Monitoring diseases based on register data: Methods and application in the Danish swine production

Lopes Antunes, Ana Carolina; Toft, Nils; Hisham Beshara Halasa, Tariq

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

DTU Vet
National Veterinary Institute

Monitoring diseases based on register data: Methods and application in the Danish swine production

PhD Thesis
Ana Carolina Lopes Antunes
December 2016

Monitoring diseases based on register data: Methods and application in the Danish swine production

Ana Carolina Lopes Antunes

PhD Thesis

December 2016

Enrolled at Section for Diagnostics and Scientific Advice, National Veterinary Institute - Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark.

Supervisors

Professor Nils Toft (Main supervisor)
Section for Diagnostics and Scientific Advice
National Veterinary Institute - Technical University of Denmark
Denmark

Senior Researcher Tariq Halasa (Co-supervisor)
Section for Diagnostics and Scientific Advice
National Veterinary Institute - Technical University of Denmark
Denmark

Assessment Committee

Researcher Vibeke Frøkjær Jensen (Chairperson)
Section for Diagnostics and Scientific Advice
National Veterinary Institute - Technical University of Denmark
Denmark

Head of contingency planning and disease control Sten Mortensen
Danish Veterinary and Food Administration
Denmark

Professor Crawford Revie
Department of Health Management
Atlantic Veterinary College, University of Prince Edward Island
Canada

"Data isn't information. ... Information, unlike data, is useful. While there's a gulf between data and information, there's a wide ocean between information and knowledge. What turns the gears in our brains isn't information, but ideas, inventions, and inspiration. Knowledge—not information—implies understanding. And beyond knowledge lies what we should be seeking: wisdom."

— Clifford Stoll

in High-Tech Heretic: Reflections of a Computer Contrarian (2000), 185-186.

Contents

Preface and acknowledgments

During my Doctor of Veterinary Medicine (DVM) degree, I was exposed to several areas of study, but the one that rapidly became my passion was epidemiology. As a result, I enrolled in several traineeships in order to develop my knowledge of different topics within this field.

My interest in disease surveillance started during my traineeship in the final year of my DVM. The continuous monitoring of animal health-related data was and is a growing field due to challenges concerning the choice of data sources and monitoring methods. Knowing "what to look for" and "how to know if a disease is spreading" depends on the context in each country.

On 10th October 2013 (the same day I got my DVM degree) I received a message from a former PhD student at Copenhagen University, advising me to look at a PhD position advertised on the National Veterinary Institute – Technical University of Denmark (DTU Vet) website. After reading the description of the position, my first thought was "I´ll send an application and see what happens!" I had no idea what I was getting into…

When I accepted the position, I knew I would have to leave my comfort zone and move to a "Viking country" for 3 years. I was naïve to think that doing a PhD in disease surveillance would involve simply plotting laboratory data from swine, cattle and poultry and using "friendly statistical methods for veterinarians" previously used in Syndromic Surveillance. How wrong I was!

I landed for the first time in Copenhagen on the evening of 10th December 2013. On 13th December - it was a Friday, yet I still don't know if that was a good or a bad sign! - I went to DTU Vet to see my future workplace and meet my supervisors and colleagues.

The journey officially started on 15th December 2013. After a few days at work, I realized that doing a PhD is not only about science, but also learning how to deal with people with very different personalities – especially our supervisors! I must confess that there were ups and downs, funny and stressful moments, and people supported me in their own way.

Firstly, I would like to thank my main supervisor Nils Toft, without whom (and in spite of his candor, criticism of my work, extremely busy schedule and tons of sarcasm!) it wouldn't have been possible to finish this PhD. Besides, he had to learn how to deal with my

"Portuguese temperament" especially when "the mustard was already getting to my nose"[1]! So I must say thank you for that, and for giving me the chance to do this PhD.

I must also thank Tariq Halasa, my co-supervisor, for his support, laughs and kind comments on my work. Thank you as well for attempting to cheer me up when I was getting frustrated during my PhD.

However, a PhD requires the involvement of more people than the supervisors.

Dan Jensen was the person who provided scientific support during the last year of my PhD. Without him, I wouldn't have been able to learn and apply "unfriendly statistical methods for veterinarians", which include terms such as priors and matrices. I would like to thank him for his support, for the pizza and beers after long hours at the office programming, and for cheering me up at times when I wasn't sure if I'd be able to finish this PhD. *Mange tak*!

I would like to thank Fernanda Dórea for her supervision both during and after my first stay at the National Veterinary Institute (SVA) in Uppsala. Thank you for teaching me, having the patience to deal with my stress and frustrations, and making me understand that it´s all part of PhD life. Thank you for your support and for boosting my confidence on bad days. Also, thank you for opening the doors of your home and providing me with shelter and food. *Muito obrigada!*

I would like to thank Annette Ersbøll for the opportunity to do my second external research at the National Institute of Public Health (SIF) - University of Southern Denmark. Thank you for your kind support during my stay at SIF.

I would like to thank all of my co-authors and colleagues: Klara Tølbøl Lauritsen, Charlotte Sonne Kristensen, Lars Erik Larsen, Mette Ely Fertner, Anna Camilla Birkegård, Anette Boklund and Kristine Bihrmann for their support and contributions to the projects included in this thesis.

I would also like to thank to the Danish Pig Research Centre – SEGES for providing the data for the project.

A PhD also requires non-scientific support.

I would like to thank all of my current and former colleagues and friends at the Section for Epidemiology. Thank you guys for your support, for the laughs, for the breakfasts, for the cakes and beers. A special thanks to Rene Bødker, Peter Lind, Carsten Kirkeby, Kaare

[1] A Portuguese idiom used when people are getting upset about something. In portuguese: *A mostarda já me estava a chegar ao nariz.*

Græsbøll, Alessandro Foddai and Birgit Kristensen for the funny chats and moments in the later afternoons – often around the mini pool table – at the office during my "fresh PhD times". Also, special thanks to Syed Sayeem Ahmed for opening the doors of his house several times and for feeding me amazing Bengali food and allowing me to enjoy time with him, his family and friends. To Ana Carolina Cuellar and Maya Gussman, thank you for the tea and cake, the shopping sessions and trips on weekends.

To the people I met at DTU Vet, SVA and SIF who were not directly involved in my project, but contributed to the good working environment, a big thank you!

To my colleagues at the Department of Large Animal Sciences at Copenhagen University, I would like to thank you for your support, laughs and beers at conferences, *A-vej*, and *Smediefesten*.

Moving out of my comfort zone forced me to adapt and create a new circle of friends. I'm thankful for all of the friends I made in Copenhagen – from all over the world – during these last 3 years. Thank you for the Friday beers, parties, brunches, dinners, funny Danish classes, *hyggelig* tea and cake and for supporting and cheering me up on stressful days. Thank you guys!

We live in a time where social apps are freely available and they can be used to keep in touch with people several flying-hours away. This helped me to keep in touch with my friends, family and boyfriend in Portugal. I would like to thank all my friends there for their support – mainly through Facebook and WhatsApp – and the warm "welcome back"s. Despite the distance, it's good to know that nothing has changed.

Finally, I thank my family and boyfriend for their support and for accepting my quarterly visits. They had to adapt to a Skype-based routine to talk about the changes in our lives. Also a big thank you for taking care of my dog Spike.

After 3 years of hard work, tons of knowledge, laughs and cries, stressful and unforgettable moments and, most surprisingly, not a single grey hair, this chapter of my life is now complete.

København, 14th December 2016

Summary

The spread of diseases is one of the most important threats to animal production and public health. Disease spread causes considerable economic losses for the agricultural sector and constitutes trade-limiting factors, as transmission to countries free from disease should be avoided.

Monitoring and surveillance systems are critical for the timely and effective control of infectious diseases. The ability of a system to detect changes in the disease burden depends on the choice of data source. Many factors can lead to inconsistent data collection among populations and it is therefore important to assess the quality of data before use in disease monitoring and surveillance.

Over the past decade, several studies have focused on using statistical control methods to detect outbreaks of (re-)emerging diseases in the context of syndromic surveillance – both in human and veterinary medicine – in an attempt to supplement traditional sentinel surveillance. However, it may not be possible to generalize the performance of these methods to the context of other countries (where data have different characteristics), or to the context of endemic diseases.

Lower incidence rates are normally expected for endemic diseases compared to highly infectious (re-emerging) diseases, due to control measures such as vaccination or health-management programs. Furthermore, the data collected differ from those obtained from traditional surveillance (generally related to incidence monitoring), due to its focus on the endemic scenario, with less frequently sampled data. This reflects the added complexity of monitoring endemic diseases, as disease burden is affected not only by the incidence, but also by the duration and recovery rate.

The aim of this thesis was to evaluate existing register data related to veterinary health, as a tool for monitoring swine diseases in Denmark. This included: i) describing and evaluating the quality of data (regarding the potential for disease monitoring and surveillance) in Danish databases related to swine health; ii) assessing the feasibility of studying changes in data records over time to detect changes that might indicate disease spread between swine herds; iii) evaluating the performance of different time-series methods for the monitoring and surveillance of endemic diseases, as well as assessing the impact of noise in the data on the results when using these methods. Some of the work presented was focused on endemic diseases, using Porcine Reproductive and Respiratory Syndrome (PRRS) as example.

Interviews were conducted with relevant stakeholders in order to assess the data quality of seven databases: the Central Husbandry Register (CHR), the swine movement database

(SMD), the national Danish database of drugs for veterinary use (VetStat), laboratory diagnostic data from the National Veterinary Institute – Technical University of Denmark (DTU-Vet lab) and the Pig Research Centre - SEGES (VSP-SEGES lab), the Specific Pathogen Free System (SPF System) and the Meat Inspection database. The guidelines from the European Centre for Disease Prevention and Control (ECDC) for monitoring data quality and surveillance systems were used. The findings showed that limitations included delayed transfer of data to databases and incomplete representation of Danish swine herds.

Laboratory submission data for testing PRRS were used to study temporal changes in data records, due to the large amount of diagnostic data available. The laboratory data proved to be useful for monitoring temporal patterns of disease occurrence. The fact that some Danish swine herds are tested monthly allows for changes in disease prevalence and incidence to be monitored, which is an example of sentinel surveillance. However, for other herds, the frequency of testing (i.e. the representativeness of the data) depends on factors such as the herd status, farmer compliance, the value of the animal, commercial purposes and ongoing control and eradication programs. This limitation did not apply to the mortality data, which is available for all Danish swine herds on a monthly basis. However, observed changes might be due to disease occurrence, or as a result of changes in herd management or a lack of accuracy in the calculation of mortality.

Several scenarios representative of changes in endemic disease sero-prevalence programs were simulated to test the performance of different monitoring methods. These included univariate process control algorithms applied directly to the simulated data, as well as using the forecast errors and trend-based methods. The performance of these methods was evaluated based on the sensitivity and time taken to detect changes, which showed that some methods were more efficient than others for specific patterns. Therefore, choosing a single temporal monitoring method is challenging, and the objectives of the monitoring program and the differing performance of the methods in detecting a specific pattern should be taken into account. Changes in the noise of the data had an impact on the univariate process control algorithms, while the trend-based methods provided a consistent approach to monitoring changes in disease or sero-prevalence.

The findings of this thesis may serve as a basis for the improvement of monitoring swine diseases in Denmark. Although the available databases have the potential for use in disease monitoring and surveillance of swine herds in Denmark, improvements are needed for accurate and real-time implementation. Further research relating to the improvement of data quality, as well as combining different data sources for monitoring endemic diseases in Denmark is needed.

Sammendrag

Sygdomsspredning er en af de vigtigste trusler mod produktion af dyr og folkesundheden. Sygdomme kan forårsage betydelige økonomiske tab for landbruget og begrænse for handel med lande, som er fri for den pågældende sygdom. Overvågnings og kontrolsystemer er kritiske for rettidig og effektiv kontrol af infektiøse sygdomme. Hvor godt et system er til at detektere ændringer i sygdomsbyrden afhænger af, hvilke datakilder der vælges. Flere faktorer kan føre til inkonsekvent dataindsamling i husdyrpopulationer. Det er derfor vigtigt at vurdere datakvaliteten, før data bruges til sygdomsovervågning.

Gennem det sidste årti har der været flere studier, der anvender statistiske kontrolmetoder til at detektere udbrud af (gen-)opståede sygdomme i forbindelse med syndromovervågning – både i human- og veterinærmedicin. Disse metoder er et forsøg på at bidrage til traditionel overvågning af sygdomsudbrud. Det har ikke været muligt at generalisere effekten af disse metoder, når de har været brugt i relation til andre lande (hvor data har andre karakteristika) eller i relation til endemiske sygdomme.

Endemiske sygdomme har lavere incidensrater end, hvad der normalt forventes, når de sammenlignes med eksotiske sygdomme. Dette skyldes anvendelsen af kontrolforanstaltninger såsom vaccination og/eller biosecurity. Data fra overvågning af endemiske sygdomme er generelt anderledes end typiske overvågningsdata for eksempel på grund af mindre hyppige dataindsamling. Dette reflekterer den øgede kompleksitet i overvågningen af endemisk sygdom, eftersom sygdomsbyrden er påvirket ikke kun af incidensen, men også af udbruddets varighed og af hvor hurtigt dyrene kommer sig.

Formålet med denne afhandling var at evaluere eksisterende veterinære sundhedsrelaterede registerdata som redskab til at overvåge svinesygdomme i Danmark. Dette inkluderer i) beskrivelse og evaluering af datakvaliteten i danske databaser med svinesundhedsrelaterede data i relation til potentialet for sygdomsovervågning, ii) detektere ændringer over tid i dataregistreringer, der muligvis indikerer spredning af sygdom mellem svinebesætninger, og iii) evaluering af forskellige metoder til overvågning af endemisk sygdomme samt påvirkningen af støjen i data ved brug af disse metoder. De fleste af studierne blev gennemført med Porcint Reproduktions- og Respirations Syndrom (PRRS) som eksempel på en typisk endemisk sygdom.

Ved hjælp af en interviewundersøgelse blev der lavet en vurdering af datakvaliteten af syv databaser: Central Husdyrregister (CHR), svineflyttedatabasen (SMD), den nationale danske database for medicin til veterinært brug (VetStat), diagnostisk laboratoriedata fra Veterinærinstituttet – Danmarks Tekniske Universitet (DTU-Vet lab) og videncentret for

svineproduktion – SEGES (VSP-SEGES lab), specifik patogen fri-systemet (SPF systemet) og kødkontroldatabasen. Vurderingen af kvaliteten af data og overvågningssystemer blev gennemført med en tilrettet version af instruktionerne fra det europæiske center for sygdomsforebyggelse og -kontrol (ECDC). Resultaterne viste, at der var begrænsninger i potentialet som følge af blandt andet forsinket dataoverføring, og hvor godt de danske svinebesætninger var repræsenteret i data.

Data for laboratorieindsendelser til test af PRRS blev brugt som et eksempel på, hvordan temporale ændringer i dataregistreringer kunne undersøges på grund af store mængder af diagnostisk data. Det blev påvist, at laboratoriedata var brugbare til at undersøge temporale mønstre i sygdomsforekomsten. Det faktum, at nogle danske svinebesætninger testes månedligt, muliggør overvågningen af ændringer i sygdomsprævalensen af disse. For andre besætninger afhænger frekvensen af test (dvs. hvor godt data repræsenterer populationen) af andre faktorer så som besætningens status, besætningsejerens accept, dyrets værdi, kommercielt formål samt kontrol- og udryddelsesprogrammer. Dødlighedsdata, var tilgængelige for alle danske svinebesætninger på månedlig basis og udgør dermed en komplet sample. De observerede ændringer i dødelighed kunne dog skyldes sygdomsforekomst, ændringer af management af besætningen eller begrænsninger i, hvordan data er udregnet.

Adskillige scenarier, der kunne repræsentere ændringer i forekomsten af endemiske sygdomme, blev simuleret for at teste forskellige overvågningsmetoder og deres effektivitet. Der blev brugt ”univariate process control” algoritmer, der blev anvendt direkte på de simulerede data, eller på prædiktionsfejlen. Desuden blev der testet forskellige trend-baserede metoder. Effektiviteten af metoderne blev evalueret på basis af hvor ofte og hvor hurtigt de detekterede de simulerede ændringer. Resultaterne viste, at nogle metoder var mere effektive end andre for specifikke mønstre. Derfor er valget af en enkelt metode til temporal overvågning vanskeligt. Formålet med overvågningsprogrammet bør indgå i overvejelserne. Ændringerne i antallet af prøver påvirkede “univariate process control”-algoritmernes effektivitet, hvorimod de trend-baserede metoder var mindre påvirkede.

Resultaterne i denne afhandling kan bruges som basis for at forbedre overvågningen af svinesygdomme i Danmark. De tilgængelige databaser har potentiale til at blive brugt til sygdomsovervågning i de danske svinebesætninger, men forbedringer er nødvendige før implementering af overvågning i real tid. Yderligere forskning i, hvordan datakvaliteten kan forbedres og forskellige datakilder kan kombineres for overvågningen af endemiske sygdomme i Danmark er nødvendig.

Sumário

O risco de propagação de doenças é uma das maiores ameaças à produção animal e à saúde pública. A propagação de doenças causa perdas económicas consideráveis no setor agrícola, levando à ocorrência de restrições comerciais na tentativa de evitar a propagação para outros países.

Os sistemas de monitorização e vigilância são fundamentais para o controlo rápido e eficaz das doenças infecciosas. A capacidade de um sistema para detectar alterações na ocorrência de doenças depende fortemente da escolha da fonte de dados. Muitos factores podem levar à recolha inconsistente de dados entre as populações e, portanto, é importante avaliar a qualidade dos dados antes da sua utilização para a monitorização e a vigilância de doenças.

Na última década, vários estudos avaliaram a capacidade de vários métodos estatísticos para detectar surtos de doenças (re-)emergentes no contexto da vigilância sindrómica - tanto na medicina humana como na medicina veterinária, numa tentativa de complementar os sistemas de vigilância tradicionais. No entanto, não é possível generalizar a adequação desses métodos para monitorizar doenças em outros países (onde os dados têm características diferentes) e no contexto de doenças endémicas.

As taxas de incidência de doenças endémicas são normalmente mais baixas quando comparadas com as taxas de incidência de doenças altamente infecciosas (re-emergentes), devido à aplicação de medidas de controlo tais como programas de vacinação ou boas práticas de gestão. Além disso, os dados são recolhidos com menor frequência do que os obtidos através da vigilância tradicional que geralmente visam a monitoração da incidência. Isto reflecte a complexidade acrescida da monitorização das doenças endémicas, uma vez que a severidade da doença é afectada não só pela incidência, mas também pela sua duração e taxa de recuperação.

O objetivo desta tese foi avaliar o potencial dos registos de saúde animal existentes como uma ferramenta para monitorizar doenças em suínos na Dinamarca. Isto incluiu: i) descrever e avaliar a qualidade dos dados nas bases de dados dinamarquesas relativamente ao seu potencial para a vigilância da doença; ii) explorar a viabilidade de monitorizar alterações temporais e espaciais nos dados para detectar mudanças que possam indicar a propagação de doenças entre varas e iii) avaliar o desempenho de diferentes métodos de análise de séries temporais quando aplicados à monitorização e vigilância de doenças endémicas e o impacto da variação dos dados nos seus resultados. A marioria dos estudos incluídos nesta tese focaram-se em doenças endémicas, usando o Síndrome Respiratório e Reprodutor Porcino (PRRS) como exemplo.

Várias entrevistas foram feitas com partes interessadas a fim de avaliar a qualidade de sete bases de dados: i) Registro Central de Pecuária (CHR), ii) movimentos de suínos (SMD), iii) registo de medicamentos para uso veterinário (VetStat), iv) s laboratórios do Instituto Nacional de Veterinária - Universidade Técnica da Dinamarca (laboratório DTU-Vet), v) do Centro de Pesquisa de Suínos - SEGES (Laboratório VSP-SEGES), vi) Sistema Específico de Patógeno Livre (SPF System) e vii) abate de suínos. As directrizes do Centro Europeu de Prevenção e Controlo de Doenças (ECDC) foram usadas para avaliar a qualidade dos dados. Os resultados revelaram que existem limitações, tais como atrasos na transferência dos dados para as bases de dados, a representatividade do efectivos suíno na Dinamarca.

Os resultados dos testes laboratoriais para a PRRS serviram de exemplo para estudar alterações temporais nos dados devido à grande quantidade de dados disponíveis. Demonstrou-se a utilidade destes dados para monitorizar padrões temporais de ocorrência de doenças. O facto de algumas varas serem testadas mensalmente permite monitorizar a prevalência e a incidência de doenças, sendo um exemplo de vigilância sentinela. No entanto, a frequência dos testes (ou seja, a representatividade da população nos dados) depende de factores como o impacto da doença, da complacência do produtor, o valor do animal, os objativos comerciais e os programas de controlo e erradicação em curso. Esta limitação não foi observada para os dados de mortalidade, que está disponível para toda a população de suínos todos os meses. No entanto, as mudanças observadas na mortalidade podem ser devidas à ocorrência de doenças, alterações na gestão da vara ou falta de precisão na forma como a mortalidade é calculada.

Foram simulados vários cenários representativos de mudanças na soroprevalência de doenças endémicas para testar a capacidade de detecção de diferentes métodos de análise de séries temporais, incluindo algoritmos para controlo estatístico de processo univariados, usados directamente nos dados simulados, bem como nos erros de previsão dos modelos e métodos baseados em tendências. O desempenho dos diferentes métodos foi avaliado com base na sua sensibilidade e tempo de detecção das alterações simuladas, mostrando que alguns métodos são mais eficientes do que outros para padrões específicos. Assim, a escolha de um único método de monitoração temporal é um desafio; os objectivos do programa de monitorização e o melhor desempenho na deteção de padrões específicos devem ser tidos em conta. As alterações na variação dos dados simulados tiveram um impacto no desempenho de algoritmos para controlo estatístico de processos univariados, ao passo que os métodos baseados na tendência da série temporal não sofreram qualquer impacto sendo um bom método para monitorar alterações na soroprevalência.

Os resultados desta tese podem servir para melhorar a monitorização de doenças em suínos na Dinamarca. Mais pesquisas são necessárias para melhorar a qualidade dos dados e para integrar diferentes bases de dados para a monitorização de doenças endémicas na Dinamarca. Os métodos de análise de séries temporais com o melhor desempenho podem ser usados para monitorizar programas de controlo e erradicação de doenças endémicas na Dinamarca.

1. Introduction

1.1 Disease monitoring and surveillance

Over the past decades, the risk of transmission and spread of new, transboundary and re-emerging diseases has become one of the most important threats to animal production and public health worldwide as a consequence of trade in a globalized world (Coker et al., 2011).

Recently, Europe experienced the emergence of new diseases such as Schmallenberg (Beer et al., 2013; Delooz et al., 2016) and re-emerging diseases such as foot-and-mouth disease (Bouma et al., 2003; McLaws et al., 2007) and African swine fever (Oļševskis et al., 2016; Sánchez-Vizcaíno et al., 2013). Endemic diseases such as porcine reproductive and respiratory syndrome (PRRS) in swine herds (Bøtner et al., 1994; Nieuwenhuis et al., 2012) or paratuberculosis in cattle herds (Garcia and Shalloo, 2015; Kreeger, 1991) also contribute to substantial economic losses for the agricultural sector, and might constitute trade-limiting factors in an attempt to avoid their spread to countries free from the diseases.

Disease monitoring describes the ongoing process of collecting data representative of the health and disease status of a given population (Salman, 2003). The main goal of disease surveillance is the early detection of changes in health status, in order to take actions to control disease spread. These actions might include control and eradication programs, where the information from monitoring and surveillance systems is combined with control and intervention strategies employed over a period of time in order to reduce and eliminate the disease occurrence (Salman, 2003).

1.2 Current trends in disease monitoring and surveillance

Monitoring and surveillance systems are critical for the timely and effective control of infectious diseases. Over the past decade, several studies have applied statistical monitoring methods for syndromic surveillance in human and veterinary medicine (Buckeridge et al., 2005; Dórea et al., 2014; Dupuy et al., 2013a; Jackson et al., 2007). In this context, different animal health data sources such as laboratory

pre-diagnostic data (Dórea et al., 2014), mortality data (Alba et al., 2015; Perrin et al., 2015) and meat inspection data (Dupuy et al., 2013b) have been used in an attempt to supplement traditional sentinel surveillance for disease outbreaks. Nevertheless, it may not be possible to generalize the performance of these methods when applied in the context of other countries (where data have different characteristics) or in the context of endemic diseases.

1.2.1. Using register data for disease monitoring and surveillance

The potential use of health-related register data for disease monitoring and surveillance is a growing field. This approach offers a cost-effective way to ensure effective resource allocation. The ability of a system to detect changes in the disease burden is dependent upon the choice of data source, its representativeness and the sampling strategy (Buckeridge, 2007).

The collection of data is influenced by many factors, including the level of awareness and knowledge about a particular disease among animal producers, the availability of a diagnostic laboratory scheme to support and confirm cases, and the extent to which farmers and veterinarians are willing to secure the flow of data. As a result, inconsistent data is collected for various diseases and among different populations (Salman, 2003). It is therefore important to assess the quality of data to ensure they are representative of the target population (European Centre for Disease Prevention and Control (ECDC), 2014; Salman, 2003) and that valid conclusions can be drawn. Other factors that might influence the quality of data and its relevance to disease surveillance include technical aspects, political requirements and stakeholder interests.

1.2.2. Disease monitoring and surveillance methods

Control and/or eradication measures are implemented whenever certain threshold levels related to the disease status have been exceeded spatially, temporally, or spatio-temporally. In some cases, it may not be obvious whether disease events have exceeded the threshold levels, and simply plotting the time-series of events will reveal these “extreme changes”. In other cases, these changes may be subtle, making

it difficult to detect changes in disease patterns based on a visual inspection of plots. In these situations, statistical techniques can be used to introduce objectivity.

Temporal monitoring methods

Recently, several studies focused on applying statistical control methods to detect outbreaks of (re-)emerging diseases in the context of syndromic surveillance – both in human and veterinary medicine (Buckeridge et al., 2005; Dórea et al., 2013; Jackson et al., 2007). Retrospective analysis is a common approach used in the literature for monitoring diseases, as it can be used to provide information on systematic patterns and to model the data. A wide range of models (such as linear, logistic, binomial, Poisson and time-series) have been implemented in syndromic surveillance to evaluate the role of a set of variables and to model trends and patterns of disease occurrence (Rodríguez-Prieto et al., 2014). The model was then used to make forecasts for each time step. The difference between the forecast and the observed data is known as the forecast error, and this is used for generating alarms. These studies applied univariate process control algorithms (UPCA) (commonly called control charts) to define detection limits for generating alarms. This approach implies the existence of historical data (collected over months or years) providing information on a systematic pattern. In the case of an intervention or change in the collection of data used by the surveillance system (e.g. a change in the law requiring testing of a larger number of individuals), it is necessary to pause the surveillance system until these new data are collected and retrospective analysis is performed before adjusting the system. The performance of the UPCA in previous studies cannot be generalized to other data sources or to endemic diseases. Although they prove useful in describing long-term and cyclical patterns and in identifying unusual changes, UPCA usually require a long series of observations (for retrospective analysis), and are unsuitable for relatively recent surveillance for which historical data is not available (Salman, 2003). Furthermore, previous studies focused on the detection of (re-)emerging disease outbreaks, rather than following up control and eradication programs for endemic diseases. The choice of specific temporal methods to detect changes is challenging since their performance depends on factors such as

the magnitude and shape of the signal and the monitored baseline (Buckeridge et al., 2005).

State-space models combine relevant prior knowledge and current information (West and Harrison, 1997). Moreover, they enable monitoring of changes in different time-series components such as trend, cyclic patterns and seasonal patterns, and can incorporate data based on different distributions. While these models have been used for disease monitoring and surveillance in humans (Cao et al., 2014; Cowling et al., 2006), how useful they are in monitoring endemic diseases remains unknown. These models have been adopted by veterinary science for use in herd-management decisions (Jensen et al., 2015; Ostersen et al., 2010; Madsen and Kristensen, 2005).

Spatial and spatio-temporal methods

In some situations, disease spread may not have a substantial temporal component, being more easily detected by its spatial distribution or the combination of both temporal and spatial components (Salman, 2003). Identifying spatial and spatio-temporal clusters has become more convenient with the recent availability of mapping tools and geographical information system (GIS) software. Traditionally, these tools were part of the digital surveillance frameworks, supporting the visualization of results or the implementation of certain spatial transformations of the data (Rodríguez-Prieto et al., 2014; Salman, 2003). They are also used to support active surveillance and design-sampling studies and to supplement other methodologies such as cluster analysis, regression models, risk assessments or simulation modelling (Rodríguez-Prieto et al., 2014). Methods such as the spatial scan statistic (Kulldorff, 1997) can be used to detect purely spatial or spatio-temporal clusters in data.

The scan statistic method (Kulldorff, 2016) is one of the most commonly used tools for spatio-temporal analysis in biosurveillance (Wagner et al., 2006). As a result, this analytical method has been incorporated into several surveillance systems (Heffernan et al., 2004; Lombardo et al., 2003). The simplicity of the method and the ease with which results can be interpreted (Robertson et al., 2010) mean that this methodology is frequently used for early warnings of events. However, several studies reported that this technique resulted in false alarms, requiring further

epidemiological research to determine the cause of any spatial or space-time clustering (Rodríguez-Prieto et al., 2014; Wagner et al., 2006).

1.2.3. Epidemiology, monitoring and surveillance of endemic diseases

Lower incidence rates are normally expected for endemic diseases when compared to highly infectious (re-emerging) diseases, due to control measures such as vaccination or health management programs (Carslake et al., 2011). Additionally, the dynamics of disease spread and previous exposure to the pathogen can lead to immunity for several individuals in a population, thus contributing to a lower incidence. As a result, we expect to observe slow and gradual increases in incidence and prevalence for endemic diseases (Carslake et al., 2011).

The frequency of testing also depends on the value of the animal and not only on the disease impact (Doherr and Audigé, 2001). In these cases, data differ from those obtained from traditional surveillance (generally related to incidence monitoring), due to its focus on the endemic scenario, with less frequently sampled data. It is also important to investigate the representativeness of the data, as well as the sampling strategies before including data in an automated system to detect changes in the disease burden (Buckeridge, 2007). This reflects the added complexity of monitoring endemic diseases, as disease burden is affected not only by the incidence, but also by the duration and recovery rate. In these cases, it is necessary to use models with a more dynamic structure, allowing the parameters to change over time.

For endemic diseases, it is also important that implemented strategies are reviewed in order to reduce and/or eliminate a specific disease as part of a control and eradication program (Doherr and Audigé, 2001). Unexpected changes in reduction, such as an increase in disease occurrence or a failure to achieve a target value of disease prevalence within a certain period of time, indicate that the implemented strategies should be revised. Failure of these programs may have a devastating economic impact on herds with susceptible animals.

1.3 Aim, goals and objectives of the thesis

The overall aim of this thesis is to evaluate the potential use and value of existing veterinary health-related register data as a tool for monitoring swine diseases in Denmark. In order to meet the overall aim, three goals were defined to drive the work presented in this thesis. These goals were to:

- Goal 1: Explore current national databases that might include data with potential for disease surveillance. Based on this, the following objectives were defined:
  - Objective 1.1: Describe the data gathered in different databases.
  - Objective 1.2: Perform a qualitative assessment of stakeholder perception of the data quality of the databases containing swine health-related data, for use in monitoring swine diseases in Denmark.
  - Objective 1.3: Suggest combinations of different databases to improve disease surveillance.
- Goal 2: Examine the feasibility of studying changes in data records over space and time to detect changes that might indicate disease spread between swine herds. The following objectives were defined:
  - Objective 2.1: Describe spatial and temporal trends present in laboratory submission data in Denmark.
  - Objective 2.2: Describe spatio-temporal clusters of mortality data in Danish Swine herds.
- Goal 3: Explore the potential of different temporal monitoring methods for monitoring control programs for endemic diseases in Danish swine herds. The following objectives were defined:
  - Objective 3.1: Compare the performance of different detection methods, including time-series modeling, time-series decomposition and UPCA, when applied to monitoring and surveillance of endemic diseases.
  - Objective 3.2: Assess the impact of the representativeness of the sample size, i.e. the noise in the data, in the performance of temporal monitoring methods.

In order to fulfill these aims and objectives, six studies were conducted, leading to six scientific manuscripts entitled:

Manuscript 1: Evaluation of the perceived utility of information routinely recorded in databases for integrated disease surveillance in swine. Submitted to Acta Veterinaria Scandinavica.

Manuscript 2: Spatial analysis and temporal trends of porcine reproductive and respiratory syndrome in Denmark from 2007 to 2010 based on laboratory submission data. Published in BMC Veterinary Research 2015; 11: 303.

Manuscript 3: Mortality in Danish Swine herds: spatio-temporal clusters and risk factors. Submitted to Preventive Veterinary Medicine.

Manuscript 4: Monitoring endemic livestock diseases using laboratory diagnostic data: A simulation study to evaluate the performance of univariate process monitoring control algorithms. Published in Preventive Veterinary Medicine 2016; 127: 15–20.

Manuscript 5: Dynamic generalized linear models for monitoring endemic diseases: moving beyond univariate process monitoring control. Published in Proceedings for the Annual Meeting of the Society for Veterinary Epidemiology and Preventive Medicine (SVEPM) 2016, 69-79.

Manuscript 6: A simulation study to evaluate the performance of statistical monitoring methods when applied to different time-series components in the context of control programs for endemic diseases. Submitted to Plos One.

2. Materials and methods

2.1 Data

2.1.1. Qualitative data

The data used in Manuscript 1 were gathered through interviews of relevant stakeholders, including the author and co-authors of the manuscript. The databases (which were selected based on their extensive use by the swine industry and research institutes) included: the Central Husbandry Register (CHR), the swine movement database (SMD), the national Danish database of drugs for veterinary use (VetStat), laboratory diagnostic data from the National Veterinary Institute – Technical University of Denmark (DTU-Vet lab) and the Pig Research Centre - SEGES (VSP-SEGES lab), the Specific Pathogen Free System (SPF System) and the Meat Inspection database.

2.1.2. Register data

Central Husbandry Register (CHR)

The CHR incorporates information on the location of animals, including farms, abattoirs, rendering plants, markets, assembly centers, agricultural shows and common pasture. Each location has its own unique ID number (CHR number) with affiliated address, Cartesian geographical coordinates, herd type and the number of swine of different age groups.

Swine Movement database (SMD)

The SMD was established in 2002 to fulfill the European legislation regarding the trade of swine in European countries. For each movement, the date, the number of swine moved, the vehicle used and the CHR numbers of sender and recipient farms are registered.

Specific Pathogen Free System (SPF System)

The SPF system defines a fixed set of rules for biosecurity, surveillance and the movement of swine between herds (Specific Pathogen Free System, n.d.). The health status is defined based on regular laboratory diagnostic results and clinical visits performed according to SPF rules. Diseases that are monitored within the SPF (known as SPF diseases) are: enzootic pneumonia, pleuropneumonia, atrophic rhinitis, dysentery, PRRS, mange and lice.

The frequency of visits and serological testing depends on the health status of the herd within the SPF system. The majority of Danish breeding herds (including nucleus and multiplier herds) is monitored on a monthly basis and has the red SPF status. Surveillance of SPF diseases that require serology testing is performed annually for production herds (including farrow-to-finisher and finisher herds), and these herds have the blue SPF status. SPF herds represent about 40% of all Danish swine herds, including 99% of Danish breeding and multiplier herds and 35% of Danish finisher herds.

Diagnostic laboratory data for PRRS

The decision was made to focus on laboratory serology results for PRRS from Danish swine herds, due to the large amount of diagnostic data stored in the DTU-Vet lab database – mainly as part the of surveillance program of SPF herds. Additionally, the importance of PRRS in the Danish pig industry was a determining factor in choosing this disease as an example.

Currently, PRRS is one of the biggest challenges for Danish swine producers (SEGES Pig Research Centre, 2015). Although control efforts are in place, PRRS continues to result in economic losses due to mortality in piglets, respiratory problems in growers and finishers, and reproductive problems in sows. It is estimated that an infection in a PRRS-negative herd costs DKK 200-500 per sow/year (SEGES Pig Research Centre, 2015). In addition, this disease is increasingly seen as an obstacle to the export of pork to several countries (SEGES Pig Research Centre, 2015).

The PRRS surveillance program is primarily based on serological testing for herds with a Specific Pathogen Free system certificate, known as SPF herds (Specific Pathogen Free System, n.d.). The frequency of testing depends on the health status of the herd within the SPF system. For non-SPF herds, veterinarians can decide whether or not to test for PRRS, and at what intervals. The decision will depend on the objective (i.e. as part of an eradication and control program, or for the diagnosis of suspected cases herds free from disease) and the costs associated with the different serology tests.

The monthly herd-level PRRS status was used to identify temporal and spatial clustering between 2007 and 2010 (Manuscript 2). The analysis was performed for both PRRSV type 1 and type 2, previously known as European (PRRS EU) and North American (PRRS US) strains, respectively (Murtaugh et al., 1995). Individual blood samples were tested for one or both PRRSV types, based on an in-house blocking enzyme-linked immunosorbent assay (ELISA) (Sørensen et al., 1998) and an immunoperoxidase monolayer assay (IPMA) (Bøtner et al., 1994). The PRRSV status of the herd was defined based on the cut-off for individual blood tests, and the herd-level cut-offs that establish the proportion of PRRSV sero-positive samples (i.e. animals) within the herd, as suggested by Mortensen et al. (2000).

Subsequently, only laboratory submissions where at least two individual blood samples underwent serological testing were included in the analyses for Manuscripts 4, 5 and 6. These serological tests included the DTU-Vet lab “in-house” ELISA (Sørensen et. al, 1997), the DTU-Vet lab “in-house” IPMA (Bøtner et al., 1994) and the IDEXX PRRS X3 Ab ELISA test (IDEXX, Ludwigsburg, Germany) used at the VSP-SEGES lab. Herds were classified as PRRS sero-positive when at least two individual blood samples in each submission tested PRRS positive (without distinguishing between the PRRS strains). The weekly between-herd sero-prevalence was calculated as the proportion of PRRS-positive herds from the total number of herds tested. It was used to define the initial sero-prevalence level in the simulation studies in Manuscripts 4, 5 and 6.

Mortality data

The swine mortality data are owned by the Danish Veterinary and Food Administration and are calculated monthly for all Danish swine herds, based on data registered in the CHR and the SMD. The information retrieved from the CHR is used as a proxy for the number of animals in a given farm every month. The SMD includes information on all movements of Danish swine herds, including movements to rendering plants. The number of dead sows and finishers transported from farms to rendering plants is registered in the database, and the number of containers in which dead weaners (up to 30 kg) are transported from farms to rendering plants is used as a proxy for the number of dead weaners per month.

Data from December 2013 to October 2015 were provided by the Danish Veterinary and Food Administration and analyzed in Manuscript 3. The information registered in the CHR database was used to identify production herds, in order to restrict the analysis to only these herds. Furthermore, only herds with ≥200 finishers, ≥50 sows or ≥200 weaners were included in the analysis. In order to ensure that the study included only active farms, herds with no dead animals over 12 consecutive months for sows and finishers and 2 consecutive months for weaners were excluded. The data were divided according to three age groups –weaners, sows and finishers – and the mortality was analyzed separately for the three age groups.

2.1.3. Simulated data

Since no information is available on what extend PRRS has been spreading and if control and eradication programmes have been implemented in Denmark during the past years, it was decided to perform several simulation studies in order to represent potential scenarios of disease spread and control programs. This was done to evaluate the performance of different temporal monitoring methods when applied to the context of endemic diseases.

The simulated PRRS sero-prevalence was defined based on the same method as described in Manuscript 4, where the number of positive herds (X) for a given week was derived from a binomial distribution ($X \sim bin(n, p)$) with a probability p and a sample size n corresponding to the number of Danish swine herds tested for PRRS

per week (Figure 1). A detailed description of the different representative scenarios of endemic disease spread, as well as eradication and control programs can be found in Manuscripts 4, 5 and 6.

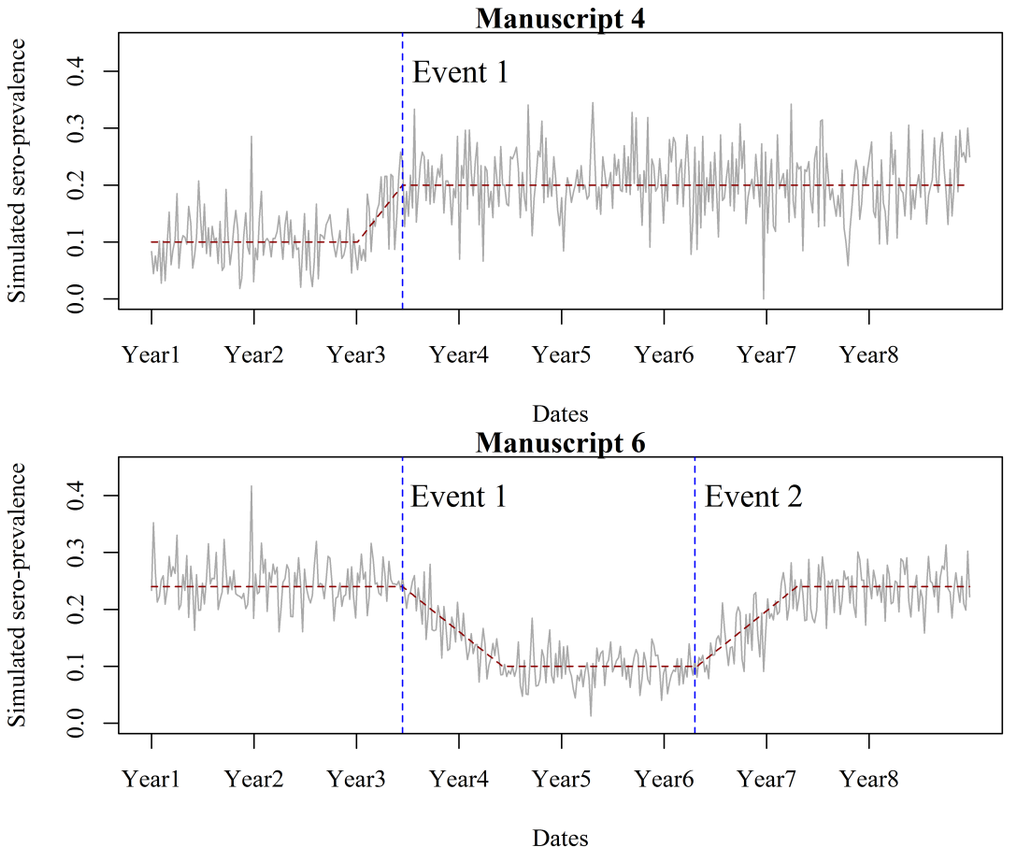


Figure 1- Example of the simulated scenarios representing changes in sero-prevalence for endemic diseases and control programs.

2.2 Methods

2.2.1. Interviews and data quality attributes

A total of 20 interviews were conducted between November 2015 and January 2016 with different stakeholders, i.e. those using and maintaining the data. This information was combined with the researchers' data-related experiences, in order to assess the data quality of seven Danish databases for monitoring swine diseases (Manuscript 1). The assessment was based on a set of qualitative data quality attributes adapted from the European Centre for Disease Prevention and Control (ECDC) guidelines for the evaluation of monitoring and surveillance systems (ECDC, 2014). The interviewees were selected based on their level of experience and involvement with the databases. A detailed description of the data quality attributes can be found in Manuscript 1 and the questionnaire can be found in the Supplementary Materials section.

2.2.2. Spatial and spatio-temporal analysis

Spatial variation

The spatial distribution of PRRSV type 1 and 2 sero-positive herds was assessed based on relative risk maps (Manuscript 2). The kernel smoothing surfaces techniques described by Berman and Diggle (1989) were used.

The Inverse Distance Weighted (IDW) interpolation technique (Huisman and de By, 2009) was used in Manuscript 3 to facilitate visualization of the spatial distribution of the monthly mortality rate for weaners, sows and finishers in Denmark.

Spatial and spatio-temporal clustering

In Manuscript 2, Retrospective Space Scan Statistics (Kulldorff, 1997) were used to identify local spatial clusters of herds positive for PRRSV type 1 and type 2 between 2007 and 2010. The data were aggregated biannually and a Bernoulli model was used, in which herds positive for PRRSV type 1 and type 2 were defined as

cases and negative herds as controls. The scanning spatial window was circular and the analysis was run using different percentages of the population at risk.

