

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

Fernanda C. Dórea<sup>1</sup>; Crawford W. Revie<sup>1</sup>; Beverly J. McEwen<sup>2</sup>, W Bruce McNab<sup>3</sup>; David Kelton<sup>4</sup>; Javier Sanchez<sup>1</sup>

<sup>1</sup>Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, Canada

<sup>2</sup>Animal Health Laboratory, University of Guelph, Canada

<sup>3</sup>Ontario Ministry of Agriculture Food and Rural Affairs, Canada

<sup>4</sup>Department of Population Medicine, University of Guelph, Canada

Corresponding Author

Fernanda Dórea

fdorea@upei.ca. Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, PE, C1A 4P3, Canada. Phone: +1(902)566-0969; Fax: +1(902)-620-5053.

Postal address:

Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, 550 University Avenue, Charlottetown, PE, C1A 4P3

## 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 unlabelled 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<sup>th</sup> percentile of daily counts, allowed for the removal of these aberrations in the absence of labelled 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<sup>th</sup> percentiles.

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

## Introduction

Surveillance has shifted in the last two decades towards systems capable of early detection of disease (Shmueli and Burkom, 2010). Modern biosurveillance 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 as syndromes. Surveillance based on these types of data is therefore referred to as *syndromic surveillance* (Centers for Disease Control and Prevention (CDC), 2006). A recent review of syndromic surveillance initiatives in veterinary medicine (Dorea, et al., 2011) 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 (Stone, 2007). Laboratory test results have been analysed retrospectively to detect temporal clustering of bacterial pathogens in public health (Dessau and Steenberg, 1993; Hutwagner, et al., 1997; Widdowson, et al., 2003) and veterinary medicine (Carpenter, 2002; Zhang, et al., 2005). 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 labelled, nor will their shape and magnitude be known (Shmueli and Burkom, 2010). Detection of abnormal behaviour 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) (Yahav and Shmueli, 2007; Shmueli and Burkom, 2010). In both cases, a baseline free of outbreaks is necessary: in the former case to create models of expected behaviour, 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 behaviour against which the presence of abnormalities can be investigated (Shmueli and Burkom, 2010).

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 behaviour” for each syndrome, separating the background behaviour from temporal aberrations present in the historical laboratory test request data.

## Methods

### 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.

### 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 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 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 Dórea et al. (*in review*).

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.

### Data characterization

All statistical analyses were performed in the R environment (<http://www.r-project.org/>) (R Development Core Team, 2011). Complete data series, with counts for every calendar day from January 1<sup>st</sup> 2008 to December 31<sup>st</sup> 2010, were generated for each syndromic group by inserting missing days and assigning to them a count of zero (R packages {timeDate}(Wuertz, et al., 2011) and {chron}(James and Hornik, 2010)). 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 was 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 (Lotze, et al., 2008).

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 (Lotze, et al., 2007), negative binomial regression, and zero-inflated versions of these methods (Zeileis, et al., 2008) (R package {pscl} (Zeileis, et al., 2008)) 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).

### Aberration removal

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

Smoothing was attempted using Holt-Winters exponential smoothing (Burkom, et al., 2007; Elbert and Burkom, 2009), 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) 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 component 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 (Burkom, et al., 2007).

The second method was based on the procedure used by Tsui et al. (2001). 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 (Serfling, 1963; Tsui, et al., 2001), 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<sup>th</sup> 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  $\lambda_i$ , 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.

## Results

### Case definition and syndromic groups

The complete list of syndromic groups is shown in Table 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 1 therefore refer to the weekday median. All the time series described are based on 5-day weeks, and 260-day years.

### Data characterization

Time series for six of the syndromic groups listed in Table 1 are shown in Figure 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.

#### *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 2-A. Mondays are labelled in the graph. Box-plots of the quartiles of daily counts for the whole *Mastitis* series, per day-of-week, are shown in Figure 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 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.

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 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.

