillance: a review of the determinants of detection. *Journal of Biomedical Informatics*, 40(4):370–379, Aug 2007.

[29] L. H. McIntyre, P. R. Davies, G. Alexander, B. D. O'Leary, R. S. Morris, N. R. Perkins, R. Jackson, and R. poland. Vetpad - veterinary practicioner aided disease surveillance system. In *Proceedings of the 10th Symposium of the International Society for Veterinary Epidemiology and Economics, 2003.*, 2003.

[30] G. Vourc'h and J. Barnouin. How to improve the detection of animal emerging diseases? A two-level (veterinarian/farmer) approach based on an Internet-Oracle database. In *Proceedings of the 10th Symposium of the International Society for Veterinary Epidemiology and Economics, 2003.*, 2003.

[31] B. DeGroot. The rapid syndrome validation project for animals - augmenting contact with the network of accredited veterinarians. *NAHSS Outlook*, April:1, 2005.

[32] C.J.M. Bartels, P. Kock, H. Middelesch, W. Wouda, L. van Wuijckhuise, and H. van der Zwaag. Cattle health surveillance in the Netherlands; how to interpret anecdotal and census data. In *Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics*, 2006.

[33] R. W. Shephard, J. A. Toribio, A. R. Cameron, P. C. Thomson, and F.C. Baldock. Incorporating the Bovine Syndromic Surveillance System (BOSSS)

within an animal health surveillance network. In *Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics, 2006*, 2006.

[34] Larry T. Glickman, George E. Moore, Nita W. Glickman, Richard J. Caldanaro, David Aucoin, and Hugh B. Lewis. Purdue University-Banfield national companion animal surveillance program for emerging and zoonotic diseases. *Vector Borne and Zoonotic Diseases*, 6(1):14–23, 2006.

[35] L. E. Shaffer. *Using pre-diagnostic data from veterinary laboratories to detect disease outbreaks in companion animals*. PhD thesis, Ohio State University, 2007.

[36] R. Maciejewski, B. Tyner, Y. Jang, C. Zheng, R. Nehme, D. S. Ebert, W. S. Cleveland, M. Ouzzani, J. Grannis, and L. T. Glickman. LAHVA: Linked animal-human health visual analytics. *Advances in Disease Surveillance*, 2010(August 23rd):1, 2007.

[37] J. C. Gibbens, S. Robertson, J. Willmington, A. Milnes, J. B. M. Ryan, J. W. Wilesmith, A. J. C. Cook, and G. P. David. Use of laboratory data to reduce the time taken to detect new diseases: VIDA to FarmFile. *Veterinary Record*, 162:771–776, 2008.

[38] A. Tierney, K. P. Coyne, S. Dawson, R. M. Gaskell, J. Bryan, R. Newton, and A. D. Radford. SAVSNET. Available at: http://www.bsava.com/News/CongressNews/SAVSNET/tabid/270/Default.aspx. Accessed on Auguts23rd, 2010.

[39] D. C. Van Metre, D. Q. Barkey, M. D. Salman, and P. S. Morley. Development of a syndromic surveillance system for detection of disease among livestock entering an auction market. *Journal of the American Veterinary Medical Association*, 234(5):658–664, Mar 1 2009.

[40] S. Checkley, J. Berezowski, J. Patel, R. Clarke, D. Peters, J. Keenliside, B. Miller, J. Bystrom, C. Annett, I. Jamal, Y. Qu, S. Turner, and T. Herntier. Emerging disease surveillance of livestock through the Alberta veterinary surveillance network. In *Proceedings of the 12th International Symposium on Veterinary Epidemiology and Economics, 2009*, 2009.

[41] M. R. Amezcua, D. L. Pearl, R. M. Friendship, and W. B. McNab. Evaluation of a veterinary-based syndromic surveillance system implemented for swine. *The Canadian Journal of Veterinary Research*, 74:241–251, 2010.

[42] R. Kosmider, L. Kelly, S. Evans, and G. Gettinby. A stastistical system for detecting Salmonella outbreaks in British livestock. *Epidemiology and infection*, 134(5):952–960, Oct 2006.

[43] A. Leblond, P. Hendrikx, and P. Sabatier. West Nile virus outbreak detection using syndromic monitoring in horses. *Vector-Borne and Zoonotic Diseases*, 7(3):403–410, Fall 2007.

[44] Tim E. Carpenter, Mariann Chrièl, and Matthias Greiner. An analysis of an early-warning system to reduce abortions in dairy cattle in Denmark incorporating both financial and epidemiologic aspects. *Preventive Veterinary Medicine*, 78(1):1–11, Jan 2007.

[45] Agricola Odoi, Craig N. Carter, Jeremy W. Riley, Jackie L. Smith, and Roberta M. Dwyer. Application of an automated surveillance-data-analysis system in a laboratory-based early-warning system for detection of an abortion outbreak in mares. *American Journal of Veterinary Research*, 70(2):247–256, Feb 2009.