The Retrospective Scan Statistic was used to detect spatial, temporal and spatio-temporal clusters of mortality in different age groups within the Danish swine herds. Monthly data from December 2013 to October 2015 were used (Manuscript 3). The Bernoulli model was applied because the number of dead animals (cases) and number of live animals (controls) were available for each herd. In this case, the scanning spatio-temporal window was circular and the analysis was run using 50%, 25% and 10% of the population at risk. The maximum temporal size of the spatio-temporal clusters was defined as 90% of the period of the study.

2.2.3. Time-series modelling

Two space-state models – a Dynamic Linear Model (DLM) based on a normal distribution and a Dynamic Generalized Linear Model (DGLM) based on binomial distributions, both with a linear growth component as described previously (West and Harrison, 1997) – were used to model the simulated PRRSV sero-prevalence data (Manuscripts 5 and 6). Briefly, these models estimate the underlying parameter vector from the observed data combined with any prior information available. The estimated value is updated each time a new value of sero-prevalence is available.

The DLM is represented as a set of two equations, defined as the observation equation (Eq. 1) and the system equation (Eq. 2):

$$Y_t = \boldsymbol{F}_t'\theta_t + v_t,\ v_t \sim N(0, V_t) \qquad \text{(Eq. 1)}$$

$$\theta_t = \boldsymbol{G}_t\theta_{t-1} + w_t,\ w_t \sim N(0, W_t) \qquad \text{(Eq. 2)}$$

where Eq. 1 describes the dependence of observation Y_t (i.e. PRRSV sero-prevalence) on an unobservable parameter θ (which was designed as a matrix) for time *t*, based on a linear function. The observational variance (V_t) was adjusted according to the number of herds tested for PRRS in a given week. Eq. 2 describes the dynamic properties of the unobservable parameter θ used to estimate the underlying values of PRRSV sero-prevalence according to Eq 1. The variance-

covariance matrix (W_t), which describes the evolution of variance and covariance of each parameter for each time step, was modeled using a discount factor.

The observation equation (Eq. 1) for the DGLM is described as:

$$p_t = F_t'\theta_t \qquad \text{(Eq. 3)}$$

and the system equation (Eq. 2) is identical for the GDLM and DLM.

More details of the estimates and a description of the models can be found in Manuscripts 4 and 5.

2.2.4. Temporal monitoring methods

Univariate process control algorithms (UPCA)

Both Manuscripts 4 and 6 explored the performance of several UPCA for monitoring endemic diseases.

In Manuscript 4, three univariate process control algorithms were used: Exponentially Weighted Moving Average (Wagner et al., 2006), Cumulative Sums (Wagner et al., 2006) and Shewhart *p* Chart (Montgomery, 2009). These methods were applied directly to the simulated weekly sero-prevalence data with simulated changes representing endemic disease spread. Alarms were generated when the observed sero-prevalence for a given week exceeded the thresholds of the algorithms. A detailed description of these algorithms is presented in Manuscript 4.

In Manuscript 6, three UPCA were used to generate alarms: Shewhart control chart (Montgomery, 2009), Tabular Cumulative Sums (Montgomery, 2009), and the V-Mask (Montgomery, 2009). The Shewhart control chart and the Tabular Cumulative Sums were applied to the normalized forecast errors obtained from the DLM and GDLM models, whereas the V-mask was applied to simple cumulative sums of the normalized forecast errors. A full description of the methods is presented in Manuscript 6.

Trend-based monitoring methods

In Manuscripts 5 and 6, the time-series was decomposed using the DLM and DGLM models. For each time step *t*, the trend-component was obtained from m_t. Alarms were generated based on the trend-component when significant differences

above and below zero were found according to 95% and 99% CI. Alarms were also generated when the absolute values of the trend-component changed the sign from positive to negative, and vice-versa (Trend Sign).

Calibration of the UPCA

The UPCA were calibrated to a false alarm rate of 1% when applied to a constant level of PRRS sero-prevalence (Manuscripts 4 and 6). Different parameters were tested for each algorithm as part of this process. This decision was made to compare the performance of the different methods, to maintain confidence in the system and to reduce the economic impact of investigating false alarms.

Performance assessment

The performance of the different temporal monitoring methods was evaluated using the cumulative sensitivity (CumSe) for week *i*, following a simulated change in PRRS sero-prevalence (Manuscripts 4 and 6). The CumSe was defined as:

$$CumSe_i = \frac{\sum_{j=1}^{i} x_j}{n.iter} \qquad \text{(Eq. 4)}$$

where x_j is the number of iterations in which an alarm was given j weeks after a change was initiated, and *n.iter* corresponds to the total number of iterations used. The sero-prevalence was considered to have increased if an alarm was generated for $i \geq 0$. This criterion was developed in order to assess the performance of the algorithms during weeks with gradual simulated changes, and during subsequent weeks with constant levels.

Assessing the impact of the sample size on the performance of temporal monitoring methods

In order to assess the impact of the representativeness of the weekly number of herds tested for diseases monitoring, the simulation study was also performed with *n* equal to 10 and 100 times the actual number of herds tested in a given week (Manuscript 4) or with a fixed number of herds tested (n=600) (Manuscript 6).

Figure 2- Collection of blood samples from a gilt in a red SPF herd for disease monitoring and surveillance. Photo: Ana Carolina Lopes Antunes

3. Results

3.1 Data quality of existing Danish swine health databases

The following paragraphs present a brief description of the databases on swine health that were evaluated in Manuscript 1, but were not used (or described) in the statistical analyses in this thesis.

The SMD is part of the CHR, and registers all movements of swine in Denmark at different levels (e.g. farms, rendering plants, slaughterhouses). The database registers information about the sender, recipient, date and time of movement, number of animals moved, and the registration number of the vehicle used. The SMD is owned by the Danish Veterinary and Food Administration (DVFA) and was established in 2002 to increase traceability.

The VetStat is the national database for registration of all prescription-only drugs used in production animals. Data include information such as the date of prescription, prescribing veterinarian, recipient farm, species, age group, and clinical indications. The database is owned by the DVFA and was established in 2000 for research purposes, to control antimicrobial usage, and to assist veterinary practitioners.

The Meat Inspection database includes meat inspection information collected in slaughterhouses. For each animal carcass, the originating farm number, slaughterhouse ID, date of slaughter, fat and meat percentage, and veterinary remarks are registered. The Meat Inspection database was implemented in 1964 with the objective of paying farmers according to the number of animals slaughtered and the meat inspection remarks, and increasing food safety, animal health, and welfare.

Further details about the evaluated databases, including the data flow, can be found in Manuscript 1. Table 1 summarizes the advantages and disadvantages (in terms of disease surveillance) of data from the seven databases, based on the ECDC guidelines. A full description of all data attributes can be found in Manuscript 1.

Table 1- Description of the advantages and disadvantages of using swine register data from several databases for monitoring diseases in Denmark.

Database	Advantages	Disadvantages
CHR	Mandatory for all geographic locations with swine. The data entered are checked and instantly become available on the database.	Irregular updates: minimum once/twice yearly. Information is updated by the farmer, creating bias in the herd type and number of animals from different age groups.
SMD	Mandatory registration of all swine movements. The data are checked when entered in the system.	Information can be registered up to 7 days after the movement, and is often registered late after this period. Based on data from 2014, 10% of swine farms did not have any registered movements.
VetStat	The system generates a warning if data are missing from a specific pharmacy. DVFA perform periodic, retrospective manual checks of the data. All herds using prescription-only drugs are present in Vetstat.	Antimicrobial prescription is influenced by many factors, not just the occurrence of diseases. Data is available no later than the 10th day of any given month. Only the prescription date is registered.
SPF	Gathers information on the health status of 99% of all breeding herds and 78% of sow herds. Warning messages are generated if the disease status for a given herd has changed based on laboratory test results. SPF herds are monitored on a regular basis.	SPF status based on serology tests might not indicate the presence of an infectious disease. Disease status is only given for a limited number of diseases.
Laboratory databases	Diagnostic test results give the disease status of each herd. Standard operating procedures are used to validate the data.	Requests for diagnostic tests are influenced by the occurrence of diseases, the value of the animal, and disease eradication and control programs. The frequency of testing differs between herds.
Meat Inspection database	Includes information on 98% of all finisher herds in Denmark.	It is only possible to monitor diseases and syndromes that cause macroscopic disease lesions. Data are entered on different terminals, with different configurations. Data entered on the abattoir terminal are not checked.

3.2 Temporal and spatio-temporal patterns of PRRS and mortality in Denmark

3.2.1. Temporal trends of PRRS based on laboratory diagnostic data

On average, 230 breeding herds and 2,776 production herds were tested for PRRS every year between 2007 and 2010 (Manuscript 2). Regarding the average time between consecutive submissions, the breeding herds were tested every month (min=1, max=37), whereas the production herds were tested every 11.33 months (min=1, max=46).

Figure 3- Number of swine herds tested for PRRS per month between 2007 and 2010 in Denmark (a), and the corresponding PRRS EU and PRRS US sero-prevalence (b).

Figure 3 shows the total number of herds tested for PRRS per month, as well as the monthly sero-prevalence for both PRRS strains (Manuscript 1). The total number of breeding herds tested for PRRS seems to be constant (Figure 3a). In contrast, the

number of production herds tested for PRRS followed a seasonal trend, with the lowest values in February and August of each year. The apparent PRRS sero-prevalence was constantly higher for the EU strain than the US strain for both types of herds (Figure 3b). A full description of the spatial distribution of both PRRS strains based on relative risk maps can be found in Manuscript 2.

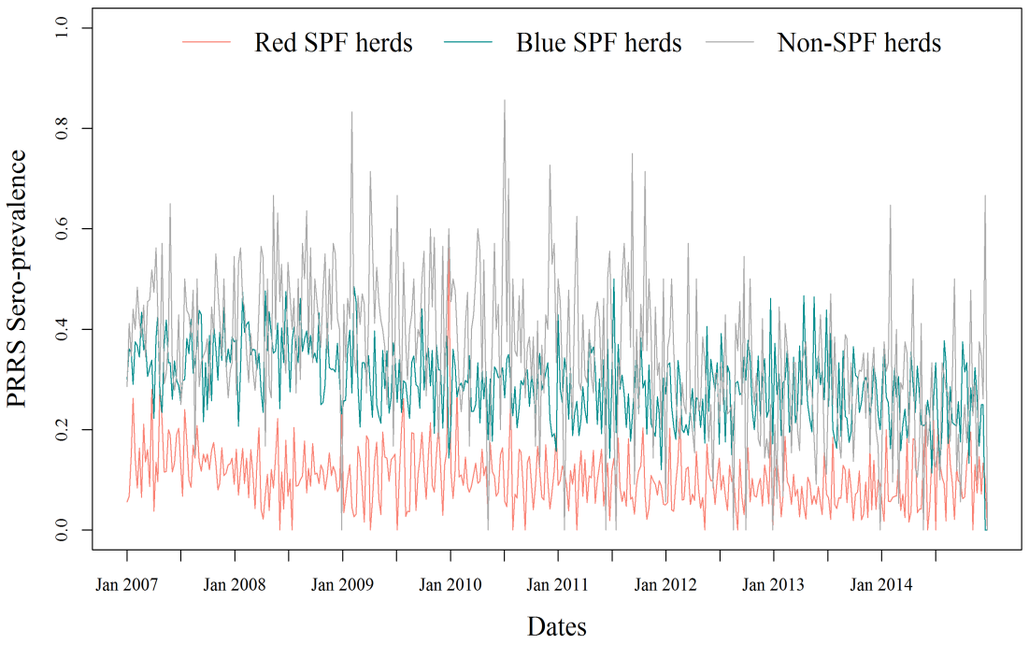


Figure 4- Weekly PRRS sero-prevalence in SPF and non-SPF herds.

The weekly apparent PRRS sero-prevalence was higher for non-SPF herds than for SPF herds (blue and red) between 2007 and 2014. The median apparent PRRS sero-prevalence was 0.10 (min=0.00, max=0.57) for red SPF herds, 0.30 (min=0.00, max=0.50) for blue SPF herds, and 0.35 (min=0.00, max=0.86) for non-SPF herds (Figure 4).

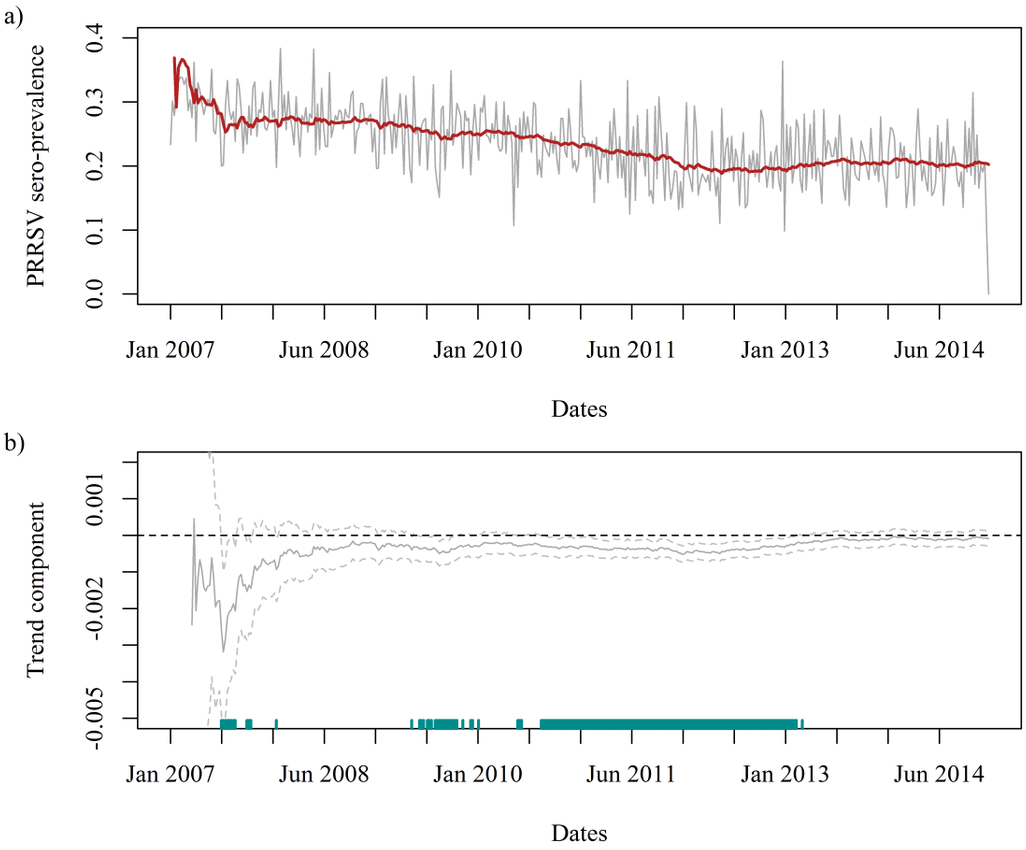


Figure 5- Weekly sero-prevalence trend in Danish swine herds from 2007 to 2014: a) A DGLM was used to model the data and the filtered mean (red); b) the corresponding trend-component was used to monitor significant decreases (based on 95% CI) from zero (blue rugs).

The weekly average PRRS sero-prevalence was 0.24, with a decrease from 0.28 in 2007 to 0.20 in 2014 (Manuscript 5). Monitoring the trend component also showed a decrease throughout this period, and significant decreases (i.e. negative growth) were detected at the end of 2007, end of 2008, early 2010, and between end of 2010 and the beginning of 2013 (Figure 5).

3.2.2. Spatio-temporal mortality trends in Danish Swine herds

Spatial and spatio-temporal patterns of mortality in 1,896 weaner herds, 1,490 sow herds and 3,839 finisher herds (from a total of 5,016 Danish swine farms) were explored in Manuscript 3. The location of the herds is shown in Figure 6a).

A detailed description of temporal and spatial patterns in the monthly mortality can be found in Manuscript 3. In summary, results showed an increase in mortality in January and July of 2014 and 2015 in all regions of the country for the three age groups. The mortality proportion in weaner herds was twice that observed in sow and finisher herds. The spatial patterns showed that a higher mortality rate was found in different areas for each age group: Southern Denmark had a higher mortality rate for sows, Zealand had a higher mortality rate for weaners, and Central Jutland had a higher mortality rate for finishers.

A full description of the spatio-temporal clusters is provided in Manuscript 3. A summary of the number of clusters found for the three age groups is presented in Table 2 and the locations are shown in Figure 6b).

Table 2- Frequency of clusters for different age groups of Danish swine herds found between December 2013 and October 2015.

Age group	Total	Cluster type	
		Single-herd cluster[1]	*Multiple-herd cluster*[2]
Sow herds	7	5	2
Weaner herds	68	57	11
Finisher herds	76	49	27

[1] Refers to clusters including only one herd.
[2] Refers to clusters including more than one herd.

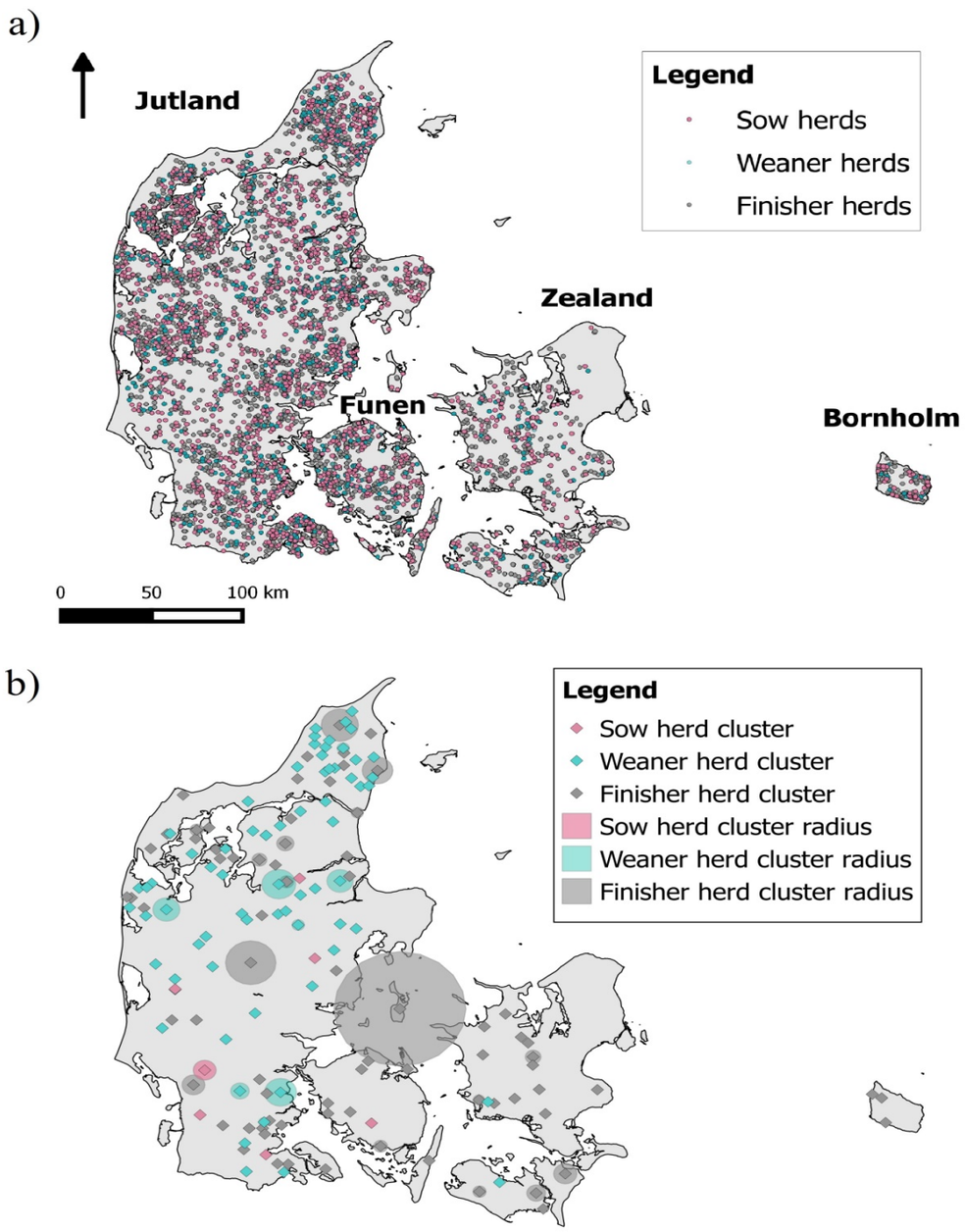


Figure 6- Map of the herds included in the mortality analysis from December 2013 to October 2015 (a), and the location of the spatio-temporal clusters found (b). The circles represent the radius of the clusters.

3.3 Performance of different temporal monitoring methods when applied in the context of endemic diseases

The results from Manuscript 4 showed that the Exponential Weighted Moving Average and Shewhart *p* Chart had higher CumSe than the Cumulative Sums when applied directly to the sero-prevalence timeline for all simulated scenarios.

Based on the DLM model, the Shewhart Control Chart performed better in detecting increases in sero-prevalence compared to decreases, while the opposite was seen for the Tabular Cumulative Sums (Manuscript 6). The trend-based methods performed well in detecting the first simulated events (increases and decreases in sero-prevalence), but performance was poor in detecting consecutive events. The method that seemed to perform most consistently was the V-mask.

When comparing the performance of the different UPCA based on DLM and GDLM, the results revealed that in general the temporal monitoring methods needed more time to achieve CumSe=0.5 based on the GDLM. The trend-based methods had a similar performance with both models.

3.4 Impact of sample size on the performance of the temporal monitoring methods

Figure 7 shows the impact of sample size on the performance of the UPCA when applied directly to the time-series, as described in Manuscript 4. For increases in sero-prevalence from 0.10 to 0.20 over 24 weeks, the number of weeks to achieve a CumSe=1.0 was halved when the sample size was increased 10-fold; increasing the sample size 100-fold resulted in CumSe=1.0 being achieved six times faster.

In Manuscript 6, reducing the noise in the data (by simulating 600 herds tested per week) resulted in achieving faster CumSe=1 for the Shewhart Control Chart and Tabular Cumulative Sums. No substantial impact was observed for the V-Mask and both trend-based methods.

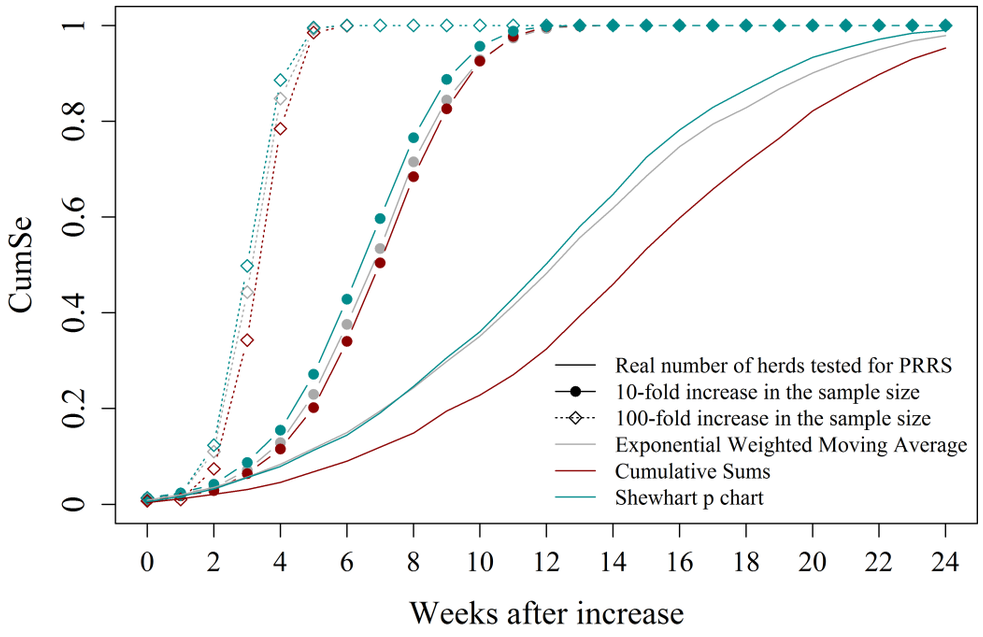


Figure 7- Impact of sample size on the cumulative sensitivity (CumSe) of the univariate process control algorithms. The results are based on an increase in sero-prevalence from 0.10 to 0.20 over 24 weeks.

4. Discussion

The overall aim of this thesis was to explore the potential uses and limitations of existing veterinary health data for monitoring swine diseases in Denmark. To fulfill this aim, the quality of data from different databases was evaluated, temporal and spatial patterns found in the data were described, and the performance of different temporal monitoring methods (in the context of endemic diseases and control programs) was assessed. This section presents some of the key findings. A more extensive discussion of each topic can be found in the corresponding manuscripts.

4.1. Potential use of current Danish databases for monitoring swine diseases

Manuscript 1 describes seven Danish swine databases and assesses stakeholders' perceptions regarding the data quality, as well as potential uses and limitations for disease monitoring and surveillance. The different databases are useful to varying degrees when it comes to monitoring swine diseases in Denmark. They all contain information that can be cautiously used in different steps, including using statistical methods for monitoring changes in data records and contingency plans for diseases. The findings from Manuscript 1 suggest that the laboratory data, VetStat and the Meat Inspection database can be used for disease monitoring, while the CHR and the SMD contain information crucial to performing risk-based surveillance and contingency planning in case of a disease outbreak.

The databases presented various disadvantages relating to different quality attributes. For example, it could take several days or weeks before the data entered into the VetStat and Meat Inspection databases became available. The time required between data entry and availability in the database (timeliness) was a limitation of using data from these databases as a (near) real-time disease-monitoring tool, since the data are not instantly available. In addition, there are limitations to using the information in VetStat and the Meat Inspection database as a proxy for disease occurrence. For instance, the data in VetStat do not provide an indication of the disease for which the antimicrobial treatment was prescribed, and there are no indications of when the treatment was applied. Also, the presence of macroscopic lesions at early stages of disease might be difficult to identify at the abattoir.

Another important data attribute for disease surveillance is how well the data represents the population over time (i.e. representativeness). The frequency of testing depends on the surveillance, eradication and control programs implemented at farm level, the value of the animals, and the farmers' awareness of disease occurrence, as only a small proportion of Danish swine herds (mainly red SPF herds) are tested weekly. The same issue is true of the Meat inspection data, where information is more representative of finisher herds.

The validity of the databases in Denmark has been improved by merging data from different databases when financial and legal consequences exist. For example, the Yellow Card legislation (Ministry of Food, 2010) monitors antimicrobial consumption based on VetStat and CHR data. Farmers and veterinarians are aware that incorrect information in these databases might push the antimicrobial level above the threshold value, leading to restrictions being imposed on the farm. Other examples include potential commercial restrictions for farmers when disease-status data are not accurate in the laboratory and SPF databases. Each farm pays for laboratory testing and SPF accreditation on a voluntary basis, with any "incorrect information" in the databases resulting in legal consequences for the laboratory and accreditation institutions. For this reason, the data are validated when they are entered, and warning messages are sent in the case of any system fails.

4.2. Spatial, temporal and spatio-temporal trends and clusters

4.2.1. PRRS sero-prevalence based on laboratory data

The overall occurrence of PRRS temporal and spatial patterns based on laboratory data was described in Manuscript 2. As previously discussed (Manuscript 1), the frequency of testing depends on the herd status (SPF herd or non-SPF herd) as well as the reason for testing (e.g. PRRS monitoring and surveillance, diagnosis of new infected herds). It is generally assumed that herds tested for PRRS have a higher health status (e.g. red SPF herds) than herds that do not submit samples (personal communication, C.S. Kristensen, 2014). This can be a limitation when monitoring temporal trends in PRRS sero-prevalence in Danish Swine herds based on laboratory data, and the same analogy can be made for other diseases.

4.2.2. Mortality data

A large number of *single-herd* and *multiple-herd* clusters were found (Manuscript 3). The higher mortality within these clusters might indicate potential welfare issues (SEGES Pig Research Centre, 2016, 2015) or the presence of infectious diseases such as PRRS (Mortensen et al., 2002) or Swine Influenza (Brown, 2000).

Before using mortality as a proxy for disease occurrence, it is important to emphasize that the mortality was calculated based on data from two different databases. The information on the number of swine in each age group in the CHR database was used as a proxy for the number of swine present in a herd for a given month. As described in Manuscript 1, there are infrequent updates of the CHR, which can potentially result in biased information on the total number of swine in the farm for each month. The movements registered in the SMD are used as a proxy for the number of dead animals per age group. The registers for weaners (up to 30 kg) are based on the number of containers (with specific dimensions) transported from a farm to the rendering plant. The number of dead weaners that can fit inside a container varies according to the weight of the animals. In addition, the total volume of dead animals placed inside the container might not be representative of the total volume of the container. This can result in a bias in the number of dead weaners presented in the database, and illustrates the challenges of using mortality data for disease monitoring purposes in Denmark, as discussed in Manuscript 1 for the CHR and SMD. As a consequence, the estimated mortality can be biased, leading to under- or overestimations, and precautions should be taken when using these data.

4.3. Temporal monitoring methods

The results in Manuscript 4 showed that the Exponentially Weighted Moving Average and Shewhart *p* Chart had similar results when detecting increases in sero-prevalence, and that their performance in this respect was better than that of the Cumulative Sums when applied directly to the time-series. One possible explanation for the poorer performance of the Cumulative Sums is that the variation in simulated sero-prevalence might have resulted in a negative cumulative sum, which resets the Cumulative Sums to zero, as verified by Dórea et al. (2013).

The aim of the simulated scenarios in Manuscript 6 was to represent relevant changes in disease occurrence for endemic diseases. Each scenario was simulated with an initial constant level of sero-prevalence followed by an increase or decrease in two different events. Both DLM and GDLM models were optimized to model a constant level, resulting in slower model-trend changes in Event 2. As a consequence, the normalized forecast errors were higher and the Tabular Cumulative Sums generated alarms earlier than the Shewhart control chart 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, which might have resulted in a higher number of alarms. This can explain why the Tabular Cumulative Sums in Manuscript 6 performed better compared to other methods for monitoring increases in sero-prevalence, in contrast to the Cumulative Sums (a similar detection method) used in Manuscript 4.

The V-Mask showed the most consistent results in relation to the number of weeks required to achieve a CumSe=50%, due to the greater flexibility in defining the control limits.

4.4. Impact of sample size on the performance of temporal monitoring methods

Increasing the number of herds that were tested reduced the noise in the simulated sero-prevalence (Manuscripts 4). As a result, the Cumulative Sums were not reset to zero, resulting in an increase in the sensitivity to detect changes to a level equal to that of other methods when 10-fold and 100-fold increases were simulated.

In Manuscript 6, decreasing the noise in the simulated sero-prevalence also resulted in higher CumSe for the Shewhart Control Chart and Tabular Cumulative Sums. Conversely, this had no impact on the V-Mask or the trend-based methods (Manuscript 6). This demonstrates the importance of choosing a suitable temporal monitoring method. The Shewhart Control Chart and Tabular Cumulative Sums techniques used in Manuscript 6 were sensitive to the intensity of noise in the data, regardless of whether they were applied to forecast errors or directly to the data. Using trend-based methods offers a way to monitor the underlying trend, usually masked by random noise in the data.

5. Conclusions

This thesis explores the potential uses and limitations of existing databases containing swine-health data in Denmark. The goals included assessing current Danish databases relevant to swine health, and their potential use in disease monitoring and surveillance. An additional aim was to examine the feasibility of studying temporal and spatial changes in data records in order to detect changes that might indicate disease spread between swine herds, and to explore the potential use of different temporal monitoring methods for monitoring endemic disease control programs. In summary, the conclusions are:

- The current databases cover different aspects of disease surveillance, including monitoring (near) real-time infectious disease status and data to support contingency planning in case of a disease outbreak. However, the limitations (such as infrequent updates, incomplete representation of Danish swine herds and delays in registering new data in databases) should be addressed in order to improve the quality of data from multiple databases for monitoring diseases in Denmark.
- The laboratory data are useful for monitoring endemic diseases. The frequency of testing depends on factors such as the SPF status, farmer compliance, the value of the animal, commercial purposes and ongoing control and eradication programs. For example, a large percentage of the laboratory diagnostic testing for PRRS is performed for red SPF herds, which might result in an underestimation of the overall disease prevalence in Denmark, as non-SPF herds are not regularly tested. This limitation did not apply to mortality data, which are available for all Danish swine herds. However, observed changes might be due to disease occurrence, or as a result of changes in herd management or the way mortality is calculated. Moreover, the data are only available on a monthly basis, which is a limitation for (near) real-time disease monitoring.
- The performance of the different temporal monitoring methods in detecting changes in sero-prevalence for endemic diseases and control programs varied.

Therefore, choosing a single temporal monitoring method is challenging, as the objectives of the monitoring program and the differing performance of the explored methods in detecting a specific pattern should be taken into account. Increasing the sample size (i.e. the number of tested herds) resulted in faster detection for the majority of UPCA, while the impact was not noticeable for the V-Mask or trend-based methods. This indicates that the V-Mask and trend-based methods provide a more consistent approach to monitoring changes in disease sero-prevalence.

- Finally, the available databases are potentially useful in disease monitoring and surveillance of swine herds in Denmark, but improvements are needed for their accurate, real-time implementation.

6. Perspectives

The findings of this thesis may serve as a basis for improving swine disease monitoring and surveillance in Denmark. The swine-health databases currently available in Demark and the temporal monitoring methods described in this thesis can be used for monitoring temporal and spatial changes in infectious diseases and for contingency planning in case of disease spread. However, due to previously discussed limitations, real-time implementation of disease monitoring and surveillance will require improvements to the databases, and their full potential should be explored through further research, as described below:

- To explore alternative data collection methods, such as smartphone apps (e.g. FARMlog, http://farmlogsvin.dk/), in order to improve the representativeness of the data and increase the frequency of updates. The information gathered using smartphones could complement the existing data in current databases, for example more accurate estimates of the number of animals present at a farm, the number of dead animals, the date of antimicrobial usage, or the occurrence of clinical symptoms.
- To combine data from different databases (such as meat inspection or antimicrobial consumption data) using a multivariate surveillance approach, in which several processes are analysed in parallel or combined. This approach is yet to be applied to monitoring diseases in veterinary science.
- To assess the performance of described trend-based methods for detecting outbreaks of (re-)emerging diseases.

To evaluate the performance of different spatio-temporal methods when applied to the context of endemic diseases and control programs.

Bibliography

Alba, A., Dórea, F.C., Arinero, L., Sanchez, J., Cordón, R., Puig, P., Revie, C.W., 2015. Exploring the Surveillance Potential of Mortality Data: Nine Years of Bovine Fallen Stock Data Collected in Catalonia (Spain). PLoS One 10, e0122547. doi:10.1371/journal.pone.0122547

Beer, M., Conraths, F.J., van der Poel, W.H.M., 2013. "Schmallenberg virus"-a novel orthobunyavirus emerging in Europe. Epidemiol. Infect. 141, 1–8. doi:10.1017/S0950268812002245

Berman, M., Diggle, P., 1989. Estimating weighted integrals of the second-order intensity of a spatial point process. J. R. Stat. Soc. 51, 81–92.

Bouma, A., Elbers, A.R.W., Dekker, A., De Koeijer, A., Bartels, C., Vellema, P., Van Der Wal, P., Van Rooij, E.M.A., Pluimers, F.H., de Jong, M.C.M., 2003. The foot-and-mouth disease epidemic in The Netherlands in 2001. Prev. Vet. Med. 57, 155–166. doi:10.1016/S0167-5877(02)00217-9

Brown, I.H., 2000. The epidemiology and evolution of influenza viruses in pigs. Vet. Microbiol. 74, 29–46.

Buckeridge, D.L., 2007. Outbreak detection through automated surveillance: A review of the determinants of detection. J. Biomed. Inform. 40, 370–379. doi:10.1016/j.jbi.2006.09.003

Buckeridge, D.L., Switzer, P., Owens, D., Siegrist, D., Pavlin, J., Musen, M., 2005. An evaluation model for syndromic surveillance: assessing the performance of a temporal algorithm. MMWR. Morb. Mortal. Wkly. Rep. 54 Suppl, 109–115. doi:su5401a18 [pii]

Bøtner, A., Nielsen, J., Bille-Hansen, V., 1994. Isolation of porcine reproductive and respiratory syndrome (PRRS) virus in a Danish swine herd and experimental infection of pregnant gilts with the virus. Vet. Microbiol. 40, 351–360. doi:10.1016/0378-1135(94)90122-8

Cao, P.-H., Wang, X., Fang, S.-S., Cheng, X.-W., Chan, K.-P., Wang, X.-L., Lu, X.,

Wu, C.-L., Tang, X.-J., Zhang, R.-L., Ma, H.-W., Cheng, J.-Q., Wong, C.-M., Yang, L., 2014. Forecasting influenza epidemics from multi-stream surveillance data in a subtropical city of China. PLoS One 9, e92945. doi:10.1371/journal.pone.0092945

Carslake, D., Grant, W., Green, L.E., Cave, J., Greaves, J., Keeling, M., McEldowney, J., Weldegebriel, H., Medley, G.F., 2011. Endemic cattle diseases: comparative epidemiology and governance. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 366, 1975–86. doi:10.1098/rstb.2010.0396

Coker, R., Rushton, J., Mounier-Jack, S., Karimuribo, E., Lutumba, P., Kambarage, D., Pfeiffer, D.U., Stärk, K., Rweyemamu, M., 2011. Towards a conceptual framework to support one-health research for policy on emerging zoonoses. Lancet Infect. Dis. 11, 326–331. doi:10.1016/S1473-3099(10)70312-1

Cowling, B.J., Wong, I.O.L., Ho, L.-M., Riley, S., Leung, G.M., 2006. Methods for monitoring influenza surveillance data. Int. J. Epidemiol. 35, 1314–1321. doi:10.1093/ije/dyl162

Delooz, L., Saegerman, C., Quinet, C., Petitjean, T., De Regge, N., Cay, B., 2016. Resurgence of Schmallenberg Virus in Belgium after 3 Years of Epidemiological Silence. Transbound. Emerg. Dis. doi:10.1111/tbed.12552

Doherr, M.G., Audigé, L., 2001. Monitoring and surveillance for rare health-related events: a review from the veterinary perspective. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 356, 1097–1106. doi:10.1098/rstb.2001.0898

Dórea, F.C., Lindberg, A., McEwen, B.J., Revie, C.W., Sanchez, J., 2014. Syndromic surveillance using laboratory test requests: A practical guide informed by experience with two systems. Prev. Vet. Med. 116, 313–24. doi:10.1016/j.prevetmed.2014.04.001

Dórea, F.C., McEwen, B.J., McNab, W.B., Sanchez, J., Revie, C.W., 2013. Syndromic Surveillance Using Veterinary Laboratory Data: Algorithm Combination and Customization of Alerts. PLoS One 8, e82183. doi:10.1371/journal.pone.0082183

Dupuy, C., Bronner, A., Watson, E., Wuyckhuise-Sjouke, L., Reist, M., Fouillet, A., Calavas, D., Hendrikx, P., Perrin, J.-B., 2013a. Inventory of veterinary syndromic surveillance initiatives in Europe (Triple-S project): Current situation and perspectives. Prev. Vet. Med. 111, 220–229.

Dupuy, C., Morignat, E., Maugey, X., Vinard, J.-L., Hendrikx, P., Ducrot, C., Calavas, D., Gay, E., 2013b. Defining syndromes using cattle meat inspection data for syndromic surveillance purposes: a statistical approach with the 2005-2010 data from ten French slaughterhouses. BMC Vet. Res. 9, 88. doi:10.1186/1746-6148-9-88

European Centre for Disease Prevention and Control (ECDC), 2014. Data quality monitoring and surveillance system evaluation- A handbook of methods and applications. Stockholm.

Garcia, A.B., Shalloo, L., 2015. Invited review: The economic impact and control of paratuberculosis in cattle. J. Dairy Sci. 98, 5019–39. doi:10.3168/jds.2014-9241

Heffernan, R., Mostashari, F., Das, D., Karpati, A., Kulldorff, M., Weiss, D., 2004. Syndromic Surveillance in Public Health Practice, New York City - Volume 10, Number 5—May 2004 - Emerging Infectious Disease journal - CDC. Emerg. Infect. Dis. J. Volume 5, 858–864.

Huisman, O., de By, R.A., 2009. Principles of Geographic Information Systems - An introductory texbook, 4th ed. The International Institute for Geo-Information Science and Earth Observation (ITC), Enschede, The Netherlands.

Jackson, M.L., Baer, A., Painter, I., Duchin, J., 2007. A simulation study comparing aberration detection algorithms for syndromic surveillance. BMC Med. Inform. Decis. Mak. 7, 6. doi:10.1186/1472-6947-7-6

Jensen, D.B., Cornou, C., Toft, N., Kristensen, A.R., 2015. A multi-dimensional dynamic linear model for monitoring slaughter pig production, in: 7th European Conference on Precision Livestock Farming. Milan, pp. 503–512. doi:10.1017/CBO9781107415324.004

Kreeger, 1991. Ruminant paratuberculosis—a century of progress and frustration. J.

vet 3, 373–381.

Kulldorff, M., 2016. SaTScan Bibliography [WWW Document]. SaTScan Softw. Spat. temporal space-time scan Stat. doi:10.1080/2330443X.2014.928245

Kulldorff, M., 1997. A spatial scan statistic. Commun. Stat. - Theory Methods 26, 1481–1496. doi:10.1080/03610929708831995

Lombardo, J., Burkom, H., Elbert, E., Magruder, S., Lewis, S.H., Loschen, W., Sari, J., Sniegoski, C., Wojcik, R., Pavlin, J., 2003. A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II). J. Urban Health 80, i32-42.

Madsen, T.N., Kristensen, A.R., 2005. A model for monitoring the condition of young pigs by their drinking behaviour. Comput. Electron. Agric. 48, 138–154. doi:10.1016/j.compag.2005.02.014

McLaws, M., Ribble, C., Stephen, C., McNab, B., Barrios, P.R., 2007. Reporting of suspect cases of foot-and-mouth-disease during the 2001 epidemic in the UK, and the herd sensitivity and herd specificity of clinical diagnosis. Prev. Vet. Med. 78, 12–23. doi:10.1016/j.prevetmed.2006.09.001

Ministry of Food, 2010. Exceptional arrangements to decrease the use of antibiotics in pig herds [in Danish].

Montgomery, D., 2009. Introduction to Statistical Quality Control., 6th Ed. ed. John Wiley & Sons, Inc., Arizona State University.

Mortensen, S., Stryhn, H., Søgaard, R., Boklund, A., Stärk, K.D.., Christensen, J., Willeberg, P., 2002. Risk factors for infection of sow herds with porcine reproductive and respiratory syndrome (PRRS) virus. Prev. Vet. Med. 53, 83–101. doi:10.1016/S0167-5877(01)00260-4

Mortesen, S., Strandbygaard, B., Bøtner, A., 2000. Herd-level tests for infection with the European and American type of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) based on serology. Herd Test. Dis. Spread Model. Swine Popul. PhD Thesis. R. Vet. Agric. Univ. Frederiksberg, Denmark. 47–69.

Murtaugh, M.P., Elam, M.R., Kakach, L.T., 1995. Comparison of the structural

protein coding sequences of the VR-2332 and Lelystad virus strains of the PRRS virus. Arch. Virol. 140, 1451–60.

Nieuwenhuis, N., Duinhof, T.F., van Nes, A., 2012. Economic analysis of outbreaks of porcine reproductive and respiratory syndrome virus in nine sow herds. Vet. Rec. 170, 225. doi:10.1136/vr.100101

Oļševskis, E., Guberti, V., Seržants, M., Westergaard, J., Gallardo, C., Rodze, I., Depner, K., 2016. African swine fever virus introduction into the EU in 2014: Experience of Latvia. Res. Vet. Sci. 105, 28–30. doi:10.1016/j.rvsc.2016.01.006

Ostersen, T., Cornou, C., Kristensen, a. R., 2010. Detecting oestrus by monitoring sows' visits to a boar. Comput. Electron. Agric. 74, 51–58. doi:10.1016/j.compag.2010.06.003

Perrin, J., Durand, B., Gay, E., Ducrot, C., Hendrikx, P., Calavas, D., Hénaux, V., 2015. Simulation-Based Evaluation of the Performances of an Algorithm for Detecting Abnormal Disease-Related Features in Cattle Mortality Records Simulation-Based Evaluation of the Performances of an Algorithm for Detecting Abnormal Disease-Related Features in. PLoS One 10. doi:10.1371/journal.pone.0141273

Robertson, C., Nelson, T.A., MacNab, Y.C., Lawson, A.B., 2010. Review of methods for space–time disease surveillance. Spat. Spatiotemporal. Epidemiol. 1, 105–116. doi:10.1016/j.sste.2009.12.001

Rodríguez-Prieto, V., Vicente-Rubiano, M., Sánchez-Matamoros, A., Rubio-Guerri, C., Melero, M., Martínez-López, B., Martínez-Avilés, M., Hoinville, L., Vergne, T., Comin, A., Schauer, B., Dórea, F., Pfeiffer, D.U., Sánchez-Vizcaíno, J.M., 2014. Systematic review of surveillance systems and methods for early detection of exotic, new and re-emerging diseases in animal populations. Epidemiol. Infect. 1–25. doi:10.1017/S095026881400212X

Salman, M., 2003. Animal Disease Surveillance and Survey Systems: Methods and Applications., 1st Editio. ed. Blackwell Publishing, Iowa.