#### *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 (Serfling, 1963; Tsui, et al., 2001) indicated that the use of non-parametric methods, such as moving percentiles, were best suited to characterise these series.

#### Aberration removal

##### *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 behaviour 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 behaviour, 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), in which most of the original data is 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<sup>th</sup> 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. Outlier substitution using the upper limit of the confidence interval (Figure 4-B) represented a better balance between removing temporal aberrations and keeping the original variation of the original data, without over-smoothing.

#### *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<sup>th</sup> 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 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 1, with results similar to those shown in Figures 4 and 5.

## **Discussion**

Syndromic surveillance operates under the assumption that anomalies indicative of disease outbreaks can be detected when information is monitored continuously (Shmueli and Burkom, 2010). 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 (Lotze, et al., 2008).

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 are 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 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 behaviour. 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 behaviour in the data to contribute to the estimates, and therefore the resulting estimates were closer to the expected baseline of normal behaviour. Regression has been used in several implemented biosurveillance systems, and it is a natural choice when this amount of historical data is available (Bradley, et al., 2005; Shmueli and Burkom, 2010). The Poisson regression, specifically, can be more robust than other linear models since it does not require constant variance (Buckeridge, et al., 2005). 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 (Schmidt and Pereira, 2011).

By reducing the model variables to key explainable factors, such as day-of-week and month, it was possible to model the baseline behaviour, 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 (Shmueli and Burkom, 2010). There are statistical methods to identify whether outbreaks are present or not (Naus and Wallenstein, 2006), and the use of diagnostic information to label outbreaks has also been suggested (Ivanov, et al., 2003). However, even if outbreaks can be identified, the problem of how to remove the outbreak signature from the background data of normal behaviour 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) 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 behaviour of the data, and that their relationship to the counts is homogeneous across the entire period (Burkom, et al., 2007).

The substitution of outliers by modelled 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<sup>th</sup> 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<sup>th</sup> 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.

## **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 behaviour. 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<sup>th</sup> percentile, it was possible to remove these aberrations in the absence of labelled 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<sup>th</sup> percentiles.

## **Acknowledgments**

This project was supported by the OMAFRA-UG Agreement through the Animal Health Strategic Investment fund (AHSI) managed by the Animal Health Laboratory of the University of Guelph.

## **Conflict of Interest**

The authors declare no conflict of interest.

## References

- Bradley, C.A., Rolka, H., Walker, D., Loonsk, J., 2005. BioSense: Implementation of a national early event detection and situational awareness system. *MMWR Morb. Mortal. Wkly. Rep.* 54 Suppl, 11-19.
- Buckeridge, D.L., Burkom, H., Campbell, M., Hogan, W.R., Moore, A.W., 2005. Algorithms for rapid outbreak detection: A research synthesis. *J. Biomed. Inform.* 38, 99-113.
- Burkom, H.S., Murphy, S.P., Shmueli, G., 2007. Automated time series forecasting for biosurveillance. *Stat. Med.* 26, 4202-4218.
- Carpenter, T.E., 2002. Evaluation and extension of the cusum technique with an application to salmonella surveillance. *J. Vet. Diagn. Invest.* 14, 211-218.
- Centers for Disease Control and Prevention (CDC), 2006. Annotated bibliography for syndromic surveillance. Available at: <http://www.cdc.gov/ncphi/diss/nndss/syndromic.htm>. Accessed on March 26th, 2010.
- Dessau, R.B., Steenberg, P., 1993. Computerized surveillance in clinical microbiology with time series analysis. *J. Clin. Microbiol.* 31, 857-860.
- Dórea, F.C., Sanchez, J., Revie, C.W., 2011. Veterinary syndromic surveillance: Current initiatives and potential for development. *Prev. Vet. Med.* 101, 1-17.
- Dórea, F.C.; Muckle, A.C.; Kelton, D.; McClure, J.T.; McEwen, B.J.; McNab, W.B.; Sanchez, J. and Revie, C.W. Exploratory analysis of methods for automated classification of laboratory test orders into syndromic groups in veterinary medicine. (*Submitted to PlosOne*)
- Elbert, Y., Burkom, H.S., 2009. Development and evaluation of a data-adaptive alerting algorithm for univariate temporal biosurveillance data. *Stat. Med.* 28, 3226-3248.
- Hutwagner, L.C., Maloney, E.K., Bean, N.H., Slutsker, L., Martin, S.M., 1997. Using laboratory-based surveillance data for prevention: An algorithm for detecting salmonella outbreaks. *Emerg. Infect. Dis.* 3, 395-400.
- Ivanov, O., Gesteland, P.H., Hogan, W., Mundorff, M.B., Wagner, M.M., 2003. Detection of pediatric respiratory and gastrointestinal outbreaks from free-text chief complaints. *AMIA. Annu. Symp. Proc.* 318-322.
- James, D., Hornik, K., 2010. Chron: Chronological objects which can handle dates and times. R package version 2.3-39. Available at: <http://CRAN.R-project.org/package=chron>. Accessed on September 8<sup>th</sup>, 2011.
- Lotze, T., Murphy, S., Shmueli, G., 2008. Implementation and comparison of preprocessing methods for biosurveillance data. *ADS* 6, 1-20.
- Lotze, T., Murphy, S.P., Shmueli, G., 2007. Preparing biosurveillance data for classic monitoring. *ADS* 2, 55.
- Naus, J., Wallenstein, S., 2006. Temporal surveillance using scan statistics. *Stat. Med.* 25, 311-324.