[46] G. Vourc'h, V. E. Bridges, J. Gibbens, B. De Groot, L. McIntyre, R. Poland, and J. Barnouin. Detecting emerging diseases in farm animals through clinical observations. *Emerging infectious diseases*, 12(2):204–210, 2006.

[47] R. G. Davis. The ABCs of bioterrorism for veterinarians, focusing on category B and C agents. *Journal of the American Veterinary Medical Association*, 224(7):1096–1104, Apr 1 2004.

[48] A.R.W. Elbers, W.L.A. Loeffen, A. Dekker, G. Koch, and E.M.A. Van Rooij. Substantial improvement of early detection of notifiable animal diseases: a call for unorthodox changes. In *Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics*, 2006.

[49] J. Berezowski, S. Checkley, R. Clarke, J. Patel, and D. Renter. The Alberta veterinary surveillance network : Part 2, veterinary practice surveillance. In *Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics, 2006*, 2006.

[50] S. Checkley, J. Berezowski, C. Annett, J. Bystrom, R. Clarke, B. Miller, J. Patel, A. Perry, D. Renter, and C. Snyder. The Alberta veterinary surveillance network : Part 1, general description/overview. In *Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics, 2006*, 2006.

[51] Richard William Shephard. *The development of a syndromic surveillance system for the extensive beef cattle producing regions of Australia.* PhD thesis, University of Sydney, 2006.

[52] P Brightling, MT Larcombe, BC Blood, and PC Kennedy. Development and use of BOVID-3, an expert system for veterinarians involved in diagnosis, treatment and prevention of diseases of cattle. *Bovine Practitioner*, 32(2):46–49, 1998.

[53] National animal disease information service. UK veterinary sentinel practice network. Available at: http://www.nadis.org.uk. Accessed on September 4th, 2010.

[54] D. A. Ashford, T. M. Gomez, D. L. Noah, D. P. Scott, and D. R. Franz. Biological terrorism and veterinary medicine in the United States. *Journal of the American Veterinary Medical Association*, 217(5):664–667, Sep 2000.

[55] R. G. Davis. The ABCs of bioterrorism for veterinarians, focusing on category A agents. *Journal of the American Veterinary Medical Association*, 224(7):1084–1095, Apr 1 2004.

[56] P. Rabinowitz, Z. Gordon, D. Chudnov, M. Wilcox, L. Odofin, A. Liu, and J. Dein. Animals as sentinels of bioterrorism agents. *Emerging infectious diseases*, 12(4):647–652, Apr 2006.

[57] Farzad Mostashari and Jessica Hartman. Syndromic surveillance: a local perspective. *Jornal of Urban Health*, 80(2 Suppl 1):i1–i7, Jun 2003.

[58] G. D. Johnson, M. Eidson, K. Schmit, A. Ellis, and M. Kulldorff. Geographic prediction of human onset of West Nile virus using dead crow clusters: an evaluation of year 2002 data in New York State. *American Journal of Epidemiology*, 163(2):171–180, Jan 15 2006.

[59] J. Shuai, P. Buck, P. Sockett, J. Aramini, and F. Pollari. A GIS-driven integrated real-time surveillance pilot system for national West Nile virus dead bird surveillance in Canada. *International Journal of Health Geographics*, 5:17, Apr 20 2006. LR: 2009111.

[60] T. McNamara. The role of zoos in biosurveillance. *International Zoo Yearbook*, 41(1):12–15, 2007.

[61] S. M. Babin, J. casper, C. Witt, S. L. H. Lewis, R. A. Wojcik, S. F. Magruder, H. S. Burkom, J. Weitzel, and J. Lombardo. Early detection of possible bioterrorist events using sentinel animals. In *The 131st Annual Meeting of APHA*, November 15-19 2003.

[62] J. Goplin and M. Benz. North Dakota electronic animal health surveillance system. *Advances in Disease Surveillance*, 4(1):8, 2007.

[63] North Carolina Department of Agriculture and Consumer Services. Multihazard threat database (MHTD). Accessed on September 9th, 2010.

[64] E. Brianti, M. Drigo, V. Zirilli, G. Poglayen, and S. Giannetto. Use of a health information system (HIS) for the epidemiological surveillance of leishmaniasis in urban areas. *Veterinary research communications*, 31 Suppl 1:213–215, Aug 2007. JID: 8100520; ppublish.

[65] Kimberly A. Smith-Akin, Charles F. Bearden, Stephen T. Pittenger, and Elmer V. Bernstam. Toward a veterinary informatics research agenda: an analysis of the PubMed-indexed literature. *International jounal of Medical Informatics*, 76(4):306–312, Apr 2007.