Sánchez-Vizcaíno, J.M., Mur, L., Martínez-López, B., 2013. African swine fever

(ASF): Five years around Europe. Vet. Microbiol. 165, 45–50. doi:10.1016/j.vetmic.2012.11.030

SEGES Pig Research Centre, 2016. Annual Report 2015. SEGES Pig Research Centre, Copenhagen.

SEGES Pig Research Centre, 2015. Annual Report 2014, Annual Report. Copenhagen. doi:87-91460-29-8

Smith, A. F., West, M., 1983. Monitoring renal transplants: an application of the multiprocess Kalman filter. Biometrics 39, 867–878.

Specific Pathogen Free System, n.d. Specific Pathogen Free System [WWW Document]. URL http://spfsus.dk/ (accessed 1.20.16).

Sørensen, K.J., Bøtner, A., Smedegaard Madsen, E., Strandbygaard, B., Nielsen, J., 1997. Evaluation of a blocking Elisa for screening of antibodies against porcine reproductive and respiratory syndrome (PRRS) virus. Vet. Microbiol. 56, 1–8. doi:10.1016/S0378-1135(96)01345-4

Sørensen, K.J., Strandbygaard, B., Bøtner, A., Madsen, E.S., Nielsen, J., Have, P., 1998. Blocking ELISA's for the distinction between antibodies against European and American strains of porcine reproductive and respiratory syndrome virus. Vet. Microbiol. 60, 169–77.

Thysen, I., 1993. Monitoring Bulk Tank Somatic Cell Counts by a Multi-Process Kalman Filter. Acta Agric. Scand. 58–64.

Wagner, M.M., Moore, A.W., Aryel, R.M., 2006. Handbook of Biosurveillance, in: Handbook of Biosurveillance. Elsevier, pp. 217–234. doi:10.1016/B978-012369378-5/50000-4

West, M., Harrison, J., 1997. Bayesian Forecasting and Dynamic Models, 2nd Ed. ed. Springer, New York, USA.

Accompanying manuscripts

Manuscript 1

Evaluation of the perceived utility of information routinely recorded in databases for integrated disease surveillance in swine. Submitted to Acta Veterinaria Scandinavica.

Manuscript 2

Spatial analysis and temporal trends of porcine reproductive and respiratory syndrome in Denmark from 2007 to 2010 based on laboratory submission data. Published in BMC Veterinary Research 2015; 11:303.

Manuscript 3

Mortality in Danish Swine herds: spatio-temporal clusters and risk factors. Submitted to Preventive Veterinary Medicine.

Manuscript 4

Monitoring endemic livestock diseases using laboratory diagnostic data: A simulation study to evaluate the performance of univariate process monitoring control algorithms. Published in Preventive Veterinary Medicine 2016; 127:15–20.

Manuscript 5

Dynamic generalized linear models for monitoring endemic diseases: moving beyond univariate process monitoring control. Published in Proceedings for the Annual Meeting of the Society for Veterinary Epidemiology and Preventive Medicine (SVEPM) 2016, 69-79.

Manuscript 6

A simulation study to evaluate the performance of statistical monitoring methods when applied to different time-series components in the context of control programs for endemic diseases. Submitted to Plos One.

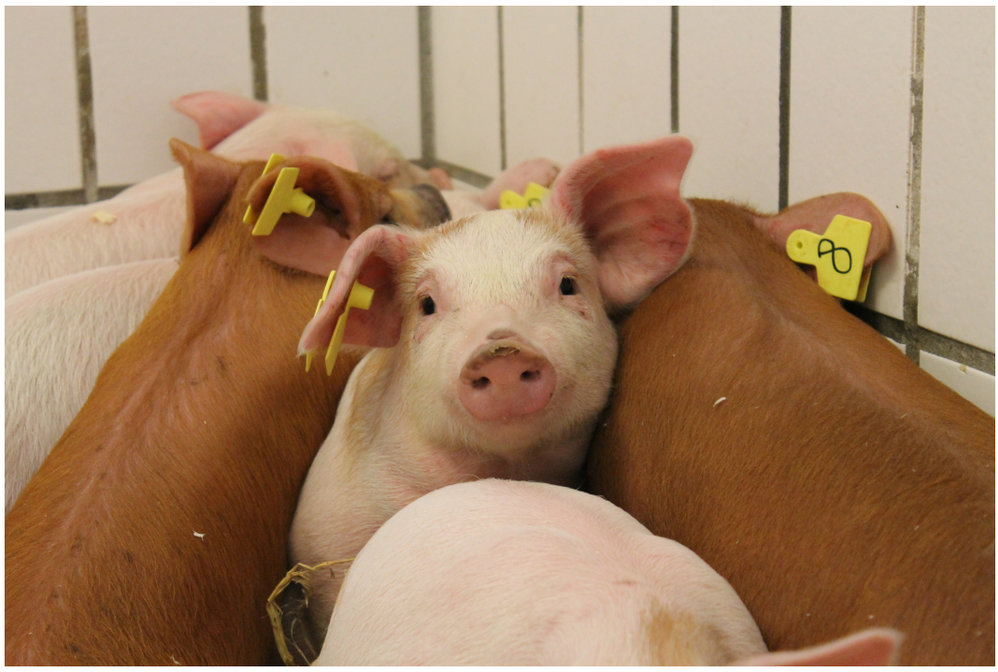


Photo: National Veterinary Institute – Technical University of Denmark.

Manuscript 1

Evaluation of the perceived utility of information routinely recorded in databases for integrated disease monitoring and surveillance in swine

Ana Carolina Lopes Antunes[1 a] *, Mette Fertner[2 a], Anna Camilla Birkegård[1], Tariq Halasa[1], Anette Boklund[1], Nils Toft[1]

[1] Section for Diagnostic and Scientific Advice, National Veterinary Institute – Technical University of Denmark, Denmark

[2] Section for Production and Health, Department of Large Animal Sciences, University of Copenhagen, Denmark.

[a] Both authors contributed equally to the manuscript.

*Corresponding author.

Submitted to Acta Veterinaria Scandinavica.

Evaluation of the perceived utility of information routinely recorded in databases for integrated disease monitoring and surveillance in swine

Ana Carolina Lopes Antunes[*1]
Section for Diagnostic and Scientific Advice, National Veterinary Institute – Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark. Email:aclan@vet.dtu.dk

Mette Fertner[*]
Section for Production and Health, Department of Large Animal Sciences, University of Copenhagen, Grønnegårdsvej 2, 1870 Frederiksberg C, Denmark. Email: mef@sund.ku.dk

Anna Camilla Birkegård
Section for Diagnostic and Scientific Advice, National Veterinary Institute – Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark. Email: acbir@vet.dtu.dk

Tariq Halasa
Section for Diagnostic and Scientific Advice, National Veterinary Institute – Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark. Email: tahbh@vet.dtu.dk

Anette Boklund
Section for Diagnostic and Scientific Advice, National Veterinary Institute – Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark. Email: anebo@vet.dtu.dk

Nils Toft
Section for Diagnostic and Scientific Advice, National Veterinary Institute – Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark. Email: ntoft@vet.dtu.dk

*Both authors contributed equally to the manuscript.
[1]Corresponding author

Abstract

Background: In Denmark, there is an ongoing collection of data regarding the swine population, productivity and health. These data are stored in several public and industry-owned databases. The databases contain information used to facilitate decision making at herd or regional/national level. The aim of this study was to assess the perceived potential of data routinely collected and registered for demographic, traceability, legislative, diagnostic and commercial purposes, as a means for integrated disease surveillance in swine. To meet this aim, the data quality of seven databases was assessed: the Central Husbandry Register (CHR), the Swine Movement Database (SMD), the national Danish database of drugs for veterinary use (VetStat), laboratory data from the National Veterinary Institute – Technical University of Denmark (DTU-Vet lab), diagnostic laboratory at the SEGES Pig Research Center SEGES (SEGES VSP lab), the Specific Pathogen Free System (SPF) and the Meat Inspection database. Furthermore, suggestions for future improvements of data quality and potential combination of databases for monitoring swine diseases were also discussed. **Results:** The extent to which the databases can be used for disease surveillance and monitoring varies. In summary, the surveillance of swine diseases was a primary objective only for the laboratory and SPF databases. There are a number of factors influencing antimicrobial use at herd level, thus questioning the utility of VetStat data for surveillance of disease. Meat Inspection data have the advantage of being recorded at animal level, but sensitivity vary between disease categories and abattoirs. In contrast, the CHR and SMD are concerned only with swine traceability, indicating the population

at risk, or to evaluate the effect of trading patterns. **Conclusions:** The data quality tended to be higher, when the databases are interrelated and are linked to economic interests of the farmer. The usefulness of the different databases covers different aspects of diseases surveillance, including disease monitoring and follow-up. Further research will be needed to address technical and methodological challenges in integrating the data from multiple databases for monitoring diseases in Danish swine.

Keywords

Data quality, health-related data, disease surveillance, swine.

Introduction

Denmark holds a number of national databases in the veterinary field, covering data on herd demographics, veterinary affiliation, animal movements, slaughter remarks, surveillance of zoonotic agents, antimicrobial use and laboratory test results [1]. Due to the availability of data, these data are widely applied for purposes such as research [2–4], legislative actions [5] and disease surveillance [6].

Disease surveillance describes the ongoing process of the assessment of health and disease status of a given population [7]. The ability of automated systems to detect changes in disease occurrence depend to a large extent upon the choice of data source [8]. The data can be associated with analytical and interpretive limitations related to data being representative of the target population [9], technical aspects, political requirements

and stakeholder interests [10], which might influence the quality of data and the acceptance of using it for disease surveillance.

In all studies on data, an evaluation of data quality is crucial [11,12]. Unlike human medicine [9,13], studies on data quality in existing veterinary databases remain scarce.

The aim of the present study was to assess the perceived potential of using data routinely collected and registered in Denmark as source for integrated disease surveillance purposes in swine. This included: i) a description of the data and structure of existing public and private databases in Denmark with swine health related data; ii) a qualitative assessment of stakeholder perception of the data quality and iii) suggestions for future improvements of data quality and potential combination of data for monitoring swine diseases.

Methods

Data quality attributes

In 2014, the European Centre for Disease Prevention and Control (ECDC) published a technical document to support processes for assessing data quality and evaluating surveillance systems for public health in European member states in order to provide accurate and timely information for decision making [9]. This document was created based on input from groups of experts on surveillance system quality combined with review of available literature in order to suggest key quality attributes to evaluate the data quality but also the system structure, i.e. the organization in an operating system.

For the present study, seven Danish swine databases were chosen by the authors and assessed based on a set of qualitative data quality attributes adapted from the ECDC guidelines to evaluate monitoring and surveillance systems [9]. The seven databases were: the Central Husbandry Register (CHR) including the Swine Movement Database (SMD), the national Danish database of drugs for veterinary use (VetStat), laboratory data from the National Veterinary Institute – Technical University of Denmark (DTU-Vet lab) and from the Pig research centre SEGES (SEGES VSP lab), the Specific Pathogen Free System (SPF) and the Meat Inspection database. These databases were selected based on their extensive use by the Danish swine industry and research institutes.

The data quality attributes and the proposed indicators used to evaluate the databases were defined by the authors (Table 1). Representative questions for each of the specific data attributes proposed by the ECDC guidelines were defined and included in a questionnaire with open questions [Additional file 1]. The questionnaire consisted of two parts. Part one focused on getting an overall description of the databases. This included the aims of the database, data sources, current use, legal accessibility as well as a description of the data flow. The second part included questions to assess the data quality for the purpose of diseases monitoring and surveillance based on the attributes presented in Table 1. In addition, the questionnaire was pre-tested on two colleagues who work with the databases.

125 Table 1: ECDC data quality attributes and proposed indicators for evaluating data
126 quality in seven Danish swine databases.

Attribute	Proposed indicator
1. Completeness	a) Warnings given by the system in case of missing information b) Examples of missing information allowed by the system
2. Validity	a) External validity: description of checks and validation of the data delivered to the database, and whether (correct) registration is related to any economic aspects b) Internal validity: description of checks and validation of the data in the database c) Examples of coding errors found in the database and how data are introduced into the system (pre-defined codes, free text)
3. Timeliness	a) How often are data updated/registered b) How much time is required between input and availability of data in the database c) How much time is required between data entry and subsequent use
4. Representativeness	a) The proportion of the population covered by the system/database
5. Usefulness	a) Use of data for control or eradication programs b) Presentation of data in e.g. reports, summary statistics or others
6. Simplicity	a) Time required to enter registrations into the system b) Time required for access to and extraction of data
7. Flexibility	a) Possibilities and timeliness of the system to adapt to changes, such as introduction of new codes/variables b) Examples of situations where new codes/variables were introduced in the database
8. Acceptability	a) Potential challenges in using the database for monitoring swine diseases and the eventual implications b) Combining different data sources for monitoring swine diseases and the eventual implications

127

128

129

130

131

132

133

Selection of database experts and interviews

A total of 19 face-to-face interviews (2 joint and 17 individual) and 1 email interview were conducted between November 2015 and January 2016.

Seventeen interviewees completed the full questionnaire for one of the 7 databases, while 3 provided input to specific questions only. Thus, each database was represented by 2-4 individual interviewees, selected based on their knowledge regarding the databases, level of experience and involvement in the databases. Hence, people maintaining and using the data, including veterinarians and IT services, were prioritized. The questionnaire and objective of the study were sent to all interviewees in advance. The duration of the interviews varied from 30 minutes to 1.5 hours. The interviewees were encouraged to express their knowledge, personal opinions and experiences with the data. Furthermore, the interviews were conducted by 1, 2 or 3 authors whom, based on their own experience with the data, raised additional questions regarding the data and the system structure for discussion with the interviewees. The interviewers took written notes of the answers.

Combining the results from the interviews with the authors personal experience

Background information on the databases was gathered from a literature search prior to the interviews, while documents and reports recommended during the interviews were retrieved afterwards for describing the databases (Part I of the questionnaire). The written notes taken during the interviews for the data quality and systems structure (Part II of the questionnaire) were summarized as presented in the results section and sent to

all interviewees to check and confirm their answers. Additionally, the authors, who have been working extensively with the included databases for research purposes, contributed with personal experience when relevant.

Results

Description of databases

In the following sections, the data evaluation framework established by the ECDC was used for the seven databases. A summary description of each database is given in Table 2. Each database was described individually, with the exception of the two laboratory databases (DTU-Vet Lab and SEGES VSP Lab), which were described together. Detailed information regarding the data flow within each database is provided in Additional file 2.

Central Husbandry Register (CHR)

The CHR is the national Danish database of farm demographics. The CHR was established in 1993 with the aim of tracing animals [14] and meeting the subsequent European legislation [15,16]. All locations where animals are gathered (e.g. farms, herds, markets, assembly centers, abattoirs, rendering plants, agricultural shows and common pasture) must be registered in the CHR. Each location has its own unique CHR number with affiliated address and Cartesian geographical coordinates. It is possible for several herds to have the same location and CHR number. A herd is defined as a group

of animals of the same species at the same location with a common aim and owner, and is identified by a unique herd number [17].

Swine Movement Database (SMD)

The SMD is technically a subset of the CHR. The original purpose of the database was to ensure traceability of all swine in Denmark, and it was established in 2002 to fulfill the European legislation regarding the trade of bovine and swine in European countries [18,19]. It is mandatory to register all swine movements in Denmark, yet these are recorded at batch level and it is not possible to trace swine movements at individual animal level.

The national Danish database of drugs for veterinary use (VetStat)

All purchases of prescription-only drugs for production animals are registered in the national database VetStat. It is mandatory to register the purchase of drugs, either passively (by pharmacies and feed mills at the point of sale) or actively (by veterinarians). Records include detailed information such as the date, prescribing veterinarian, receiving farm ID, species, age group, and clinical indication [20]. The database was originally implemented for research purposes in the year 2000, but has since been expanded to assist health advisory services provided by veterinarians for decision making and in keeping track of developments in drug consumption. Since 2010, VetStat data have also been used by the authorities to restrict antimicrobial use at farm level in the Yellow Card program [5]. VetStat has been presented to foreign delegations

on several occasions, and in relation to export, VetStat may enhance trade agreements with other countries by documenting antimicrobial control.

DTU-Vet Lab and SEGES *VSP Lab*

The DTU-Vet lab and SEGES VSP lab conduct extensive diagnostic examinations on a wide range of swine diseases in Denmark. Both laboratories have collaborative protocols and perform diagnostic testing in parasitology, immunology, virology, bacteriology, histopathology and necropsies. Furthermore, the DTU-Vet lab is the reference laboratory for all notifiable swine diseases in Denmark, including brucellosis, tuberculosis, swine vesicular disease, foot-and-mouth disease, classical swine fever and African swine fever [21]. For both labs, the frequency of incoming samples depends on the national monitoring and surveillance programs, the SPF status of the herd, outbreak investigations and eradication programs. Laboratory tests beyond regulative rules depend on the decision of the farmer and guidance by the affiliated veterinarian. Both laboratories have systems for recording information to track samples during the process and send results and invoices to clients. In both systems, the data are extracted and the diagnostic tests results are used for disease monitoring and surveillance by the SPF System and for research purposes with special authorization.

The Specific Pathogen Free system (SPF System)

The SPF system was created in 1971 to combine health information with commercial interests [22]. The SPF system defines a fixed set of rules for biosecurity, surveillance

and swine movement between herds (SPF-SuS). The SPF status of a farm is based on regular laboratory diagnostics and clinical visits performed according to SPF rules. The majority of Danish breeding herds (including nucleus and multiplier herds) are tested every month while the production herds (including farrow-to-finisher, fattener herds and finisher herds) are tested annually.

Meat Inspection database

Since 1964, data on meat inspection of swine slaughtered in Denmark have been registered in the Meat Inspection database [23]. The original aim of the database was to ensure food safety and correct payments to the farmer. It has been expanded to ensure animal health and welfare, in order to fulfill EU legislation [24]. The database is owned by the Classification Authority [25] while data additionally are stored by DVFA, which is responsible for the veterinary control in the abattoirs [26].

234 Table 2: Summarized features of seven Danish databases that may be used in the monitoring and surveillance of swine diseases
235 in Denmark.

Feature	CHR	SMD	VetStat	Laboratory databases	SPF	Meat Inspection database
Year of implementation	1993	2002	2000	SEGES VSP lab: 1988 DTU-Vet lab: 1908 (year in which the laboratory was stablished)	1971	1964
Current objectives	Retrieve demographical information at farm level.	Tracing swine.	Research, assist veterinary practitioners, to control antimicrobial usage.	Diagnostic, monitoring and surveillance of livestock diseases in Denmark.	Manage the health status of participating farms.	Payment of farmers, Food safety, Animal health and welfare.
Data providers	Farmers, VSP-SEGES[1] staff.	Abattoirs, Export stables, Pick up places, Rendering plants, Farmers.	Pharmacies, Veterinarians, Feed mills.	Farmers, Veterinarians, Abattoirs, Research institutes.	Veterinarians, Farmers, Laboratories, Ear-tag database, Zoonosis register, SPF haulage contractors.	Abattoirs.
Data entry	Online or through VSP-SEGES, Aarhus N.	Receiving farm, Exporting farm (for exports only), Transport company.	Apothecary, Veterinarian, Veterinary secretary, Employee in the feed mill.	Laboratory technician or automatic system depending on the diagnostic test performed.	Automatic data entry from the different data providers, Staff at SPF-SuS.	Technician or veterinarian working in the abattoir.
Database administrators	DVFA[2]	DVFA	DVFA	DTU Vet[3]/ VSP-SEGES	SPF SuS[4]	Classification Authority
Case definition	Geographic locations where swine are gathered at herd level.	Movement of a batch of swine at herd level.	Purchase records of prescription-only drugs at farm level.	Laboratory submission including sample(s) collected at individual and herd level.	Farm level	Carcass, Animal level

Information available by case	Farm number, Date of establishment and closure, Contact details of the owner, Veterinary praxis number, Animal species, Production type, Numbers of animals.	Farm number of sender and recipient, Date and time of movement, Number of swine moved, Registration number of vehicle, Number of the trade certification for the movement For rendering plants only: number of dead finisher moved, number of dead sow numbers and number of containers (including dead pigs) move.	Recipient (farm number) Date of purchase, Prescribing veterinarian/ Practice, Product information, Amount of drug, Targeted animal species, age group and diagnostic group.	Farm number, Veterinarian, Biological material, Date of collection, Date the sample was received and analyzed, Anamnesis (for necropsies and histopathology), Analysis codes, Test results.	Farm number, Laboratory results, Danish Standard, Movement data, SPF health status, Salmonella.	Originating farm number, Gambrel number, Abattoir ID, Date of slaughter, Slaughter number, Delivery number, Weight, Meat percentage, Sex, Veterinary remarks, Measure of fat and meat depth.
Surveillance programs for specific diseases	NA	NA	NA	SPF diseases, Salmonella level, Notifiable diseases (OIE listed diseases).	Enzootic pneumonia, Pleuropneumonia, Atrophic rhinitis, Dysentery, Porcine Reproductive and Respiratory Syndrome, Mange, Lice.	Notifiable diseases (OIE listed diseases).
Geographic coverage	National	National	National	Farms sending samples	Participating farms	Farms sending swine for slaughter in Denmark.
Data collection	Compulsory	Compulsory	Compulsory	Compulsory for Salmonella and Notifiable diseases (OIE listed diseases).	SPF farms	Compulsory post mortem inspection at slaughter.

Legal access to the data	Public access on the website [45].	Public access on the website [46].	Farmers and veterinarians have permission to affiliated farms. Open-access for registered users.	Special authorization to have access to the laboratory diagnostic results.	Public access on the website [6].	Farmers receive a report on slaughter remarks along with the account. Others need to apply for permission.
Deliverables and outputs	SEGES yearly report, DVFA Animal Health annual reports.		Yellow Card program (farm level), DVFA monthly statistics (national level) [47], DANMAP yearly report (national level) [48].	Quarterly and yearly reports on the number of diagnostic tests for specific pathogens and tests available online [49].	Annual statistics reports (internal)	Yearly report by the Classification Inspection

[1] SEGES VSP-SEGES: SEGES Pig research Centre.

[2] DVFA: Danish Veterinary and Food Administration

[3] DTU Vet: National Veterinary Institute – Technical University of Denmark.

[4] SPF SuS: Specific Pathogen Free system - SEGES

The table summarizes the background information on the databases gathered from a literature search prior to the interviews, documents and reports recommended during the interviews for describing the databases.

243 Database attributes

244 The data quality and system evaluation is presented in Tables 3 and 4. The results represent the
245 assessment by the interviewees and the authors as well as an evaluation of the attributes as defined in
246 Table 1.

247 As an example, the CHR database (Table 3) records information from all farms (representativeness),
248 requires that all variables are entered (completeness) using pre-coded or free text fields, and the data
249 delivered to the online platform are checked (validity). The number of animals in each farm is updated
250 at least once or twice per year, whereas other changes, such as change of ownership, are recorded
251 within 7 days (timeliness). The CHR data are currently used for animal traceability, risk-based disease
252 and welfare controls, manure reports and the control of antimicrobial use (usefulness). The information
253 is updated by farmers or SEGES directly to the online platform (simplicity). The introduction of new
254 variables or changes requires modification of Danish ministerial orders and agreements with IT
255 companies (flexibility). The interviewees mentioned that the CHR database is widely used in
256 combination with other databases to retrieve information on the swine herd (acceptability).

257

258

259

260

261

262

263 Table 3: Data quality and system evaluation of three public databases for swine herds in Denmark.

Data quality attribute		CHR	SMD	VetStat
Completeness	1)	All variables must be filled in.	All variables must be filled in before data can be sent to the database except for specific cases (see Additional file 2).	A warning is generated if all information from a specific pharmacy is missing in the monthly registrations.
	2)		Specific possible missing variables: vehicle number, number of dead swine/containers.	Complete missing cases possible (for registrations by veterinarians); ID of the drug.
Validity	1)	Pre-coded fields where possible; Retrieves information from the official road register; Registrations as perceived by the farmer.	Retrieves information from CHR; Partly economic: movement to abattoirs and rendering plants.	Free-typed text; Double check of purchase in pharmacies; Majority economic: purchase from pharmacies.
	2)	Computer-generated checks, letters of notification are sent out.	Computer-generated checks, letters of notification are sent out; For export: the registration is validated against the Danish Transport Standard.	Retrospective manual checks periodically made by DVFA employees.
	3)	Number of weaners and finishers registered tends to be less precise than the number of sows registered. Farmers may tend to register the number of pen places available instead of the actual number of present swine.		There can be discrepancies between the actual age / disease group treated and the one reported in VetStat. Incorrect CHR or species may appear due to the lack of pre-coded fields.
Timeliness	1)	Existing herds: updates at minimum once/twice yearly; Establishment of new herd, change of ownership, arrival of new type of swine: register within 7 days; Cessation of herd: register within 6 months	Movement of swine must be recorded within 7 days of movement; Late registrations are often found [a].	Data are registered at the time of purchase (feedmills, veterinarians and pharmacies), or shortly after (veterinarians).
	2)	Online registrations are available instantly.	Online registrations are available instantly.	All registrations for a given month are available no later than the 10^{th}.
	3)	Instantly.	Instantly.	Up to two months: summary statistics are generated using data from the month before the previous one to ensure all data are present in the database at the time of calculation.

Representa-tiveness		Mandatory for all geographic locations with at least one swine; There exist holdings with pigs not registered in the CHR.	It is mandatory to register all movements of swine. Underreport of swine movements were verified, especially in small farms[a].	All herds using prescription-only drugs are present in VetStat.
Usefulness		Traceability of animals; Risk-based selection of herds for diseases and welfare controls; Manure reports.	Tracing back swine in an outbreak situation; Eradication and control programs at herd level.	Research; Control of antimicrobial usage; Assist Veterinary Health Advisory Services.
Simplicity	1)	Farmers or SEGES VSP staff update information online.	Farmers or SEGES VSP staff update information online; Abattoirs upload data once a day.	Pharmacists, veterinarians and feed mills introduce the data manually;
	2)	Registrations are instantly available online (herd-level).	Registrations are instantly available.	1-2 minutes (herd level); Hours (national level)
Flexibility		Introducing new variables requires a change in the Danish regulations and an agreement with the IT company maintaining the system.	Introducing new variables requires a change in the Danish regulations and an agreement with the IT company maintaining the system.	Introducing new variables requires a change of the Danish regulations.
Acceptability	1)	Infrequent updates; Herd type defined by the farmer; Number of weaners and finishers (in particular) may deviate from actual number; Precautions using CHR data as disease-measuring tool.	Precautions using SMD data as disease-measuring tool.	Incongruence between original aim and current use of the database; Precautions using VetStat data as disease-measuring tool.
	2)	Widely used in combination with other databases to retrieve information on herd demographics.	Used in combination with other databases to retrieve information on animal movement, e.g. may be used to track spread of disease.	

[a] Based on authors personal experience with the data.

Data quality is assessed based on eight attributes adapted from the ECDC guidelines. The table summarizes the opinion of the interviewees and the authors' personal experience with the data.

271 Table 4: Data quality and system evaluation of four private databases for swine herds in Denmark.

Data quality attribute		SPF	Laboratory	Meat Inspection database
Completeness	1)	The system generates emails reminding the farmer and veterinarians to collect samples.	Missing information is allowed for some of the variables, but the systems will give warning messages if variables such as ID and results are missing, preventing the journal from being closed.	The amount of data received weekly by the Classification Inspection is compared to the expected number of entries; Missing information on e.g. meat percentage may occur in up to 0.5% of the cases before a warning is given.
	2)	Free text field to enter comments/information regarding the herd.	For submissions made by other institutes (including experimental studies) the herd/farm ID is not required. This represents very few submissions.	Complete missing cases possible: unreadable delivery number tattooed on the ham or separation of the carcass from the gambrel; A maximum of six remarks can be registered per carcass.
Validity	1)	Automatically retrieves information from CHR and laboratory data; Economic: SPF status influences the price of the swine sold by the farmer.	Extra checks at insertion of data; Retrieves information from CHR; Pre-coded fields for data entry.	Machine-generated values (meat quality); Pre-coded fields or free-typed text (veterinary remarks) depending on the system in the abattoir.
	2)	No validation of the data.	Integral quality documents including standard operating procedures (SOPs).	Data are not double checked.
	3)	Free text typing of owner names when emitting notifications.	Free text for some variables, especially for pathology results.	Difference in sensitivity among abattoirs [24].
Timeliness	1)	Overnight.	Continuously as the laboratory results are available.	Daily.
	2)	Once a day, but may be corrected instantly during working hours.	Instantly.	The database receives registrations from all abattoirs daily/weekly depending on the abattoir.
	3)	Instantly.	The samples can be tested on the same day or can take weeks; The bills with the results are sent in the same day or it can take several weeks, depending on the diagnostic test performed.	Up to 1 week from slaughter until the farmer is paid.

272

Representativeness		Results gathered for all SPF herds on a regular basis: 99% of breeding animals, 78% of sows and 34% of finishers in Denmark.	80-90% of the total number of Danish commercial swine herds (estimated only by the interviewees and considering all the different diagnostic tests); A large majority of the herds tested based on serology results are SPF herds, whereas the number of submissions from non-SPF herds represents a lower percentage of the total of submissions and depends of several factors [a].	98% of all swine slaughtered in Denmark (estimated by the interviewees).
Usefulness		Health declarations; Eradication and control programs at herd level	Diseases diagnostic, monitoring and surveillance; Outbreaks detection; Eradication and control programs at herd level.	Provide information for stakeholders to make decisions on relevant political issues.
Simplicity	1)	Serology results from laboratories are introduced automatically in the system; Health status can be manually changed when needed and is available on the website.	Serology results from SEGES VSP are introduced automatically in the system; Serology results from DTU Vet entered manually or automatically; Results from necropsies and pathology results need some time to type the results.	Once registered at the slaughter line, remarks are automatically transferred to the database in the Classification Inspection; No further data handling.
	2)	Access to a small number of cases: 1-2 minutes. Access to a large number of cases: several minutes.	Access to a small number of cases: 1-2 minutes. Access to a large number of cases: several minutes.	Hours for download of a large amount of data.
Flexibility		Easy to include new diseases and pathogens.	Database managed by DTU vet: easy to change the information system to include new variables; Database managed by a private company for VSF-SEGES: difficult and costly (assuming the system implemented in 2015).	Requires a change of the Danish regulations and agreement with the IT company maintaining the system, which is costly; Demanding to change the system at the abattoir and to instruct technicians and inspectors.
Acceptability	1)	Monitoring diseases and plan control and eradication programs at herd level; Precautions using SPF data as a disease-measuring tool.	Monitoring diseases, plan control and eradication programs and enables monitoring of specific herds; Precaution should be taken when using the laboratory submission data as a proxy of diseases occurrence.	Estimate true prevalence, sensitivity and specificity where possible; Enables evaluation of macroscopic disease lesions on a large proportion of Danish finishers/sows; Precautions using meat inspection data as disease-measuring tool.
	2)	Input data provided by different databases.	Laboratory data has the potential to be merged with other databases, such as movement data and CHR (GIS systems).	

[a] Based on authors personal experience with the data.

Data quality is assessed based on eight attributes adapted from the ECDC guidelines. The column "Laboratory" covers information from the laboratory at the National Veterinary Institute, Technical University of Denmark (public), as well as the laboratory at the SEGES Pig Research Centre (private). The table summarizes the opinion of the interviewees and the authors' personal experience with the data.

Discussion

In this study, we assessed possibilities of using routinely collected and registered swine health related data for integrated disease surveillance purposes. In addition, we presented a qualitative approach derived from existing guidelines to evaluate data quality in existing databases, exemplified in the study of seven Danish swine databases.

CHR

CHR data are extensively used in combination with other databases, for example swine movements, retrieval of information on herd demographics for laboratory data, standardization of antimicrobial use [5] and the selection of farms for risk-based farm visits evaluating welfare [27]. The integration of databases may have the advantage of minimizing the risk of typographical errors, yet it also means that incorrect information in one database is transferred to others. It is therefore of paramount importance that data stored in the CHR are correct. A number of automated procedures have been implemented to ensure the completeness and validity of CHR data (Table 3). However, not all variables are equally valid, e.g. the classification of production type is registered as perceived by the farmer, which may lead to misclassification bias. One farmer may define the herd as a production herd, while another would call it a hobby herd. As a consequence, using this classification to identify target farms for disease monitoring and surveillance can be problematic.

The frequency of updates in the CHR is irregular (timeliness, Table 3). According to legislation, updates should be performed at least once or twice per year, depending on the farm size. However, the farmer is able to update more often, which may lead to diversity in the precision of registration between farms. According to the interviewees and the authors, the number of sows may be more reliable than

the number of weaners and finishers. In addition, CHR data are related to the manure reports, where the number of animals must correspond to the area of land use [28] favouring a low number of registered swine. However, the CHR data are also used in the yellow card scheme, where a high number of swine may favour a reduced estimation of the average antimicrobial use per each of the three age groups; sows, finishers and weaners. Despite that all farms should be registered in CHR, it was not possible to estimate the actual completeness of the database. However, there is a general belief in authorities in Denmark, including an acceptance of the need to register in public databases. Furthermore, since CHR data are used for many purposes, we must assume that farms not registered in CHR do exist; they are few, with a low number of swine and of limited importance.

SMD

As for the CHR, no studies have been performed quantifying the completeness of the SMD. However, the authors verified that some inward movements are underreported, especially in small farms. This means that either the registration in the CHR or the lack of registration of inward movements were incorrect. Furthermore, a considerable number of farms do not register the movements of swine within the required seven days (based on authors personal experience), thus the timeliness of the database is not always reliable. A delay in the timeliness is mostly important for traceability during outbreaks of disease. In such a situation, all farmers are obliged to update their registrations at once, and therefore the influence of this delay is expected to be limited.

VetStat

At the time of purchase, antimicrobial products are registered in VetStat for a specific age group and the diagnostic group of primary interest, as estimated by the veterinarian. However, in practice antimicrobial products may be used for other age and/or diagnostic groups than the ones stated (personal communication Laura Mie Jensen, DVFA). Diagnostic groups are defined as reproduction/urogenital, udder, gastrointestinal, respiratory, joints/limbs/CNS/skin or metabolic disorders [20]. As such, one may expect the amount of antimicrobials prescribed for a specific diagnostic group to indicate the level of clinical disease for that particular group. Likewise, VetStat holds registrations on purchase of vaccines and pain killers, which may also indicate clinical disease. However, four issues bias this expected correlation between use of antimicrobials and presence of clinical disease. Firstly, antimicrobials may be used as methaphylaxis and thereby targeting the treatment of animals prior to clinical disease [29]. Secondly, some categories of registration in VetStat are more diverse (e.g. joints/limbs/CNS/skin) than others (e.g. respiratory). Thirdly antimicrobials may in practice be used for other disease groups than the one registered. And finally we do not know when the animals are treated, except assuming that the antimicrobials have been used in the time frame between two purchases of a similar product.

As antimicrobials are quantified as standardized dosages (in Denmark, ADDs) based on standard weights of the animals, the variation in weights within the specific age groups influences the calculated ADDs. As sows and piglets are defined as one single group called “breeding animals/piglets”, the estimated number of treated animals (ADDs) in this group is expected to deviate more from the actual number of treated animals (standard weight 200 kg) than in the groups “weaners” (15 kg) and “finishers” (50 kg). Additionally, antimicrobial use by the farmer is influenced by factors not related to the level of disease, such as changes in legislation [4], prices of products [30], campaigns run by the

pharmaceutical companies and his personal threshold for initiation of treatment. As such the use of VetStat in disease surveillance has its limitations, when data simultaneously are used as a tool of control, e.g. in the Yellow card system.

DTU-Vet lab and SEGES *VSP lab*

For laboratory data, validity may be compromised by limitations in the system, specific changes or untrained individuals entering registrations. In particular, a wide range of issues related to coding may occur, such as errors, variation throughout time, incompleteness [12] and changes in the sensitivity and specificity of diagnostic tests. Furthermore, ongoing national monitoring and surveillance programs, compulsory as well as voluntary, outbreak investigations and eradication programs will affect representativeness over time. As an example of a voluntary monitoring program, the frequency of samples sent to the laboratory for testing PRRS in Denmark varies with the SPF status of the herd [31]. Moreover, the frequency of testing depends on the value of the animal and not only on the disease impact [32]. Also in case of a new disease occurring, veterinarians tend to send more diagnostic material to the laboratory. As a result, the veterinary practitioners focus on specific diseases may vary over time and thereby affect the number of submissions sent to the laboratory. Another limitation of using these data is that a herd tested positive based on a serological diagnostic test might be an indication of previous rather than active infection, as antibodies against a specific disease might persist long time post-infection, as for example with PRRS [33]. Furthermore, data from the two laboratory databases might underestimate the level of disease, as some veterinarians might choose to send the samples to laboratories abroad.

SPF system

The SPF system is a good example of how laboratory diagnostic information can be automatically integrated with variables from other databases. However, in relation to representativeness, it only covers herds registered in the SPF System. SPF herds have often been characterized by good health and biosecurity, while the opposite is not necessarily true for non-SPF herds, since farms can adopt SPF biosecurity rules, while disregarding the serological testing for SPF diseases. Moreover, as discussed for the laboratory data, positive results based on serology might not be an indication of active infection, but a result from previous exposure to the disease. Furthermore, the farmer may choose to accept a positive status for a specific disease on the farm in order to be able to import swine from another positive farm, or to accept a positive status once the farm has got it, because it is too expensive to regain negative test status.

Meat Inspection database

Meat inspection data have previously been shown to have low sensitivity [3], possibly due to variation in the stage of lesions presented at slaughter [34]. Since the reformation of slaughter codes in 2009 [35], courses have been held targeting a standardized assessment of carcasses in different abattoirs. However, follow-up courses may be needed regularly to maintain similar assessment criteria. The variation in registrations among abattoirs might also be explained by variation in the configuration of terminals, where some abattoirs use pre-typed codes, while digits must be typed in separately in other abattoirs. Furthermore, validity can be influenced by changes in staff, abattoir procedures [3] and may differ between disease categories due to a low sensitivity [36,37]. In addition, the speed of the slaughter line leaves no time for double-checking, or retrospective updates. Changes in disease codes [35,38] and

389 legislative actions can influence the content of the database as seen in Switzerland [39] and complicate
390 the comparison of meat inspection data throughout time.
391
392 *General discussion*
393 The comparison of observations over time is a key feature of incorporating these data into a
394 surveillance system, since retrospective analysis using historical data is required by many of the
395 statistical quality control methods used for disease surveillance in human and veterinary sciences
396 [2,40,41]. It is therefore of the utmost importance to be aware of changes in the databases that may
397 affect the attributes described in Table 3 and 4.
398 Completeness and timeliness have definitely improved over time in the CHR, as data collection has
399 been updated from a written mailed questionnaire of several pages to an electronic online version.
400 Despite of this improvement, the frequency of updates in the CHR still a limitation. For other databases
401 such as VetStat and the Meat Inspection database, the time between data entrance and its availability
402 (timeliness) is a limitation for using disease data as a (near-)real time disease monitoring tool.
403 Likewise, the time between the samples are received and the results are available in the database also
404 depends on the type and number of laboratory diagnostic tests requested, and it can be a weakness
405 when using these data for disease monitoring and surveillance.
406 Validity has been improved by merging data from several sources and linking the results to the
407 economic interests of the farmer. For example, VetStat and CHR have been combined and used in the
408 Yellow Card legislation [5]. Farmers and veterinarians are now aware that incorrect information in one
409 of these registers may result in the herd exceeding the antimicrobial threshold value and being put
410 under restrictions. In general, the validity of databases tends to improve, when the advantages are
411 apparent to the farmer (such as having an SPF certificate for commercial purposes), or when incorrect

412 registrations have negative consequences (*e.g.* the influence of both VetStat and CHR registrations on
413 the Yellow Card scheme, while errors in laboratory diagnosis can result on commercial restrictions for
414 farmers). Another means of improving validity could be to increase the quality of data entered into the
415 system, for example by means of specific automated data checks.
416 The representativeness of the data is a key attribute for data quality. It is important that the data are
417 representative of a population over time in order to monitor disease occurrence. For the laboratory data
418 and the Meat Inspection database, only a proportion of the population is registered daily/weekly,
419 mainly representing red SPF herds and finisher herds, respectively. This is a limitation when
420 monitoring disease occurrence and trends on a national scale.
421
422 *Methodology*
423 The ECDC guidelines [9] were adapted to meet the requirements of veterinary databases, in order to
424 monitor data quality and to evaluate surveillance systems. However, the ECDC guidelines include the
425 assessment of quantitative measures, which would require extensive resources (including IT experts) to
426 quantify each ECDC attribute for a large number of variables. Therefore, we opted for a qualitative
427 approach to standardize the evaluation of seven diverse databases. The qualitative approach resulted in
428 an overall assessment of pros and cons for the individual databases. Although interviews were
429 standardized by using a questionnaire, bias of the results could not be entirely avoided due to influence
430 of personal experience of the interviewees with the data. To overcome this issue, we included
431 stakeholders from all institutions with experience and interest in the databases. However, inclusion of
432 quantitative measures could have reduced the effect of personal agendas, but it would have required
433 substantial resources or limited the number of databases evaluated. An alternative qualitative approach
434 is the analysis of strengths, weaknesses, opportunities and threats (SWOT analysis), which has been

applied by Stärk and Nevel [42] to evaluate four databases used to monitor swine health in England. However, the ECDC guidelines were designed specifically to evaluate disease surveillance systems and their data quality, including specific quality key attributes, which can be used for both qualitative and quantitative data quality assessment. Since the ECDC guidelines describe a systematic approach specifically designed to evaluate data quality, we regard this to be an optimal approach for data evaluation.

Suggestions for data quality improvement

Improving data quality for disease monitoring and surveillance requires that the identified challenges are addressed. The CHR and SMD are restricted to the traceability of swine and to predict how the disease might spread. To improve data quality in CHR register, clear guidelines regarding the categorization of herds on the registration page might reduce misclassifications. Also, the development of mobile phone applications in order to continuously update the number of animals within the herds could be worth considering to improve the accuracy of the CHR. An improvement of the SMD could include the age group of the moved swine. This variable is included in the CHR and VetStat, hence the farmer is used to asses which age group his swine belong to. For disease surveillance purposes, the data registered in VetStat should register the date of treatment and not the date of prescription, requiring direct reports from farmers to VeStat and as consequence new IT solutions. This would allow us to improve the quality of the data used with the potential to monitor changes in consumption. In order to use antimicrobial treatment as a proxy for disease occurrence the population and time at risk must be known. Therefore, an improvement of VetStat would be to include information on the duration of treatment, the number of swine treated and an average weight at the time of treatment. However, these registrations would need to be done by the farmer himself, and would most likely lead to a reduction in

both completeness and validity as the demand for registration increases. To improve validity of VetStat data in its current state, it would be recommendable to separate the age group "breeding animals/piglets" into two separate categories, as there are major weight differences among piglets and breeding animals. Another suggestion to improve validity would be to have precoded-fields for data entry or to link data entry instead of free-typing text (VetStat, table 3) with information from the CHR register, to avoid registration of drugs for non-existing CHR numbers.

Perspectives of using the databases for monitoring and surveillance swine diseases in Denmark

To what extent the seven databases can be used as indicators of swine diseases varies; these databases contain information, which can be used in different steps of disease monitoring and surveillance, including 1) data monitoring, 2) defining control and preventive measures, and 3) follow-up.

Over the last decade, the increasing availability of electronic records collected actively or passively led to the development of new analytical and modelling tools. It is believed that using large volumes of data will improve the timeliness of epidemiological information, resulting in more accurate disease surveillance [43].

The combination of multiple data streams from different databases can be used as a multivariate surveillance approach, in which several processes are analysed in parallel or combined [44]. To our knowledge, this approach has not been applied in veterinary sciences. Further epidemiological research will be needed to evaluate and purpose the best approach for monitoring changes in the data that could indicate spread of disease. This depends on the targeted disease (emerging vs endemic) and on the choice of the study unit of disease monitoring and surveillance (herd, regional, national level).

Conclusions

The present study describes and evaluates data routinely collected and registered in seven Danish swine databases and discuss suggestions for improvement and integration of data. In addition, we present a qualitative approach derived from existing guidelines to evaluate data quality in existing databases.