R Development Core Team. A Language and Environment for Statistical Computing. ISBN 3-900051-07-0. R Foundation for Statistical Computing, Vienna, Austria.

Schmidt, M.A, Pereira, J.B.M., 2011. Modelling Time Series of Counts in Epidemiology. *Int. Stat. Rev.*, 79, 1, 48–69.

Serfling, R.E., 1963. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Rep.* 78, 494-506.

Shmueli, G., Burkom, H., 2010. Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics* 52, 39-51.

Stone, M., 2007. The potential for exotic disease syndromic surveillance within veterinary laboratory submissions data. *Proceedings of the Epidemiology and Animal Health Management Branch of the NZVA.* 91-102.

Tsui, F.C., Wagner, M.M., Dato, V., Chang, C.C., 2001. Value of ICD-9 coded chief complaints for detection of epidemics. *Proc. AMIA. Symp.* 711-715.

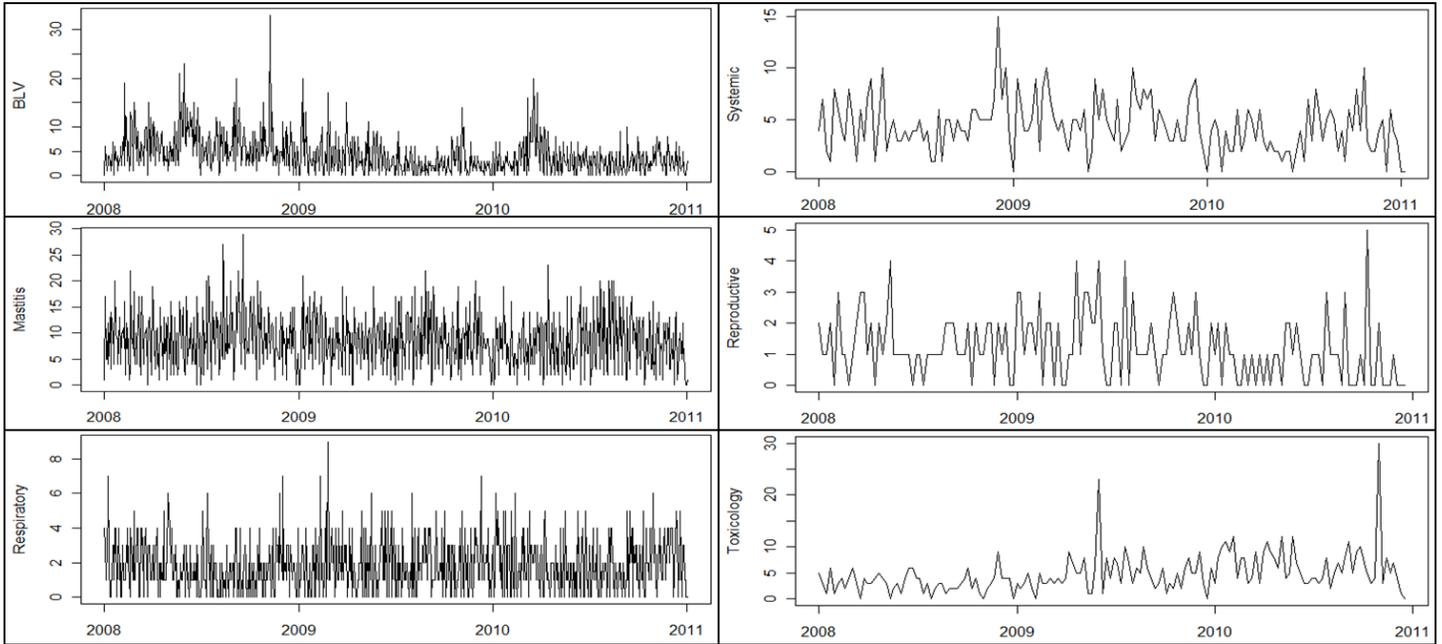
Widdowson, M.A., Bosman, A., van Straten, E., Tinga, M., Chaves, S., van Eerden, L., van Pelt, W., 2003. Automated, laboratory-based system using the internet for disease outbreak detection, the netherlands. *Emerg. Infect. Dis.* 9, 1046-1052.

Wuertz, D., Maechler, Yohan Chalabi with contributions from Martin, Byers, J.W., others, a., 2011. *timeDate: Rmetrics - chronological and calendarical objects.* R package version 2130.93. Available at: <http://CRAN.R-project.org/package=timeDate>. Accessed on September 28<sup>th</sup>, 2011.

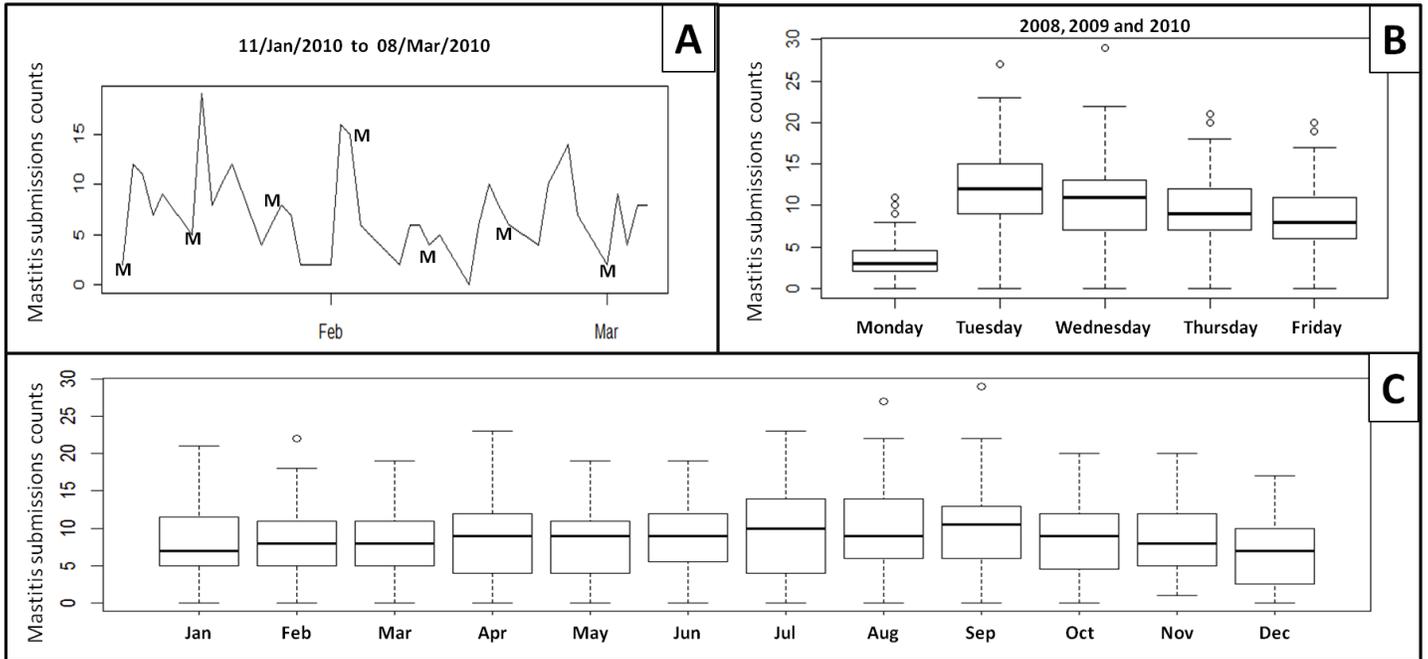
Yahav, I., Shmueli, G., 2007. Algorithm Combination for Improved Performance in Biosurveillance Systems. In: Zeng, D. (Ed.), *BioSurveillance 2007.* , pp. 79-90.