[66] L. Hoinville, J. Ellis-Iversen, D. Vink, E. Watson, L. Snow, and J. Gibbens. Discussing the development and application of methods for effective surveillance in livestock populations. Technical report, Report of a workshop held prior to the ISVEE conference, Durban, South Africa, August 2009, December 2009.

[67] G. M. Gobar, J. T. Case, and P. H. Kass. Program for surveillance of causes of death of dogs, using the internet to survey small animal veterinarians. *Journal of the American Veterinary Medical Association*, 213(2):251–256, Jul 15 1998.

[68] G.E. Moore, Ward, M.P., J. Dhariwal, C.C. Wu, N.W. Glickman, H.B. Lewis, and L.T. Glickman. Development of a national companion animal syndromic surveillance sytem for bioterrorism. In *Proceedings of GISVET 2004. Guelph, Ontario.*, 2004.

[69] J. Robotham and L. E. Green. Pilot study to investigate the feasibility of surveillance of small animals in the UK. *The Journal of small animal practice*, 45(4):213–218, Apr 2004.

[70] P. C. Bartlett, J. B. Kaneene, J. H. Kirk, M. A. Wilke, and J. V. Martenuik. Development of a computerized dairy herd health data base for epidemiologic research. *Preventive Veterinary Medicine*, 4:3–14, 1986.

[71] P. C. Bartlett, J. W. Van Buren, M. Neterer, and C. Zhou. Disease surveillance and referral bias in the veterinary medical database. *Preventive Veterinary Medicine*, 94(3-4):264–271, May 1 2010.

[72] A. De Vries and J. K. Reneau. Application of statistical process control charts to monitor changes in animal production systems. *Journal of Animal Science*, 88(13 Suppl):E11–24, Apr 2010.

[73] M. Mork, A. Lindberg, S. Alenius, I. Vagsholm, and A. Egenvall. Comparison between dairy cow disease incidence in data registered by farmers and in data from a disease-recording system based on veterinary reporting. *Preventive Veterinary Medicine*, 88(4):298–307, Apr 1 2009.

[74] J. W. Buehler, R. S. Hopkins, J. M. Overhage, D. M. Sosin, V. Tong, and CDC Working Group. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. *MMWR.Recommendations and reports : Morbidity and mortality weekly report.Recommendations and reports / Centers for Disease Control*, 53(RR-5):1–11, May 7 2004. ID: 1285; LR: 20041117; JID: 101124922; ppublish.

[75] Vitali Sintchenko and Blanca Gallego. Laboratory-guided detection of disease outbreaks: three generations of surveillance systems. *Archives of Pathology and Laboratory Medicine*, 133(6):916–925, Jun 2009.

[76] Kathy Zurbrigg and Tim Blackwell. Do submissions to a veterinary diagnostic laboratory accurately reflect disease incidence or prevalence in the catchment area? In *Proceedings of the 8th Annual Conference of the International Society for Disease Surveillance*, 2009.

[77] Canadian Food Inspection Agency. The Canadian animal health surveillance network (CAHSN), March 10 2009. Accessed on August23rd, 2010.

[78] Gluck Equine Reseacrh Center. Research in the department of veterinary science. Accessed on August 23rd, 2010.

[79] A. Egenvall, B. N. Bonnett, P. Olson, and A. Hedhammar. Validation of computerized Swedish dog and cat insurance data against veterinary practice records. *Preventive Veterinary Medicine*, 36(1):51–65, Jul 17 1998. LR: 20061115; JID: 8217463; ppublish.

[80] J. C. Penell, A. Egenvall, B. N. Bonnett, and J. Pringle. Validation of computerized Swedish horse insurance data against veterinary clinical records. *Preventive Veterinary Medicine*, 82(3-4):236–251, Dec 14 2007.

[81] M. Engle. The value of an "early warning" surveillance system for emerging diseases. accessed at: http://www.pork.org/newsandinformation/news/docs/mengle.pdf. In: Weber, w.d. development of an animal haalth monitoring system based on salughter condemnation data. Proceedings of the 8th conference of the International Society for Disease Surveillance, 2009. Miami-FL.

[82] W. Weber. Development of an animal health monitoring system based on slaughter condemnation data. In *Proceedings of the 8th Annual Conference of the International Society for Disease Surveillance*, 2009.

[83] Jacqueline Benschop. *Epidemiological investigation of surveillance strategies for zoonotic Salmonella*. PhD thesis, Massey University, 2009.

[84] K. Smith, A. Martinez, R. Craddolph, H. Erickson, D. Andresen, and S. Warren. An integrated cattle health monitoring system. *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society.IEEE Engineering in Medicine and Biology Society.Conference*, 1:4659–4662, 2006.

[85] William B. Lober, L. Trigg, and B. Karras. Information system architectures for syndromic surveillance. *Morbidity and Mortality Weekly Report*, 53 Suppl:203–208, Sep 2004.