A number of limitations and potentials for disease monitoring and surveillance were identified and described for the databases. A general finding was that the validity of the databases tended to improve, when registrations were interrelated with other databases and are of economic interests to the farmer. Completeness and timeliness of the data have improved with the use of electronic registrations. However, infrequent updates and delays between data entrance and its availability (timeliness) is a limitation for using some of the databases as (near-)real time disease monitoring tool. Additionally, the population coverage (representativeness) varies over time which is a limitation when monitoring disease occurrence and trends on a national scale.

Despite of the limitations, the combination of these databases has potential to improve diseases surveillance in Denmark. Further research is needed to explore and evaluate different statistical monitoring methods, which allows to include different data sources.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Not applicable.

Competing interests

None of the authors have any financial or personal relationships that could inappropriately influence or bias the content of the paper.

Funding

Not applicable.

Authors' contributions

ACLA performed the study design, conducted the interviews, carried out the data management and data analysis and wrote the article; MF performed the study design, conducted the interviews, carried out the data management and data analysis and wrote the article; ACB contributed to the study design, conducted the interviews, carried out the data management and critically reviewed the manuscript; TH performed the study design and critically reviewed the manuscript; AB critically reviewed the manuscript; NT contributed to the study design and critically reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors would like to thank all the interviewees in the study for contributing their expert opinions and time.

References

1. Houe H, Gardner I, Nielsen L. Use of information on disease diagnoses from databases for animal health economic, welfare and food safety purposes: strengths and limitations of recordings. Acta Vet. Scand. 2011;53:S7.

2. Lopes Antunes AC, Dórea F, Halasa T, Toft N. Monitoring endemic livestock diseases using laboratory diagnostic data: A simulation study to evaluate the performance of univariate process monitoring control algorithms. Prev. Vet. Med. 2016;127:15–20.

3. Enøe C, Christensen G, Andersen S, Willeberg P. The need for built-in validation of surveillance data so that changes in diagnostic performance of post-mortem meat inspection can be detected. Prev. Vet. Med. 2003;57:117–25.

4. Jensen VF, de Knegt LV, Andersen VD, Wingstrand A. Temporal relationship between decrease in antimicrobial prescription for Danish pigs and the "Yellow Card" legal intervention directed at reduction of antimicrobial use. Prev. Vet. Med. 2014;117:554–64.

5. Ministry of Food. Exceptional arrangements to decrease the use of antibiotics in pig herds [in Danish]. 2010 p. 5–8.

6. SPF-Sus. SPF-SuS [Internet]. [cited 2016 Jan 20]. Available from: http://spfsus.dk/

7. Salman M. Animal Disease Surveillance and Survey Systems. Methods and Aplications. 1st Ed. Salman M, editor. Ames, Iowa, USA: Wiley-Blackwell; 2003.

8. Buckeridge DL. Outbreak detection through automated surveillance: A review of the determinants of detection. J. Biomed. Inform. 2007;40:370–9.

9. European Centre for Disease Prevention and Control (ECDC). Data quality monitoring and surveillance

548 system evaluation- A handbook of methods and applications. Stockholm; 2014.

549 10. Stärk KDC, Häsler B. The value of information: Current challenges in surveillance implementation. Prev.
550 Vet. Med. Elsevier B.V.; 2015;122:229–34.

551 11. Emanuelson U, Egenvall A. The data – Sources and validation. Prev. Vet. Med. 2014;113:298–303.

552 12. Sørensen HT, Olsen J, Sabroe S. A Framework for Evaluation of Secondary Data Sources for
553 Epidemiological Research. Int. J. Epidemiol. 1996;25:435–42.

554 13. Tayi GK, Ballou DP. Examining data quality. Commun. ACM. 1998;41:54–7.

555 14. Danish Veterinary and Food Administration. Protocols on databases in the Danish Veterinary and Food
556 Administration [in Danish]. 2012.

557 15. European Commission. COMMISSION DECISION of 23 October 2000 laying down detailed rules for
558 registration of holdings in national databases for porcine animals as foreseen by Council. 2000.

559 16. European Council. Council directive 2008/71/EC of 15 July 2008 on the identification and registration of
560 pigs. 2008.

561 17. Ministry of Food. Registration of herds in CHR [in Danish]. 2013 p. 1–6.

562 18. European Union. Directive 2000/15/EC of the European Parliament and the Council of 10 April 2000
563 amending Council Directive 64/432/EEC on health problems affecting intra-Community trade in bovine animals
564 and swine. 2000 p. 34–5.

565 19. Ministry of Food. Tagging, registration and movement of cattle, swine, sheep and goats [in Danish]. 2014.

566 20. Stege H, Bager F, Jacobsen E, Thougaard A. VETSTAT - The Danish system for surveillance of the
567 veterinary use of drugs for production animals. Prev. Vet. Med. 2003;57:105–15.

21. Danish Veterinary and Food Administration. List of national reference laboratories [Internet]. [cited 2016 Jan 20]. Available from: http://www.foedevarestyrelsen.dk/Kontrol/Autorisation/Lister-over-veterinaere-virksomheder/Sider/nationale_referencelaboratorier.aspx

22. Holmegaard H. The beginning of the Danish SPF-system [in Danish]. 2005.

23. Willeberg P, Gerbola MA, Petersen BK, Andersen JB. The Danish pig health scheme: Nation-wide computer-based abattoir surveillance and follow-up at the herd level. Prev. Vet. Med. 1984;3:79–91.

24. European Union. Corrigendum to Regulation (EC) No 854/2004 of the European Parliament and of the Council of 29 April 2004 laying down specific rules for the organisation of official controls on products of animal origin intended for human consumption. Off. J. Eur. Union 2004.

25. Ministry of Food. Weighing, classification and reporting of market prices on swine, cattle and sheep carcasses [in Danish]. 2012.

26. Ministry of Food. Circular on meat inspection [in Danish]. 2011 p. 1–72.

27. Danish Veterinary and Food Administration. Animal welfare in Denmark [in Danish]. 2014.

28. Ministry of Food. Commercial holding of animals, manure, ensilage etc [in Danish]. 2015 p. 1–49.

29. Ministry of Food. Order on the veterinary use, dispensing and prescription prescription of drugs for animals [in Danish]. 2015 p. 1–22.

30. Gibbons JF, Boland F, Buckley JF, Butler F, Egan J, Fanning S, et al. Influences on antimicrobial prescribing behaviour of veterinary practitioners in cattle practice in Ireland. Vet. Rec. 2013;172:14.

31. Lopes Antunes AC, Halasa T, Lauritsen KT, Kristensen CS, Larsen LE, Toft N. Spatial analysis and temporal trends of porcine reproductive and respiratory syndrome in Denmark from 2007 to 2010 based on laboratory submission data. BMC Vet. Res. 2015;11:303.

589 32. Doherr MG, Audigé L. Monitoring and surveillance for rare health-related events: a review from the
590 veterinary perspective. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 2001;356:1097–106.

591 33. Yoon K-J, Zimmerman JJ, Swenson SL, McGinley MJ, Eernisse KA, Brevik A, et al. Characterization of the
592 Humoral Immune Response to Porcine Reproductive and Respiratory Syndrome (PRRS) Virus Infection. J. Vet.
593 Diagnostic Investig. 1995;7:305–12.

594 34. Stärk KDC, Alonso S, Dadios N, Dupuy C, Ellerbroek L, Georgiev M, et al. Strengths and weaknesses of
595 meat inspection as a contribution to animal health and welfare surveillance. Food Control. 2014;39:154–62.

596 35. Ministry of Food. Circular on performance of meat inspection [in Danish]. 2009.

597 36. Bonde M, Toft N, Thomsen PT, Sørensen JT. Evaluation of sensitivity and specificity of routine meat
598 inspection of Danish slaughter pigs using Latent Class Analysis. Prev. Vet. Med. 2010;94:165–9.

599 37. Nielsen SS, Nielsen GB, Denwood MJ, Haugegaard J, Houe H. Comparison of recording of pericarditis and
600 lung disorders at routine meat inspection with findings at systematic health monitoring in Danish finisher pigs.
601 Acta Vet. Scand. 2015;57:1–8.

602 38. Ministry of Food. Circular on changes in the performance of meat inspection [in Danish]. 2010.

603 39. Vial F, Reist M. Comparison of whole carcass condemnation and partial carcass condemnation data for
604 integration in a national syndromic surveillance system: the Swiss experience. Meat Sci. 2015;101:48–55.

605 40. Dupuy C, Morignat E, Dorea F, Ducrot C, Calavas D, Gay E. Pilot simulation study using meat inspection
606 data for syndromic surveillance: use of whole carcass condemnation of adult cattle to assess the performance of
607 several algorithms for outbreak detection. Epidemiol. Infect. Cambridge University Press; 2015;143:2559–69.

608 41. Cowling BJ, Wong IOL, Ho L-M, Riley S, Leung GM. Methods for monitoring influenza surveillance data.
609 Int. J. Epidemiol. 2006;35:1314–21.

610 42. Stärk KDC, Nevel A. Strengths, weaknesses, opportunities and threats of the pig health monitoring systems
611 used in England. Vet. Rec. 2009;165:461–5.

612 43. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big Data for Infectious Disease Surveillance
613 and Modeling. J. Infect. Dis. Oxford University Press; 2016;214:S375–9.

614 44. Frisén M, Andersson E, Schiöler L. Evaluation of multivariate surveillance.

615 45. Danish Veterinary and Food Administration. CHR [in Danish] [Internet]. 2016 [cited 2016 Feb 10].
616 Available from: https://chr.fvst.dk/

617 46. Ministry of Food. Swine movement database [Internet]. [cited 2016 Feb 23]. Available from:
618 svineflyt.fvst.dk

619 47. Danish Veterinary and Food Administration. VetStat [in Danish] [Internet]. 2016 [cited 2016 Feb 10].
620 Available from: http://www.foedevarestyrelsen.dk/Leksikon/Sider/VetStat.aspx

621 48. DANMAP. DANMAP [Internet]. 2015 [cited 2016 Feb 10]. Available from: http://www.danmap.org/

622 49. National Veterinary Institute - Technical University of Denmark. Annual reports for diagnosis, monitoring
623 and emergency response [in Danish]. [Internet]. [cited 2016 May 26]. Available from:
624 http://www.vet.dtu.dk/Diagnostik/Aarsrapporter-for-diagnostik_overvaagning_beredskab

625 50. Ministry of Food. Health advisory agreements in pig farms [in Danish]. 2016 p. 1–18.

626

Additional file 1- Questionnaire used during the interviews

Part I - System/database overview

1. What is (/was) the (original) objective of the database?
    a. Where is it described?
2. Which data sources are used? (veterinarians, farmers, laboratory, etc.)
3. For what purpose/s is/are these data being used?
4. Which disease surveillance programs are based on this database? (N/A in some cases)
5. Is the data gathered from all Danish pig herds?
6. Is the data compulsory or collected voluntarily?
    a. If voluntary: What type of farmers / economic advantages or costs are related to participation?
7. Who is responsible for gathering the information? (data sources)
8. Who is responsible for entering the data into the system? (data entry)
9. What specific information is being recorded? (where is it described?)
10. Who administers the database? (data operators)
11. Who has access to the database and who extracts the data? Is it the same person?
12. How is the data stored in the system? Integrated or relational? Is it connected to other databases (e.g. GLR-CHR-VetStat (and movement?) are apparently part of the same database)?
13. Are there any reports produced based on the data?
    a. If so, how often are they produced?

Part II - Data quality and system structure evaluation

Completeness

14. Does the system give a warning if information is missing? For example, if you should collect information from 18 herds and you only have information for 4 herds?
15. Does the system allow "missing information", when a registration is typed in?
16. Can you describe and give examples of missing information from different variables for cases registered in the database?

Validity

17. Is anyone responsible for checking and validating the data delivered to the database compared with the case (external validity)?
    a. If yes, who? Data operator / manager? Same person every time?
    b. How often is this performed?
    c. What does this data check include (random or same check every time)?
    d. What actions are taken if an error is found?
18. Is anyone responsible for checking and validating the data in the database (coding errors: internal validity)?
    a. If yes, who? Data operator / manager? Same person every time?
    b. How often is this performed?
    c. What does this data check include (random, or same check every time)?
    d. What actions are taken if an error is found?

19. Which coding errors can be found in the database? Please give examples of variables with coding errors.
20. When data is entered into the database, is it recorded using pre-defined codes used in the system (words) for all variables, or is it "free" typing text/numbers?

Timeliness

21. How often is the database updated?
    a. Does this happen before / after the eventual data checks?
22. How much time passes between the data becoming available and being uploaded to the database?
23. How much time passes between data entry and its subsequent use?
24. Has the database been exposed to any major changes, or is it possible to compare data throughout time?

Representativeness

25. What proportion of the population is covered by the system (can be expressed in numbers or percentages)?

Usefulness

26. Can you indicate action plans taken (such as disease control/eradication programs) based on information originating from the system/database?
27. Are the data being used for specific purposes such as reports, research or other?
    a. Are these in agreement with the original purpose of the database?

Simplicity

28. How much time does it take to load the data into the system?
29. How much time does it take to have access to the data in the system?

Flexibility

30. How easy/time-consuming is it to adjust information in the system?
31. Is it possible to add new codes/variables into the system?
32. Can you please give examples of situations where new codes/variables were introduced in the database? What implications did this have (e.g. was it necessary to create a completely new system)?
33. How easy/time-consuming is it to expand the system, for instance to include new data (new variables)?

Acceptability

34. Do you think that these data can be used for monitoring pig diseases?
35. Do you think that is it possible to combine these data with other databases for monitoring pig diseases? What are the eventual implications of combining with other databases?

Additional file 2 – Dataflow in the databases

CHR

It is the responsibility of the herd owner to register the herd and keep the records up-to-date. Depending on their size, existing herds must update their records once or twice (for large herds, i.e. swine herds >300 sows, >3,000 finishers or >6,000 weaners [17]) a year. Establishment of a new swine herd or the inclusion of an additional age group must be registered within 7 days, and closure of a herd no later than 6 months after removal of the last pig.

Farmers can register online [45] or ask SEGES (Aarhus N) to do it for them. The majority of data are instantly available to the public on the website. Computer-generated logical checks and follow-up letters are generated on a daily/weekly basis using registrations in the database. Examples of such checks include: swine herds holding swine, but without movements, or swine herds with no registered swine or movements.

SMD

Data are entered either by automatic upload or manually through an interface. Data are available instantly when entered through an interface, while uploading occurs twice a day. All movements must be recorded within 7 days. The data can be corrected by the person that entered them, or by the authorities. The receiving farm is responsible for registering the movements. However, in cases where swine are exported outside of Denmark, the farm of origin is responsible for the registrations.

The number of swine moved, the date and time of movement, the CHR and herd number of sender and recipient, the registration number of vehicle, and the number of the trade certificate for the movement

731 are registered in the SMD [19]. The trade certificate is an official document that follows the swine.
732 Furthermore, movements of dead swine to rendering are recorded as number of containers (primarily
733 used for weaners, swine from 7-30kg) or number of dead sows or dead finishers (swine from
734 30-100 kg) [19].
735 Reports on the movement of specific herd numbers are publicly available via the website [46].
736
737
738 **VetStat**
739 All purchased prescription-only veterinary drugs are registered in VetStat.
740 Upon arrival at the pharmacy or feed mill, the prescription is typed in free text and is automatically
741 recorded in the system at the same time as payment is made. Pharmacy procedures include thorough
742 checks of all delivered drugs. All registrations made by pharmacies are automatically transferred to the
743 Danish Health Authorities, who ensure that human and veterinary registrations are separated.
744 Veterinary registrations are forwarded to the IT company handling VetStat, but the number of records
745 is not checked, and some registrations may not be delivered to VetStat until the 10th day of the
746 following month. A warning appears if VetStat has not received any registrations from the largest
747 pharmacies within a month. In contrast, veterinarians and feed mills must actively register records on
748 purchased veterinary drugs, which is possible in 4 different ways: 1) Through the IT system of the
749 veterinary practice; 2) Registration on paper sent to DVFA; 3) Uploading a file on the VetStat website;
750 4) Registration directly on the VetStat website. Veterinarians are able to use methods 1-4 and feed
751 mills can use methods 3 and 4. Registrations by veterinarians and feed mills must be registered no later
752 than the 10th day of the month following purchase. Initially, all data arrive at the IT company handling
753 VetStat and are then merged with CHR data for use in Yellow Card calculations [5]. Data are

subsequently reloaded into VetStat. Only herds with a Veterinary Health Advisory Contract (VAC) [50] are included in the Yellow Card program [5]. Summary statistics are generated using data from the month before the previous one to ensure that all data are present in the database at the time of calculation.

DTU-Vet lab and SEGES VSP lab

After receiving the samples (organic material for analysis), the laboratory technicians are responsible for creating the journal (case file), including registration of the sample ID, CHR number, farmer identification and veterinarian. The information is then checked by a second individual. Depending on both the diagnostic test performed and the system, the results are transferred automatically to the system or are entered manually. Academic staff is responsible for validating the results. Depending on the type of diagnostic test performed, the results are available on the same day (i.e. serology) or within several days or weeks (i.e. histopathology, bacteriology). The results are reported to clients by email or letter. In cases where individual samples are missing results, the system gives a warning and will not allow for closure of the journal.

SPF system

Laboratory diagnostic results from SPF herds are retrieved automatically on working days from DTU-Vet and VSP-SEGES. The system generates an alarm if SPF herds are classified as positive for a disease that does not correspond to the current SPF herd status. In these cases, the results are checked manually and decisions to change or keep the health status are made. During working hours, changes to the herd health status are updated immediately on the SPF website.

777 The SPF system also includes data on *Salmonella* for all Danish swine herds. For breeding and
778 multiplier herds, the *Salmonella* index is calculated based on serological testing. For herds delivering
779 more than 200 finishers annually, the Salmonella level is retrieved from the Zoonosis Register. The
780 salmonella level is calculated each month by serological testing of "meat juice" (drip fluid released
781 from meat after freezing and thawing).
782 The system also gathers data related to swine movements. A Danish animal trade company, SPF-
783 Denmark (SPF-DANMARK), plans and performs swine movements, taking into account the SPF status
784 of the herds. The company provides information on the farmer's name, address and the number of
785 finishers and sows to the SPF register[1]. The number of weaners is retrieved from the ear-tag register.
786 The system generates an alarm if animals with a certain health status are sold to herds with a higher
787 health status.
788 Furthermore, information about the Danish Standard (SEGES - Videncenter for Svineproduktion) is
789 also available in the SPF system. The requirements for Danish swine farmers should correspond to the
790 regulatory and industry requirements. These requirements are described in the Danish Product
791 Standard. An independent company carries out audits (inspections) in the herds. This information is
792 also available on the SPF website and includes 100% of all Danish swine herds.
793
794 **Meat Inspection database**
795 Upon arrival at the abattoir, all swine are checked *ante mortem* by the official veterinarian to determine
796 cases of welfare violation or signs of OIE-listed diseases [24].
797 After slaughter, each carcass is associated with a specific gambrel number, to which the following
798 information is registered: abattoir ID, date of slaughter, slaughter number, originating CHR number,

[1] Information on movements is also registered in the SMD.

delivery number, carcass weight, meat percentage, price, sex, up to ten (six in practice) different veterinary remarks, measure of fat and meat depth, meat and if necessary skatole (boars only). The delivery number is used to identify the herd of origin at all times during the slaughtering process.

The majority of registrations are measured and recorded automatically. Only veterinary remarks are registered manually by the technician on the slaughter line or the veterinarian at the site of re-examination.

822

Manuscript 2

Spatial analysis and temporal trends of porcine reproductive and respiratory syndrome in Denmark from 2007 to 2010 based on laboratory submission data

Ana Carolina Lopes Antunes[1]*, Tariq Halasa[1], Klara Tølbøl Lauritsen[2], Charlotte Sonne Kristensen[3], Lars Erik Larsen[4], Nils Toft[1]

[1]Section for Epidemiology, National Veterinary Institute - Technical University of Denmark, Denmark

[2]Section for Diagnostic and Scientific Advice, National Veterinary Institute - Technical University of Denmark, Denmark

[3]Danish Pig Research Centre, SEGES P/S, Denmark

[4]Section for Virology, National Veterinary Institute - Technical University of Denmark, Denmark

*Corresponding author.

Published in BMC Veterinary Research 2015; 11: 303.

Lopes Antunes et al. BMC Veterinary Research (2015) 11:303
DOI 10.1186/s12917-015-0617-0

BMC Veterinary Research

RESEARCH ARTICLE Open Access

Spatial analysis and temporal trends of porcine reproductive and respiratory syndrome in Denmark from 2007 to 2010 based on laboratory submission data

Ana Carolina Lopes Antunes[1*], Tariq Halasa[1], Klara Tølbøl Lauritsen[2], Charlotte Sonne Kristensen[3], Lars Erik Larsen[4] and Nils Toft[1]

Abstract

Background: Porcine reproductive and respiratory syndrome (PRRS) has been a cause for great concern to the Danish pig industry since it was first diagnosed in 1992. The causative agent of PRRS is an RNA virus which is divided into different genotypes. The clinical signs, as well as its morbidity and mortality, is highly variable between herds and regions. Two different genotypes of PRRS virus (PRRSV) are found in Denmark: type 1 and type 2. Approximately 40 % of Danish swine herds are seropositive for one or both PRRSV types. The objective of this study was to describe the temporal trend and spatial distribution of PRRSV in Danish swine herds from 2007 to 2010, based on type-specific serological tests from the PRRS surveillance and control program in Denmark using the results stored in the information management system at the National Veterinary Institute, Technical University of Denmark (DTU Vet).

Results: The average monthly seroprevalence of PRRSV type 1 was 9 % (minimum of 5 %; maximum of 13 %) in breeding herds, and 20 % (minimum of 14 %; maximum of 26 %) in production herds; PRRSV type 2 had an average seroprevalence of 3 % (minimum of 1 %; maximum of 9 %) in breeding herds and of 9 % (minimum of 5 %; maximum of 13 %) within production herds. The seroconversion rate followed a similar and consistent pattern, being higher for type 1 than for type 2 for both PRRSV types. Regarding the spatiotemporal results, the relative risk distribution maps changed over time as a consequence of the changes in PRRSV seroprevalence, suggesting a general decline in the extent of areas with higher relative risk for both type 1 and 2. Local spatial analysis results demonstrated the existence of statistically significant clusters in areas where the relative risk was higher for both herds.

Conclusions: PRRSV type 1 seroprevalence was constantly higher than for PRRSV type 2 in both herd types. Significant spatial clusters were consistently found in Denmark, suggesting that PRRSV is endemic in these areas. Furthermore, relative risk distribution maps revealed different patterns over time as a consequence of the changes in seroprevalence.

Keywords: PRRSV, Laboratory submission, Spatiotemporal, Seroprevalence, Serocovertion rate

Background

Porcine reproductive and respiratory syndrome (PRRS) causes significant financial losses for the pig industry in Europe, United States (US) and Asia [1-5].

The causative agent of PRRS is an RNA virus [6, 7], the PRRS virus (PRRSV), which is divided into genotypes: type 1 and type 2, previously known as European and North American strains, respectively [8]. The severity of the diseases is highly variable between herd as a result from immunological factors, herd management and the pathogenicity resulting in different clinical signs, morbility and mortality rates [9, 10].

The first Danish case of PRRSV type 1 was diagnosed in March 1992 in a sow herd located in southern Denmark [11]. A voluntary PRRSV control program was established in 1996 by the Federation of Danish Pig

* Correspondence: aclan@vet.dtu.dk
[1]Section for Epidemiology, National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark
Full list of author information is available at the end of the article

Producers and Slaughterhouses in order to reduce the spread of the virus. Initially, a national serological screening based on an DTU Vet (National Veterinary Institute, Technical University of Denmark) "in-house" Blocking Enzymed-Linked Immunosorbent Assay (ELISA) and Immunoperoxidase monolayer assay (IPMA) was carried out, which demonstrated that the seroprevalence of PRRSV type 1 in Danish herds was 33 % [12]. This screening did not reveal the presence of PRRSV type 2. The second step included vaccination of 1100 herds with a modified-live PRRSV type 2 vaccine, between 1 July and 1 October 1996. The vaccine was approved by the Danish Health authorities from 1 July 1996 and licensed for use in pigs between 3 and 18 weeks old. However, this vaccine had already been used in October 1995 to vaccinate all boars entering artificial insemination stations [13]. This procedure was performed in quarantine units with special permission from the Danish authorities. Following approval in 1996, the vaccination was not only carried out in PRRSV seropositive herds, but also in many herds that had no clinical symptoms of PRRS.

In 1997, PRRSV type 2 was isolated for the first time in Denmark from fetuses, dead piglets and sows, suggesting transplacental infection had occurred after PRRSV infection of pregnant sows. In addition, non-vaccinated Danish herds previously uninfected with PRRSV type 1 had become infected with the vaccine-like PRRSV. The PRRSV type 2 virus was also spread from artificial insemination centers in semen, by introducing vaccinated animals to herds, and by airborne transmission to PRRS-free and non-vaccinated herds [13].

Despite disease control efforts in Denmark, PRRS continues to contribute towards the economic losses associated with mortality in piglets, respiratory problems in growers and finishers, and reproductive problems in sows. Furthermore, previously full sequencing of PRRSV type 1 and type 2 [14, 15], demonstrated a high variance in several genomic regions in the PRRSV type 1 strains circulating in Denmark, further complicating the control of the disease. Currently, the between-herd seroprevalence of PRRSV in the Danish pig population is considered to be around 40 %, based on the number of herds with a known status (unpublished data). Spatial and temporal analysis can be used to identify the location, shapes and sizes of potential diseases outbreaks [16].

The spatiotemporal description of PRRS based on laboratory data might help decision makers to re-evaluate their conclusions on the spread of the disease and assess the efficiency of the implemented control strategies. DTU Vet was the only laboratory in Denmark to perform serological tests for PRRS virus from 2007 to 2010. Using only the data from 2007 to 2010 would therefore allow us to study the spatiotemporal occurrence of PRRS. This analysis will allow us to characterize changes in the PRRSV seroprevalence and seroconversion rate, and to assess the spatial distribution of PRRSV seropositive herds, facilitating control of the disease on local and regional basis, e.g. by changing management routines, trade customs etc. and make a descriptive analysis and find patterns, clusters, etc to make help prioritize funds for controlling these diseases.

The objective of the present study was to describe the temporal trend and spatial distribution of both PRRSV types in Danish breeding and production herds from 2007 to 2010.

Methods

Data description

The Specific Pathogen Free (SPF) System was implemented in Denmark in 1971. It is a voluntary health program with established rules for monitoring Enzootic pneumonia, Porcine pleuropneumonia, Swine dysentery, Atrophic rhinitis, PRRS, mange and lice [17]. This program is primarily based on serological testing performed on a regular basis according to the herd type: the breeding herds (including nucleus and multiplier herds) are tested on a monthly basis and are classified as "red" herds; the production herds (including farrow-to-finisher and finisher herds) are tested every 12 months and classified as "blue" herds. The "red" and "blue" are designation used within the SPF system to classify the herds according to its herd health status. For each testing is necessary to take individual blood samples from 10 animals (5 gilts and 5 sows) and from 20 animals randomly selected within the herd for the red and blue herds respectively (personal communication, C.S. Kristensen, 2014). The SPF herds represent about 40 % of all Danish swine herds, but since many large farms are enrolled, 73 % of the Danish sows are included [18].

The laboratory submissions are requested according to the SPF status of the herd. For non-SPF herds, the veterinarians can decide which serological test to request, and at what interval. The outcome of this decision will depend on the overall objective and the costs associated with the different serology tests.

Laboratory submissions stored in the DTU Vet information management system in the period from 1 January 2007 to 31 December 2010 were extracted. Each laboratory submission consisted of individual blood samples collected from the same herd on the same day. Only submissions with between 2 and 60 individual blood samples tested by serological tests ELISA and/or IPMA for one or both PRRSV strains were included in the analysis. A total of 27,854 laboratory submissions tested for PRRS at the National Veterinary Institute were included in the analysis, representing a total of 879,327 serological tests performed on 404,029 individual blood samples collected from a total of 4702 Danish swine herds.

The laboratory submissions were merged with the SPF system database in order to classify the herds into

breeding and production herds. All red herds in the SPF system database were classified as breeding herds ($N = 264$); the remaining herds (blue SPF herds and non-SPF herds) were classified as production herds ($N = 4438$). The herd classification was year-specific since the SPF health status can change over time.

Ethics approval

The study was conducted using surveillance data and did not involve experiments on animals. The serum samples used for the study were obtained from blood samples voluntarily collected for monitoring PRRS. From an ethical perspective, all of the material collected and used as part of this study was outside the scope of Directive 2010/63.

PRRSV status

The herd PRRSV status in each laboratory submission was defined based on the cut-off for individual blood tests, in addition to the herd-level cut-off, which establishes the proportion of PRRSV seropositive samples (i.e. animals) within the herd. This approach was performed due to the recognized cross-reactivity between serological tests for the two PRRSV types [19], and the possible co-existence of PRRSV type 1 and 2 within herds [20].

For herds with more than one submission per month, the latest submission within the month was used to classify the herd.

Individual blood samples classification

At DTU Vet, in-house ELISAs and IPMAs were used to test for PRRSV antibodies, enabling us to distinguish between PRRSV type 1 and PRRSV type 2 specific antibodies.

The blocking ELISAs were performed according to [19, 21]; ELISA plates were separately coated with either PRRSV type 1 or 2 antigens, and the individual test serum samples were added to both the type 1 and the type 2 ELISA-plates. After incubation with the samples, biotinylated polyclonal swine-IgG directed against either PRRSV type 1 or 2, respectively, was added to the plates. For final development, peroxidase conjugated streptavidine and TMB were used, and colorimetric reactions were then measured based on optical density (OD). Results were considered positive if the OD% ≤44. Both ELISAs were run in parallel for the same sample and if the test result was positive for at least one type, the type 1/type 2 ratio was determined based on the obtained OD values in order to distinguish between the two PRRSV types. Ratios below 1.3 indicated the presence of type 1 PRRSV whereas ratios above 1.9 was an indication of type 2 [19]. Ratios between these values did not allow for distinction between the two PRRSV types.

The IPMA technique is described by [11]. In summary, the IPMA plates were prepared with MARC-145 cell lines, fixed with either PRRSV type 1 or 2 [21]. These plates were then incubated with serial sample dilutions from 50 to 6250. The enzyme peroxidase was used to catalyse a chemical reaction to color PRRSV specifically stained cells, and the plates were examined under a microscope. Specific staining of infected cells indicated the presence of PRRSV antibodies.

Serological tests with missing results in the database were excluded from the analysis ($N = 6202$).

Each individual blood sample was classified as PRRSV type 1 seropositive, PRRSV type 2 seropositive, PRRSV type 1 and 2 seropositive or seronegative according to the following criteria:

- Samples only tested by IPMA were classified according to [21];
- Samples tested by both ELISAs were classified based on the ratio type 1/type 2 according to [19];
- Samples with ratios between 1.3 and 1.9, were classified as both PRRSV type 1 and 2 seropositive;
- For samples tested by ELISA and IPMA, the IPMA results were prioritized in order to identify the PRRSV type;
- Samples tested only against one PRRSV strain by ELISA or IPMA were classified based only on those results.

Herd-level PRRSV classification

The herd-level PRRSV status was defined based on the number of PRRSV seropositive samples as suggested by [22]. The number of individual blood samples tested by ELISA and IPMA to classify the herd PRRS status varied according to the total number of individual blood samples tested per herd.

For herds with animals tested for both strains by IPMA, the classification was made following a comparison of titers in IPMA-PRRSV type 1 and IPMA-PRRSV type 2 in each individual sample. Herds were defined as PRRSV type 2 seropositive if the number of individual blood samples with IPMA-PRRSV type 2 ≥ IPMA-PRRSV type 1 per submissions [number of individual blood samples tested per submission] was equal or higher than 2 [2–5], 3 [6–15], 4 [16–18], 5 [29–35], 6 [36–45] and 7 [46–60]. Herds were defined PRRSV type 1 seropositive if the number of individual blood samples with IPMA-PRRSV type 1 > IPMA-PRRSV type 2 was equal or higher than 2 [2–7], 3 [8–15], 4 [16–29], 5 [30–45] and 6 [46–60]. For those submissions where IPMA was used to test for only one PRRSV type, the herds were considered to be PRRSV type 2 seropositive when IPMA-PRRSV type 2 titers ≥ 1250 was equal or higher than 2 [2–5], 3 [6–10], 4 [11–15], 5 [16–21], 6 [22–28], 7 [29–36], 9 [37–46] and 11 [47–60]. In addition, the cut-off points to classify herds as PRRSV type 1 seropositive were 2 [2–5], 3 [6–10], 4 [11–15], 5 [16–22], 6

[23–29], 7 [30–36], 8 [37–44], 9 [45–53] and 10 [54–61] IPMA-PRRSV type 1 titers ≥ 1250 in individual samples.

For laboratory submissions with individual blood samples tested only by ELISA, the herds were classified as PRRSV type 1 seropositive if they had at least 2 [2–19], 3 [20–39], 4 [40–59] or 5 [60] individual blood samples with a ratio <1.2. If the herds had at least 2 samples with a ratio ≥ 2 at any sample size, these herds were classified as PRRSV type 2 seropositive.

For a herd to be classified as PRRSV seronegative, all individual blood samples must test seronegative for both PRRSV types in both tests (ELISA and IPMA).

Statistical analysis

PRRSV seroprevalence in herds submitting samples

The seroprevalence of PRRSV type 1 and 2 in herds submitting samples was calculated on a monthly basis, where the number of PRRSV positive herds was divided by the total number of herds tested for PRRSV in that specific month.

Seroconversion rate in breeding herds

According to 23], PRRSV antibody titers reach the lower limits of detection at around 324 days post-inoculation (PI). Therefore, the breeding herds were classified as newly PRRSV seropositive if they had been seronegative in the previous 12 months. The number of new positive herds was modelled assuming a negative binomial distribution according to the following model:

$$Y \sim \mu + offset\ (\log(tar)) \quad (1)$$

where Y is the number of new positive herds per month from January 2008 to December 2010, μ is the intercept of the model and *tar* is the average time at risk in the previous 12 months. The average time at risk was calculated for each month based on the average number of previous months in which the herds were PRRS seronegative (i.e. classified as susceptible).

Herd identification

The herd identification number was used to obtain the geographic coordinates (UTM EUREF89, zone 32) from the CHR (Central Husbandry Register) database. Herds with missing location data ($N = 107$) were omitted from the spatial analysis.

PRRSV relative risk maps

PRRSV relative risk maps were made to facilitate visualization of the spatial distribution of PRRSV type 1 and 2 seropositive and seronegative herds biannually from 2007 to 2010.

The odds of a herd at a given location c being PRRS positive were calculated as $p(c) = \lambda_1(c) / (\lambda_1(c) + \lambda_0(c))$, where λ_1 and λ_0 are the intensity functions of positive and negative herds respectively. The risk surfaces were created by calculating the ratio of intensity functions for positive and negative herds on a grid of 2 × 2 km cells. The kernel smoothing surfaces were calculated based on a Gaussian model [24]; no edge-correction was performed.

The specification of the bandwidth is more important than the choice of kernel function [25]. Therefore, the median of specific biannual bandwidths were calculated for each PRRSV type and used to perform kernel smoothing, in order to identify any temporal differences.

Cluster analysis

Retrospective Space Scan Statistics [26] were used to identify local spatial clusters of PRRSV type 1 and type 2 seropositive herds biannually from 2007 to 2010. This method has been used in veterinary medicine to identify PRRSV outbreaks in United States [27] and Canada [28]. The Bernoulli model was used since the herds were classified as either PRRSV type 1 and 2 seropositive (cases) or seronegative (controls). The scanning window was circular and no overlapping clusters were permitted. The analysis was repeated five times using different maximum population sizes (i.e. herds) at risk, including 5, 15, 25, 35 and 50 %. The p-value was obtained using 999 Monte Carlo simulations and a 5 % significance level was used based on a likelihood ratio test.

All analyses were performed in R version 3.1.1 [29]. Kernel smoothing densities were made using the'sm package' [30] for estimating the bandwidth and'spatialkernel package' [31] for kernel estimation. Spatial cluster analysis was based on SatScan version 9.3.1 32].

Results

Data description

The total number of herds, laboratory submissions and blood samples tested per year during the period from January 2007 to December 2010 are listed in Table 1. On average, 2776 production and 230 breeding herds were tested annually; the median number of annual submissions was 12 for breeding herds and 1 for production herds. The average time between two consecutive submissions was 1 month (maximum of 37) for breeding herds and 11.33 months (minimum of 1 and maximum of 46) for production herds. The descriptive statistics of PRRS serological diagnostic tests performed are described in Table 2.

The total number of breeding herds submitting samples on a monthly basis between 2007 and 2010 did not vary from year to year. In contrast, the total number of tested production herds followed a seasonal trend (Fig. 1). In general, the number of positive herds followed the same trend as the total number of herds tested. The number of herds testing seropositive was higher for PRRSV type 1

Table 1 Descriptive statistics by frequency of laboratory submissions sent to DTU Vet laboratory for testing PRRSV during the period from 2007 to 2010 for breeding (Breed) and production (Prod) herds. Each laboratory submission consisted of individual blood samples collected from the same herd on the same day

Year	2007		2008		2009		2010	
Herd type	Breed	Prod	Breed	Prod	Breed	Prod	Breed	Prod
Total number of tested herds	237	2982	233	2729	228	2720	220	2673
Median number of submissions per herd (Q1 – Q3)	12 (12–13)	1 (1–1)	12 (12–13)	1 (1–1)	12 (12–13)	1 (1–1)	12 (12–13)	1 (1–1)
Total number of samples	31,505	73,561	33,430	69,233	30,572	67,640	33,420	64,668
Median number of samples per herd (Q1–Q3)	10 (10–15)	20 (17–20)	10 (10–15)	20 (16–20)	10 (10–10)	20 (15–20)	10 (10–15)	20 (15–20)

than for PRRSV type 2. This applied to both production and breeding herds, the only exceptions being in April 2007 and June 2010, when the number of PRRSV type 2 seropositive production herds increased to the same value as PRRSV type 1 seropositive production herds.

No herds were classified as positive for both PRRSV types simultaneously in the same month.

Figure 2 shows the distribution of all herds tested for PRRSV based on serology from 2007 to 2010. The majority of these herds were located in Jutland, reflecting the higher pig density in this region.

PRRSV seroprevalence

The apparent PRRSV seroprevalence in tested herds appears to be higher for PRRSV type 1 than for PRRSV type 2 from 2007 to 2010 (Fig. 3). There appeared to be an overall decrease in the seroprevalence for both PRRSV types (though this was not tested for statistical significance). The monthly average PRRSV type 1 seroprevalence was 0.09 (minimum of 0.05; maximum of 0.13) in breeding herds and 0.20 (minimum of 0.14; maximum of 0.26) in production herds; PRRSV type 2 had an average of 0.03 (minimum of 0.01; maximum of 0.09) in breeding herds and 0.09 (minimum of 0.05; maximum of 0.13) in production herds.

PRRSV seroconversion rate in breeding herds

The total number of new PRRSV seropositive breeding herds per month is presented in Fig. 4. The monthly seroconversion rate followed a constant pattern for both PRRSV types, being higher for type 1 (average of 0.65 herds per 100 herds) than type 2 (average of 0.21 herds per 100 herds).

Smoothed relative risk surfaces

The smoothed relative risk surface of the probability of swine herds being positive for both PRRS-strains changed spatiotemporally (Fig. 5). The median values for the biannual bandwidths were h = (29,576.49; 31,069.79) and h = (30,885.97; 31,401.67) for PRRSV type 1 and 2, respectively.

In general, the extent of areas with higher relative risk decreased from 2007 to 2010 for both PRRSV types.

Regarding PRRSV type 1 relative risk distribution between July 2007 and December 2008, the areas with the highest relative risk were located in the west of Denmark.

Table 2 Descriptive statistics by frequency (N) and percentage (%) of PRRS serological diagnostic tests performed from 2007 to 2010

Year		2007		2008		2009		2010	
	Serological test	*N*	%	*N*	%	*N*	%	*N*	%
Total number of serological tests performed	ELISA-type 1	101,925	44.1	100,172	44.2	95,133	44,7	94,493	45.2
	ELISA-type 2	101,924	44.1	100,174	44.2	95,133	44.7	94,495	45.2
	IPMA-type 1	14,307	6.2	14,426	6.4	12,421	5.8	10,830	5.2
	IPMA-type 2	12,804	5.5	11,775	5.2	10,225	4.8	9040	4.3
Total number of samples		105,066	-	102,663	-	98,212	-	98,088	-
Number of samples only tested by:	ELISA-type 1	1	<0.01	1	<0.01	0	0.00	0	0.00
	ELISA-type 2	0	0.00	3	<0.01	0	0.00	0	0.00
	IPMA-type 1	783	0.8	605	0.6	1021	1.0	1095	1.1
	IPMA-type 2	402	0.4	168	0.2	532	0.5	784	0.8
Number of samples tested by doubled ELISA		89,529	85.2	87,156	84.9	84,569	86.1	85,659	87.3
Number of samples tested by ELISA and IPMA		12,395	11.8	13,015	12.7	10,564	10.8	8837	9.0

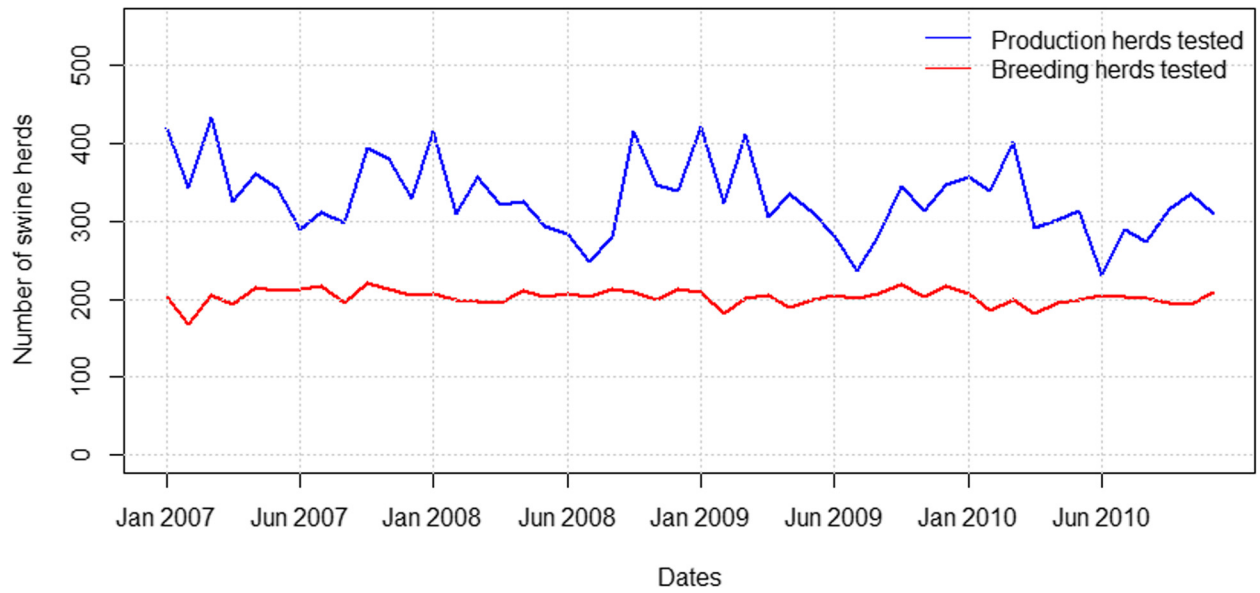


Fig. 1 Number of production and breeding herds tested for PRRSV per month from January 2007 to December 2010

During the remaining periods, the same areas covered a smaller geographic area and were located in the northwestern and the southwestern parts of Denmark.

The overall relative risk was lower for PRRSV type 2 when compared to PRRSV type 1. In this case, the highest relative risk areas had a larger extent in 2007, which later decreased. In the following years, these areas remained in the western part of the country.