Zeileis, A., Kleiber, C., Jackman, S., 2008. Regression models for count data in R. *J. Stat. Softw.* 27, 8.

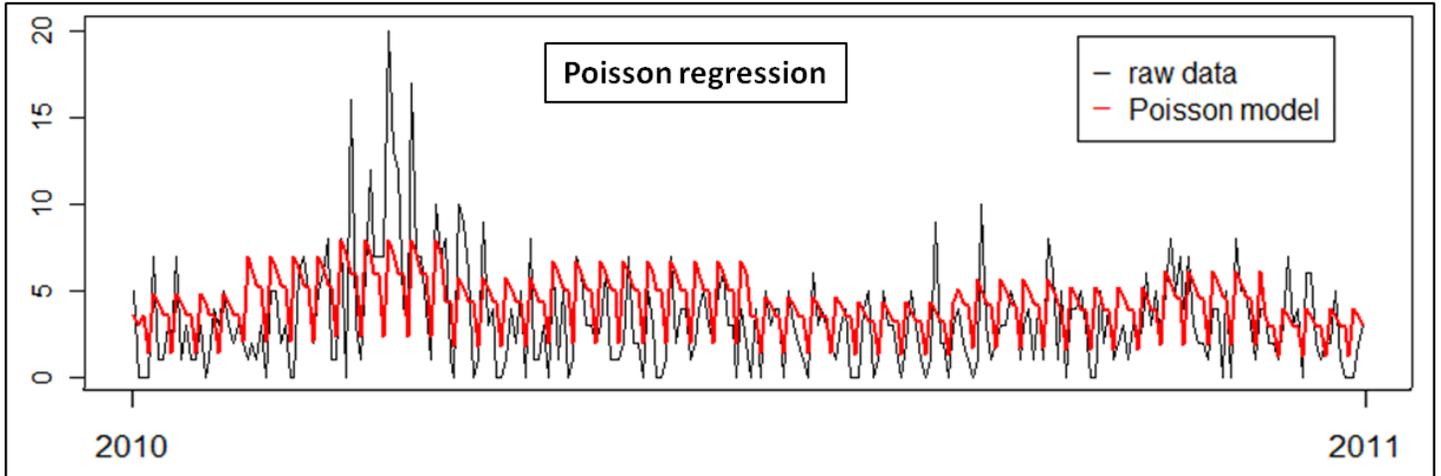
Zhang, X., McEwen, B., Mann, E., Martin, W., 2005. Detection of clusters of salmonella in animals in ontario from 1991 to 2001. *Can. Vet. J.* 46, 517-9, 522-3.