[86] O. Ivanov, M. M. Wagner, W. W. Chapman, and R. T. Olszewski. Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance. *AMIA Annual Symposium Proceedings*, pages 345–349, 2002.

[87] R. Wurtz and M. Popovich. Animal disease surveillance: A framework for supporting disease detection in public health. White Paper: Animal Disease Surveillance, WHP027-A, March 2002.

[88] David L. Buckeridge, Howard Burkom, Murray Campbell, William R. Hogan, and Andrew W. Moore. Algorithms for rapid outbreak detection: a research synthesis. *Journal of Biomedical Informatics*, 38(2):99–113, Apr 2005.

[89] Michael Höhle, Michaela Paul, and Leonhard Held. Statistical approaches to the monitoring and surveillance of infectious diseases for veterinary public health. *Preventive Veterinary Medicine*, 91(1):2–10, Sep 2009.

[90] M. P. Ward and T. E. Carpenter. Techniques for analysis of disease clustering in space and in time in veterinary epidemiology. *Preventive Veterinary Medicine*, 45(3-4):257–284, Jun 2000.

[91] M. P. Ward and T. E. Carpenter. Analysis of time-space clustering in veterinary epidemiology. *Preventive Veterinary Medicine*, 43(4):225–237, Feb 29 2000.

[92] T. E. Carpenter. Methods to investigate spatial and temporal clustering in veterinary epidemiology. *Preventive Veterinary Medicine*, 48(4):303–320, Mar 2001.

[93] J. C. Benneyan. Statistical quality control methods in infection control and hospital epidemiology, part I: Introduction and basic theory. *Infection Control and Hospital Epidemiology*, 19(3):194–214, Mar 1998.

[94] T. Lotze, S. Murphy, and G. Shmueli. Implementation and comparison of preprocessing methods for biosurveillance data. *Advances in Disease Surveillance*, 6(1):1–20, 2008.

[95] J. Benschop, M. A. Stevenson, J. Dahl, R. S. Morris, and N. P. French. Temporal and longitudinal analysis of Danish swine salmonellosis control programme data: implications for surveillance. *Epidemiology and Infection*, 136(11):1511–1520, Nov 2008.

[96] A. M. Perez, D. Zeng, C. J. Tseng, H. Chen, Z. Whedbee, D. Paton, and M. C. Thurmond. A web-based system for near real-time surveillance and space-time cluster analysis of foot-and-mouth disease and other animal diseases. *Preventive Veterinary Medicine*, 91(1):39–45, Sep 1 2009.

[97] M. Kulldorff. A spatial scan statistics. *Communications in Statistics - Theory and Methods*, 26(6):1481–1496, 1997.

[98] M. Kulldorff. SatScan(tm) user guide, 2010. Available at http://www.satscan.org/.

[99] David L. Buckeridge, Anna Okhmatovskaia, Samson Tu, Martin O'Connor, Csongor Nyulas, and Mark A. Musen. Predicting outbreak detection in public health surveillance: quantitative analysis to enable evidence-based method selection. *AMIA Annual Symposium Proceedings*, 1:76–80, 2008.

[100] Ken P. Kleinman and Allyson M. Abrams. Assessing surveillance using sensitivity, specificity and timeliness. *Statistical Methods in Medical Research*, 15(5):445–464, Oct 2006.

[101] Ken P. Kleinman and Allyson M. Abrams. Assessing the utility of public health surveillance using specificity, sensitivity, and lives saved. *Statistics in Medicine*, 27(20):4057–4068, Sep 2008.

[102] J. B. Kaneene, M. Saffell, D. J. Fedewa, K. Gallagher, and H. M. Chaddock. The Michigan equine monitoring system I. design, implementation and population estimates. *Preventive Veterinary Medicine*, 29(4):263–275, Feb 1997.

[103] P. A. Durr and S. Eastland. Use of web-enabled databases for complex animal health investigations. *Revue scientifique et technique (International Office of Epizootics)*, 23(3):873–884, Dec 2004.

[104] L. Smith, G. Paiba, R. Lysons, and J. Gibbens. UK surveillance: Rapid analysis & detection of animal-related risks (RADAR) - from concept to reality. In *Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics, 2006*, 2006.

[105] G. A. Paiba, S. R. Roberts, C. W. Houston, E. C. Williams, L. H. Smith, J. C. Gibbens, S. Holdship, and R. Lysons. UK surveillance: provision of quality assured information from combined datasets. *Preventive Veterinary Medicine*, 81(1-3):117–134, Sep 14 2007.

[106] A. R. Cameron. Data management and analysis systems for bluetongue virus zoning in Australia. *Veterinaria Italiana*, 40(3):365–368, Jul-Sep 2004.

[107] A. Conte, P. Colangeli, C. Ippoliti, C. Paladini, M. Ambrosini, L. Savini, F. Dall'Acqua, and P. Calistri. The use of a web-based interactive geographical information system for the surveillance of bluetongue in Italy. *Revue scientifique et technique (International Office of Epizootics)*, 24(3):857–868, Dec 2005.