Spatial cluster analysis

The significant spatial clusters of PRRSV type 1 and 2 are shown in Fig. 6. The descriptive statistics of these

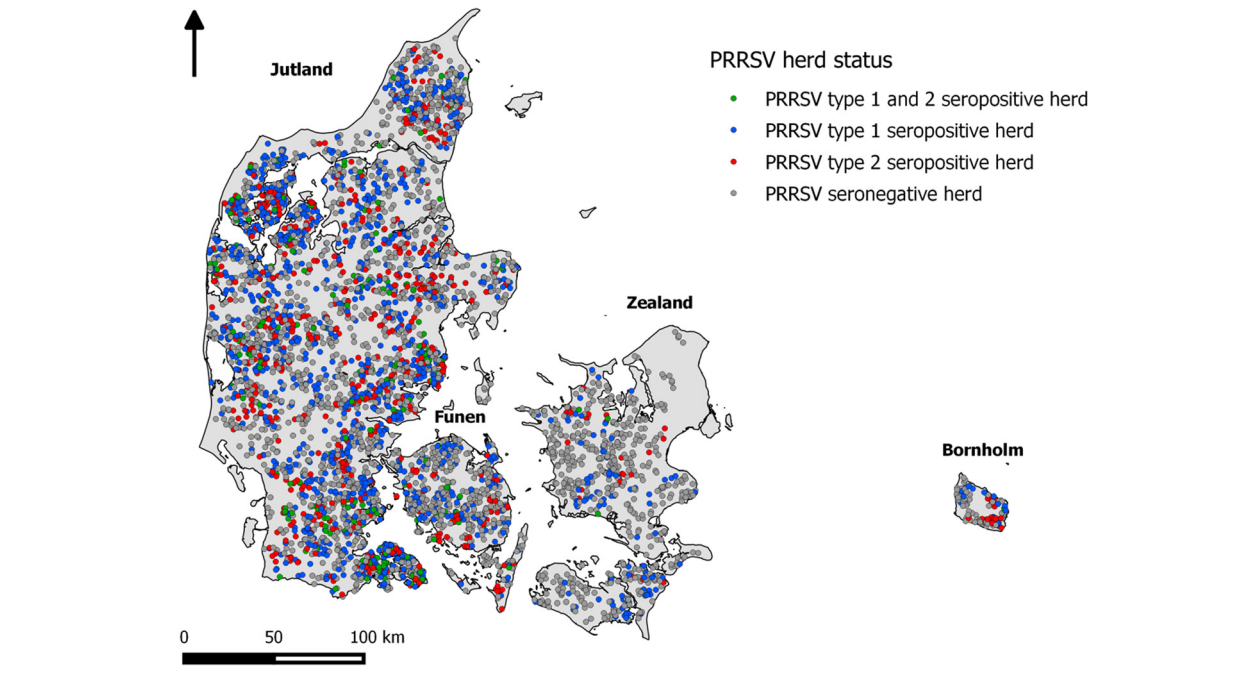


Fig. 2 PRRSV herd status distribution from 2007 to 2010, including only herds submitting samples. Herds were classified as PRRS seropositive if they were positive during a minimum of 1 month between 2007 and 2010; herds classified as seropositive for both strains during this period were labeled in green; negative herds (*grey*) were not classified as PRRS positive during the period of study

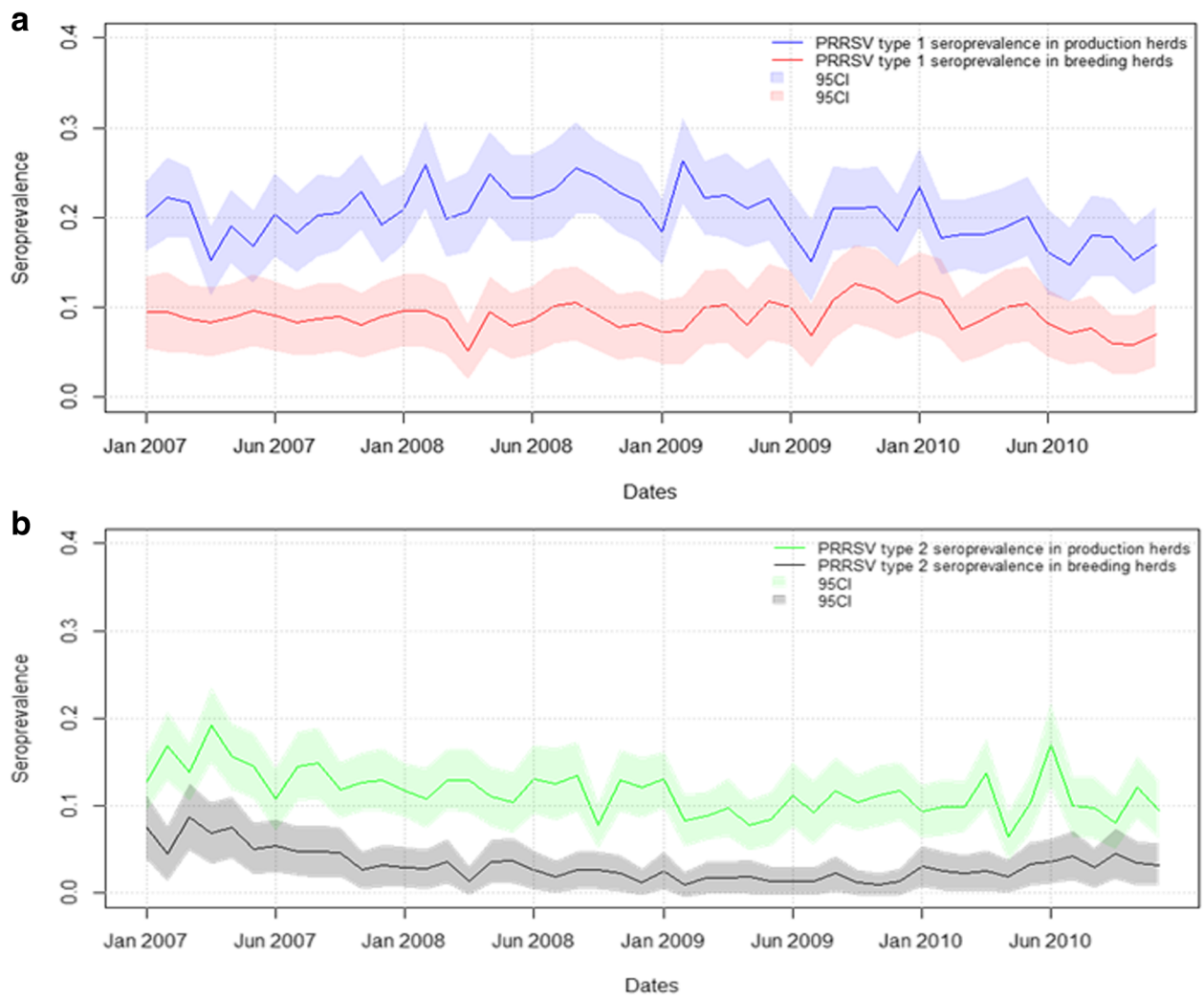


Fig. 3 Monthly PRRSV seroprevalence in Danish pig herds. The figure illustrated the monthly PRRSV type 1 (**a**) and type 2 (**b**) seroprevalence in production and breeding herds from January 2007 to December 2010

significant clusters are presented in Additional file 1. Increasing the maximum spatial window size from 5 to 15 % of the population at risk resulted in the aggregation of two or more secondary clusters for some 6-month periods. For higher percentages of populations at risk, the number and size of clusters did not change (results not shown). Several clusters were found for each 6-month period. The spatiotemporal pattern for PRRSV type 1 clusters changed over time, except for those located in the northwest of Jutland. Similarly, the locations and sizes of PRRSV type 1 clusters also altered over time from January 2007 to December 2010. In this case, there was a constant cluster in the central eastern part of Jutland.

Discussion

This is the first study to use surveillance data from laboratory submissions to describe the occurrence of PRRSV in Denmark. The use of laboratory submission records was essential in order to gather previous information and assess the spatial distribution of PRRSV seropositive herds. Such information might be used to evaluate the efficiency of control strategies implemented on a local or regional basis. Using laboratory submission data from a surveillance program might help to identify and record new PRRSV cases in a more reliable way than other sources of information.

The frequency of testing and the type of serological test requested depends on the Danish herd status (SPF or non SPF) and the purpose (PRRSV surveillance or diagnostic). For example, if the objective is to detect infection early, IPMA is normally requested, because high IPMA values are indicative of recent infection as ELISA titers tend to persist for a longer time period [19]. Animals can also be tested for trading purposes, to maintain/gain an SPF certificate, and prior to being

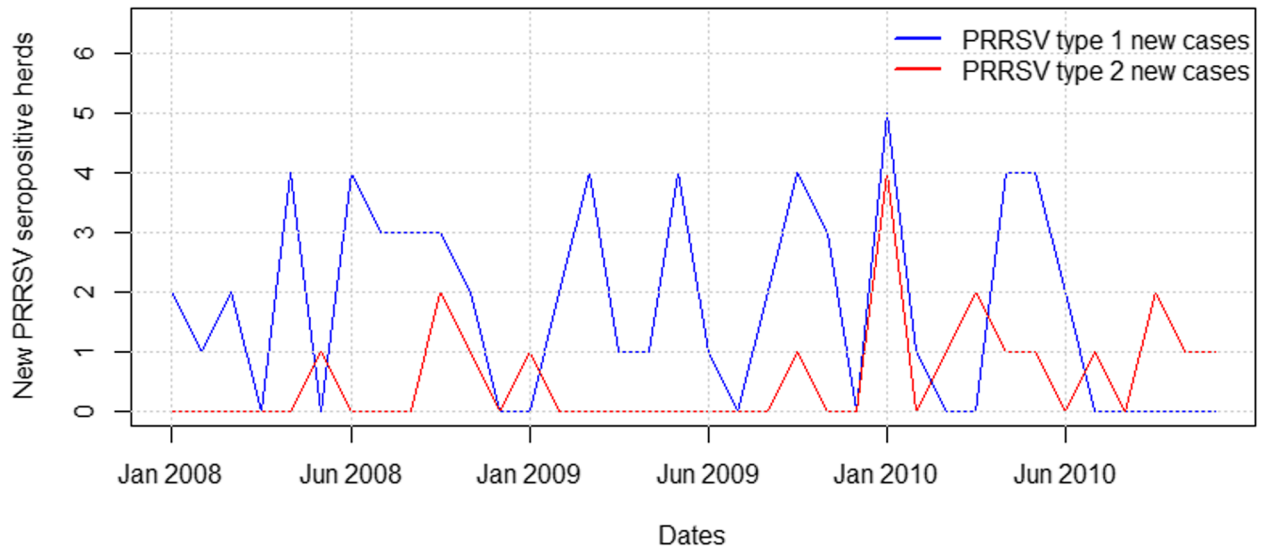


Fig. 4 Number of new PRRSV type 1 and 2 seropositive breeding herds from 2008 to 2010

introduced in a farm after sanitation procedures. Different reasons for testing might explain the variation in frequency of laboratory submissions in both herd types over the study period. It is our general assumption that herds submitting samples for surveillance or diagnosis have a higher health status compared to those herds that never submit samples (personal communication, C.S. Kristensen, 2014). Therefore, the overall seroprevalence of PRRSV in Danish swine herds may be underestimated based on the submission data used in the present study.

The serological tests used in this study do not differentiate between antibodies from the naturally infected pigs, and those that have been vaccinated against PRRS. However, it is reasonable to assume that an observed seroconversion will be related to a preceding natural infection with PRRSV of homologous type, since vaccination is unlikely in a PRRSV-negative herd. In this study, we therefore argue that an observed seroconversion must initially have been caused by a natural infection, yet we are aware that we might have been measuring vaccine antibodies at the time of sampling.

Herds were classified as seropositive for PRRSV type 1 or type 2 based on the number of seropositive samples per submission. Individual blood samples tested by double ELISAs and IPMAs were classified based on the latest serological results, in order to focus on the most recent PRRSV status [33] demonstrated that high titers in IPMA are indicative of new PRRSV infections. In addition, [19] demonstrated that detection in the IPMA decreases after 3–4 months post infection, therefore making ELISA a more sensitive test to detect late immune responses. In our study, seroprevalence and the seroconversion rate were calculated based on both types of serological tests in order to have the maximum information available over time for each herd.

In this study, the seroprevalence was calculated on a monthly basis to describe the occurrence of PRRSV type 1 and 2 in Denmark. Variation in the seroprevalence for both types might be explained by variation in the number of herds tested per month and the SPF status. Figure 3 indicates an overall decrease in prevalence for both PRRSV strains in both herd types. A recent study by [18] based on information available from the SPF system database estimated that 65 % of sow herds and 60 % of finisher herds in Denmark are PRRSV negative. In our study, these types of herds were classed as production herds, and our results agreed with these findings. The high biosecurity and monthly surveillance of the breeding herds might explain the relatively constant seroconversion rate in Fig. 5.

The information available to us from the SPF system database only provided the herd status on the 31 December of each year. It is therefore unknown whether these herds were under sanitation controls, or if their SPF status changed over time, resulting in possible variation in the frequency of PRRSV testing, which in turn could have influenced the number of new PRRSV seropositive herds. For example, if the SPF status of a PRRSV seropositive breeding herd changed, the herd would be included as a different type in the analysis. This would result in an unknown PRRS status for a period of time, and classification as newly PRRS seropositive when gaining the same SPF status. This happened when a red herd in the SPF

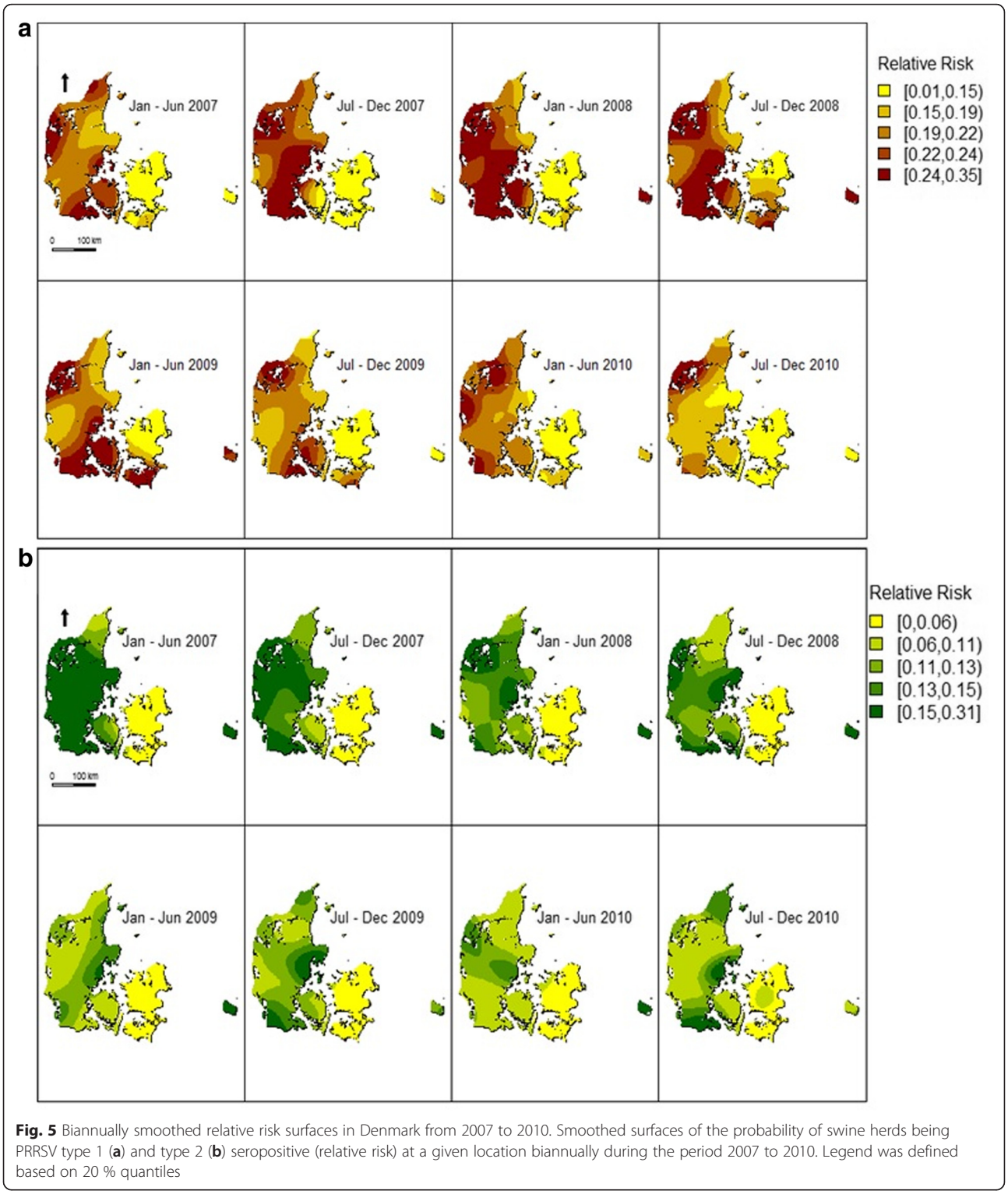


Fig. 5 Biannually smoothed relative risk surfaces in Denmark from 2007 to 2010. Smoothed surfaces of the probability of swine herds being PRRSV type 1 (**a**) and type 2 (**b**) seropositive (relative risk) at a given location biannually during the period 2007 to 2010. Legend was defined based on 20 % quantiles

database (i.e. a breeding herd) lost their SPF status for a period of time. It was not possible to establish the seroconversion rate for production herds due to the long period of time between consecutive laboratory submissions.

The relative risk distribution maps changed over time as a consequence of the changes in the seroprevalence. The general decline in the extent of areas with higher relative risk for both PRRSV types followed the same trend as observed for the seroprevalence.

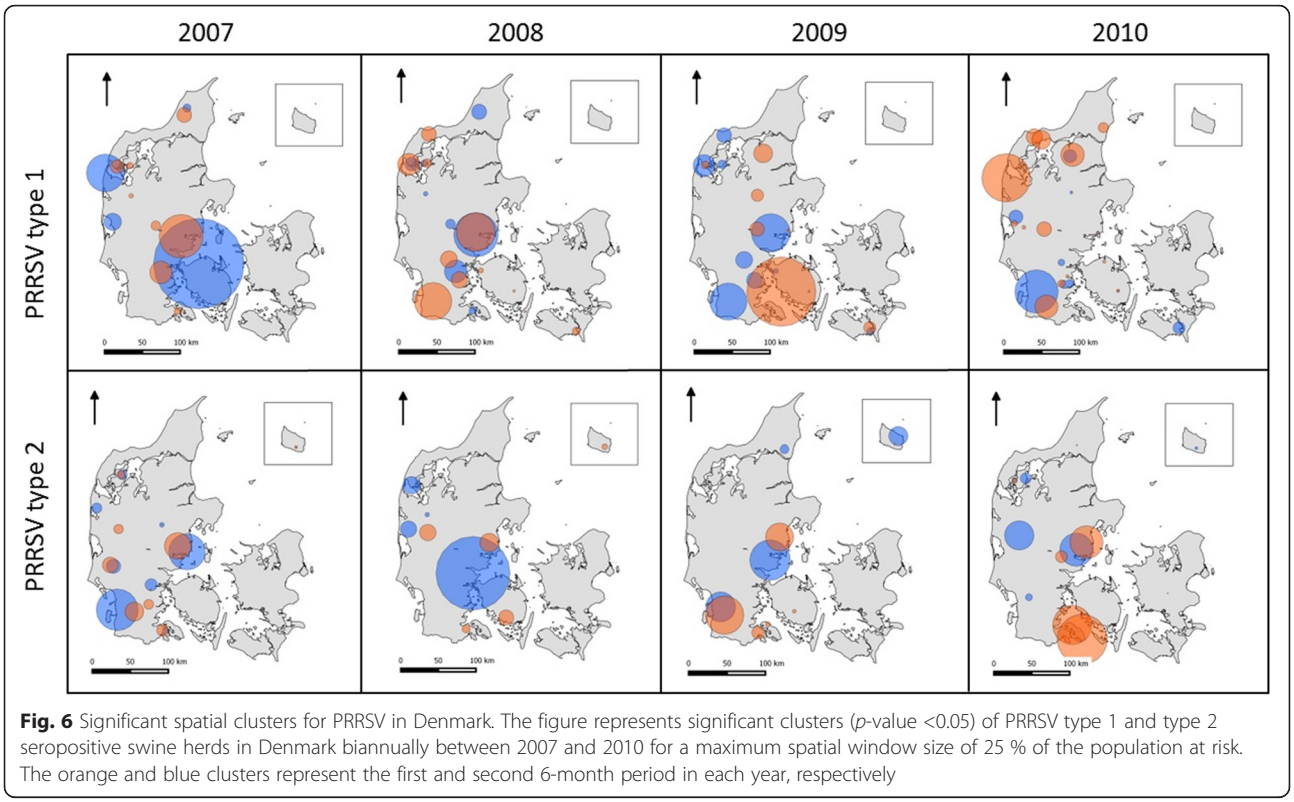


Fig. 6 Significant spatial clusters for PRRSV in Denmark. The figure represents significant clusters (*p*-value <0.05) of PRRSV type 1 and type 2 seropositive swine herds in Denmark biannually between 2007 and 2010 for a maximum spatial window size of 25 % of the population at risk. The orange and blue clusters represent the first and second 6-month period in each year, respectively

Conclusions

This study described the occurrence of PRRSV in Denmark from 2007 to 2010, based on laboratory submission data. PRRSV type 1 seroprevalence was consistently higher than type 2 seroprevalence in both production and breeding herds. The relative risk maps showed changes in the spatial distribution of both PRRSV types over time. Significant spatial clusters were consistently found in Denmark, suggesting that PRRSV is endemic in these areas. Furthermore, relative risk distribution maps revealed different patterns over time as a consequence of the changes on the seroprevalence.

Our findings might help decision makers to re-evaluate their conclusions on the spread of the disease and assess the efficiency of the implemented control strategies.

Additional file

Additional file 1: Descriptive statistics of significant spatial clusters for PRRSV type 1 and 2. (DOCX 78 kb)

Abbreviations

ELISA: enzyme-linked immunosorbent assay; IPMA: immunoperoxidase monolayer assay; PRRS: porcine reproductive and respiratory syndrome; PRRSV: porcine reproductive and respiratory syndrome virus; SPF: specific pathogen free system.

Competing interests

None of the authors have any financial or personal relationships that could inappropriately influence or bias the content of the paper.

Authors' contributions

ACLA performed the data management, data analysis and wrote the article; TH contributed to the data management, coordinated the study and critically reviewed the manuscript; KTL and LAEL provided advice on interpretation of serological test results, contributed to understanding the contents of the DTU Vet database and critically reviewed the manuscript; CSK provided additional data, PRRSV surveillance and swine herd management advice, and critically reviewed the manuscript; NT coordinated the study and critically reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Access to the data used in this study was provided by the National Veterinary Institute – Technical University of Denmark. The authors would like to thank all participants in the research group for the valuable comments and input at all stages of the research. Access to the SPF data was provided by the Danish Pig Research Centre.

Author details

[1]Section for Epidemiology, National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark. [2]Section for Diagnostic and Scientific Advice, National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark. [3]Danish Pig Research Centre, SEGES P/S, Vinkelvej 11, 8620 Kjellerup, Denmark. [4]Section for Virology, National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark.

Received: 19 May 2015 Accepted: 16 December 2015
Published online: 21 December 2015

References

1. Balka G, Hornyák A, Bálint A, Kiss I, Kecskeméti S, Bakonyi T, et al. Genetic diversity of porcine reproductive and respiratory syndrome virus strains circulating in Hungarian swine herds. Vet Microbiol. 2008;127:128–35.

2. Kim HK, Park SJ, Rho SM, Han JY, Nguyen VG, Park BK. One year's study of dynamic and evolution of types I and II PRRSV in a swine farm. Vet Microbiol. 2011;150:230–8.
3. Nieuwenhuis N, Duinhof TF, van Nes A. Economic analysis of outbreaks of porcine reproductive and respiratory syndrome virus in nine sow herds. Vet Rec. 2012;170:225.
4. Holtkamp DJ, Kliebenstein JB, Neumann EJ, Zimmerman JJ, Rotto HF, Yoder TK, et al. Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers, vol. 21. 2013. p. 72–84.
5. Thanapongtharm W, Linard C, Pamaranon N, Kawkalong S, Noimoh T, Chanachai K, et al. Spatial epidemiology of porcine reproductive and respiratory syndrome in Thailand. BMC Vet Res. 2014;10:174.
6. Wensvoort G. Mystery swine disease in the Netherlands: The isolation of Lelystad virus. Vet Q. 1991;13:121–30.
7. Benfield D. Characterization of swine infertility and respiratory syndrome (SIRS) virus (isolate ATCC VR-2332). J Vet Diagn Investig. 1992;4:127–33.
8. Murtaugh MP, Elam MR, Kakach LT. Comparison of the structural protein coding sequences of the VR-2332 and Lelystad virus strains of the PRRS virus. Arch Virol. 1995;140:1451–60.
9. Goldberg TL, Weigel RM, Hahn EC, Scherba G. Associations between genetics, farm characteristics and clinical disease in field outbreaks of porcine reproductive and respiratory syndrome virus. Prev Vet Med. 2000;43:293–302.
10. Han D, Hu Y, Li L, Tian H, Chen Z, Wang L, et al. Highly pathogenic porcine reproductive and respiratory syndrome virus infection results in acute lung injury of the infected pigs. Vet Microbiol. 2014;169:135–46.
11. Bøtner A, Nielsen J, Bille-Hansen V. Isolation of Porcine Reproductive and Respiratory Syndrome (PRRS) virus in a Danish swine herd and experimental infection of pregnant gilts with the virus. Vet Microbiol. 1994;40:351–60.
12. Bøtner A, Toft PS. Porcin Reproduktions - Og Respirations sygdom (PRRS) - En oversigtsartikel med fokus på den danske situation og erfaringer med anvendelse af en levende vaccine. Dansk Veterinærtidsskrift - Videncenter Svineproduktion. 1999;82(5):175–81.
13. Bøtner A, Strandbygaard B, Sørensen KJ, Have P, Madsen KG, Madsen ES, et al. Appearance of acute PRRS-like symptoms in sow herds after vaccination with a modified live PRRS vaccine. Vet Rec. 1997;141:497–9.
14. Kvisgaard LK, Hjulsager CK, Kristensen CS, Lauritsen KT, Larsen LE. Genetic and antigenic characterization of complete genomes of Type 1 Porcine Reproductive and Respiratory Syndrome viruses (PRRSV) isolated in Denmark over a period of 10 years. Virus Res. 2013;178:197–205.
15. Kvisgaard LK, Hjulsager CK, Brar MS, Leung FCC, Larsen LE. Genetic dissection of complete genomes of Type 2 PRRS viruses isolated in Denmark over a period of 15 years. Vet Microbiol. 2013;167:334–44.
16. Wagner MM, Moore AW, Aryel RM. Handbook of biosurveillance. London: Elsevier Academic press ; 2006:217-234.
17. Specific Pathogen Free System (SPF - Danmark) [https://www.spf.dk/].
18. Kristensen C, Jensen P, Christiansen M. Udbredelse af PRRS-negative besætninger i Danmark 2013. Videcenter for Svineproduktion; 2014:1–10.
19. Sørensen KJ, Strandbygaard B, Bøtner A, Madsen ES, Nielsen J, Have P. Blocking ELISA's for the distinction between antibodies against European and American strains of porcine reproductive and respiratory syndrome virus. Vet Microbiol. 1998;60:169–77.
20. Chaikhumwang P, Tantituvanont A, Tripipat T, Tipsombatboon P, Piriyapongsa J, Nilubol D. Dynamics and evolution of highly pathogenic porcine reproductive and respiratory syndrome virus following its introduction into a herd concurrently infected with both types 1 and 2. Infect Genet Evol. 2015;30:164–74.
21. Sørensen KJ, Bøtner A, Smedegaard Madsen E, Strandbygaard B, Nielsen J. Evaluation of a blocking Elisa for screening of antibodies against porcine reproductive and respiratory syndrome (PRRS) virus. Vet Microbiol. 1997;56:1–8.
22. Mortesen S, Strandbygaard B, Bøtner A. Herd-level tests for infection with the European and American type of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) based on serology. Herd Test Dis Spread Model Swine Popul PhD Thesis R Vet Agric Univ Frederiksberg, Denmark. 2000. p. 47–69.
23. Yoon K-J, Zimmerman JJ, Swenson SL, McGinley MJ, Eernisse KA, Brevik A, et al. Characterization of the Humoral Immune Response to Porcine Reproductive and Respiratory Syndrome (PRRS) Virus Infection. J Vet Diagn Invest. 1995;7:305–12.
24. Berman M, Diggle P. Estimating weighted integrals of the second-order intensity of a spatial point process. J R Stat Soc. 1989;51(Series B (Methodological),No 1):81–92.
25. Berke O. Exploratory spatial relative risk mapping. Prev Vet Med. 2005;71:173–82.
26. Kulldorff M. A spatial scan statistic. Commun Stat - Theory Methods. 1997; 26:1481–96.
27. Tousignant SJP, Perez AM, Lowe JF, Yeske PE, Morrison RB. Temporal and spatial dynamics of porcine reproductive and respiratory syndrome virus infection in the United States. Am J Vet Res. 2015;76:70–6.
28. Rosendal T, Dewey C, Friendship R, Wootton S, Young B, Poljak Z. Spatial and temporal patterns of porcine reproductive and respiratory syndrome virus (PRRSV) genotypes in Ontario, Canada, 2004–2007. BMC Vet Res. 2014;10:83.
29. R Core Team. R. A Language and environment for statistical computing. 2014.
30. Bowman A, Azzalini A. R package sm: Smoothing methods for nonparametric regression and density estimation (version 2.2–4). 2013.
31. Zheng P, Diggle P. Package Spatialkernel: Nonparameteric estimation of spatial segregation in a multivariate point process (version 0.4–19). 2013.
32. Kulldorff M. SaTScan TM 9.3.1: Software for the spatial and space-time scan statistics. 2014.
33. Bøtner A. Diagnosis of PRRS. Vet Microbiol. 1997;55:295–301.

Manuscript 3

Mortality in Danish Swine herds: spatio-temporal clusters and risk factors

Ana Carolina Lopes Antunes[1*], Annette Kjær Ersbøll[2], Kristine Bihrmann[2], Nils Toft[1]

[1]Section for Diagnostics and Scientific Advice, National Veterinary Institute - Technical University of Denmark, Denmark

[2]National Institute of Public Health, University of Southern Denmark, Denmark

*Corresponding author.

Submitted to Preventive Veterinary Medicine.

Mortality in Danish Swine herds: spatio-temporal clusters and risk factors.

Ana Carolina Lopes Antunes

Section for Diagnostics and Scientific Advice, National Veterinary Institute, Technical University of Denmark.

Annette Kjær Ersbøll

National Institute of Public Health, University of Southern Denmark, Denmark.

Kristine Bihrmann

National Institute of Public Health, University of Southern Denmark, Denmark.

Nils Toft

Section for Diagnostics and Scientific Advice, National Veterinary Institute, Technical University of Denmark.

Corresponding author: Ana Carolina Lopes Antunes, Section for Diagnostics and Scientific Advice, National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870, Frederiksberg C, Denmark. Email: aclan@vet.dtu.dk

Abstract

The aim of this study was to explore spatio-temporal patterns in swine mortality from Danish swine herds from December 2013 to October 2015 and discuss the use of mortality for syndromic surveillance in Denmark. The potential of using mortality for disease monitoring have previously been explored in the context of syndromic surveillance. The value of using mortality generated on a regular and mandatory basis for all swine herds remains unexplored for swine surveillance in Denmark. A total of 5016 farms were included in the analysis, corresponding to 1896 weaner herds, 1490 sow herds and 3839 finisher herds. The spatio-temporal analysis included data description on space and time and cluster analysis for three age groups: weaners (up to 30 kg), sows and finishers. Logistic regression models were used to assess the association of potential factors for herds being included inside a cluster.

A large number of *single-herd* clusters i.e. clusters with only one herd, and fewer *multiple-herd* clusters, i.e. clusters with at least 2 herds included, were found. Factors such as herd size, farm type, SPF status and Atrophic rhinitis had an impact on herds being inside vs outside *multiple-herd* clusters.

The presence of *single-herd* clusters might indicate welfare and disease issues, while *multi-herd* clusters could be suggestive of the presence of infectious diseases within the cluster area.

There is potential of using mortality for disease surveillance. However, detected clusters might not be due to disease, but the result of changes such as herd management. Further

analysis to explore other spatio-temporal monitoring methods is needed before incorporating mortality in a Danish disease monitoring system.

Keywords: Mortality, swine, spatio-temporal analysis, risk factors.

1. Introduction

Over the last years an increased number of studies in veterinary syndromic surveillance have been performed. Data used for syndromic surveillance relate to non-specific health indicators that enable the early identification of the impact (or absence of impact) of animal health threats (Triple S Project, 2011). Such data are often recorded for other purposes than disease surveillance. Previous studies explored the potential of using different animal health register data such as laboratory submissions (Dórea et al., 2013), meat inspection (Dupuy et al., 2015; Vial and Reist, 2015) and mortality (Alba et al., 2015; Backer et al., 2011; Morignat et al., 2014; Perrin et al., 2010, 2012). Although mortality is routinely collected through national registers and movements to rending plants, it is rare to find syndromic surveillance systems implemented based on such data (Dupuy et al., 2013). The advantage of using mortality is that they are recorded to fulfill the European Commission requirements. All farmers are obliged to report their cadavers and have them removed to rendering plants for purposes of food safety and traceability in all Member States (European Union, 2000). This regulation ensures a continuous data flow and constitutes a strong basis for a surveillance system.

65 The value of using these data, generated on a regular basis and covering the entire swine
66 population, remains unexplored for swine disease surveillance in Denmark. The (near)
67 real- time monitoring of swine mortality could be used to identify potential changes in
68 temporal, spatial and spatio-temporal patterns, reflecting some underlying problem or
69 cause for concern.
70 Performing retrospective analysis of historical data might help decision makers to re-
71 evaluate their conclusions on the spread of the diseases and assess the efficiency of the
72 implemented control strategies. For example, knowledge of the spatial distribution of
73 herds with higher mortality can be used to facilitate control of diseases on local and
74 regional basis, by changing management routines, trade customs etc. Additionally,
75 exploring available data provides a cost-effective way to assess the potential and
76 limitations of using such data for real-time disease monitoring and surveillance in
77 Denmark.
78 Spatio-temporal methods have been used regularly in veterinary epidemiology. The
79 focus has been on retrospective analysis of spatial clusters and related risk factors (Lian
80 et al., 2007; Pfeiffer et al., 2007; Themudo et al., 2011). Methods such as Spatial scan
81 statistics (Kulldorff et al., 1998) can be used to detect purely spatial, temporal or spatio-
82 temporal clusters in the context of biosurveillance (Wagner et al., 2006). This method is
83 commonly used due to the freely available SaTScan software (Kulldorff, 2015). This
84 tool have been previously proved to identify disease outbreaks of (new) emerging
85 diseases, based on laboratory submission data for cattle in Great Britain (Hyder et al.,
86 2011).

The aim of the present study was to explore spatio-temporal patterns in mortality from Danish swine herds from December 2013 to October 2015. This included: i) descriptive analysis of spatial and temporal changes in the data; ii) identification of spatio-temporal clusters of increased mortality for Danish finishers, sows and weaners herds; and iii) examination of possible herd-level factors associated with being inside vs outside a cluster.

2. Materials and methods

Data

The Danish Veterinary and Food Administration calculate the Danish swine mortality. The swine mortality is calculated based on public data registered in the Central Husbandry Register (CHR) and the Swine Movement Database (SMD). The CHR is the Danish national database on farm demographics. A farm is defined as a single location with its own unique CHR number. For each farm, the postal address, the Cartesian geographical coordinates and the type of farm (i.e. production farm, hobby farm, organic farm, etc) are given. At a pig farm, the number of pen-places for animals in up to three different age groups (weaners (up to 30 kg), sows and finishers) is registered. We will refer to each age group individually as herds. All movements of swine in Denmark, including movements to rendering plants must be registered to the SMD. For each movement registered in the SMD, the date and the number of dead finishers and/or sows

are recorded. For weaners, the number of small and large containers with room for 7 or 9 weaners, respectively, is registered.

Swine mortality from December 2013 to October 2015 were provided by the Danish Veterinary and Food Administration. The data included variables such as the monthly total number of dead animals and the monthly total number of swine for each age group per farm. Information retrieved from the CHR database was used as a proxy of the monthly number of animals present in each farm for each of the three age groups. Information about the movements from farms to rending plants was used to estimate the number of dead animals for each farm. The monthly mortality was calculated for each age group as a proportion: the number of dead animals in a given age group divided by the number of animals recorded in the CHR for that same age group. We will refer to this proportion as mortality throughout the manuscript.

The Specific Pathogen Free (SPF) System is a voluntary health program with established rules for monitoring Enzootic pneumonia, Porcine pleuropneumonia, Swine dysentery, Atrophic rhinitis, Porcine Reproductive and Respiratory Syndrome (PRRS), mange and lice within farms with an SPF certificate (SPF farm) ("SPF-DANMARK," 2015). The disease monitoring is primarily based on clinically examination of a representative number of swine, as well as blood samples and nasal swabs for the relevant SPF diseases. The visits are conducted by veterinarians from the Pig Research Centre performed on a regular basis according to the herd type. Based on the results the herd is assigned to a given SPF status declaring disease freedom or infection status.

Data management

The mortality were merged with the CHR data in order to identify and include only production herds in the analysis and obtain the geographic coordinates (UTM EUREF89, zone 32) for each farm. Furthermore, only farms reporting ≥200 finishers, ≥50 sows or ≥200 weaners were included in the analysis. Data were split into three datasets by age groups and all analyses were performed independently for each age group. A farm could have different age groups and in this case, the same farm was included in different analyses. Sows and finisher herds with zero mortality in 12 consecutive months were excluded from the analysis. Weaner herds with zero mortality in 2 consecutive months were also excluded from the analysis. This decision was made to ensure that only active herds were included in the study.

Data extracted from the SPF System database on the 31st of December 2014 were merged with the mortality to define the SPF status and the presence, absence or unknown status of the diseases monitored within the SPF system. Due to the low number of positive herds for some diseases, only Enzootic pneumonia, Porcine pleuropneumonia, Atrophic rhinitis, and PRRS were included in the analysis.

The herd size (categorized for each age group), the Specific Pathogen Free status (SPF vs. non-SPF), farm type (categorized by the different age groups present in the farm) were defined for all herds included in the study. Additionally, the disease status, e.g. the absence or presence of Enzootic pneumonia, Porcine pleuropneumonia, Atrophic rhinitis, and PRRS was defined for all SPF herds, whereas non-SPF herds were classified as unknown disease status.

153

154 *Spatial interpolation*

155 The inverse distance weighted interpolation technique (IDW) was used to visualize the
156 spatial distribution of the mortality for each age group. The mortality at a given location
157 was estimated as the weighted average of the mortality within a certain distance with a
158 weighting function with the power parameter equal 2 (Huisman and By, 2009). The
159 maximum distance used was varied from 5km to 15km to evaluate the impact of the
160 maximum distance on the results: less spatial heterogeneity was found when using larger
161 distances while the opposite was found for smaller distances. Therefore, a maximum
162 distance of 7.5 km from the prediction location was defined based on the average
163 distance between the farms and the results were plotted using grid cells of 1km×1km.
164 The analysis was performed in R (version 3.1.1) (R Core Team, 2014) using the 'gstat
165 package' (version 1.1-2) (Pebesma and Graeler, 2016).

166

167 *Spatio-temporal local clustering analysis*

168 The Scan statistics is a powerful method for detecting spatial, temporal and spatio-
169 temporal clusters (Kulldorff, 2016). Retrospective Space-time Scan Statistics (Kulldorff
170 et al., 1998) was used to identify local spatial-temporal clusters of mortality for each age
171 group from December 2013 to October 2015. The Bernoulli model was used with the
172 number of dead animals from an age group for a given farm (i.e. unique location) as
173 cases and the number of animals reported in CHR from the same age group in the same
174 farm as controls. The scanning spatial window was circular and no overlapping clusters

were permitted. The analysis was performed defining the maximum spatial cluster size as 10%, 25% and 50% of the population at risk and a minimum of 1 month and a maximum of 90% of the study period for the temporal cluster size. The *p*-value was obtained using 999 Monte Carlo simulations and a 5% significance level was used based on a likelihood ratio test. The analysis was made in SatScan version 9.4.2 (Kulldorff, 2015).

When examining the clusters, it was noticed that a large number had zero km radius (purely temporal clusters). The spatio-temporal clusters in which only one farm was included (zero km radius) were named *single-herd clusters* and clusters with several herds included were named *multiple-herd* clusters throughtout the manuscript. A *post hoc* description of temporal patterns of mortality in herds included in the *single-herd* clusters was made to inspect which changes triggered the clusters using Scan Statistics.

Factors associated with the risk of a herd being inside vs ouside multiple-herd clusters

It was decided to look only for risk factors in herds included in *multiple-herd* clusters because it might indicate the presence (i.e. spread) of infectious diseases, whereas *single-herd* clusters might indicate problems with herd management or diseases within the herd. Herds were classified as being inside a *multiple-herd* cluster, if they belonged to a multiple-herd cluster in at least one month during the study period.

Logistic regression was used to examine possible factors associated with the probability of a herd being inside a *multiple-herd* cluster. First, a univariable logistic regression was

carried out for each factor. Then forward selection was used to build a multivariable model, by adding the most significant variable to the model using a p-value of 0.05 as the threshold. The overall significance of each variable was tested using Chi-square test in the anova() function. For significant variables with more than 2 levels, post hoc test of pairwise comparison between levels was done using the 'lsmeans' package (Lenth, 2015) (version 2.20-23). The analysis was also performed in R (version 3.1.1) (R Core Team, 2014).

3. Results

Data description

A total of 5010 farms (i.e. unique locations) were included in the analysis, divided among the five regions corresponding to 1057 farms in North Jutland, 1765 in Central Jutland, 1548 in Southern Denmark, 126 in Capital Region of Denmark and 514 in Zealand (Figure 1). This corresponded to 1896 weaner herds, 1490 sow herds and 3839 finisher herds. The mortality for different farm types (i.e. farms with a single or multiple age groups) is represented in Table 1.

The monthly median number of herds included in the study was 1776 (mininum: 1391, maximum:1834) for weaner herds, 1462 (minimum:1391; maximum: 1818) for sow herds and 3679 (minimum: 3557, maximum: 3710) for finisher herds. From December 2013 to October 2015, the median mortality observed were 0.017 (minimum=0,

maximum=0.913), 0.008 (minimum=0, maximum=0.696) and 0.008 (minimum=0, maximum=0.805) for weaners, sows and finishers, respectively.

The average mortality for the 5 different regions of Denmark is shown in Figure 2. No major differences were seen for the mortality among the different regions. There appeared to be an increased mortality in January 2014, July 2014, January 2015 and July 2015 in all regions for the three age groups. In general, the mortality in weaner herds was double the mortality observed in sow and finisher herds.

The spatial distribution of the mortality for the three age groups changed over time (Figures 3, 4 and 5) from December 2013 to October 2015. The overall distribution of the mortality in weaners showed higher values in Central Jutland. For sows, the highest mortality occurred mainly in North Jutland, Southern Denmark and Zealand. Finisher herds located in Central Jutland and Zealand presented higher mortality. The areas with higher mortality were mainly present in January and July of 2014 and 2015.

241

242 Table 1- Farms included in the study with the corresponding mortality (proportion)

243 observed from December 2013 to October 2015 for each of the age groups.

