What tools are useful for monitoring endemic diseases? A simulation study based on different time-series components

AC LOPES ANTUNES[1]*, D JENSEN[2], T HALASA[1], N TOFT[1]
[1]Section for Epidemiology, National Veterinary Institute – DTU, Bülowsvej 27, 1870 Frederiksberg C, Denmark; [2]Section for Production and Health, Department of Large Animal Science – KU, Grønnegårdsvej 8, 2000 Frederiksberg C, Denmark

*aclan@vet.dtu.dk

Abstract
Control and eradication programs play an important role in disease monitoring and surveillance. It is important to follow up on implemented strategies to reduce and/or eliminate a specific disease. The objectives of this study were to investigate the performance of different detection methods, including methods commonly used in biosurveillance as well as state space models, for monitoring the effect of endemic disease control and eradication programs.

We simulated 16 different scenarios of changes in disease sero-prevalence, inspired by real-world data from the Danish PRRS (Porcine Reproductive and Respiratory Syndrome) monitoring program. The changes included increases, decreases and/or constant sero-prevalence levels in different combinations. Two state space models were used to model the simulated data and different monitoring methods, such as univariate process control algorithms (UPCA) and monitoring of the trend component were tested. The performance was evaluated as the proportion of iterations with an alarm for a given week.

Results revealed that the different UPCA performed differently with respect to detecting increasing and decreasing changes in sero-prevalence. The trend-based methods performed well for detecting the first event but its performance was poorer in adapting to several consecutive events.

The different monitoring methods had different performances in monitoring increasing and decreasing changes in disease sero-prevalence, showing that the objectives of the monitoring program should be taken into account when choosing which methods to use. The principles used in this study can also be applied in disease surveillance of (re-)emerging diseases.

Keywords: *surveillance, endemic diseases, time-series components*

Introduction
Surveillance and monitoring systems are critical for timely and effective detection of diseases. These systems are also important to follow up on implemented strategies to reduce and/or eliminate a specific disease (1).

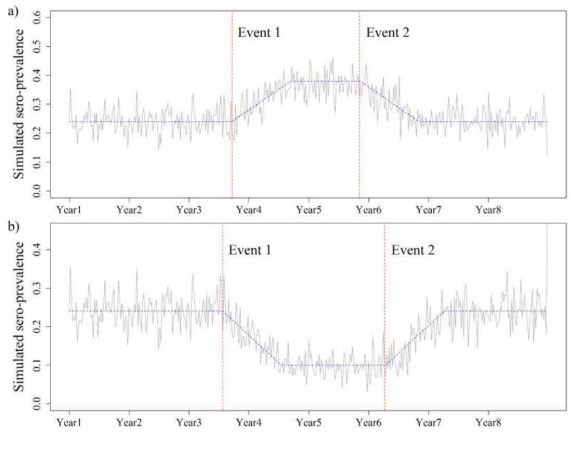
Over the last decade, several studies have applied statistical control methods to detect outbreaks of (re-)emerging diseases in the context of (syndromic) surveillance in human and veterinary medicine (2–4). Still, it may not be possible to generalise the performance of these methods for monitoring endemic diseases, where changes are expected to happen in more gradual patterns (5).

The Danish monitoring program for PRRS is an example of a monitoring program for an endemic disease. Despite control efforts, PRRS continues to contribute to economic losses (6). Danish PRRS surveillance and monitoring is primarily based on regular serological testing of herds.

In this study, we investigated the performance of different detection methods through a simulation study of an endemic disease with known changes. Data from the Danish PRRSV monitoring program was used to inspire and perform the simulation study. Two state space models were chosen for this study based on their ability to monitor changes in different time-series components (7). For generating alarms, five different approaches were evaluated for each model: three univariate process control algorithms (UPCA) (8) and two alternative methods for monitoring changes based on the trend component of a time series.