[108] P. R. Davies, S. R. Wayne, J. L. Torrison, B. Peele, B. D. de Groot, and D. Wray. Real-time disease surveillance tools for the swine industry in Minnesota. *Veterinaria Italiana*, 43(3):731–738, Jul-Sep 2007.

[109] A. C. Clements, D. U. Pfeiffer, M. J. Otte, K. Morteo, and L. Chen. A global livestock production and health atlas (GLiPHA) for interactive presentation, integration and analysis of livestock data. *Preventive Veterinary Medicine*, 56(1):19–32, Nov 29 2002.

[110] M. Egbert. Web-based disease tracking: A West Nile virus example, 2004. Available at: http://proceedings.esri.com/library/userconf/proc04/docs/pap1131.pdf. Accessed on September 10th, 2010.

[111] M. M. Wagner and G. Wallstrom. *Methods for Algorithm Evaluation*, pages 301–310. Handbook of biosurveillance. Academic Press, UK, 2006.

[112] Fu-Chiang Tsui, Jeremy U. Espino, Virginia M. Dato, Per H. Gesteland, Judith Hutman, and Michael M. Wagner. Technical description of RODS: a real-time public health surveillance system. *Journal of the American Medical Informatics Association*, 10(5):399–408, 2003.

[113] J. S. Lombardo and D. L. Buckeridge. *Disease surveillance - a public health informatics approach*. John Wiley & Sons, New Jersey, 1st edition, 2007.

[114] D. Zeng, H. Chen, C. Castillo-Chavez, W. B. Lober, and M. Thurmong. *Infectious Disease Informatics and Biosurveillance*. Springer, 1st edition, 2011.

[115] Michael M. Wagner. Models of computer-based outbreak detection. *The Reference Librarian*, 38:79:343–363, 2003.

[116] World Organization for Animal Health (OIE). Terrestrial Animal Health Code. Accessed on September 14th, 2010.

[117] L. Vrbova, C. Stephen, N. Kasman, R. Boehnke, M. Doyle-Waters, A. Chablitt-Clark, B. Gibson, M. FitzGerald, and D. M. Patrick. Systematic review of surveillance systems for emerging zoonoses. *Transboundary and emerging diseases*, 57(3):154–161, Jun 2010.

[118] Jean-Paul Chretien, Nancy E. Tomich, Joel C. Gaydos, and Patrick W. Kelley. Real-time public health surveillance for emergency preparedness. *American Joumal of Public Heatth*, 99(8):1360–1363, Aug 2009.

[119] Kenneth D. Mandl, B. Reis, and C. Cassa. Measuring outbreak-detection performance by using controlled feature set simulations. *Morbidity and Mortality Weekly Report*, 53 Suppl:130–136, Sep 2004.

[120] W. W. Chapman, L. M. Christensen, M. M. Wagner, P. J. Haug, O. Ivanov, J. N. Dowling, and R. T. Olszewski. Classifying free-text triage chief complaints into syndromic categories with natural language processing. *Artificial Intelligence in Medicine*, 33(1):31–40, Jan 2005.

[121] J. Dara, J. N. Dowling, D. Travers, G. F. Cooper, and W. W. Chapman. Evaluation of preprocessing techniques for chief complaint classification. *Journal of Biomedical Informatics*, 41(4):613–623, Aug 2008.

[122] Ben Y. Reis and Kenneth D. Mandl. Syndromic surveillance: the effects of syndrome grouping on model accuracy and outbreak detection. *Annals of Emergency Medicine*, 44(3):235–241, Sep 2004.

[123] Michael M. Wagner, J. Espino, F. C. Tsui, P. Gesteland, W. Chapman, O. Ivanov, A. Moore, W. Wong, J. Dowling, and J. Hutman. Syndrome and outbreak detection using chief-complaint data–experience of the real-time outbreak and disease surveillance project. *Morbidity and Mortality Weekly Report*, 53 Suppl:28–31, Sep 2004.

[124] Ozlem Uzuner. Recognizing obesity and comorbidities in sparse data. *Journal of the American Medical Informatics Association*, 16(4):561–570, 2009.

[125] O. Maimon and L. Rokach. *Data Mining and Knowledge Discovery Handbook*, chapter Decision Trees. Springer Science /&/ Business Media, Inc., 2005.

[126] P.G. Zhang. *Data Mining and Knowledge Discovery Handbook*, chapter Neural Networks. Springer Science /&/ Business Media, Inc., 2005.

[127] A. Shmilovici. *Data Mining and Knowledge Discovery Handbook*, chapter Support Vector Machines. Springer Science /&/ Business Media, Inc., 2005.

[128] Prashant K. Rohatgi. Radiological evaluation of interstitial lung disease. *Current opinion in pulmonary medicine*, 17(5), 2011.