		Age groups present on the farm						
		Only weaner herds	Only sow herds	Only finisher herds	Weaner and sow herds	Weaner and finisher herds	Sow and finisher herds	Weaner, sow and finisher herds
Total number of farms		**283**	**276**	**2746**	**618**	**497**	**98**	**498**
Mortality for weaners	Min	0.000	-	-	0.000	0.000	-	0.000
	Median	0.011	-	-	0.023	0.014	-	0.024
	Max	0.370	-	-	0.556	0443	-	0.936
Mortality for sows	Min	-	0.000	-	0.000	-	0.000	0.000
	Median	-	0.008	-	0.008	-	0.008	0.008
	Max	-	0.696	-	0.155	-	0.130	0.113
Mortality for finishers	Min	-	-	0.000	-	0.000	0.000	0.000
	Median	-	-	0.008	-	0.008	0.010	0.016
	Max	-	-	0.805	-	0.225	0.388	0.468

244

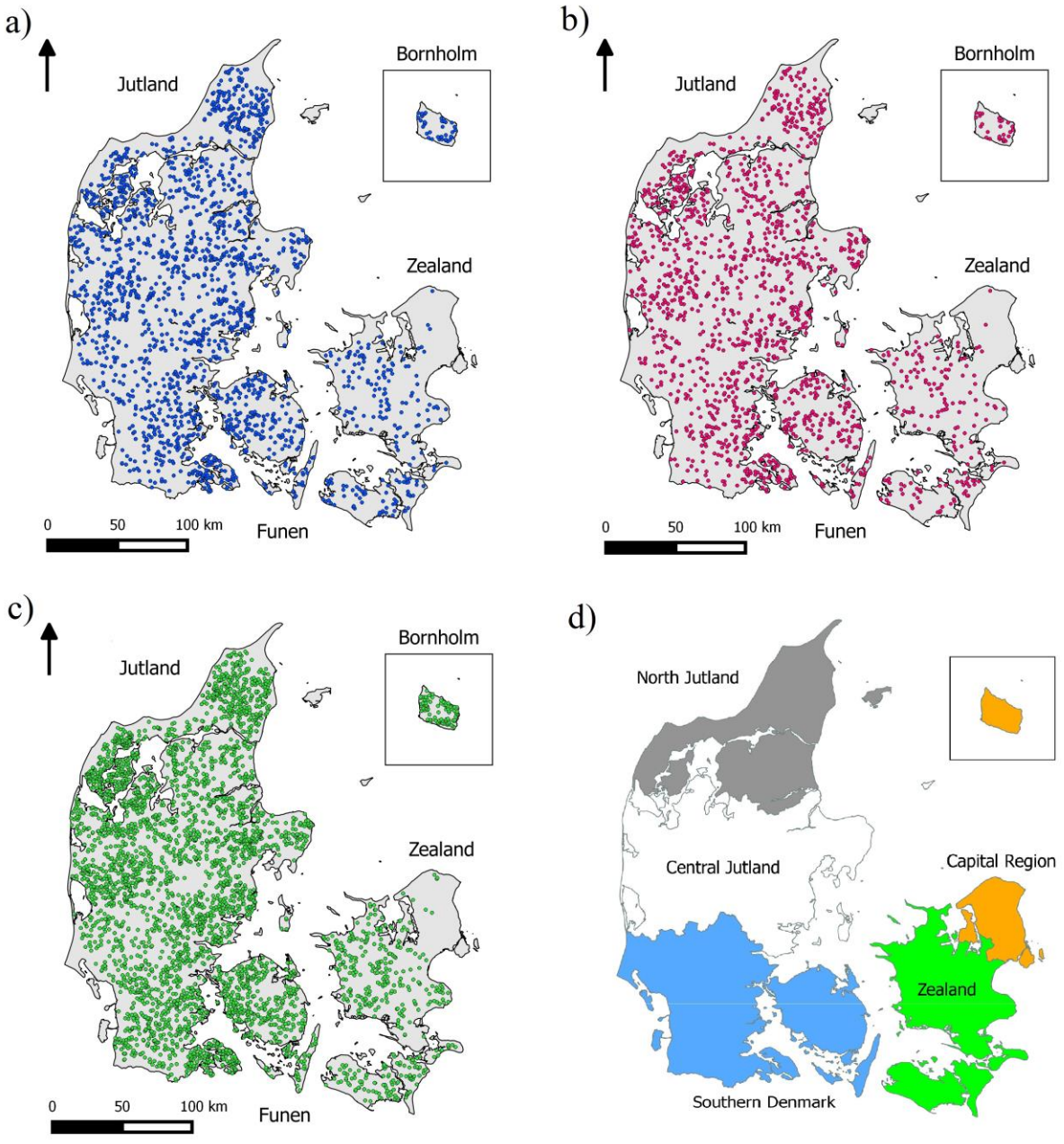
245

246

247

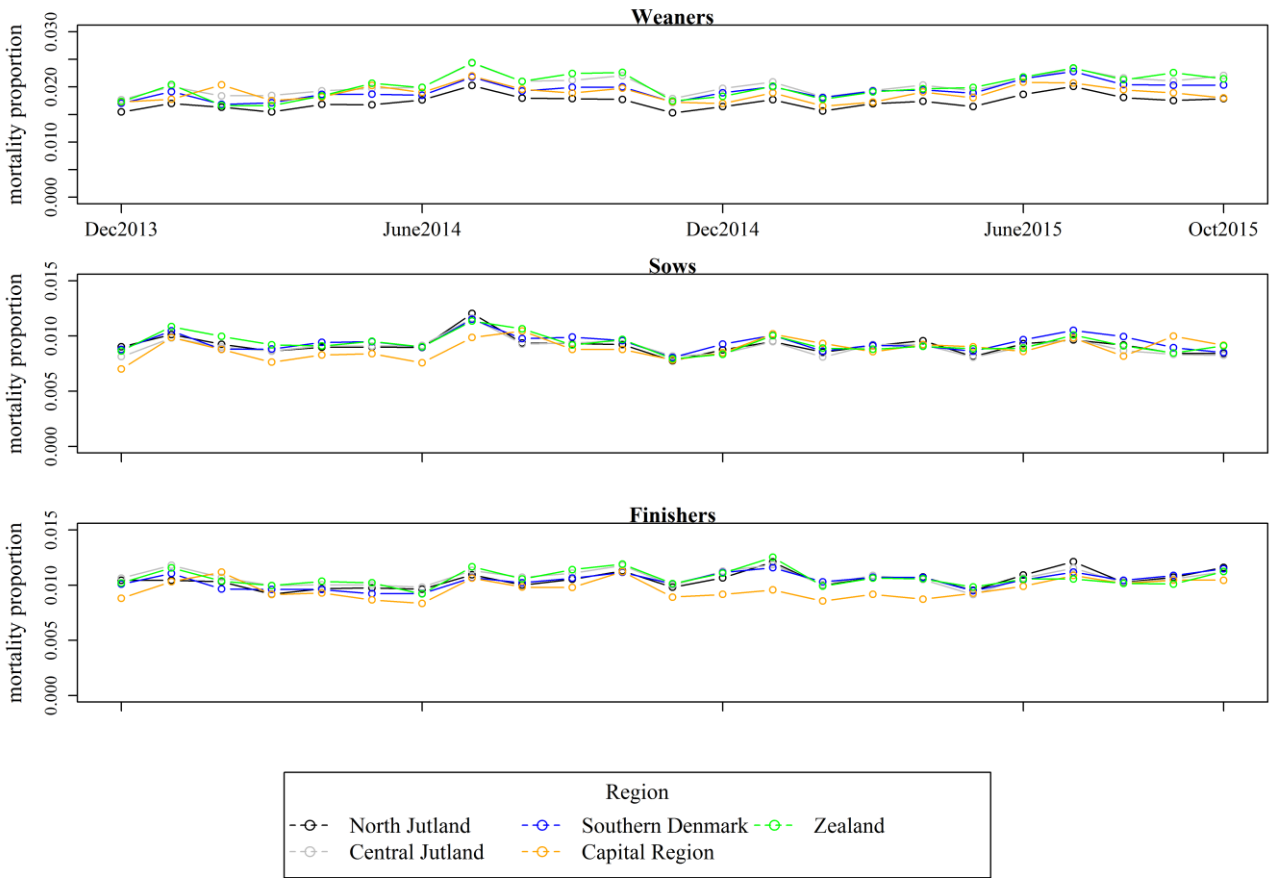
248

249 Figure 1- Location of the Danish swine herds included in the study by age group, (a)
250 weaners, (b) sows, (c) finishers, and (d) description of the five administrative regions in
251 Denmark.



252

253 Figure 2- Monthly average aggregated mortality for the three age groups of swine in the
254 5 administrative regions of Denmark.



255

256

257

258

259

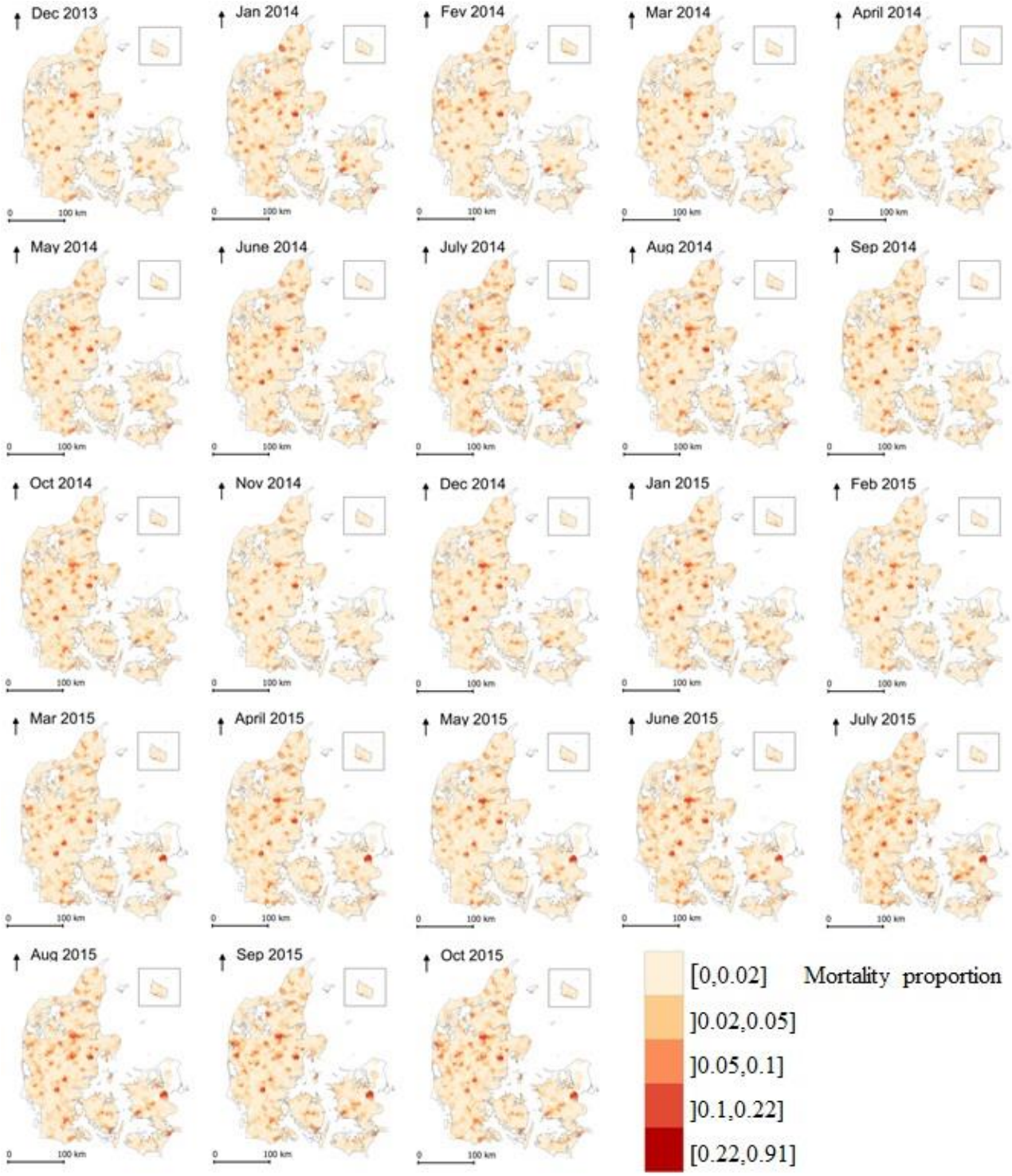
260

261

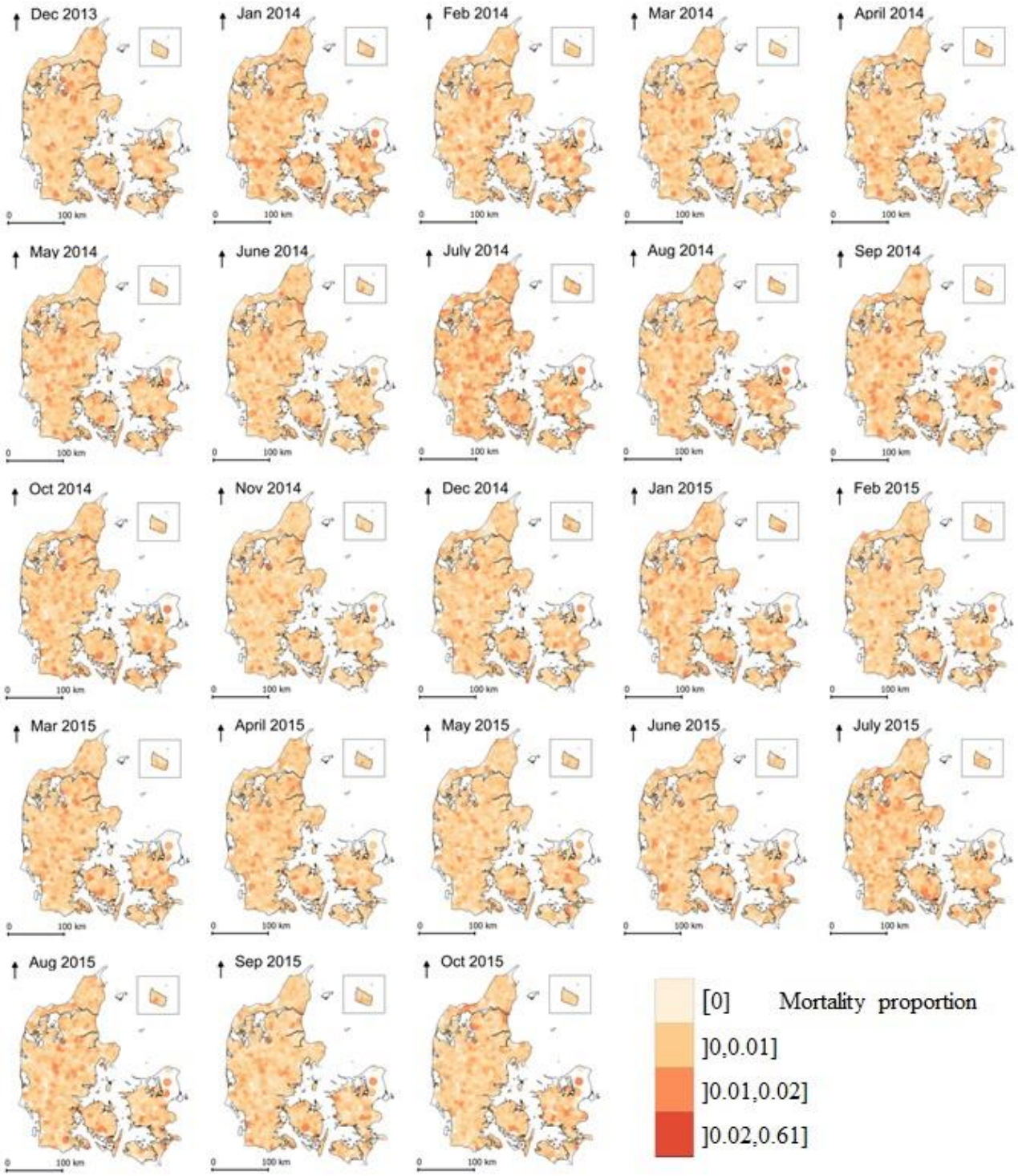
262

263

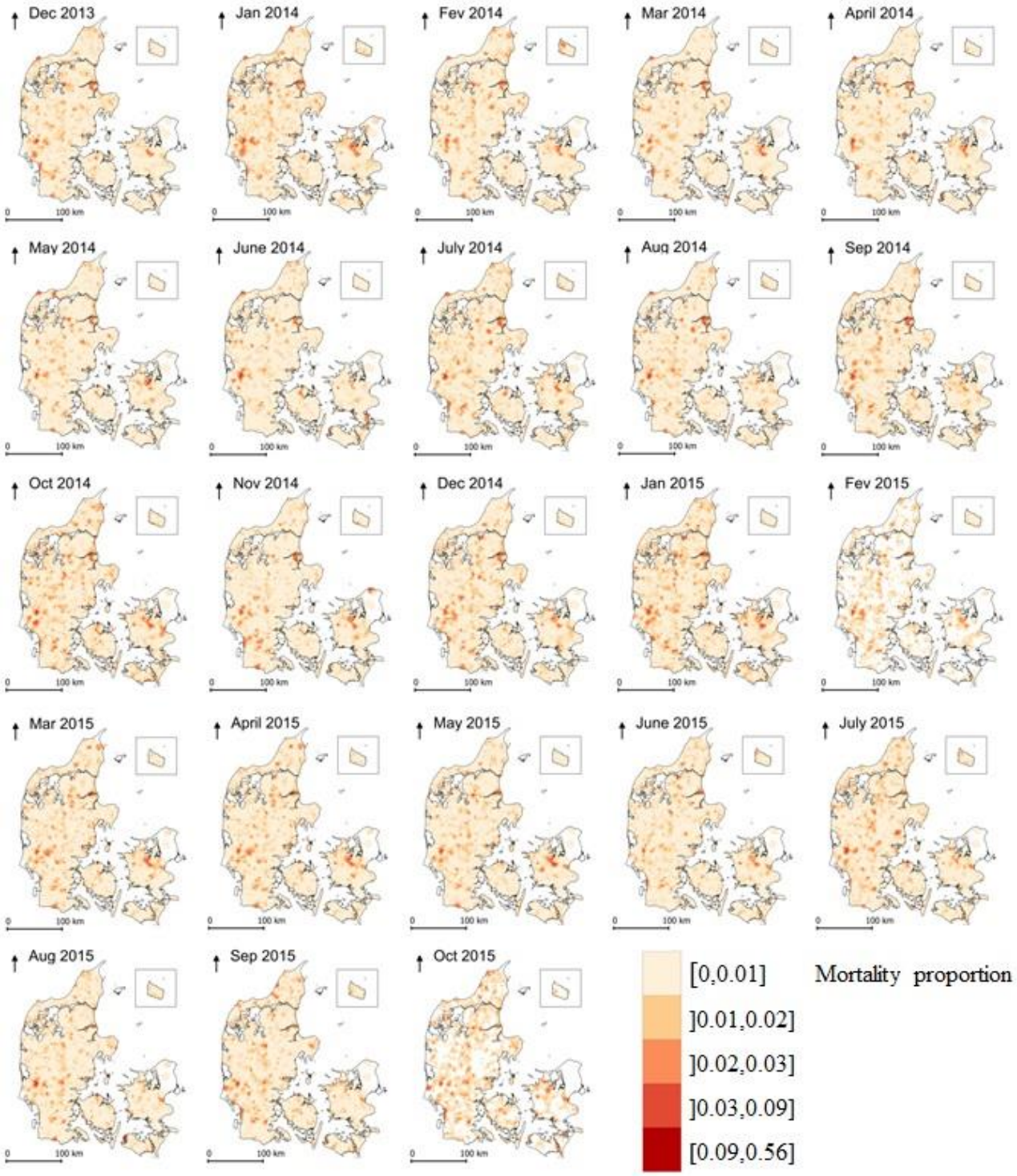
264 Figure 3- Monthly mortality in weaner herds from December 2013 to October 2015.



265 Figure 4- Monthly mortality in sow herds from December 2013 to October 2015.



266 Figure 5- Monthly mortality in finisher herds from December 2013 to October 2015.



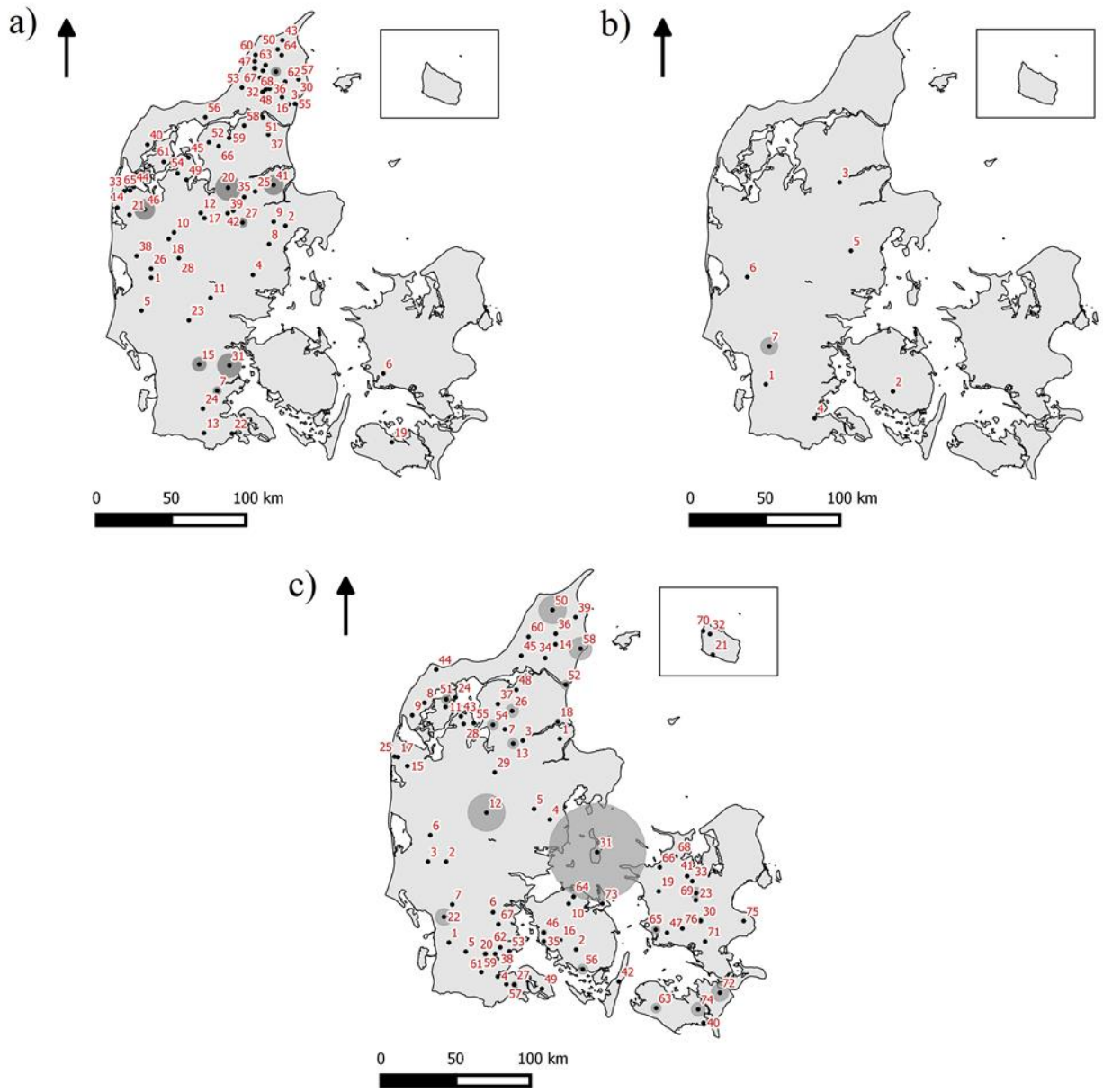
Spatio-temporal clustering

Figure 6 shows the location of significant high-mortality spatio-temporal clusters based on a maximum of 50% of the population at risk in weaners, sows and finishers. The descriptive statistics of these significant clusters and the starting and ending months are presented in Appendix A. Moreover, interactive videos with the location and duration of spatio-temporal clusters for each age group are available as Supplementary Material. A total of 68 spatio-temporal clusters, corresponding to 57 *single-herd* and 11 *multiple-herd* clusters, were found for weaners. For the sows, 5 *single-herd* and 2 *multiple-herd* (total =7) clusters were found. For finishers, 49 *single-herd* and 27 *multiple-herd* (total = 76) clusters were found (Appendix A). The spatio-temporal clusters in weaners were mainly located in the North and Central Jutland and had a higher occurrence between June 2014 and August 2015 where 67 clusters were simultaneously observed. For sows, the clusters were mainly located in Central Jutland and Southern Denmark and the highest simultaneous presence (5 clusters), was observed from March 2014 to September 2014. The spatio-temporal clusters for finishers were located all over Denmark and the highest number of clusters, corresponding to 68 clusters, was observed between September 2014 and October 2015.

For the analysis using smaller maximum cluster sizes (10% and 25% of population at risk), the number of clusters increased up to three times. As a result of this increase, the size and duration of the clusters decreased. The clusters were present in the same areas as the clusters based on a maximum cluster size of 50% of the population at risk (results not shown).

A visual assessment of the temporal changes of mortality in *single-herd* clusters suggested several different patterns (Appendix B). In some cases, a cluster consisted only of few months with a distinctive "peak" in mortality. Other clusters had a longer duration, with a smaller increase in mortality.

Figure 6- Location of the spatio-temporal clusters of high mortality in Danish swine herds. Each spatio-temporal cluster found for (a) weaners, (b) sows and (c) finishers is identified by a number corresponding to the cluster ID (see Appendix A) and the circles represent the cluster size.



Logistic regression models

Tables 2, 3 and 4 describe the frequency distribution of the herds inside vs outside the clusters stratified by the herd size, farm type, SPF status and the disease status for Enzootic pneumonia, Porcine pleuropneumonia, Atrophic rhinitis, and PRRS.

The results from univariable logistic regression analyses of the herd being inside vs outside *multiple-herds* clusters are presented in Table 5. The variables farm type, herd size, SPF status, and Atrophic rhinitis were significant ($p<0.05$) when analyzed based on univariable models for weaners and finisher herds. It was not possible to perform the analysis for sow herds only 2 *multiple-herd* clusters were found (including 7 farms in total).

In the multivariable analysis for both weaner and finisher herds, only farm type was included in the final model, as no other variables were significant.

328 Table 3-Frequency distribution (number and percentage) of herds inside versus outside the spatio-temporal high risk clusters
329 stratified by farm type.

		Farm type						
		Only weaners herds	Only sows herds	Only finishers herds	Weaners and sows herds	Weaners and finishers herds	Sows and finishers herds	Weaners, sows and finishers herds
Weaners	**Total**	**283**	**-**	**-**	**618**	**497**	**-**	**498**
	Single-herd cluster (%)	3 (1.06)	-	-	33 (5.34)	2 (0.40)	-	19 (3.82)
	Multiple-herd clusters (%)	3 (1.06)	-	-	20 (3.24)	16 (3.22)	-	20 (4.02)
	Non clusters (%)	277 (97.88)	-	-	565 (91.42)	479 (96.38)	-	459 (92.17)
Sows	**Total**	**-**	**276**	**-**	**618**	**-**	**98**	**498**
	Single-herd cluster (%)	-	2 (0.72)	-	2 (0.32)	-	1 (1.02)	0 (0.00)
	Multiple-herd clusters (%)	-	2 (0.72)	-	0 (0.00)	-	0 (0.00)	5 (1.00)
	Non clusters (%)	-	272 (98.55)	-	616 (99.68)	-	97 (98.98)	493 (99.00)
Finishers	**Total**	**-**	**-**	**2746**	**-**	**497**	**98**	**498**
	Single-herd cluster (%)	-	-	19 (0.69)	-	13 (2.62)	1 (1.02)	16 (3.21)
	Multiple-herd clusters (%)	-	-	155 (5.64)	-	45 (9.05)	9 (9.18)	45 (9.04)
	Non clusters (%)	-	-	2572 (93.66)	-	439 (88.33)	88 (89.80)	437 (87.75)

330

331 Table 4- Frequency distribution (number and percentage) of herds inside versus outside the spatio-temporal high risk clusters
332 stratified by the disease status for Enzootic pneumonia, Porcine pleuropneumonia, Atrophic rhinitis, and Porcine Reproductive
333 and Respiratory Syndrome (PRRS).

			Enzootic pneumonia			*Porcine pleuropneumonia*			*Atrophic rhinitis*			*PRRS*		
Herd type	*Cluster type*	*Total number of herds*	*Positive* (%)	*Negative* (%)	*Unknown* (%)	*Positive* (%)	*Negative* (%)	*Unknown* (%)	*Positive* (%)	*Negative* (%)	*Unknown* (%)	*Positive* (%)	*Negative* (%)	*Unknown* (%)
Weaners	*Single-herd*	57	28 (49.1)	13 (22.8)	16 (28.1)	27 (47.4)	14 (24.6)	16 (28.0)	1 (1.8)	40 (70.2)	16 (28.0)	14 (24.6)	27 (47.4)	16 (28.0)
	Multiple-*herd*	59	24 (40.7)	16 (27.1)	19 (32.2)	18 (30.5)	22 (37.3)	19 (32.2)	0 (0.0)	40 (67.8)	19 (32.2)	9 (15.3)	31 (52.5)	19 (32.2)
	Non Cluster	1780	824 (46.3)	402 (22.6)	554 (31.1)	730 (41.0)	496 (27.9)	554 (31.1)	19 (1.1)	1207 (67.8)	554 (31.1)	415 (23.3)	840 (47.2)	525 (29.5)
Sows	*Single-herd*	5	5 (100.0)	0 (0.0)	0 (0.0)	3 (60.0)	2 (40.0)	0 (0.0)	5 (100.0)	0 (0.0)	0 (0.0)	3 (60.0)	2 (40.0)	0 (0.0)
	Multiple-herd	7	2 (28.6)	2 (28.6)	3 (42.8)	2 (28.6)	2 (28.6)	3 (42.8)	0 (0.0)	4 (57.1)	3 (42.9)	2 28.6)	3 (42.8)	2 (28.6)
	Non Cluster	1478	739 (50.0)	375 (25.4)	364 (24.6)	672 (45.5)	442 (29.9)	364 (24.6)	21 (1.4)	1093 (74.0)	364 (24.6)	345 (23.3)	796 (53.9)	337 (22.8)
Finishers	*Single-herd*	49	21 (42.9)	8 (16.3)	20 (40.8)	13 (26.5)	16 (32.7)	20 (40.8)	0 (0.0)	29 (59.2)	20 (40.8)	10 (20.4)	19 (38.8)	20 (40.8)
	Multiple-herd	254	79 (31.1)	31 (12.2)	144 (56.7)	73 (28.7)	37 (14.6)	144 (56.7)	2 (0.8)	108 (42.5)	144 (56.7)	44 (17.3)	68 (26.8)	142 (55.9)
	Non Cluster	3536	845 (23.9)	402 (11.4)	2289 (64.7)	731 (20.7)	516 (14.6)	2289 (64.7)	19 (0.6)	1228 (34.7)	2289 (64.7)	456 (12.9)	815 (23.0)	2265 (64.1)

334

335 Table 5- Univariable logistic regression on the association between significant variables and the probability of the herd (weaner
336 or finisher herds) being inside *multiple-herds* clusters.

Herd	*Variables*	*Categories*	*Estimates**	*Std. Err.*	*P > Z*	*Odds Ratio*	*95%CI Odds Ratio*	
							Lower limit	*Upper limit*
Weaners	*Herd size (number of animals per herd)*	*200-1799*	Reference group		<0.0001[1]			
		1800-2849	-1.43	0.33	<0.0001	0.23	0.12	0.44
Finishers	*SPF status*	*Non SPF herd*	Reference group		0.0107[1]			
		SPF herd	0.3380	0.13	0.0101	1.40	1.08	1.81
	Farm type	*Only finisher herds* [a]	Reference group		0.0016[1]			
		Finishers and weaner herds [bc]	0.53	0.18	0.0027	1.70	1.19	2.39
		Finishers and sow herds [abc]	0.53	0.36	0.1413	1.70	0.05	3.26
		Finishers, sows and weaners herd [c]	0.54	0.18	0.0025	1.71	1.20	2.40
	Atrophic rhinitis	*Negative* [a]	Reference group		0.0376 [1]			
		Positive [ab]	0.18	0.75	0.8107	1.20	0.19	4.20
		Unknown [b]	-0.34	0.13	0.0112	0.72	0.55	0.93

337 [1]Overall significance of the variable.

338 *Estimates with different letters as superscript are significantly different at a 5% significance level.

339

4. Discussion

This study was performed to explore the spatio-temporal patterns of mortality in Danish swine herds. Two types of spatio-temporal clusters were found including a large number of *single-herd* and fewer *multiple-herd* clusters. Further analysis was conducted in order to investigate potential risk factors and temporal trends for these clusters.

Previous studies have demonstrated the potential of using mortality for disease detection (Backer et al., 2011; Perrin et al., 2012), monitoring animal health status (Alba et al., 2015) and assess the impact of unexpected environmental events (Morignat et al., 2014). However, due to the large variability of mortality (and the way it is reported) in animal populations among countries, is not possible to extrapolate the usefulness and challenges from previous studies to the Danish context.

In this study, it was decided to analyze the data by age group due to the physiological difference among the groups. For example, a higher mortality is expected in weaners compared to other age groups due to parturition, nutrition, thermal stress and diseases such as post-weaning diarrhea (SEGES Danish Pig Research Centre, 2014).

Applying Scan Statistics techniques at herd level allowed us to detect small changes at herd level. However, it is important to take into account that mortality is based on two different databases. The information on the number of animals for different age groups in the CHR database was used as a proxy of the number of animals present in a herd for a given month. This information is updated in the database minimum once/twice yearly by farmers or SEGES Pig Research Centre ("Pig Research Centre (VSP- SEGES)," 2016). Thus, a dynamic herd with variability in its herd size might be misrepresented in the CHR. As consequence, changes in mortality can be biased. The movements registered in the SMD are used as proxy of the number of dead animals for different age groups. The registration of

dead weaners (up to 30 Kg) are based on the number of containers (with specific dimensions) transported from a farm to the rending plant. In the mortality calculation, a fixed number is used for each container size, regardless of how many animals are actually in the container. This may also bias the mortality, thus illustrating the challenges of using mortality for monitoring purposes in Denmark.

Spatial and temporal changes of mortality

The spatial distribution of farms with higher mortality in weaner and sow herds concurred with higher farm density. This might be explained by the higher prevalence of certain infectious diseases in areas with higher animal density (Mortensen et al., 2002; Poljak et al., 2008).

The temporal patterns found in mortality for the three age groups suggested increases in January and July for each year. These increases do not appear to be biologically justified, but can probably be linked to infrequent updates on the CHR database.

Local spatio-temporal clusters

The analysis identified a large number of *single-herd* clusters, i.e. herds with a higher than expected mortality, where the neighbors did not experience an increased mortality. These farms may deal with welfare issues (SEGES Danish Pig Research Centre, 2014; SEGES Pig Research Centre, 2015) or the presence of diseases where good biosecurity and herd management are in place, so that the infection does not transmit to neighbors. Still, for infectious diseases such as PRRS (Mortensen et al., 2002), Swine Influenza Virus (Brown, 2000), or Porcine Circo Virus type 2 (Baekbo et al., 2012) airborne transmission to neighboring farms would be highly likely. Thus resulting in *multiple-herd clusters*, especially in areas with high farm density. Transmission between neighboring farms by (mechanical)

vectors also increases diseases transmission for some diseases, such as *Actinobacillus pleuropneumoniae* (Kristensen et al., 2004).

Risk factors for herds being inside multiple-herd clusters

While several potential risk factors where identified in the univariable analysis, only farm type remained in a multivariable analysis, due to a strong correlation between variables. The effect for both weaners and finishers was as expected: the specialized farms with only one age group had lower mortality than farms with more age groups. In general, these larger specialized farms have high biosecurity and working with only one age group allows more specialization of the staff.

5. Conclusions

This study explored spatio-temporal patterns in mortality from Danish swine herds and its potential for syndromic surveillance.

This study shows the presence of a large number of significant *single-herd* and *multiple-herd* clusters for weaners, sows and finisher herds. The *single-herd* clusters represent potential isolated welfare and disease problems, while *multi-herd* clusters could be indicative of local spread of an infectious disease.

There is potential of using mortality for disease surveillance. However, detected clusters might not be due to disease, but the result of changes in herd management, legislative rules and climatic factors. Hence, follow-up of detected clusters is necessary. Further analysis to explore and select the appropriate spatial, temporal and spatio-temporal monitoring methods is needed in order to incorporate mortality in Danish disease monitoring system.

Acknowledgements

The authors would like to acknowledge Danish Veterinary and Food Administration for providing the data used in this study.

References

Alba, A., Dórea, F.C., Arinero, L., Sanchez, J., Cordón, R., Puig, P., Revie, C.W., 2015. Exploring the Surveillance Potential of Mortality Data: Nine Years of Bovine Fallen Stock Data Collected in Catalonia (Spain). PLoS One 10, e0122547. doi:10.1371/journal.pone.0122547

Backer, J.A.A., Brouwer, H., van Schaik, G., van Roermund, H.J.W.J.W., Thulke, H.-H., 2011. Using mortality data for early detection of Classical Swine Fever in The Netherlands. Prev. Vet. Med. 99, 38–47. doi:10.1016/j.prevetmed.2010.10.008

Baekbo, P., Kristensen, C.S., Larsen, L.E., 2012. Porcine circovirus diseases: a review of PMWS. Transbound. Emerg. Dis. 59 Suppl 1, 60–7. doi:10.1111/j.1865-1682.2011.01288.x

Brown, I.H., 2000. The epidemiology and evolution of influenza viruses in pigs. Vet. Microbiol. 74, 29–46.

Dórea, F., Muckle, A., Kelton, D., McClure, J.T., McEwen, B.J., McNab, W.B., Sanchez, J., Revie, C.W., 2013. Exploratory analysis of methods for automated classification of laboratory test orders into syndromic groups in veterinary medicine. PLoS One 8, e57334. doi:10.1371/journal.pone.0057334

Dupuy, C., Bronner, A., Watson, E., Wuyckhuise-Sjouke, L., Reist, M., Fouillet, A., Calavas, D., Hendrikx, P., Perrin, J.-B., 2013. Inventory of veterinary syndromic surveillance initiatives in

Europe (Triple-S project): Current situation and perspectives. Prev. Vet. Med. 111, 220–229.

Dupuy, C., Morignat, E., Dorea, F., Ducrot, C., Calavas, D., Gay, E., 2015. Pilot simulation study using meat inspection data for syndromic surveillance: use of whole carcass condemnation of adult cattle to assess the performance of several algorithms for outbreak detection. Epidemiol. Infect. 143, 2559–2569. doi:10.1017/S0950268814003495

European Union, 2000. Directive 2000/15/EC of the European Parliament and the Council of 10 April 2000 amending Council Directive 64/432/EEC on health problems affecting intra-Community trade in bovine animals and swine.

Huisman, O., By, R. a De, 2009. Principles of Geographic Information Systems - An introductory texbook, 4th ed. The International Institute for Geo-Information Science and Earth Observation (ITC), Enschede, The Netherlands.

Hyder, K., Vidal-Diez, A., Lawes, J., Sayers, A.R., Milnes, A., Hoinville, L.J., Cook, A.J.C., 2011. Use of spatiotemporal analysis of laboratory submission data to identify potential outbreaks of new or emerging diseases in cattle in Great Britain. BMC Vet. Res. 7, 14. doi:10.1186/1746-6148-7-14

Kristensen, C.S., Angen, Ø., Andreasen, M., Takai, H., Nielsen, J.P., Jorsal, S.E., 2004. Demonstration of airborne transmission of Actinobacillus pleuropneumoniae serotype 2 between simulated pig units located at close range. Vet. Microbiol. 98, 243–9. doi:10.1016/j.vetmic.2003.10.026

Kulldorff, M., 2016. SaTScan Bibliography [WWW Document]. SaTScan Softw. Spat. temporal space-time scan Stat. doi:10.1080/2330443X.2014.928245

Kulldorff, M., 2015. SaTScan TM 9.4.2: Software for the spatial and space-time scan statistics.

Kulldorff, M., Athas, W.F., Feurer, E.J., Miller, B.A., Key, C.R., 1998. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico., American journal of public health.

Lenth, R., 2015. Package “lsmeans”: Least-Squares Means.

Lian, M., Warner, R.D., Alexander, J.L., Dixon, K.R., 2007. Using geographic information systems and spatial and space-time scan statistics for a population-based risk analysis of the 2002 equine West Nile epidemic in six contiguous regions of Texas. Int. J. Heal. Geogr. 174, 302–9. doi:10.1186/1476-072X-6-42

Morignat, E., Perrin, J.-B., Gay, E., Vinard, J.-L., Calavas, D., Hénaux, V., 2014. Assessment of the impact of the 2003 and 2006 heat waves on cattle mortality in France. PLoS One 9, e93176. doi:10.1371/journal.pone.0093176

Mortensen, S., Stryhn, H., Søgaard, R., Boklund, A., Stärk, K.D.., Christensen, J., Willeberg, P., 2002. Risk factors for infection of sow herds with porcine reproductive and respiratory syndrome (PRRS) virus. Prev. Vet. Med. 53, 83–101. doi:10.1016/S0167-5877(01)00260-4

Pebesma, E., Graeler, B., 2016. Package “gstat”:Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation.

Perrin, J.-B., Ducrot, C., Vinard, J.-L., Morignat, E., Gauffier, A., Calavas, D., Hendrikx, P., 2010. Using the National Cattle Register to estimate the excess mortality during an epidemic: application to an outbreak of Bluetongue serotype 8. Epidemics 2, 207–14. doi:10.1016/j.epidem.2010.10.002

Perrin, J.-B.B., Ducrot, C., Vinard, J.-L.L., Morignat, E., Calavas, D., Hendrikx, P., 2012. Assessment of the utility of routinely collected cattle census and disposal data for syndromic surveillance. Prev. Vet. Med. 105, 244–252. doi:10.1016/j.prevetmed.2011.12.015

Pfeiffer, D.U., Minh, P.Q., Martin, V., Epprecht, M., Otte, M.J., 2007. An analysis of the spatial and temporal patterns of highly pathogenic avian influenza occurrence in Vietnam using national surveillance data. Vet. J. 174, 302–9. doi:10.1016/j.tvjl.2007.05.010

Pig Research Centre (VSP- SEGES) [WWW Document], 2016. URL https://www.seges.dk/en

Poljak, Z., Dewey, C.E., Martin, S.W., Christensen, J., Carman, S., Friendship, R.M., 2008. Prevalence of and risk factors for influenza in southern Ontario swine herds in 2001 and 2003. Can. J. Vet. Res. = Rev. Can. Rech. vétérinaire 72, 7–17.

R Core Team, 2014. R. A Language and Environment for Statistical Computing.

SEGES Danish Pig Research Centre, 2014. Annual Report 2014, Annual Report. Copenhagen. doi:10.1016/j.parkreldis.2015.02.017

SEGES Pig Research Centre, 2015. Annual Report 2015. SEGES Pig Research Centre, Copenhagen.

SPF-DANMARK [WWW Document], 2015. URL https://www.spf.dk/ (accessed 5.12.15).

Themudo, G.E., Østergaard, J., Ersbøll, A.K., 2011. Persistent spatial clusters of plasmacytosis among Danish mink farms. Prev. Vet. Med. 102, 75–82. doi:10.1016/j.prevetmed.2011.06.003

Triple S Project, 2011. Assessment of syndromic surveillance in Europe. Lancet.

Vial, F., Reist, M., 2015. Comparison of whole carcass condemnation and partial carcass condemnation data for integration in a national syndromic surveillance system: the Swiss experience. Meat Sci. 101, 48–55. doi:10.1016/j.meatsci.2014.11.002

Wagner, M.M., Moore, A.W., Aryel, R.M., 2006. Handbook of Biosurveillance, in: Handbook of Biosurveillance. Elsevier, pp. 217–234. doi:10.1016/B978-012369378-5/50000-4

Appendix A: Description of statistical significant clusters.

Statistically significant spatio-temporal clusters (p<0.05) of high mortality in Danish weaner herds between December 2013 and October 2015. For each cluster, the radius as well as starting and ending months (month 1 corresponds December 2013 and month 23 corresponds to October 2015) are given along with descriptive statistics of the number farms in the cluster (Number of locations), cluster significance (p-value) based on 999 Monte Carlo replications, Log Likelihood Ratio (LLR) and Relative Risk (RR) for a maximum spatial window size of 50% of the population at risk and a minimum of 1 month and maximum of 90% of the study period.