Materials and methods
It was decided to perform a simulation study to assess the performance of different monitoring methods in detecting temporal changes in disease sero-prevalence. The simulated number of positive herds for a given week was derived from a binomial distribution with a probability p and a sample size n corresponding to the number of Danish swine herds tested for PRRS per week as described earlier (9). For all simulated scenarios, the first 104 weeks were simulated with a constant initial sero-prevalence of 0.24. This corresponded to the average PRRS sero-prevalence in Danish herds observed using laboratory diagnostic data from 2007 to 2014. Then, this constant level was followed by an increase/decrease in the weekly sero-prevalence (Event 1), followed by a second constant level and then a second increase/decrease (Event 2) (Figure 1). Different combinations of events in sero-prevalence (increase/decrease) and durations were tested for each scenario resulting in a total of 16 simulated scenarios. Each scenario was simulated with 2000 repetitions.

Figure 1. Example of the simulated scenarios representing an endemic disease with underlying changes in prevalence. An initial constant sero-prevalence of 0.24 was used for at least 104 weeks, followed by an increase to 0.38 (a) or decrease to 0.10 (b) over a period of 52 weeks at two different times (Events 1 and 2).



A DLM (Dynamic Linear Model) and a DGLM (Dynamic Generalised Linear Model) based on normal and binomial distributions respectively, both with a linear growth component, were used to model the different simulated scenarios. The general objective of these models is to estimate an underlying parameter vector from observed data combined with any prior information available before an observation is made at time t. Different monitoring methods were applied for monitoring the simulated changes. Three UPCA were used to generate alarms based on normalised forecast errors from both models: Shewart control chart, Tabular Cumulative Sums, and the V-Mask (8). Additionally, 99% confidence intervals (CI) and changes in the sign of the trend component were used as trend-based methods (Figure 2).

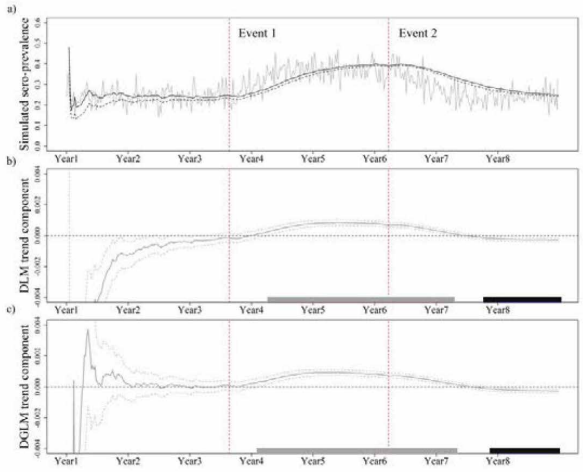
The performance of these methods was assessed based on the cumulative sensitivity (CumSe), in which the proportion of alarms generated for each week after both events were started was calculated.

Results

Results revealed that the use of a Shewart Control Chart resulted in better detection of an increasing sero-prevalence compared to the detection of a decreasing sero-prevalence for all simulated scenarios, while the opposite was observed for the Tabular cumulative sums. The trend-based methods detected the first event well, but its performance in adapting to several consecutive events was inferior. As an example, to detect an increase in the sero-prevalence from 0.24 to 0.38 over a period of 104 weeks 37 weeks were needed to achieved a CumSe=50% based on 99% CI for the first event. The same CumSe was achieved in two weeks when monitoring the Trend Sign, for the first event. For the second event, a CumSe=50% was only achieved after 74 weeks based on 99% and 59 weeks based on the Trend Sign weeks. The most consistently performing method seemed to be the V-mask for both events, achieving a CumSe=50% between six and 20 weeks after the events.

Additionally, results showed that the univariate process control algorithms needed more time to achieve CumSe=50% when applied to DGLM compared to DLM while the trend-based methods had identical performances in both models.