[129] Junji Shiraishi, Qiang Li, Daniel Appelbaum, and Kunio Doi. Computer-aided diagnosis and artificial intelligence in clinical imaging, 11/01 2011.

[130] M. Wesolowski and B. Suchacz. Artificial neural networks: theoretical background and pharmaceutical applications: a review. *J AOAC Int*, 95(3):652–668, 2012.

[131] T. K. Kelly, P. Chalk, J. Bonomo, J. Parachini, B. A. Jackson, and G. Cecchine. The office of science and technology policy blue ribbon planel on the threat of biological terrorism directed against livestock. In *Conference Proceedings*, April 2004.

[132] S. Babin, C. Witt, J. Casper, R. A. Wojcik, S. L. H. Lewis, and J. Lombardo. Syndromic animal surveillance in the electronic surveillance system for the early notification of community-based epidemics. In *National Multi-Hazard Symposium: "One Medicine" Approach to Homeland Security*, December 11-12 2003.

[133] Loren E. Shaffer, Julie A. Funk, Paivi Rajala-Schultz, Michael M. Wagner, Thomas E. Wittum, and William J.A. Saville. Evaluation of microbiology orders from a veterinary diagnostic laboratory as a potential data source for early outbreak detection. *Advances in Disease Surveillance*, 6:2:1–7, 2008.

[134] Fernanda C. Dórea, Javier Sanchez, and Crawford W. Revie. Veterinary syndromic surveillance: Current initiatives and potential for development. *Preventive Veterinary Medicine*, 101(1-2):1–17, Aug 2011. http://dx.doi.org/10.1016/j.prevetmed.2011.05.004.

[135] Illés Solt, Domonkos Tikk, Viktor Gál, and Zsolt T. Kardkovács. Semantic classification of diseases in discharge summaries using a context-aware rule-based classifier. *Journal of the American Medical Informatics Association*, 16(4):580–584, 2009.

[136] R. Farkas, G. Szarvas, I. Hegedus, A. Almasi, V. Vincze, R. Ormandi, and R. Busa-Fekete. Semi-automated construction of decision rules to predict morbidities from clinical texts. *Journal of the American Medical Informatics Association : JAMIA*, 16(4):601–605, Jul-Aug 2009.

[137] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley &Sons, Inc., 2001.

[138] F. Sebastiani. Machine learning in automated text categorization. *ACM Computer Surveys*, 34(1):1–47, 2002.

[139] R. B. Dessau and P. Steenberg. Computerized surveillance in clinical microbiology with time series analysis. *Journal of Clinical Microbiology*, 31(4):857–860, Apr 1993.

[140] L. C. Hutwagner, E. K. Maloney, N. H. Bean, L. Slutsker, and S. M. Martin. Using laboratory-based surveillance data for prevention: an algorithm for detecting Salmonella outbreaks. *Emerging Infectious Diseases*, 3(3):395–400, 1997.

[141] T. E. Carpenter. Evaluation and extension of the CUSUM technique with an application to Salmonella surveillance. *Journal of Veterinary Diagnostic and Investigation*, 14(3):211–218, May 2002.

[142] Fernanda C. Dórea, C. Anne Muckle, David Kelton, JT. McClure, Beverly J. McEwen, W. Bruce McNab, Javier Sanchez, and Crawford W. Revie. Exploratory analysis of methods for automated classification of laboratory test orders into syndromic groups in veterinary medicine. *PLoS ONE*, 8(3):e57334, 03 2013.

[143] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

[144] Diethelm Wuertz, Yohan Chalabi with contributions from Martin Maechler, Joe W. Byers, , et al. timedate: Rmetrics - chronological and calendarical objects, 2011. R package version 2130.93.

[145] David James and Kurt Hornik. chron: Chronological objects which can handle dates and times, 2010. R package version 2.3-39. S original by David James, R port by Kurt Hornik.

[146] T. Lotze, S. P. Murphy, and G. Shmueli. Preparing biosurveillance data for classic monitoring. *Advances in Disease Surveillance*, 2:55, 2007.

[147] A. Zeileis, C. Kleiber, and S. Jackman. Regression models for count data in R. *Journal of Statistical Software*, 27(8), 2008.

[148] Howard Burkom, Sean Murphy, and Galit Shmueli. Automated time series forecasting for biosurveillance. *Statistics in Medicine*, 26(22):4202–4218, Sep 2007.

[149] Yevgeniy Elbert and Howard S. Burkom. Development and evaluation of a data-adaptive alerting algorithm for univariate temporal biosurveillance data. *Statistics in Medicine*, 28(26):3226–3248, Nov 2009.

[150] F. C. Tsui, M. M. Wagner, V. Dato, and C. C. Chang. Value of ICD-9 coded chief complaints for detection of epidemics. *AMIA Annual Symposium Proceedings*, 1:711–715, 2001.