Cluster	Cluster type	Radius (m)	Start month	End month	Number of locations	LLR	*p* value	RR
1	Single-herd	0	2	21	1	7860.71	<0.001	18.85
2	Single-herd	0	4	23	1	4821.56	<0.001	19.10
3	Single-herd	0	3	22	1	4738.63	<0.001	24.61
4	Single-herd	0	2	21	1	4449.98	<0.001	10.05
5	Single-herd	0	3	22	1	4373.95	<0.001	9.42
6	Single-herd	0	2	21	1	4305.49	<0.001	9.20
7	Multiple-herd	3000	4	23	3	4043.55	<0.001	9.00
8	Single-herd	0	1	20	1	3780.79	<0.001	19.38
9	Single-herd	0	4	23	1	3754.97	<0.001	14.12
10	Single-herd	0	2	21	1	3272.74	<0.001	16.73
11	Single-herd	0	4	23	1	2968.67	<0.001	18.14
12	Single-herd	0	3	22	1	2900.29	<0.001	15.50
13	Single-herd	0	1	20	1	2868.94	<0.001	8.00
14	Single-herd	0	4	23	1	2817.72	<0.001	13.56
15	Multiple-herd	5214	4	23	3	2787.90	<0.001	3.63
16	Single-herd	0	3	22	1	2723.79	<0.001	7.42
17	Single-herd	0	3	22	1	2714.60	<0.001	14.86
18	Single-herd	0	4	23	1	2574.72	<0.001	15.58
19	Single-herd	0	4	23	1	2557.14	<0.001	9.43
20	Multiple-herd	9506	4	23	6	2541.39	<0.001	2.55
21	Single-herd	0	3	22	1	2432.72	<0.001	13.86
22	Single-herd	0	1	20	1	2348.49	<0.001	13.56
23	Single-herd	0	3	22	1	2296.20	<0.001	15.39
24	Single-herd	0	3	22	1	2256.59	<0.001	9.27
25	Single-herd	0	4	23	1	2219.90	<0.001	11.70
26	Single-herd	0	4	23	1	2190.49	<0.001	5.11

27	Multiple-herd	3473	4	23	5	2095.00	<0.001	2.46
28	Single-herd	0	4	23	1	2069.82	<0.001	14.41
29	Single-herd	0	2	21	1	2068.02	<0.001	9.43
30	Single-herd	0	1	20	1	2059.97	<0.001	11.18
31	Multiple-herd	8962	4	23	21	1824.65	<0.001	1.62
32	Single-herd	0	2	21	1	1784.66	<0.001	13.13
33	Single-herd	0	2	21	1	1771.68	<0.001	8.62
34	Multiple-herd	1940	2	21	2	1719.22	<0.001	7.87
35	Single-herd	0	4	23	1	1607.22	<0.001	7.19
36	Single-herd	0	2	21	1	1454.06	<0.001	6.80
37	Single-herd	0	1	20	1	1445.61	<0.001	10.88
38	Single-herd	0	2	21	1	1426.39	<0.001	3.68
39	Single-herd	0	3	22	1	1407.50	<0.001	8.91
40	Single-herd	0	4	23	1	1379.12	<0.001	5.99
41	Multiple-herd	7360	4	23	7	1322.39	<0.001	1.90
42	Single-herd	0	4	23	1	1302.71	<0.001	8.52
43	Single-herd	0	4	23	1	1260.43	<0.001	4.43
44	Single-herd	0	2	21	1	1241.02	<0.001	8.29
45	Single-herd	0	1	20	1	1226.88	<0.001	7.00
46	Multiple-herd	7569	4	23	5	1220.95	<0.001	2.16
47	Single-herd	0	4	23	1	1033.46	<0.001	7.47
48	Single-herd	674	1	20	2	1023.58	<0.001	3.22
49	Single-herd	0	4	23	1	1021.36	<0.001	6.38
50	Single-herd	0	1	20	1	999.14	<0.001	5.57
51	Single-herd	0	2	21	1	986.77	<0.001	5.05
52	Single-herd	0	4	23	1	907.81	<0.001	8.74
53	Single-herd	0	2	21	1	834.83	<0.001	7.69
54	Single-herd	0	3	22	1	803.94	<0.001	6.50
55	Single-herd	0	7	23	1	762.21	<0.001	6.35
56	Single-herd	0	3	22	1	744.68	<0.001	7.80
57	Single-herd	0	1	8	1	738.10	<0.001	7.76
58	Single-herd	0	3	22	1	721.19	<0.001	4.33
59	Single-herd	0	1	20	1	718.37	<0.001	4.03
60	Single-herd	0	10	23	1	608.11	<0.001	2.53
61	Multiple-herd	1167	2	21	2	583.06	<0.001	2.56
62	Single-herd	0	2	21	1	580.69	<0.001	5.48
63	Single-herd	0	5	23	1	553.41	<0.001	2.37
64	Single-herd	0	4	23	1	543.91	<0.001	3.81
65	Single-herd	0	4	23	1	486.90	<0.001	1.71

66	Single-herd	0	4	23	1	482.74	<0.001	1.91
67	Single-herd	0	4	23	1	472.29	<0.001	3.26
68	Multiple-herd	3424	4	23	3	466.30	<0.001	1.83

Statistically significant spatio-temporal clusters ($p<0.05$) of high mortality in Danish sow herds between December 2013 and October 2015. For each cluster, the radius as well as starting and ending months (month 1 corresponds December 2013 and month 23 corresponds to October 2015) are given along with descriptive statistics of the number farms in the cluster (Number of locations), cluster significance (p-value) based on 999 Monte Carlo replications, Log Likelihood Ratio (LLR) and Relative Risk (RR) for a maximum spatial window size of 50% of the population at risk and a minimum of 1 month and maximum of 90% of the study period.

Cluster	Cluster type	Radius (m)	Start month	End month	Number of locations	LLR	*p* value	RR
1	Single-herd	0	19	19	1	663.29	<0.001	75.83
2	Single-herd	0	4	23	1	512.14	<0.001	9.81
3	Single-herd	0	1	20	1	434.43	<0.001	3.85
4	Single-herd	0	21	21	1	326.75	<0.001	14.24
5	Single-herd	0	4	23	1	221.99	<0.001	3.85
6	Multiple-herd	327	3	22	2	180.42	<0.001	1.90
7	Multiple-herd	6361	1	10	5	173.27	<0.001	2.25

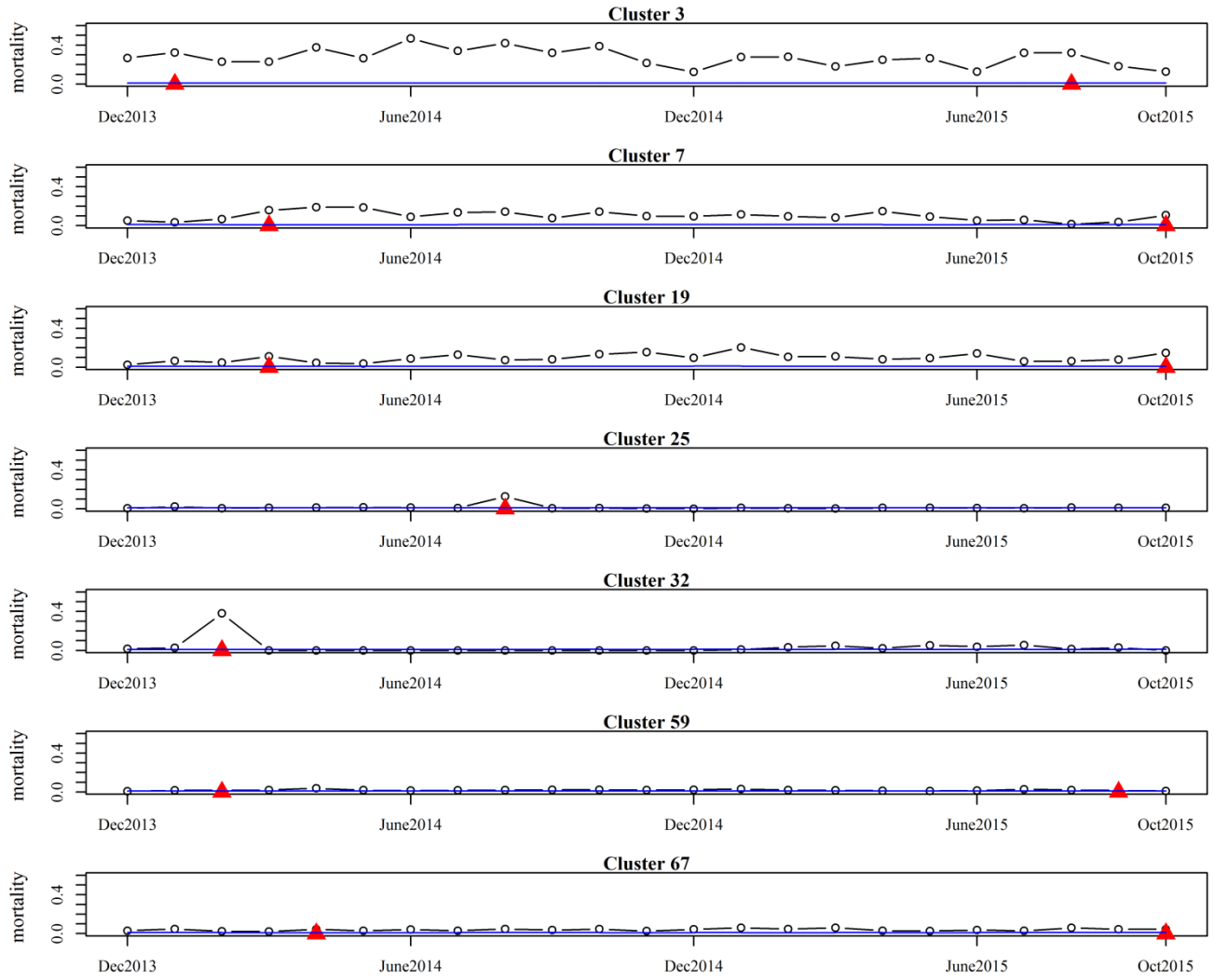
Statistically significant spatio-temporal clusters ($p<0.05$) of high mortality in Danish finisher herds between December 2013 and October 2015. For each cluster, the radius as well as starting and ending months (month 1 corresponds December 2013 and month 23 corresponds to October 2015) are given along with descriptive statistics of the number farms in the cluster (Number of locations), cluster significance (p-value) based on 999 Monte Carlo replications, Log Likelihood Ratio (LLR) and Relative Risk (RR) for a maximum spatial window size of 50% of the population at risk and a minimum of 1 month and maximum of 90% of the study period.

Cluster	Cluster type	Radius (m)	Start month	End month	Number of locations	LLR	*p* value	RR
1	Single-herd	0	1	13	1	4410.21	<0.001	30.71
2	Single-herd	0	2	21	1	3611.51	<0.001	5.92
3	Single-herd	0	2	21	1	3551.23	<0.001	27.17
4	Single-herd	0	20	20	1	1905.00	<0.001	76.53
5	Single-herd	0	1	20	1	1852.43	<0.001	7.49
6	Single-herd	0	1	20	1	1314.19	<0.001	4.36
7	Single-herd	0	4	23	1	1233.35	<0.001	10.08
8	Single-herd	0	1	20	1	1204.50	<0.001	9.94
9	Single-herd	0	10	23	1	1168.39	<0.001	14.56
10	Single-herd	0	1	13	1	1145.42	<0.001	16.97
11	Single-herd	0	18	23	1	1145.34	<0.001	15.87
12	Multiple-herd	14037	1	20	43	988.69	<0.001	1.43
13	Multiple-herd	3915	2	21	2	907.41	<0.001	3.24
14	Multiple-herd	706	7	23	2	838.51	<0.001	4.51
15	Multiple-herd	758	1	20	2	838.24	<0.001	2.81
16	Multiple-herd	1550	4	23	2	805.13	<0.001	6.34
17	Single-herd	0	1	20	1	790.02	<0.001	7.76
18	Multiple-herd	2295	1	20	2	781.64	<0.001	6.31
19	Single-herd	0	4	23	1	778.93	<0.001	9.68
20	Single-herd	0	4	23	1	751.48	<0.001	8.91
21	Single-herd	0	2	17	1	728.69	<0.001	7.87
22	Multiple-herd	6360	5	23	13	645.26	<0.001	1.64
23	Multiple-herd	4357	1	20	4	611.40	<0.001	3.43
24	Single-herd	0	1	10	1	605.94	<0.001	5.93
25	Single-herd	0	9	9	1	578.02	<0.001	12.05

26	Multiple-herd	4800	8	23	4	564.79	<0.001	2.30
27	Multiple-herd	2193	7	23	3	556.97	<0.001	3.00
28	Single-herd	0	6	23	1	544.99	<0.001	4.99
29	Single-herd	0	3	22	1	527.10	<0.001	6.20
30	Multiple-herd	2115	3	18	2	498.72	<0.001	2.83
31	Multiple-herd	36767	8	23	63	493.80	<0.001	1.30
32	Single-herd	0	3	3	1	482.91	<0.001	36.12
33	Single-herd	0	3	20	1	480.24	<0.001	9.24
34	Single-herd	0	2	21	1	461.68	<0.001	6.21
35	Single-herd	0	5	23	1	460.22	<0.001	8.73
36	Multiple-herd	1180	7	23	3	450.55	<0.001	2.29
37	Single-herd	0	4	23	1	431.08	<0.001	2.87
38	Single-herd	0	3	14	1	420.96	<0.001	6.42
39	Single-herd	0	3	22	1	413.98	<0.001	6.17
40	Single-herd	0	16	22	1	409.98	<0.001	4.34
41	Single-herd	0	2	19	1	379.82	<0.001	6.43
42	Single-herd	0	1	19	1	378.83	<0.001	3.48
43	Single-herd	0	1	14	1	364.08	<0.001	4.22
44	Single-herd	0	1	9	1	353.50	<0.001	6.75
45	Single-herd	0	3	22	1	349.64	<0.001	3.84
46	Multiple-herd	2208	2	21	3	345.87	<0.001	2.13
47	Single-herd	0	4	23	1	338.96	<0.001	2.15
48	Single-herd	0	1	20	1	330.88	<0.001	3.64
49	Multiple-herd	576	1	13	3	330.21	<0.001	2.07
50	Multiple-herd	10350	6	23	38	325.08	<0.001	1.28
51	Multiple-herd	3732	8	23	6	319.74	<0.001	1.90
52	Multiple-herd	3323	4	23	2	319.55	<0.001	2.172
53	Single-herd	0	15	21	1	306.01	<0.001	4.58
54	Multiple-herd	4101	1	15	4	299.51	<0.001	2.67
55	Single-herd	0	2	9	1	298.21	<0.001	4.25
56	Multiple-herd	3818	4	23	4	297.77	<0.001	1.97
57	Single-herd	0	1	20	1	294.55	<0.001	3.71
58	Multiple-herd	8545	3	20	18	292.00	<0.001	1.38
59	Single-herd	0	3	22	1	282.84	<0.001	1.97
60	Single-herd	0	4	23	1	278.10	<0.001	2.70

61	Multiple-herd	1964	11	16	2	276.63	<0.001	3.79
62	Multiple-herd	9356	2	21	2	268.76	<0.001	1.86
63	Multiple-herd	3648	2	14	4	263.07	<0.001	1.67
64	Single-herd	0	7	23	1	262.83	<0.001	2.62
65	Multiple-herd	3154	20	23	4	257.56	<0.001	2.69
66	Single-herd	0	7	23	1	252.50	<0.001	4.16
67	Single-herd	0	5	23	1	245.45	<0.001	3.90
68	Single-herd	0	4	23	1	217.06	<0.001	3.61
69	Single-herd	0	2	21	1	206.69	<0.001	2.69
70	Single-herd	0	1	14	1	205.96	<0.001	4.14
71	Single-herd	0	8	22	1	185.51	<0.001	2.68
72	Multiple-herd	6478	2	21	11	183.18	<0.001	1.39
73	Single-herd	0	1	14	1	167.86	<0.001	2.25
74	Multiple-herd	5237	2	21	8	167.82	<0.001	1.38
75	Single-herd	0	1	20	1	165.10	<0.001	2.57
76	Single-herd	0	4	23	1	116.18	<0.001	3.46

Appendix B: Temporal component description of 7 randomly selected single-herd temporal clusters in finisher herds. The mortality (proportion) is described for each herd, with the corresponding average mortality for all finisher herds (blue line) and the starting and ending period of the cluster (red arrows).



Manuscript 4

Monitoring endemic livestock diseases using laboratory diagnostic data: A simulation study to evaluate the performance of univariate process monitoring control algorithms

Ana Carolina Lopes Antunes[1]*, Fernanda Dórea[2], Tariq Halasa[1], Nils Toft[1]

[1] Section for Epidemiology, National Veterinary Institute –Technical University of Denmark, Denmark

[2] Department of Disease Control and Epidemiology, National Veterinary Institute, Sweden

*Corresponding author.

Published in Preventive Veterinary Medicine 2016; 127: 15–20.

Preventive Veterinary Medicine 127 (2016) 15–20

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Monitoring endemic livestock diseases using laboratory diagnostic data: A simulation study to evaluate the performance of univariate process monitoring control algorithms

Ana Carolina Lopes Antunes[a,*], Fernanda Dórea[b], Tariq Halasa[a], Nils Toft[a]

[a] Section for Epidemiology, National Veterinary Institute, Technical University of Denmark, Denmark
[b] Department of Disease Control and Epidemiology, National Veterinary Institute, Sweden

ARTICLE INFO

Article history:
Received 29 January 2016
Received in revised form 1 March 2016
Accepted 7 March 2016

Keywords:
Endemic disease
Laboratory results
Univariate process monitoring control algorithms

ABSTRACT

Surveillance systems are critical for accurate, timely monitoring and effective disease control. In this study, we investigated the performance of univariate process monitoring control algorithms in detecting changes in seroprevalence for endemic diseases. We also assessed the effect of sample size (number of sentinel herds tested in the surveillance system) on the performance of the algorithms.

Three univariate process monitoring control algorithms were compared: Shewart *p* Chart[1] (PSHEW), Cumulative Sum[2] (CUSUM) and Exponentially Weighted Moving Average[3] (EWMA). Increases in seroprevalence were simulated from 0.10 to 0.15 and 0.20 over 4, 8, 24, 52 and 104 weeks. Each epidemic scenario was run with 2000 iterations. The cumulative sensitivity[4] (CumSe) and timeliness were used to evaluate the algorithms' performance with a 1% false alarm rate. Using these performance evaluation criteria, it was possible to assess the accuracy and timeliness of the surveillance system working in real-time.

The results showed that EWMA and PSHEW had higher CumSe (when compared with the CUSUM) from week 1 until the end of the period for all simulated scenarios. Changes in seroprevalence from 0.10 to 0.20 were more easily detected (higher CumSe) than changes from 0.10 to 0.15 for all three algorithms. Similar results were found with EWMA and PSHEW, based on the median time to detection. Changes in the seroprevalence were detected later with CUSUM, compared to EWMA and PSHEW for the different scenarios. Increasing the sample size 10 fold halved the time to detection (CumSe = 1), whereas increasing the sample size 100 fold reduced the time to detection by a factor of 6.

This study investigated the performance of three univariate process monitoring control algorithms in monitoring endemic diseases. It was shown that automated systems based on these detection methods identified changes in seroprevalence at different times. Increasing the number of tested herds would lead to faster detection. However, the practical implications of increasing the sample size (such as the costs associated with the disease) should also be taken into account.

1. Introduction

Surveillance systems are critical for the effective and timely control of infectious diseases. Surveillance based on the continuous monitoring of secondary animal health data sources is a growing field, but the ability of automated systems to detect changes in the disease burden depends upon the choice of data source, their representativeness and sampling strategy (Buckeridge, 2007).

Sentinel surveillance systems are used when the health status of a population is periodically assessed based on a limited number of herds. These systems can be used for monitoring trends in diseases, in order to identify outbreaks and monitor the burden of disease in a population, providing a more cost-effective

Abbreviations: PRRSV, Porcine Reproductive and Respiratory Syndrome Virus; SPF System, Specific Pathogen Free System; EWMA, Exponentially Weighted Moving Average; CUSUM, Cumulative Sums; PSHEW, Shewart *p* Chart; UCL, Upper control limit; CumSe, Cumulative sensitivity.

* Corresponding author at: Section for Epidemiology, National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark.
E-mail address: aclan@vet.dtu.dk (A.C. Lopes Antunes).

[1] Shewart *p* Chart (PSHEW).
[2] Cumulative Sum (CUSUM).
[3] Exponentially Weighted Moving Average (EWMA).
[4] Cumulative sensitivity (CumSe).

http://dx.doi.org/10.1016/j.prevetmed.2016.03.005
0167-5877/

alternative to other disease surveillance methods. The testing frequency and sample size required for sentinel surveillance are dependent upon several factors, such as the goal of the surveillance, the etiology of the infectious agent, and the diagnostic test sensitivity and specificity (McCluskey, 2003).

One example of sentinel surveillance is the Danish monitoring program for Porcine Reproductive and Respiratory Syndrome Virus[5] (PRRSV), targeting swine breeding herds. Despite disease control efforts, PRRSV has continued (since its first diagnosis in 1992) to contribute to economic losses due to mortality in piglets, respiratory problems in growers and finishers, and reproductive problems in sows (Kvisgaard et al., 2011). The surveillance is primarily based on serological testing in order to maintain the Specific Pathogen Free System[6] (SPF System) certificate (Specific Pathogen Free System (SPF-SuS), 2015). The frequency of testing is dependent upon the health status defined as "red", "blue" and "green". The majority of Danish breeding herds have the "red" health status and therefore are tested on a monthly basis. In order to gain or maintain the SPF status, farmers must participate in a voluntary control program, for which they must provide health declarations and information on their herd health status for seven diseases, including PRRSV.

During the past decade, several studies have applied statistical quality control methods for syndromic surveillance in human and veterinary medicine (Dupuy et al., 2015; Dórea et al., 2013a,b; Jackson et al., 2007; Mandl et al., 2004). These studies mainly focused on detecting simulated outbreaks representing different scenarios of disease spread sometimes associated with emerging or re-emerging diseases. However, it may not be possible to generalize the performance of these algorithms in detecting disease outbreaks when monitoring changes in the burden of endemic diseases. Due to the availability of control measures such as vaccination or health management programs, the dynamic of disease spread is expected to be different for endemic diseases, resulting in a lower incidence rate when compared to exotic diseases. Moreover, the natural immunity developed from previous exposure to the agent also reduces an animal's susceptibility to endemic diseases. Therefore, it is unlikely that "extreme" changes in incidence and prevalence would be observed for diseases already present (and controlled) in the population. The dynamics of within-herd and between-herd endemic disease transmission also depends on the nature of the pathogen (Carslake et al., 2011), and can contribute to different temporal patterns of endemic disease spread.

For endemic diseases, it might be beneficial to monitor changes in disease prevalence rather than incidence. In these cases, the data differ from that obtained from traditional biosurveillance (generally related to incidence monitoring), as it is focused on the endemic scenario with less frequently sampled data. Moreover, this reflects the added complexity of monitoring endemic diseases, as disease burden is affected not only by the incidence but also by the disease duration and recovery rate.

In this study, we investigated the performance of three univariate process monitoring control algorithms commonly used in biosurveillance (Wagner and Moore, 2006) when applied to endemic disease monitoring. The algorithms were chosen for this study based on the simulated scenarios and on the type of simulated data (proportion data). The aim was to demonstrate that monitoring based on the weekly seroprevalence of a subset of the population for an endemic disease could be used to detect changes in disease occurrence in an accurate and timely manner. In addition, the impact of sample size (i.e. the number of sentinels) was explored. The design of our study was based on the Danish PRRSV monitoring program.

2. Methods

2.1. The Danish PRRSV monitoring program

Compulsory serological testing is performed on a monthly basis for all herds certified as SPF, which includes almost all Danish breeding herds (Specific Pathogen Free System (SPF-SuS), 2015).

Laboratory submission data stored in the National Veterinary Institute—Technical University of Denmark (DTU Vet) information management system and in the Laboratory for Swine Diseases-SEGES Pig Research Centre (VSP-SEGES) were used to determine the weekly number of Danish breeding herds tested for PRRSV and the corresponding between-herd seroprevalence from January 2007 to December 2014. Each laboratory submission consisted of individual blood samples collected from different animals in the same herd on the same day. Only submissions where at least two individual blood samples were subject to serological tests, including Blocking Enzyme-Linked Immunosorbent Assay (ELISA) (Sørensen et al., 1997; IDEXX, Ludwigsburg, Germany) and/or Immunoperoxidase monolayer assay (IPMA) (Bøtner et al., 1994), for one or both PRRSV strains were included in the analysis. Results from experimental studies were not included in the analysis.

Herds were classified as PRRSV seropositive when at least two individual blood samples in each submission tested PRRSV positive, independently of the PRRSV strain. The between-herd PRRSV seroprevalence was calculated weekly as the proportion of PRRSV positive herds within the total number of herds tested. The average between-herd PRRSV seroprevalence was 0.10 and the median weekly number of herds tested for PRRSV was 54 (minimum = 4, maximum = 85, standard deviation = 12.7).

2.2. Simulation experiment

As no additional knowledge of the true PRRSV seroprevalence was available, a simulation experiment was devised to derive the number of seropositive herds over a week, in order to control the development of changes. A baseline scenario of PRRSV seroprevalence of 0.1 was defined based on the data. In this scenario, the number of positive herds (X) per week from 2007 to 2014 were drawn from a binomial distribution ($X \sim bin(n, p)$) with a probability (p) of 0.1 and a sample size (n) equal to the number of Danish breeding herds tested for PRRSV in a given week. The weekly seroprevalence was calculated as the simulated number of seropositive herds, divided by the total number of herds tested that week. This simulation produced a stationary process representing an endemic disease under control.

The seroprevalence was increased from $p = 0.1$ to $p = 0.15$ and $p = 0.20$, over 4, 8, 24, 52 and 104 weeks. These 10 scenarios were designed to represent possible seroprevalence increases for the disease and population under study, considering the control measures in place. The final week of the simulated increase corresponded to the maximum increase. Following this, seroprevalence was maintained at the increased level (0.15 and 0.20). Two of these scenarios are illustrated in Fig. 1. The simulated increases in seroprevalence were started in random weeks between 2009 and 2012, and the weeks preceding this increase were used to train the algorithms.

2.3. Univariate process monitoring control algorithms

Three univariate process monitoring control algorithms used in previous studies in veterinary science (Dórea et al., 2013a,b; Dupuy et al., 2015) were investigated: Exponentially Weighted Moving Average (EWMA), Cumulative Sums (CUSUM) and Shewart p Chart

[5] Porcine Reproductive and Respiratory Syndrome Virus (PRRSV).
[6] Specific Pathogen Free System (SPF System).

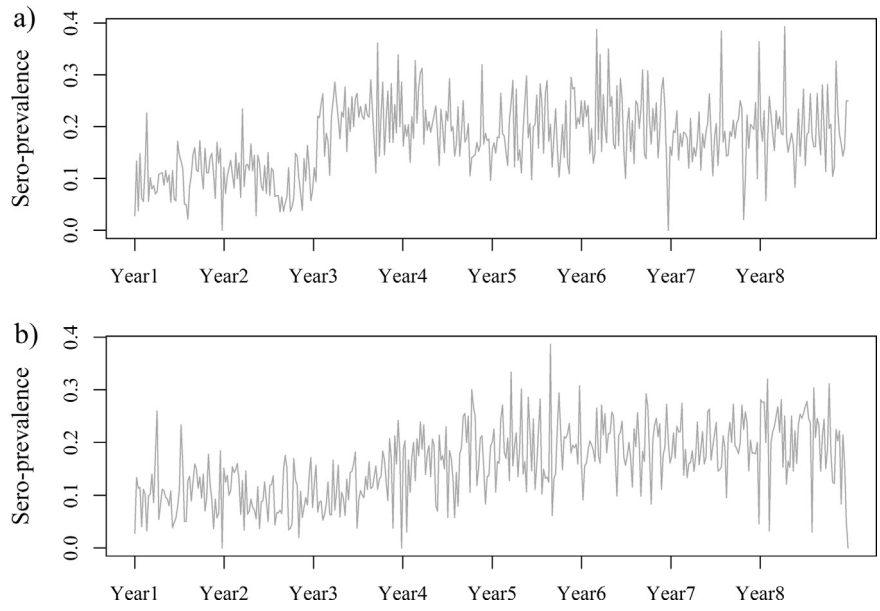


Fig. 1. Simulated between-herd weekly seroprevalence. The seroprevalence was simulated based on a binomial distribution with *n* equal to the number of Danish breeding herds tested for PRRS. Simulated changes in seroprevalence from 0.1 to 0.2 over (a) 4 weeks and (b) 104 weeks are represented.

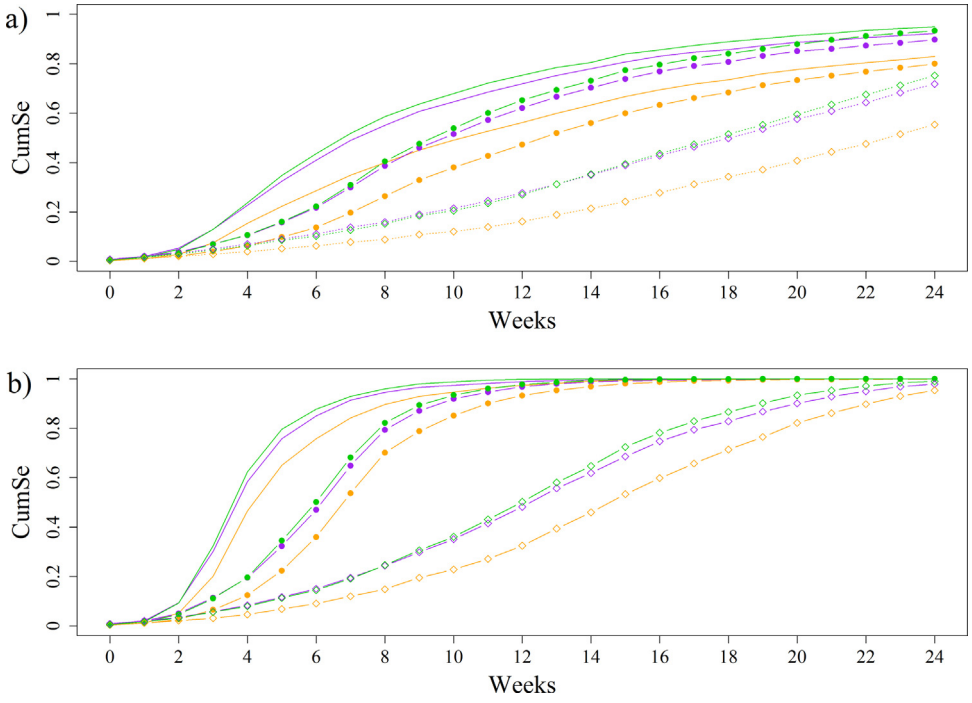


Fig. 2. Cumulative sensitivity (CumSe) of the univariate process control algorithms for different scenarios. Results for EWMA (purple), CUSUM (orange) and PSHEW (green) are represented in each scenario. Increases in the seroprevalence from 0.1 to (a) 0.15 and (b) 0.20 over 4 weeks (straight lines), 8 weeks (circles) and 24 weeks (diamonds) are represented. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(PSHEW). The PSHEW chart is used when the sampling fraction is nonconforming, i.e., the sample size is not fixed but is taken into account (Montgomery, 2009). The CUSUM and EWMA are commonly used to detect small shifts in the process (Montgomery, 2009).

The EWMA uses all previous time points, with a weighting for previous observations that reduces exponentially. The EWMA statistic Z and the upper control limit[7] (UCL) were obtained from Wagner and Moore (2006):

$$Z_t = \lambda p_t + (1-\lambda) Z_{t-1} \quad (1)$$

$$\mathrm{UCL}(Z)_t = \bar{Z}_t + L\sigma_{Z_t} \quad (2)$$

[7] Upper control limit (UCL).

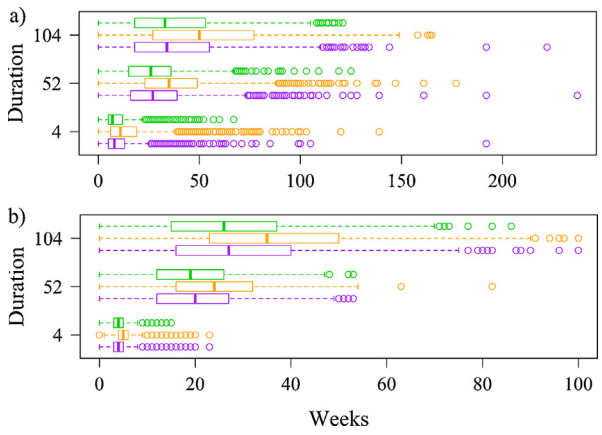


Fig. 3. Timeliness of univariate process control algorithms for the different scenarios. Comparative timeliness for changes in seroprevalence from 0.10 to (a) 0.15 and (b) 0.20 for EWMA (purple), CUSUM (orange) and PSHEW (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

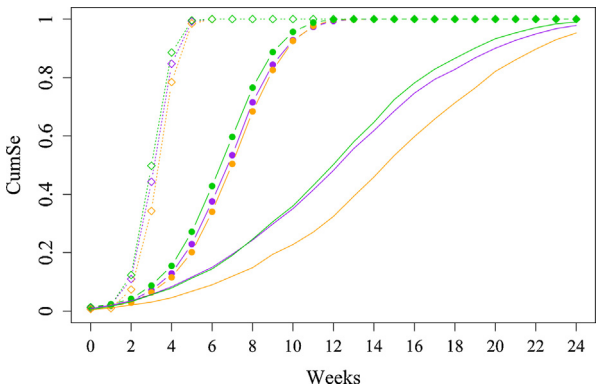


Fig. 4. Impact of the sample size on the algorithms cumulative sensitivity (CumSe). The results are represented for increases in the seroprevalence from 0.10 to 0.20 over 24 weeks. The CumSe of the EWMA (purple), CUSUM (orange) and PSHEW (green) are shown based on the real number of herds tested for PRRSV (straight lines), and a 10 (dots) versus 100 (diamonds) fold increase in the sample size. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where λ was the smoothing parameter, p_t was the seroprevalence for week t, L was the magnitude above the expected value, and $\overline{Z_t}$ was the average value of Z_t and Z_{t-1},

$$\sigma_{Z_t}{}^2 = var\left(Z_{<t}\right)\left(\frac{\lambda}{2-\lambda}\right)\left[1-\left(1-\lambda\right)^{2t}\right] \quad (3)$$

The CUSUM for week t was calculated as described in Wagner and Moore (2006):

$$CUSUM_t = max\left\{0, p_t - \overline{p_t} + CUSUM_{t-1}\right\} \quad (4)$$

where p_t was the seroprevalence in week t, and $\overline{p_t}$ was the mean seroprevalence in previous weeks. Alarms were raised if $CUSUM_t$ exceeded a threshold H, expressed in terms of the standard deviation of the control process.

The PSHEW for each week t (current time point) was calculated based on the average seroprevalence (p_t) and its UCL, as described in Montgomery (2009):

$$p_t = \frac{\Sigma_{i=1}^{t-1} x_i}{\Sigma_{i=1}^{t-1} n_i} \quad (5)$$

$$UCL(p)_t = p_t + k\sqrt{\frac{p_t\left(1-p_t\right)}{n_t}} \quad (6)$$

where x_i was the number of positive herds in previous weeks, n_i was the number of herds tested in previous weeks, n_t was the number of herds tested in week t, and K was the number of standard deviations. An alarm was raised for week t if the seroprevalence was higher than $UCL(p)_t$.

2.4. Calibration of algorithms

The three algorithms were calibrated to a false alarm rate of 1% when applied to the baseline (simulated constant seroprevalence of 0.10), using the following approach: the first 2 years (104 weeks) of the simulated seroprevalence were used to train the algorithm, and the following 6 years were used to calculate the false alarm rates. This process was simulated 2000 times for each parameter of the algorithm under calibration. The PSHEW was calibrated with K between 1 and 3; the CUSUM with H between 0.06 and 0.2 (corresponding to between 0.5 and 2 standard deviations); and the EWMA calibration explored λ from 0.1 to 0.9 and L between 2 and 4.

2.5. Performance assessment

Two performance indicators were defined to evaluate the accuracy and timeliness of the univariate process monitoring control algorithms.

The accuracy was evaluated using the cumulative sensitivity (CumSe) for week i after an initiated increase. The CumSe was defined as:

$$CumSe_i = \frac{\sum_{j=1}^{i} x_j}{n.iter}$$

where x_j was the number of iterations in which an alarm was given j weeks after an increase was started, and $n.iter$ corresponded to the total number of iterations used. It was considered that an increase in the seroprevalence was detected if an alarm was generated for $i \geq 0$. This method was chosen to assess the algorithms' performance for each week, including weeks where the gradual increases were simulated, in addition to the following weeks (after the end of the period).

Timeliness was defined as the number of weeks between the start of an increase and an alarm by the algorithm.

2.6. Convergence rate

A total of 10,000 iterations were simulated, with an increase in the seroprevalence from 0.10 to 0.15 achieved over 104 weeks. The number of iterations required to reach a stable timeliness and sensitivity (convergence) was determined visually by plotting the variance of the average timeliness or sensitivity with a stepwise increase of 100 iterations up to 10,000 iterations. Stable results were observed using 2000 iterations and hence all further simulations were run with 2000 iterations.

2.7. Assessing the effect of the number of sentinel herds

In order to assess the importance of the weekly number of sentinel herds (i.e. herds tested), the simulation experiment was repeated with n equal to 10 or 100 times the actual number of herds tested in a given week. For these larger samples, the three algorithms were again calibrated for a 1% false alarm rate based on 2000 iterations. The same scenarios and performance indicators were used to evaluate the algorithms with higher samples sizes.

All methods were implemented in R (version 3.1.1) (R Core Team, 2014).

Table 1
Performance evaluation of different detection algorithms. The timeliness was calculated as the median time to detect simulated changes in seroprevalence for all simulated scenarios for the Exponentially Weighted Moving Average (EWMA), Cumulative Sum (CUSUM) and Shewart *p* Chart (PSHEW).

			Algorithms	EWMA			CUSUM			PSHEW		
			Sample size	1×	10×	100×	1×	10×	100×	1×	10×	100×
Changes in seroprevalence	From 0.10 to 0.15	Duration	4	8	3	2	11	3	2	7	3	2
			8	10	5	3	13	5	3	10	5	3
			24	19	11	5	23	11	6	18	10	5
			52	27	18	8	35	19	9	26	17	8
			104	34	26	12	50	28	14	33	25	12
	From 0.10 to 0.20	Duration	4	4	2	2	5	3	2	4	2	2
			8	7	4	2	7	4	2	6	4	2
			24	13	7	4	15	7	4	12	7	3
			52	20	11	6	24	12	6	19	11	5
			104	27	18	8	35	19	9	26	17	8

3. Results

The selected values used to calibrate the algorithms for a false alarm rate of 1% corresponded to $\lambda = 0.8$ and $L = 3$ for the EWMA, $H = 0.12$ for the CUSUM and $K = 2.5$ for the PSHEW. The same parameters were retained in order to calibrate the EWMA for a 1% false alarm rate when the sample size increased 10 fold and 100 fold. For these bigger sample sizes and 1% false alarm rate, the CUSUM was re-calibrated with $H = 0.035$ and 0.011; the PSHEW was set up with $K = 2.5$ and 2.3.

The results for the CumSe are presented in Fig. 2. EWMA and PSHEW had higher CumSe when compared with the CUSUM from week 1 until the end of the period for all simulated scenarios. Changes in seroprevalence from 0.10 to 0.20 (Fig. 2b) were easier to detect (higher CumSe) when compared with changes from 0.10 to 0.15 (Fig. 2a) for all three algorithms. As an example, the CumSe of the EWMA for changes from 0.10 to 0.15 over 4 weeks was 0.23 at the end of the period, and 0.92 at 24 weeks. These values corresponded to 0.58 and 1.0 for changes from 0.10 to 0.20.

The final achieved CumSe was the same 24 weeks after the simulated increase for changes in the seroprevalence from 0.10 to 0.20 for the three algorithms (Fig. 2b). The increase over 24 weeks reached the same final CumSe as in other scenarios, showing that at the end of the period, the event would most likely already have been detected. The CumSe achieved at end of the period was lower for an increase from 0.10 to 0.15, and it resulted in a lower final CumSe after 24 weeks (Fig. 2a).

The highest CumSe for increases in seroprevalence from 0.10 to 0.15 over 52 weeks (not shown) corresponded to 0.72, 0.52 and 0.74 for EWMA, CUSUM and PSHEW, respectively, at the end of the period. These values rose to 0.98, 0.93 and 0.99 for EWMA, CUSUM and PSHEW for increases simulated over 104 weeks. Increases in the seroprevalence from 0.10 to 0.20 over 52 weeks resulted in CumSe of 0.99 for the three algorithms at the end of the period. This value was 1.0 for increases over 104 weeks at the end of the period.

Regarding the timeliness (Fig. 3), EWMA and PSHEW showed similar results based on the median time to detection. CUSUM detected changes in the seroprevalence later compared to EWMA and PSHEW for the different scenarios.

For a seroprevalence increase from 0.10 to 0.20, over 24 weeks, a 10-fold increase in sample size reduced the time to achieve CumSe = 1.0 from 24 weeks to 12 weeks, as shown in Fig. 4. A 100-fold increase in sample size reduced the time further to only 4 weeks. Therefore, increasing the sample size 10 fold halved the time to detection (CumSe = 1.0), whereas increasing the sample size 100 fold resulted in the time to detection being obtained six times faster.

Increasing the sample size also resulted in faster detections for the three algorithms in the different simulated scenarios (Table 1). For the 100 fold larger sample size, changes in seroprevalence were detected 36 weeks earlier when compared with the baseline for the CUSUM for an increase from 0.1 to 0.15 over 104 weeks.

4. Discussion

We investigated the performance of three univariate process monitoring control algorithms in monitoring data related to the burden of endemic diseases. The particular case study was inspired by seroprevalence data from the Danish PRRSV monitoring program, which uses breeding herds as sentinels.

The simulated increases in seroprevalence were meant to reflect expected scenarios of disease burden change in an endemic case, differing from previous studies which focused on scenarios of disease introduction (Dórea et al., 2013a,b; Dupuy et al., 2015). Changes in endemic diseases are likely to have different characteristics than emerging and re-emerging disease outbreaks (Dicker, 2012). Therefore, gradual increases in seroprevalence were simulated. The simulated scenarios were chosen based on the Danish pig production context, where almost all breeding herds are SPF-herds and farmers must follow rules concerning biosecurity, health control and transportation. Based on this, it is unlikely that the seroprevalence would increase above 0.20 in breeding herds.

A predefined acceptable false alarm rate of 1% was used to calibrate the algorithms. This decision was made to maintain confidence in the system and to reduce the economic impact of investigations due to false alarms.

EWMA and PSHEW had similar performance, and both were better in terms of accuracy (CumSe) and timeliness than the CUSUM for all scenarios. The general lower performance of CUSUM can be attributed to the noise in the simulated baseline—weeks with negative cumulative sum reset the algorithm to zero, reducing the chances of detecting small but sustained increases in the seroprevalence. This was also verified by earlier work (Dórea et al., 2013a,b). However, Dupuy et al. (2015) demonstrated that the CUSUM had a higher sensitivity when compared with the EWMA for detecting outbreaks in proportion data. This could be explained by the fact that this author simulated outbreaks representing scenarios potentially associated with emerging or re-emerging diseases, whereas the EWMA is known for detecting gradual increases in the mean, as simulated in this study.

As expected, our results revealed that faster increases resulted in a more rapid rise in CumSe. However, it is interesting to note that the final CumSe after 24 weeks is similar, showing that at the end of the period, the event would most likely have been detected for the simulated scenarios. The lower CumSe achieved at the end of fast increases (4 and 8 weeks) could be linked to the fact that all observations (simulated seroprevalence) until week *i* were used to fit the algorithm and to calculate the threshold, including weeks with simulated increases in seroprevalence. In this case, faster increases will reproduce higher thresholds, resulting in a lower sensitivity (i.e. less detections). This could be a limitation for a monitoring and

surveillance system working in real-time, resulting in an excessive number of alarms missed due to the low sensitivity. In order to overcome this issue, the approach suggested by Dórea et al. (2013b) – in which the baseline will be auto-corrected in case of an alarm – can be adopted in order to obtain a higher sensitivity.

The time to detection was shorter for faster increases in seroprevalence. Similar results were found in earlier work (Stoto and Schonlau, 2004), where simulated signals that increased in magnitude quickly over time tended to be detected more rapidly than slowly rising signals. The balance between the sensitivity, timeliness and specificity of the surveillance system is essential; the decision regarding which attribute to prioritize depends on the objectives of the system, the communication strategy between all surveillance stakeholders, and the financial resources used in investigations (Dupuy et al., 2015). Furthermore, the epidemiology of the disease, including the incubation time, the transmission mode, the current context (Wang et al., 2010) and its economic impact, should also be considered when deciding which attributes to prioritize.

Changing the number of herds tested had an impact on the simulated seroprevalence, contributing to the variation of the noise in the baseline. Increasing the sample size reduced the underlying variance in the seroprevalence, which justifies the need to re-calibrate the algorithms using different parameters. Furthermore, this reduced variation might also explain the almost identical performance of the CUSUM when increasing the sample size.

A 10-fold increase in sample size resulted in certain detection in half the time, whereas a 100-fold increase reduced this time by a factor of 6. However, the practical implications of increasing the number of tested herds, in particular the associated costs, should also be taken into account. An economic assessment of the impact of the disease and the cost of changing the current disease surveillance protocol would be needed to evaluate the gain in days/weeks of early detection.

5. Conclusions

This study investigated the performance of three univariate process monitoring control algorithms in monitoring endemic diseases. It was shown that automated systems based on these detection methods would eventually detect most changes in seroprevalence. Increasing the number of tested herds provided faster detection. However, the practical implications of increasing the sample size, such as the costs associated with the disease also need to be taken into account.

Acknowledgments

The authors would like to acknowledge Charlotte Sonne Kristensen for providing essential input on the Danish Pig Production context needed to design the simulation study.