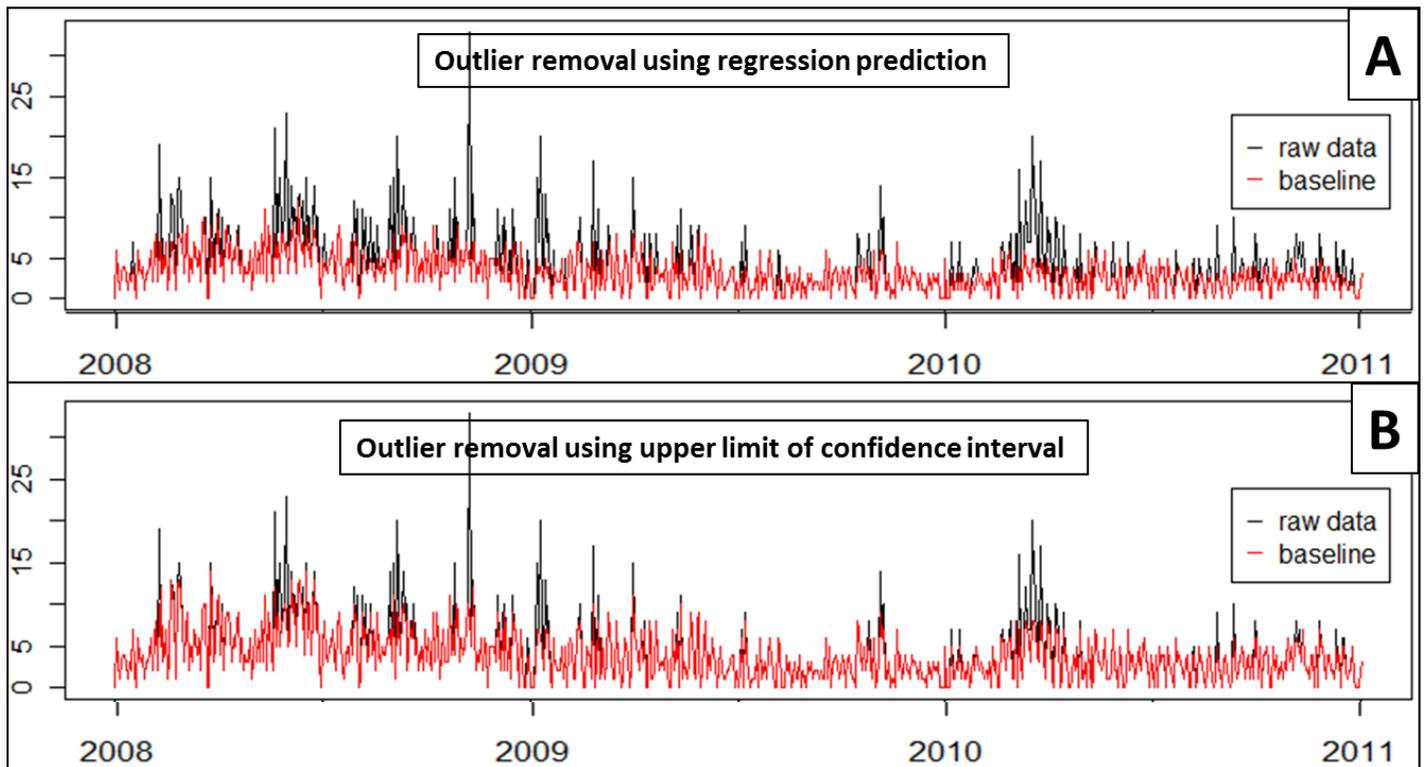
**Figure 1.** Examples of time series of daily (left) and weekly (right) counts of requests for tests associated with various syndromic groups.