[151] R. E. Serfling. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Reports*, 78:494–506, 1963.

[152] Colleen A. Bradley, H. Rolka, D. Walker, and J. Loonsk. BioSense: implementation of a national early event detection and situational awareness system. *Morbidity and Mortality Weekly Report*, 54 Suppl:11–19, Aug 2005.

[153] Alexandra M. Schmidt and João Batista M. Pereira. Modelling time series of counts in epidemiology. *International Statistical Review / Revue Internationale de Statistique*, 79(1):pp. 48–69, 2011.

[154] Joseph Naus and Sylvan Wallenstein. Temporal surveillance using scan statistics. *Statistics in Medicine*, 25(2):311–324, Jan 2006.

[155] W.H. Woodall. Use of control charts in health-care and public-health surveillance. *Journal of Quality Technology*, 38:2:89–104, 2006.

[156] Michael L. Jackson, Atar Baer, Ian Painter, and Jeff Duchin. A simulation study comparing aberration detection algorithms for syndromic surveillance. *BMC Med Inform Decis Mak*, 7:6, 2007.

[157] Fernanda C. Dórea, Crawford W. Revie, Beverly J. McEwen, W. Bruce McNab, David Kelton, and Javier Sanchez. Retrospective time series analysis of veterinary laboratory data preparing a historical baseline for cluster detection in syndromic surveillance. *Preventive Veterinary Medicine*, 109:219–227, 2013.

[158] J. C. Gibbens, J. W. Wilesmith, C. E. Sharpe, L. M. Mansley, E. Michalopoulou, J. B. M. Ryan, and M. Hudson. Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months. *Veterinary Record*, 149:729–743, 2001.

[159] A Picado, FJ Guitian, and DU Pfeiffer. Space-time interaction as an indicator of local spread during the 2001 FMD outbreak in the UK. *Preventive Veterinary Medicine*, 79(1):3–19, 2007.

[160] I.M.G.A. Santman-Berends, J.A. Stegeman, P. Vellema, and G. van Schaik. Estimation of the reproduction ratio (Ro) of bluetongue based on serological field data and comparison with other BTV transmission models. *Preventive Veterinary Medicine*, In press(0):–, 2013.

[161] L. C. Hutwagner, W. W. Thompson, G. M. Seeman, and T. Treadwell. A simulation model for assessing aberration detection methods used in public health surveillance for systems with limited baselines. *Statistics in Medicine*, 24(4):543–550, Feb 2005.

[162] Z. G. Stoumbos, M. R. Reynolds, T. P. Ryan, and W. H. Woodall. The state of statistical process control as we proceed into the 21st century. *Journal of the American Statistical Association*, 95(451):992–998, 2000.

[163] Howard S. Burkom. Development, adaptation, and assessment of alerting algorithms for biosurveillance. *Johns Hopkins Apl Technical Digest*, 24(4):335–342, 2003.

[164] Sean Patrick Murphy and Howard Burkom. Recombinant temporal aberration detection algorithms for enhanced biosurveillance. *Journal of the American Medical Informatics Association*, 15(1):77–86, 2008.

[165] John L Szarka, 3rd, Linmin Gan, and William H. Woodall. Comparison of the early aberration reporting system (EARS) W2 methods to an adaptive threshold method. *Statistics in Medicine*, 30(5):489–504, Feb 2011.

[166] F. F. Nobre and D. F. Stroup. A monitoring system to detect changes in public health surveillance data. *International Journal of Epidemiology*, 23(2):408–418, Apr 1994.

[167] L. Ngo, I. B. Tager, and D. Hadley. Application of exponential smoothing for nosocomial infection surveillance. *American Journal of Epidemiology*, 143(6):637–647, Mar 1996.

[168] C. Chatfield. The Holt-Winters forecasting procedure. *Applied Statistics*, 27(3):264–279, 1978.

[169] Kathy J. Hurt-Mullen and J. Coberly. Syndromic surveillance on the epidemiologist's desktop: making sense of much data. *Morbidity and Mortality Weekly Report*, 54 Suppl:141–146, Aug 2005.

[170] T. H. Lotze, G. Shmueli, and I. Yahav. *Simulatin and Evaluating Biosurveillance Datasets*, pages 23–51. Biosurveillance: Methods and Case Studies. CRC Press, United States, 2011.

[171] David L. Buckeridge, P. Switzer, D. Owens, D. Siegrist, J. Pavlin, and M. Musen. An evaluation model for syndromic surveillance: assessing the performance of a temporal algorithm. *Morbidity and Mortality Weekly Report*, 54 Suppl:109–115, Aug 2005.

[172] Ben Y. Reis, Marcello Pagano, and Kenneth D. Mandl. Using temporal context to improve biosurveillance. *Proceedings of the National Academy of Sciences of the United States of America*, 100(4):1961–1965, Feb 2003.