Figure 2. Results show the simulated weekly sero-prevalence and the filtered mean obtained from the DLM (black dashed line) and GDLM (black solid line) (a) and the corresponding DLM (b) and DGLM trend component (c). The rugs indicate were the trend component is significantly above (grey) and below (black) zero.



Discussion

We demonstrated that the Shewart control chart was able to detect decreasing changes faster than increases. This is in accordance with (8), who mentioned that the Shewart control charts can detect decreasing changes earlier when compared with other methods such as the Tabular cumulative sums. Still, the Tabular cumulative sum was faster in detecting increases. Both DLM and GDLM models were optimised to model a constant level, resulting in slower model-trend changes in Event 2. As a consequence, the normalised forecast errors were higher and the Tabular Cumulative Sums generated alarms earlier for increases in the sero-prevalence. In addition, the variation (noise) in the simulated data was higher when simulating increases in sero-prevalence for Event 2 (higher p), which might have resulted in a higher number of alarms. This can explain the better performance of the Tabular cumulative sums. The most consistent results for all simulated scenarios were verified for the V-Mask, possible as a result of its flexibility, i.e. its position changes according to the magnitude of the forecast errors for each time step t - when compared to other monitoring methods. These results highlight that UPCA were sensitive to the intensity of noise in the data. The results for the trend-based methods showed that despite of their good performance in detecting the first event, both models needed time to adapt to Event 2 of both scenarios. This can result from the fact that the models are forced to adapt to three consecutive simulated levels of the sero-prevalence to Event 2.

In this study, we showed that there is no robust method for all scenarios. Similar conclusions were drawn in previous studies in the context of syndromic surveillance for (re-)emerging diseases (2,10) where authors concluded the inexistence of "the one method which fits all" outbreak shapes. In this context, the choice of a specific monitoring method will be a challenge and should take into the account the objectives of the monitoring program and the performance of the monitoring methods in different time patterns (11). Moreover, the type of control program implemented should also be taken into account and it depends on the nature of the disease, political and economic considerations and the infrastructures present in the country where it is to be implemented (12).

In this study, state space models were used for monitoring endemic disease and control programs using two distinctive monitoring approaches for the time-series components. The principles stated might also be applied in general modelling, monitoring and surveillance of (re-)emerging diseases in human and veterinary sciences.

References

1. **Doherr MG, Audigé L.** *Philos Trans R Soc Lond B Biol Sci* 356(1411), 1097–106, 2001
2. **Dórea FC *et al.*** *PLoS One* 8(12), e82183, 2013
3. **Jackson ML *et al.*** *BMC Med Inform Decis Mak* 7(1), 6, 2007
4. **Buckeridge DL *et al.*** *Morb Mortal Wkly Rep* 54 (Suppl), 109–15, 2005
5. **Pradhan AK.** *J Dairy Sci* 92(4), 1811–25, 2009
6. **Kvisgaard L.** *Understanding and combating PRRS in Europe.* Novi Sad: Serbia, p60, 2011
7. **West M *et al.*** *Bayesian Forecasting and Dynamic Models. 2nd Ed.* New York, USA: Springer, pp32-67, 97-142, 270-316, 516-560, 1997
8. **Montgomery D.** *Introduction to Statistical Quality Control. 6th Ed.* Arizona State University: John Wiley & Sons, Inc., pp234-446, 2009.
9. **Lopes Antunes AC *et al.*** *Prev Vet Med* 127, 15–20, 2016
10. **Yahav I *et al.*** *Intelligence and Security Informatics: Biosurveillance.* Springer, pp91–102, 2007
11. **MM Wagner *et al.*** *Handbook of Biosurveillance.* Elsevier, pp217–34, 2006
12. **Christensen J.** *Animal Disease Surveillance and Survey Systems Methods and Aplications. Ist Ed.* Blackwell Publishing, pp15–34, 2003
13. **Dórea FC *et al.*** *Prev Vet Med* 13, 116(3), 313–24, 2014