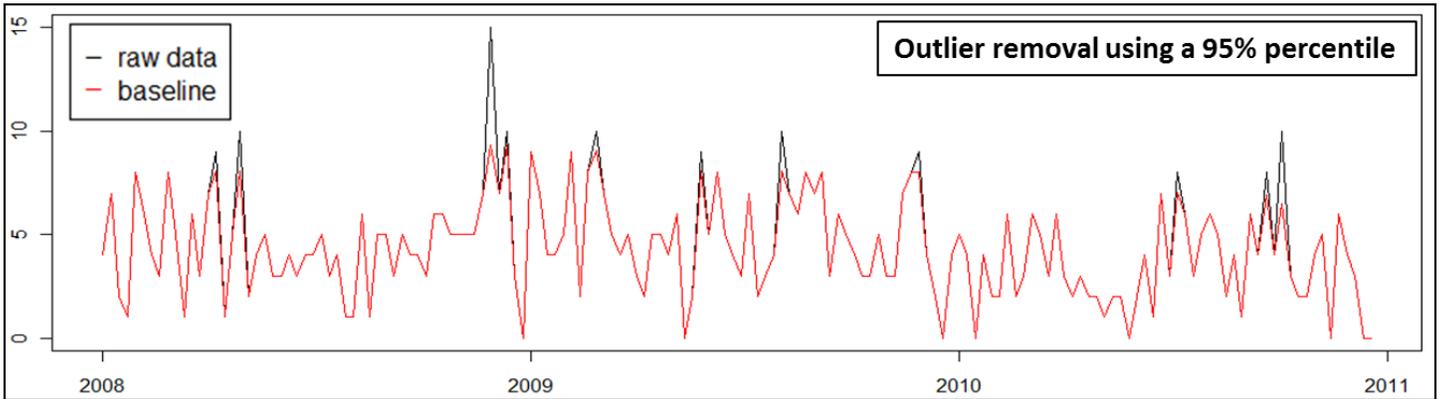
**Figure 2.** Day-of-week and month effects exemplified using the mastitis series. A- 7 week zoomed view of the series. Mondays are labeled with “M”. B-Box-plots of all counts in each day of the week for the entire series (2008 through 2010). C- Box-plots of all counts in each month for the entire series (2008 through 2010).



**Figure 3.** Poisson model predictions (red) and the original BLV time series used to fit the model (black). Model variables were day-of-the-week and month.



**Figure 4.** Raw data of daily counts for the BLV series (black), superimposed by a baseline constructed after removal of temporal aberrations (red) using an iterative process based on Poisson regression. In the top graph, data points greater than the 95<sup>th</sup> 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 95<sup>th</sup> percentile interval in each iteration.



**Figure 5.** Raw data of weekly counts for the systemic syndrome series (black) superimposed by a baseline constructed after removal of temporal aberrations (red) using an iterative process based on moving 95<sup>th</sup> percentiles, in windows of 26 weeks.

**Table 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.

<b>DAILY monitoring</b>	<b>Days with 0 counts</b>	<b>Percentiles</b>			
		<b>25%</b>	<b>50%</b>	<b>75%</b>	<b>100%</b>
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
<i>Neospora caninum</i> 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

<b>WEEKLY monitoring</b>	<b>Weeks with 0 counts</b>	<b>Percentiles</b>			
		<b>25%</b>	<b>50%</b>	<b>75%</b>	<b>100%</b>
Circulatory, hepatic and hematopoietic <sup>1</sup>	39.2%	0	1	2	6
Nervous	37.8%	0	1	2	5
Reproductive and Abortion <sup>2</sup>	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 <sup>3</sup>	64.1%	0	0	1	5

<sup>1</sup>This group merges syndromic groups which initially contained very small numbers of submissions: “circulatory”, “hepatic” and “hematopoietic”; <sup>2</sup>merges “reproductive” and “abortion”; <sup>3</sup>merges “skin”, “eyes” and “ears”.

\*Test requests that could not be classified into any of the other groups.