[173] Lori Hutwagner, Timothy Browne, G Matthew Seeman, and Aaron T. Fleischauer. Comparing aberration detection methods with simulated data. *Emerging Infectious Diseases*, 11(2):314–316, Feb 2005.

[174] F. C. Dórea, A. R. Vieira, C. Hofacre, D. Waldrip, and D. J. Cole. Stochastic model of the potential spread of highly pathogenic avian influenza from an infected commercial broiler operation in Georgia. *Avian Diseases*, 54(1 Suppl):713–719, Mar 2010.

[175] Neil M. Ferguson, Christl A. Donnelly, and Roy M. Anderson. The Foot-and-Mouth Epidemic in Great Britain: Pattern of Spread and Impact of Interventions. *Science*, 292(5519):1155–1160, 2001.

[176] MJ Keeling, ME Woolhouse, DJ Shaw, L Matthews, M Chase-Topping, DT Haydon, SJ Cornell, J Kappey, J Wilesmith, and Grenfell BT. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*, 294(5543):813–817, 2001.

[177] R.R. Kao. Landscape fragmentation and foot-and-mouth disease transmission. *Veterinary Record*, 148:746–747, 2001.

[178] Shannon E. Fraker, William H. Woodall, and Shabnam Mousavi. Performance metrics for surveillance schemes. *Quality Engineering*, 20(4):451–464, 2008.

[179] Christian Sonesson and David Bock. A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 166:1:52–21, 2003.

[180] Ronald D. Fricker Jr., Benjamin L. Hegler, and David A. Dunfee. Comparing syndromic surveillance detection methods: EARSŠ versus a CUSUM-based methodology. *Statistics in Medicine*, 27:3407–3429, 2008.

[181] Ling Wang, Marco F. Ramoni, Kenneth D. Mandl, and Paola Sebastiani. Factors affecting automated syndromic surveillance. *Artificial Intelligence in Medicine*, 34(3):269–278, Jul 2005.

[182] Fernanda C. Dórea, Crawford W. Revie, Beverly J. McEwen, W. Bruce McNab, and Javier Sanchez. Syndromic surveillance using veterinary laboratory data: data pre-processing and algorithm performance evaluation. *Journal of the Royal Society Interface*, In press.

[183] Joseph S. Lombardo, H. Burkom, and J. Pavlin. ESSENCE II and the framework for evaluating syndromic surveillance systems. *Morbidity and Mortality Weekly Report*, 53 Suppl:159–165, Sep 2004.

[184] T. O Sullivan, R. Friendship, D.L. Pearl, B. McEwen, A. Ker, and C. Dewey. The association between submission counts to a veterinary diagnostic laboratory and the economic and disease challenges of the Ontario swine industry from 1998 to 2009. *Preventive Veterinary Medicine*, in press, 2012.

[185] Ian Dohoo, Wayne Martin, and Henrik Stryhn. *Veterinary Epidemiologic Research*. VER Inc., 2010.

[186] C. C. van den Wjingaard, W. van Pelt, N. J. Nagelkerke, M. Kretzschmar, and M.P. Koopmans. Evaluation of syndromic surveillance in the Netherlands: its added value and recommendations for implementation. *Eurosurveillance*, 16(9):http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19806, 2011.

[187] D. L. Buckeridge. Comments on 'some methodological issues in biosurveillance'. *Statistics in Medicine*, 30:420–422, 2011.

[188] Ronald D. Fricker Jr. Rejoinder: Some methodological issues in biosurveillance. *Statistics in Medicine*, 30:434–441, 2011.

[189] W. W. Chapman. *Natural Langugage Processing for Biosurveillance*, pages 255–271. Handbook of Biosurveillance. Academic Press, USA, 2006.

[190] G. D. Alton, D. L. Pearl, K. G. Bateman, W. B. McNab, and O. Berke. Factors associated with whole carcass condemnation rates in provincially-inspected abattoirs in Ontario 2001-2007: implications for food animal syndromic surveillance. *BMC Veterinary Research*, 6(1):42, Aug 12 2010.

[191] Taj Azarian, Sarah Winn, Saad Zaheer, James Buehler, and Richard S. Hopkins. Utilization of syndromic surveillance with multiple data sources to enhance public health response. *Advances in Disease Surveillance*, 7:1:1–7, 2009.

[192] Terri L. O Sullivan, Robert M. Friendship, David L. Pearl, Beverly McEwen, and Catherine E. Dewey. Identifying an outbreak of a novel swine disease using test requests for porcine reproductive and respiratory syndrome as a syndromic surveillance tool. *BMC Vet Res*, 8(1):192, Oct 2012.

[193] B. Richards, M. Kabay, and J. Hutchison. Analysis of laboratory data for general disease surveillance in australia. report to the australian biosecurity

crc for emerging infectious diseases. Technical Report PROJECT AL.122R, AB-CRC Laboratory Surveillance Project, 2010.