References

Bøtner, A., Nielsen, J., Bille-Hansen, V., 1994. Isolation of porcine reproductive and respiratory syndrome (PRRS) virus in a Danish swine herd and experimental infection of pregnant gilts with the virus. Vet. Microbiol. 40, 351–360, http://dx.doi.org/10.1016/0378-1135(94) 90122-8.

Buckeridge, D.L., 2007. Outbreak detection through automated surveillance: a review of the determinants of detection. J. Biomed. Inform. 40, 370–379, http://dx.doi.org/10.1016/j.jbi.2006.09.003.

Carslake, D., Grant, W., Green, L.E., Cave, J., Greaves, J., Keeling, M., McEldowney, J., Weldegebriel, H., Medley, G.F., 2011. Endemic cattle diseases: comparative epidemiology and governance. Philos. Trans. R. Soc. Lond. B Biol. Sci. 366, 1975–1986, http://dx.doi.org/10.1098/rstb.2010.0396.

Dórea, F.C., McEwen, B.J., Mcnab, W.B., Revie, C.W., Sanchez, J., 2013a. Syndromic surveillance using veterinary laboratory data: data pre-processing and algorithm performance evaluation. J. R. Soc. Interface 10, 20130114, http://dx.doi.org/10.1098/rsif.2013.0114.

Dórea, F.C., McEwen, B.J., McNab, W.B., Sanchez, J., Revie, C.W., 2013b. Syndromic surveillance using veterinary laboratory data: algorithm combination and customization of alerts. PLoS One 8, e82183, http://dx.doi.org/10.1371/journal.pone.0082183.

Dicker, R.C., 2012. Principles of Epidemiology in Public Health Practice, 3rd ed. Cdc. U.S. Department of health and human services—Centers for Disease Control and Prevention (CDC), Atlanta.

Dupuy, C., Morignat, E., Dorea, F., Ducrot, C., Calavas, D., Gay, E., 2015. Pilot simulation study using meat inspection data for syndromic surveillance: use of whole carcass condemnation of adult cattle to assess the performance of several algorithms for outbreak detection. Epidemiol. Infect. 143, 2559–2569, http://dx.doi.org/10.1017/S0950268814003495.

Jackson, M.L., Baer, A., Painter, I., Duchin, J., 2007. A simulation study comparing aberration detection algorithms for syndromic surveillance. BMC Med. Inform. Decis. Making 7, 6, http://dx.doi.org/10.1186/1472-6947-7-6.

Kvisgaard, L., Hjulsager, C., Rathkjen, P., Breum, S., Trebbien, R., Larsen, L.E., 2011. PRRSV outbreak with high mortality in northen part of Denmark, in: EuroPRRS2011-Understanding and Combating PRRS in Europe. Novi Sad, Serbia, p. 60.

Mandl, K.D., Reis, B., Cassa, C., 2004. Measuring outbreak-detection performance by using controlled feature set simulations. MMWR. Morb. Mortal. Wkly. Rep. 53 (Suppl), 130–136.

McCluskey, B.J., 2003. Use of sentinel herds in monitoring and surveillance systems. In: Salman, M.D. (Ed.), Animal Disease Surveillance and Survey Systems: Methods and Applications. Blackwell Publishing, pp. 119–133.

Montgomery, D., 2009. Introduction to Statistical Quality Control, 7th ed. John Wiley & Sons, Inc., Arizona State University.

R Core Team, 2014 n.d. R. A Language and Environment for Statistical Computing.

Sørensen, K.J., Bøtner, A., Smedegaard Madsen, E., Strandbygaard, B., Nielsen, J., 1997. Evaluation of a blocking Elisa for screening of antibodies against porcine reproductive and respiratory syndrome (PRRS) virus. Vet. Microbiol. 56, 1–8, http://dx.doi.org/10.1016/S0378-1135(96)01345-4.

Specific Pathogen Free System (SPF-SuS), 2015. Specific Pathogen Free System [WWW Document] (accessed 5.12.15) URLhttps://www.spf.dk/.

Stoto, M., Schonlau, M.A., 2004. Syndromic surveillance: is it worth the effort? Chance 17, 19–24.

Wagner, M.M., Moore, A.W., 2006. Classical Time-Series Methods for Biosurveillance. In: Wagner, M.M., Moore, A.W., Aryel, R.M. (Eds.), Handbook of Biosurveillance. Elsevier Inc., San Franscisco, pp. 217–234, http://dx.doi.org/10.1016/B978-012369378-5/50000-4.

Wang, X., Zeng, D., Seale, H., Li, S., Cheng, H., Luan, R., He, X., Pang, X., Dou, X., Wang, Q., 2010. Comparing early outbreak detection algorithms based on their optimized parameter values. J. Biomed. Inform. 43, 97–103, http://dx.doi.org/10.1016/j.jbi.2009.08.003.

Manuscript 5

Dynamic generalized linear models for monitoring endemic diseases: moving beyond univariate process monitoring control

Ana Carolina Lopes Antunes[1]*, Dan Jensen[2], Tariq Halasa[1], Nils Toft[1]

[1] Section for Epidemiology, National Veterinary Institute – Technical University of Denmark, Denmark

[2] Section for Production and Health, Department of Large Animal Science – University of Copenhagen, Denmark

*Corresponding author.

Published in Proceedings for the Annual Meeting of the Society for Veterinary Epidemiology and Preventive Medicine (SVEPM) 2016, 69-79.

DYNAMIC GENERALIZED LINEAR MODELS FOR MONITORING ENDEMIC DISEASES: MOVING BEYOND UNIVARIATE PROCESS MONITORING CONTROL ALGORITHMS

A.C. LOPES ANTUNES*, D. JENSEN, T. HALASA, N. TOFT

SUMMARY

The objective was to use a Dynamic Generalized Linear Model (DGLM) based on a binomial distribution with a linear trend, for monitoring the PRRS (Porcine Reproductive and Respiratory Syndrome sero-prevalence in Danish swine herds. The DGLM was described and its performance for monitoring control and eradication programmes based on changes in PRRS sero-prevalence was explored. Results showed a declining trend in PRRS sero-prevalence between 2007 and 2014 suggesting that Danish herds are slowly eradicating PRRS. The simulation study demonstrated the flexibility of DGLMs in adapting to changes in trends in sero-prevalence. Based on this, it was possible to detect variations in the growth model component. This study is a proof-of-concept, demonstrating the use of DGLMs for monitoring endemic diseases. In addition, the principles stated might be useful in general research on monitoring and surveillance of endemic and (re-)emerging diseases.

INTRODUCTION

New methods for monitoring animal diseases continue to be an active area of research. In the past decade, several studies applied statistical quality control methods for syndromic surveillance in human and veterinary medicine (Buckeridge et al., 2005; Jackson et al., 2007; Dórea et al., 2013). Many of these studies applied univariate process monitoring control algorithms to detect outbreaks of re-emerging diseases. In these cases, control and/or eradication measures are implemented whenever certain threshold levels related to the infection or disease status have been exceeded. However, the term "monitoring" can also be used to describe actions, where a continuous process of collecting data on animal diseases is ongoing, but without any instant control activities (Salman, 2003).

For endemic diseases, it is common to implement control and eradication programmes at herd and regional levels to reduce the economic impact of diseases. Often, these programmes are based on laboratory diagnostics. One example is the Danish monitoring programme for Porcine Reproductive and Respiratory Syndrome (PRRS).

* Ana Carolina Lopes Antunes, National Veterinary Institute, Technical University of Denmark, Bülowsvej 27, 1870 Frederiksberg C, Denmark. Email: aclan@vet.dtu.dk

Despite disease control efforts in Denmark, PRRS continues to contribute towards the economic losses of the industry since its first diagnosis in 1992. PRRS monitoring is primarily based on serological testing performed on regular basis from herds that have the Specific Pathogen Free System (SPF) certificate (Specific Pathogen Free System (SPF-SuS), 2015). The frequency of testing depends on the SPF herd type, being performed once a month for breeding herds and once a year for finisher herds. The SPF herds represent about 40% of all Danish swine (SPF-SuS, 2015). For non-SPF herds, PRRS diagnostic test are not mandatory and different reasons might explain the variation in frequency of laboratory testing. Thus, diagnostic laboratory submissions of PRRS are collected based on different purposes and frequencies in Denmark.

For disease monitoring, the resulting time series are characterized by observational noise as a result of the variation in the disease prevalence and of the number of samples and herds tested over time. Furthermore, its randomness and non-stationary nature are difficult to model. In these cases, it is necessary to use models with a more dynamic structure, where it is possible to add trends, cyclic patterns and also allow the parameters to change over time. State space models are one possible approach in which relevant prior knowledge and current information are combined. While state space models have been adopted in herd management (Jensen et al., 2015; Madsen & Kristensen, 2005; Ostersen et al., 2010), their use has been underutilized in veterinary sciences for diseases surveillance purposes. In the literature, there are few studies using these type of models for disease monitoring and surveillance in humans (Cao et al., 2014; Cowling et al., 2006).

The objective was to use a state space model for monitoring the PRRS sero-prevalence in Danish swine herds. The binomial DGLM with a linear growth was described and its performance for monitoring control and eradication programmes based on changes in PRRS sero-prevalence was explored. This study is a proof of concept, demonstrating the use of DGLMs for monitoring endemic disease, but the principles stated might also be useful in general research on monitoring and surveillance of endemic and (re-)emerging diseases.

MATERIALS AND METHODS

Data source

Laboratory submission data stored in the National Veterinary Institute – Technical University of Denmark (DTU Vet) information management system and in the Laboratory for Swine Diseases-SEGES Pig Research Centre (VSP-SEGES) were used to determine the weekly PRRS sero-prevalence in Danish swine herds from January 2007 to December 2014.

Each laboratory submission consisted of individual blood samples collected from the same herd on the same day from different animals. Only submissions where at least 2 individual blood samples were tested by serological tests including Blocking Enzyme-Linked Immunosorbent Assay (ELISA) and/or Immunoperoxidase monolayer assay (IPMA) for one or both PRRSV (Porcine Reproductive and Respiratory Syndrome Virus)strains were included in the analysis. These serological tests used were a DTU Vet “in-house” ELISA (Sørensen et. al, 1997) and IPMA (Bøtner et al., 1994). Furthermore, diagnostic test results performed at VSP-SEGES were based on IDEXX PRRS X3 Ab ELISA test (IDEXX, Ludwigsburg, Germany). Results from experimental studies were excluded from the analysis.

Herds were classified as PRRS sero-positive when at least 2 individual blood samples in each submission tested PRRS positive, independently of the PRRS strain. The between-herd PRRS sero-prevalence was calculated weekly as the proportion of PRRS positive herds from the total number of herds tested for PRRS.

Modelling

A binomial DGLM with a linear growth as described by West and Harrison (1997) was used to model the data. The general purpose of the DGLM is to estimate the underlying parameter vector from the observed data (θ) combined with any prior information available at time 0 (D_0) before any observation is made. This can be achieved sequentially where the estimated value is updated each time a new value (PRRS sero-prevalence) is obtained. In this case, the conditional distribution of θ_t given by D_t ($\theta_t|D_t$) was estimated. These models can be used to estimate a one-step forecast of the mean, allowing for a comparison with the actual observed PRRS sero-prevalence. Moreover, the linear growth component includes a time-varying slope (or local linear trend), allowing the system to adapt to a possible positive or negative growth for each *t*.

In general, the DGLM consists of an observation equation (Eq. 1) and a system equation (Eq. 2):

$$g(p_t) = F_t'\theta_t \quad (1)$$

$$\theta_t = G_t\theta_{t-1} + W_t \quad (2)$$

Equation 1 describes how the values of an observation (PRRS sero-prevalence) derive from $g(p_t)$, depends on an unobservable parameter vector (θ) for time *t* based on a linear function. For the model specification, $g()$ is the identity function. Equation 2 describes the dynamic properties of the unobservable parameter vector θ. In this study, the transposed design matrix (F_t') has the structure presented in Table 1, in order to estimate underlying values of PRRSV sero-prevalence according to Eq 1. The system matrix (G) used to update the mean of the PRRSV sero-prevalence for each time step taking into account the trend. Both matrix structures were constant for each *t* (week). The variance-covariance matrix (W_t) describes the evolution of variance and covariance of each parameter for each time step. Rather than estimating (W_t), the system variance was modelled using a discount factor (see Eq. 4).

Table 1. Matrices structure used in Eq. 1 and 2.

F_t'	G_t
$[1 \quad 0]$	$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

The DGLM update for each time step t was performed as follows:

a) the posterior distribution for Θ_{t-1} was expressed by a prior mean (m_{t-1}) and a variance (C_{t-1}), $(\Theta_{t-1}\,|D_{t-1})\sim[m_{t-1}, C_{t-1}]$;

b) the prior distribution for θ_t $(\Theta_t\,|D_{t-1})\sim[a_t, R_t]$ was made based on the prior mean (a_t) and prior variance (R_t) which were calculated as described in Eq. 3 and 4. The specification of the variance components was specified using a discount factor (δ);

$$a_t = G_t m_{t-1} \tag{3}$$

$$R_t = \frac{1}{\delta} G_t C_{t-1} G_t' \tag{4}$$

c) the prior distribution for Y_t $(Y_t\,|D_{t-1})\sim[f_t, q_t]$ was calculated based on the forecast mean (f_t) and forecast variance (q_t) (Eq. 5 and 6);

$$f_t = F_t' a_t \tag{5}$$

$$q_t = F_t' R_t F_t \tag{6}$$

d) the posterior mean (f_t^*) and variance (q_t^*) were calculated as described in Eq. 7 and 8. In this case, it was assumed that the prior probability p (PRRS sero-prevalence) of a binomial distribution was Beta(α, β). If κ successes (PRRS positive herds) out of n trials (number of herds tested for PRRS) were observed, the posterior p, given the new observation was Beta($\alpha_t + \kappa_t\,, \beta_t + n_t - \kappa_t$). The parameters α_t and β_t were calculated according to Eq. 9 and 10.

$$f_t^* = \frac{\alpha_t + \kappa_t}{\alpha_t + \beta_t + n_t} \tag{7}$$

$$q_t^* = \frac{f_t^*(1 - f_t^*)}{\alpha_t + \beta_t + n_t + 1} \tag{8}$$

$$\alpha_t = f_t \left(\frac{f_t(1 - f_t)}{q_t} - 1\right) \tag{9}$$

$$\beta_t = (1 - f_t) \left(\frac{f_t(1 - f_t)}{q_t} - 1\right) \tag{10}$$

e) the posterior distribution for Θ_{t-1} in a) was calculated based on its mean matrix m_t and its variance-covariance matrix C_t as demonstrated in Eq. 11 and 12.

$$m_t = a_t + R_t F_t (f_t^* - f_t)/q_t \tag{11}$$

$$C_t = R_t - R_t F_t F_t' R_t (1 - q_t^*/q_t)/q_t \tag{12}$$

Model initialization: Reference analysis was used to estimate the initial parameters $D_0 \sim [m_0, C_0]$ as described by West and Harrison (1997). We defined the matrices K_t and H_t and the vectors k_t and h_t for the first two observations $p_{1:2}$.

For $t = 1$, the initial parameters were defined as H_1=0, h_1=0, $K_1 = H_1 + F_1 F_1'$ and $k_1 = h_1 + F_1 p_1$. For $t = 2$, the vectors and matrices were updated as described in Eq. 13 to Eq. 16.

$$H_2 = G_2^{-1\prime} K_1 G_2^{-1} \tag{13}$$

$$h_2 = G_2^{-1\prime} k_1 \tag{14}$$

$$K_2 = H_2 + F_2 F_2' \tag{15}$$

$$k_2 = h_2 + F_2 p_2 \tag{16}$$

Then, the prior distribution for $t = 3$ was calculated according to Eq. 17 and 18.

$$m_2 = K_2^{-1} k_2 \tag{17}$$

$$C_2 = K_2^{-1} \tag{18}$$

System variance: The DGLM model was run based on different discount factors (δ) ranging from 0.1 up to 1 by increments of 0.01. The discount factor which minimized the sum of the squared forecast errors based on the first two years of the data was chosen for the analysis.

Monitoring model components: The values obtained from the m vector for each time step *t* were used to decompose the time series and obtain the model growth (PRRS sero-prevalence trend). The variance on the growth parameter was calculated from the C matrix and used to calculate 95% confidence intervals (CI).

Simulated scenarios: PRRS sero-prevalence baseline was simulated for 8 years, in which the number of positive herds (X) per week was drawn from a binomial distribution ($X \sim bin(\mathrm{n}, \mathrm{p})$) with a probability (p) (PRRS sero-prevalence) and a sample size (n) equal to the number of Danish herds tested for PRRS per week between 2007 and 2014. The weekly sero-prevalence was calculated as the simulated number of sero-positive herds divided by the total weekly number of herds tested. The first 104 weeks were simulated with a constant initial prevalence of 0.24, corresponding to the average PRRS sero-prevalence in Danish herds observed based on the laboratory diagnostic data from 2007 to 2014. In the first scenario (Scenario A), a constant decrease from p=0.24 to p=0.10 during 4 years followed by constant sero-prevalence was simulated. The second scenario (Scenario B) represented a decrease in the sero-prevalence from p=0.24 to p=0.10 during 2 years, followed by an increase to p=0.18 during the subsequent 2 years.

The sensitivity (Se) and timeliness were used to evaluate the performance of the DGLM to detect significant changes in the simulated scenarios. The Se was defined as the proportion of

simulations in which significant changes in the model growth component from zero were found. Timeliness was defined as the number of weeks between a change in the PRRS sero-prevalence (decrease, increase, constant) was simulated and detected.

Convergence rate: A total of 20,000 simulations of weekly PRRS sero-prevalence with a constant decrease from 0.24 to 0.05 over 5 years were carried out. The number of iterations needed to reach a stable variance in the average time to detect significant changes (convergence) was determined visually by plotting the variance of the average timeliness with a stepwise increase of 100 iterations up to 20,000 iterations against the number of iterations. Stable results were observed when using only 10,000 iterations and hence all further simulations were run with 10,000 iterations.

All analyses were performed using R (version 3.1.1) (R Core Team, 2014).

RESULTS

Data description

A total of 56,341 laboratory submissions from 5,390 Danish swine were included in the analysis. The average weekly number of herds tested for PRRSV was 130 (min=9, max=206); the mean weekly number of PRRS positive herds was 31 herds (min=0, max= 60). The weekly average PRRS sero-prevalence was 0.24 (min=0, max=0.38). The yearly average of PRRS sero-prevalence declined from 0.28 in 2007 to 0.20 in 2014, with an average decrease of 0.01 per year.

Model initialization and discount factor

Table 2 shows the posterior C_2 and m_2 matrices obtained from the reference analysis and used as priors for the DGLM model for $t = 3$. The discount factor which minimized the sum of forecast errors for the data was δ=0.98.

Table 2. Priors for *t* = 3 obtained from the reference analysis.

m_2	C_2
$\begin{bmatrix} 0.30 \\ 0.07 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$

Modelling and decomposing DGLM

Results show a declining trend of PRRS sero-prevalence between 2007 and 2014. Significant decreases (95% CI excluding zero) were detected mainly in the last 6 months of 2007; end of 2008 to the first semester of 2010 and from the last quarter of 2010 until the beginning of 2013 (Fig. 1). No significant increases in PRRS sero-prevalence were observed and all values for the growth component were below 0.

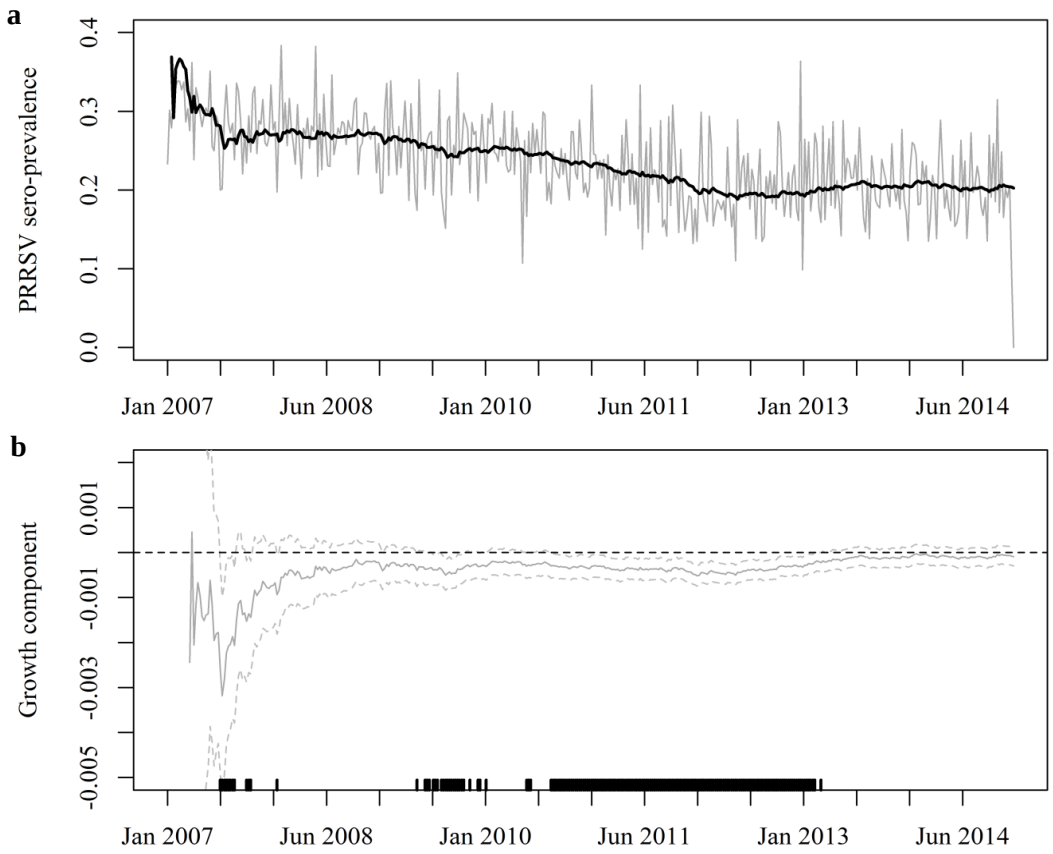


Fig. 1 Using a DGLM to monitor PRRS sero-prevalence in Danish swine herds from 2007 to 2014. Results show the weekly PRRS sero-prevalence and the filtered mean (black) (a) and the corresponding DGLM growth component (b). The black rugs indicate were the growth component is significantly different from zero.

Simulated scenarios

The simulated scenarios are represented in Fig. 2 and 3. The results for the simulation study are presented in Table 3. Significant changes in the model growth component from zero were found in both scenarios. However, the DGLM detected changes in the growth with a higher sensitivity for decreasing changes when compared to constant growth in the time series. The lowest sensitivity was found for Scenario A when the PRRS sero-prevalence became constant after the decrease, with the DGLM growth component being non-significantly different from zero in 39.02% of the simulations.

Table 3. Timeliness (weeks) and Se for the simulated scenarios.

	Scenario A		Scenario B	
Intervention	Decrease	Constant	Decrease	Increase
Timeliness (median)	47	96	27	146
(min-max)	(0-89)	(57-106)	(0-56)	(110-257)
Se (%)	100	39.02	100	99.64

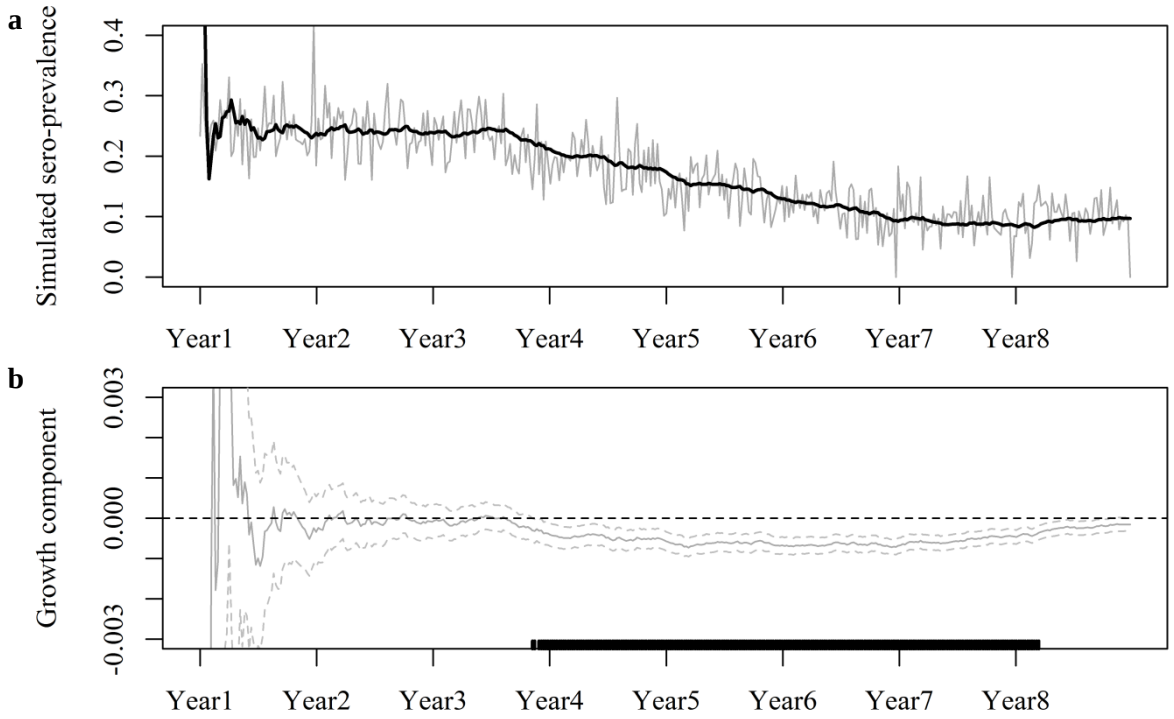


Fig. 2 Simulated control program Scenario A. PRRS sero-prevalence was constant during 104 weeks, followed by a decrease to 0.10 during 208 weeks and then a constant prevalence. The DGLM filtered mean (black line) (a) and the corresponding DGLM growth component (b) (grey lines) are presented. The black rugs indicate a significant negative the growth component based on 95% CI.

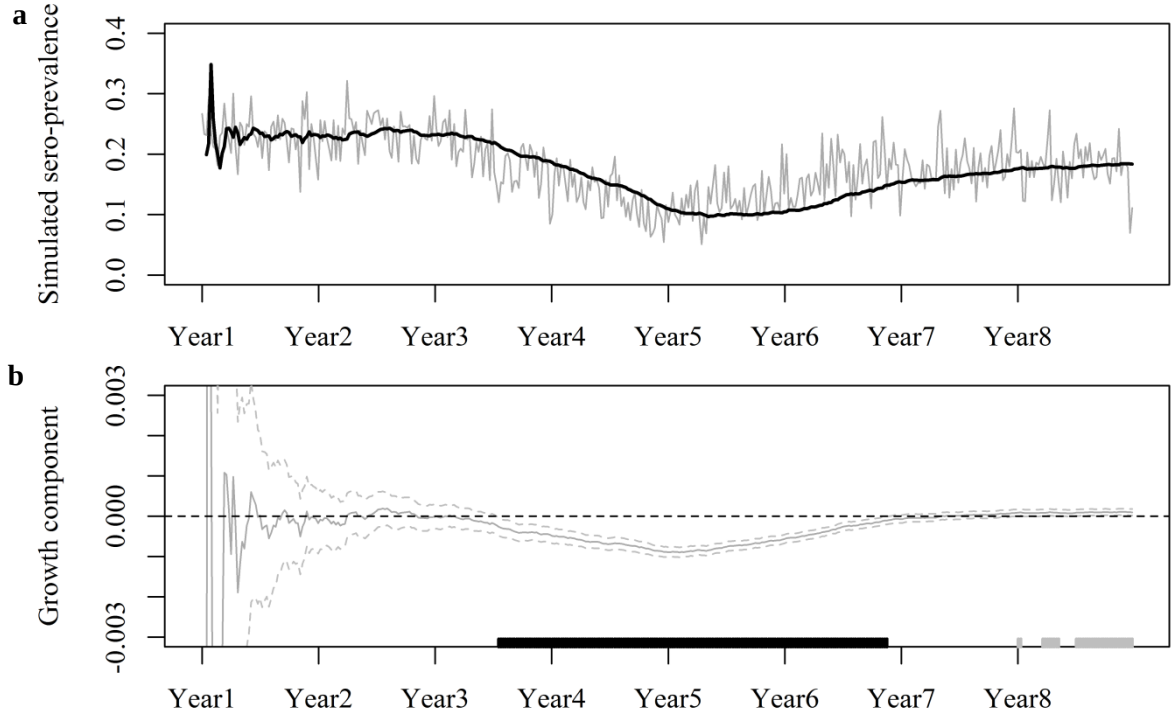


Fig. 3 Simulated control program Scenario B. PRRS sero-prevalence was constant during 104 weeks, followed by a decrease to 0.10 during 104 weeks and an increase up to 0.18 during 104 weeks. The DGLM filtered mean (black line) (a) and the corresponding DGLM growth component (b) (grey lines) are presented. The black and grey rugs indicate significant declines and increase in the growth component based on 95% CI, respectively.

DISCUSSION

The objective of this study was to use a binomial DGLM with a linear growth component for monitoring PRRS sero-prevalence in similar contexts to the Danish Pig Industry. The same model can be used for monitoring other prevalence data. These types of models can also be derived for Poisson distribution for monitoring count data, such as the number of samples submitted for analysis etc. Moreover, an ordinary Dynamic Linear Model (DLM) can be used if the data are normally distributed. They also allow for modelling interventions as well as changes in level shift through multi-process models (Thysen, 1993). The DGLM provide a flexible framework in which it is possible to include different data sources in a multivariate process as shown by (Jensen et al., 2015). Moreover, the use of this method allows monitoring of trends and also other components of time series such as seasonal, regression and autoregressive effects components which have a wide interest in biomedical time series applications (West & Harrison, 1997).

As no information on PRRS outbreaks and eradication programmes is available for Danish swine herds, a simulation study was conducted. One limitation of this study is related to the simulation approach used; the simulated sero-prevalence was based on a binomial distribution. The variation in the number of herds tested had an impact on the simulated prevalence contributing to the variation (noise in the baseline). As a consequence, the timeliness to detect interventions showed a wide range of values and the sensitivity was not similar for all interventions. One approach to overcome this issue could be to aggregate the data on a monthly basis, thus reducing the noise in the baseline and possibly improve the performance of the model to adapt to changes in the trend.

The DGLM model was able to detect changes in both scenarios. However, it is important to notice that decreases were larger compared to the increases, corresponding to an absolute decay in sero-prevalence of 0.145 and absolute increase of 0.08. For scenario B, significant positive changes in the model growth component were found after a period in which non-significant changes were found. These justify the longer time needed to detect increases. The variation in the growth parameter was monitored based on 95% CI's. Different approaches could be, *e.g.* Shewart control charts, cumulative sensitivities, V-mask (Montgomery, 2013) or target values, which might yield improved the performances.

In a Bayesian framework the choice of priors is critical for making inference. Reference analysis was used to initiate the DGLM model. From a practical point of view, when a system is set up, the number of observations is low to make the influence of the priors significant. In this case, the use of "non-informative" priors can be used. This method offers an easily applied default analysis (West & Harrison, 1997) when running a DGLM. However, it can be seen from the simulated scenarios that the DGLM takes 3 months to adapt to the data. For this reason, it is important to have historical data (retrospective analysis) to train the model when setting up a monitoring system.

The systems variance was defined based on a discount factor, expressing the decay of information in the system. Defining δ=0.98 implies a small systems variance with a very slow adaptation to new observations. This value was defined using the same method described in Kristensen et al. (2010), where δ should optimized for the performance of the model in making forecasts, *i.e.*, minimizing the forecast errors for the first two years of data (retrospective analysis). In recent literature (Bono et al., 2012; Jensen et al., 2015), the Expectation-Maximization algorithm (Dempster et al., 1977) was used to define the W variance-covariance matrix. This approach offers a general approach to iterative computation

of maximum-likelihood estimates when the observations can be viewed as incomplete data. The use of a discount factor provides a parsimonious approach when compared to the full estimation of W.

In summary, results show a declining trend on PRRS sero-prevalence between 2007 and 2014 suggesting more Danish herds are eradicating PRRS. The simulation study highlighted that DGLM are flexible models able to adapt to changes in the time series. It was possible to detect variations in the growth component of simulated scenarios. This study is a proof of concept, demonstrating the use of DGLMs for monitoring endemic disease, but the principles stated might also be useful in general modelling, monitoring and surveillance of (re)emerging diseases. Further analysis to compare the performance of the DGLM, including different components, to other models will be investigated in future studies.

ACKNOWLEDGEMENTS

The authors would like to thank to VSP-SEGES for providing the serological test results.

REFERENCES

Bono, C., Cornou, C. and Kristensen, A.R. (2012). Dynamic production monitoring in pig herds I: Modeling and monitoring litter size at herd and sow level. Livestock Science 149, 289–300. doi:10.1016/j.livsci.2012.07.023

Buckeridge, D. L., Switzer, P., Owens, D., Siegrist, D., Pavlin, J. and Musen, M. (2005). An evaluation model for syndromic surveillance: assessing the performance of a temporal algorithm. MMWR. Morbidity and Mortality Weekly Report 54 Suppl, 109–115. doi:su5401a18

Bøtner, A., Nielsen, J. and Bille-Hansen, V. (1994). Isolation of porcine reproductive and respiratory syndrome (PRRS) virus in a Danish swine herd and experimental infection of pregnant gilts with the virus. Veterinary Microbiology 40(3-4), 351–360. doi:10.1016/0378-1135(94)90122-8

Cao, P.-H., Wang, X., Fang, S.-S., Cheng, X.-W., Chan, K.-P., Wang, X.-L., Lu, X.,Wu, C.-L.,Tang, X.-J., Zhang, R.-L., Ma, H.-W., Cheng, J.-Q., Wong, C.-M. and Yang, L. (2014). Forecasting influenza epidemics from multi-stream surveillance data in a subtropical city of China. PloS One 9(3), e92945. doi:10.1371/journal.pone.0092945

Cowling, B.J., Wong, I. O.L., Ho, L.-M., Riley, S. and Leung, G.M. (2006). Methods for monitoring influenza surveillance data. International Journal of Epidemiology 35(5), 1314–1321. doi:10.1093/ije/dyl162

Dempster, A.P.A., Laird, N.M.N. and Rubin, D.D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B Methodological, 39(1), 1–38. doi:http://dx.doi.org/10.2307/2984875

Dórea, F.C., McEwen, B.J., Mcnab, W.B., Revie, C.W. and Sanchez, J. (2013). Syndromic surveillance using veterinary laboratory data: data pre-processing and algorithm

performance evaluation. Journal of the Royal Society, Interface / the Royal Society 10(83), 20130114. doi:10.1098/rsif.2013.0114

Dupuy, C., Morignat, E., Dorea, F., Ducrot, C., Calavas, D. and Gay, E. (2015). Pilot simulation study using meat inspection data for syndromic surveillance: use of whole carcass condemnation of adult cattle to assess the performance of several algorithms for outbreak detection. Epidemiology and Infection 143(12), 2559–2569. doi:10.1017/S0950268814003495

Jackson, M.L., Baer, A., Painter, I. and Duchin, J. (2007). A simulation study comparing aberration detection algorithms for syndromic surveillance. BMC Medical Informatics and Decision Making 7(1), 6. doi:10.1186/1472-6947-7-6

Jensen, D.B., Cornou, C., Toft, N. and Kristensen, A.R. (2015). A multi-dimensional dynamic linear model for monitoring slaughter pig production. In 7th European Conference on Precision Livestock Farming (pp. 503–512). Milan. doi:10.1017/CBO9781107415324.004

Kristensen, A.R., Jørgensen, E. and Toft, N. (2010). Herd Management Science II. Advanced topics. p92

Madsen, T.N. and Kristensen, A.R. (2005). A model for monitoring the condition of young pigs by their drinking behaviour. Computers and Electronics in Agriculture 48(2), 138–154. doi:10.1016/j.compag.2005.02.014

Montgomery, D. (2013). Introduction to statistical quality control. (7th Ed.). Arizona State University: John Wiley and Sons, Inc. 234-276,416-447p

Ostersen, T., Cornou, C. and Kristensen, R. (2010). Detecting oestrus by monitoring sows' visits to a boar. Computers and Electronics in Agriculture 74(1), 51–58. doi:10.1016/j.compag.2010.06.003

R Core Team. (2014). R. A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.r-project.org

Specific Pathogen Free System (SPF-SuS). (2015). Specific Pathogen Free System. Retrieved May 12, 2015, from https://www.spf.dk/

Sørensen, K. J., Bøtner, A., Smedegaard Madsen, E., Strandbygaard, B. and Nielsen, J. (1997). Evaluation of a blocking Elisa for screening of antibodies against porcine reproductive and respiratory syndrome (PRRS) virus. Veterinary Microbiology 56(1-2), 1–8. doi:10.1016/S0378-1135(96)01345-4

Thysen, I. (1993). Monitoring Bulk Tank Somatic Cell Counts by a Multi-Process Kalman Filter. Acta Agriculture Scandinavia (43), 58 – 64

West, M. and Harrison, J. (1997). Bayesian Forecasting and Dynamic Models (2nd Ed.). New York, USA: Springer. 136-137,p516-560

Manuscript 6

A simulation study to evaluate the performance of statistical monitoring methods when applied to different time-series components in the context of control programs for endemic diseases

Ana Carolina Lopes Antunes[1]*, Dan Jensen[2], Tariq Halasa[1], Nils Toft[1]

[1] Section for Epidemiology, National Veterinary Institute – Technical University of Denmark, Denmark

[2] Section for Production and Health, Department of Large Animal Science – University of Copenhagen, Denmark

*Corresponding author.

Submitted to Plos One.

A simulation study to evaluate the performance of statistical monitoring methods when applied to different time-series components in the context of control programs for endemic diseases

Ana Carolina Lopes Antunes[1*], Dan Jensen[2], Tariq Halasa[1], Nils 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

*Corresponding author. Email:aclan@vet.dtu.dk

Abstract

Disease monitoring and surveillance play a crucial role in control and eradication programs, as it is important to track implemented strategies in order to reduce and/or eliminate a specific disease. The objectives of this study were to assess the performance of different statistical monitoring methods for endemic disease control program scenarios, and to explore what impact of variation (noise) in the data had on the performance of these monitoring methods.

We simulated 16 different scenarios of changes in weekly sero-prevalence. The changes included different combinations of increases, decreases and constant sero-

prevalence levels (referred as events). Two space-state models were used to model the time series, and different statistical monitoring methods (such as univariate process control algorithms and monitoring of the trend component) were tested. Performance was evaluated based on the number of iterations in which an alarm was raised for a given week after the changes were introduced.

Results revealed that the Shewhart Control Chart was better at detecting increases over decreases in sero-prevalence, whereas the opposite was observed for the Tabular Cumulative Sums. The trend-based methods detected the first event well, but performance was poorer when adapting to several consecutive events. The V-Mask method seemed to perform most consistently, and the impact of noise in the baseline was greater for the Shewhart Control Chart and Tabular Cumulative Sums than for the V-Mask and trend-based methods.

The performance of the different statistical monitoring methods varied when monitoring increases and decreases in disease sero-prevalence. Combining two of more methods might improve the potential scope of surveillance systems, allowing them to fulfill different objectives due to their complementary advantages.

Introduction

Surveillance and monitoring systems are critical for the timely and effective detection of changes in disease status. Over the last decade, several studies have applied different statistical monitoring methods for detecting outbreaks of (re-)emerging diseases in the context of syndromic surveillance in both human and veterinary medicine

[1–3]. Different types of models (such as linear models, logistic regression and time-series models) have been implemented in the context of syndromic surveillance in order to evaluate the performance and implementation of these methods [4].

However, it may not be possible to make generalizations about the performance of these methods when used for monitoring endemic diseases and control programs. In this case, the availability of control measures (such as vaccination or health-management programs) results in lower incidence rates for endemic diseases than for (re)-emerging diseases. The dynamics of disease spread and immunity within a population from previous exposure also contribute to a lower incidence, resulting in slow and gradual changes in incidence and prevalence for endemic diseases [5]. It is important to follow-up on implemented control strategies in order to reduce and/or eliminate a specific disease [6]. Unexpected changes (such as an increase in disease prevalence or a failure to achieve a target value of disease prevalence within a certain period of time) indicate that the implemented strategies should be revised. When a control program fails to achieve certain goals, it can have a devastating impact on herds with susceptible animals.

In previous work, we assessed the performance of univariate process control algorithms (UPCA) in monitoring changes in the burden of endemic diseases based on sentinel surveillance [7]. However, these methods were not tested in the context of voluntary disease control and monitoring programs. In such cases, the frequency of testing depends on the monetary value of the animal and not just on the impact of the disease [6]. Programs for monitoring endemic diseases include the Danish Porcine

Reproductive and Respiratory Syndrome Virus (PRRSV) monitoring program. Despite disease-control efforts, PRRSV has contributed to economic losses since its first diagnosis in 1992 [8]. Monitoring of PRRSV is primarily based on serological testing within the Specific Pathogen Free System (SPF System) [9]. The frequency of testing depends upon the health status of the herd within this system. As a consequence, the number of samples is not constant and it is necessary to use methods with a more dynamic structure, allowing the parameters to change over time, thus taking into account the variation in sample size. Previous studies have also discussed the influence of variation in the number of samples (i.e. the noise present in data) on the performance of different monitoring methods [7,10].

State-space models have a flexible structure, allowing parameters to be updated for each time step [11]. In addition, they can be decomposed, and changes in the components (such as trends and seasonal patterns) can be monitored for inference [12]. While state-space models have been used to monitor influenza in humans [13–15] as well as and for herd-management decisions [16–19], it has not yet been determined how useful these techniques are for monitoring endemic diseases.

The objectives of this study were to assess the performance of different statistical monitoring methods for endemic disease control programs, and to explore what impact of variation (noise) in the data had on the performance of these statistical monitoring methods. The simulation study was motivated by the Danish PRRSV monitoring program.

Two state-space models were chosen for this study based on their ability to monitor changes in different time-series components [11]. Five different statistical monitoring methods were evaluated for each model: three UPCA used in process-control monitoring [20], and two methods for monitoring changes based on the trend component of the time series.

Materials and Methods

All methods described in this section were implemented using R version 3.1.1 [22].

Data

Laboratory submission data stored in the National Veterinary Institute – Technical University of Denmark (DTU Vet) information management system and in the Laboratory for Swine Diseases –SEGES Pig Research Centre (VSP-SEGES) were used to determine the weekly PRRS sero-prevalence in Danish swine herds between January 2007 and December 2014 (418 weeks in total). The weekly PRRS sero-prevalence was calculated using the same method described in a previous study [7]. A total of 51,639 laboratory submissions from 5,095 Danish swine herds were included. The average between-herd PRRS sero-prevalence was 0.24 (minimum=0, maximum=0.38) and the median number of herds tested for PRRS was 122 (minimum=8, maximum=191) per week.

Simulation study

A baseline scenario for sero-prevalence was defined based on the method described by Lopes Antunes et al. [7], where the number of positive herds per week was derived from a binomial distribution with probability (*p*) and sample size (*n*) equal to the number of Danish herds tested for PRRS in a given week. The data is publicly available at the following link: https://figshare.com/s/8760d1be0d738e57292b. The weekly sero-prevalence was calculated as the simulated number of sero-positive herds divided by the total number of herds tested per week.

There was a constant initial sero-prevalence of 0.24 for the first 104 weeks of all simulated scenarios, corresponding to the average PRRS sero-prevalence observed in Danish herds in the diagnostic laboratory data from 2007 to 2014 (Fig 1). In Scenario A, this period was followed by an increase in the weekly sero-prevalence (Event 1), a constant level, and then a decrease (second event). Scenario B consisted of a decrease in the sero-prevalence (Event 1) followed by a constant level, then an increase during the subsequent weeks (Event 2). Each scenario was simulated with changes in the weekly sero-prevalence, including gradual increases to 0.33 and 0.38 (for Scenario A) and gradual decreases to 0.15 and 0.10 (for Scenario B) over 52 and 104 weeks. Different combinations and durations of events (increases/decreases in sero-prevalence) were tested for each scenario, resulting in a total of 16 simulated scenarios (Table 1). Event 1 of each scenario was started at a random time between weeks 104 and 156, and Event 2 was started after a random interval of between 52 and 104 weeks following the end of Event 1.

Fig 1. Simulated scenarios representing endemic disease monitoring. The between-herd weekly sero-prevalence was simulated using a binomial distribution based on the Danish herds tested for PRRSV during the corresponding week. An initial sero-prevalence of 0.24 was maintained for at least 104 weeks. This was followed by either an increase to 0.38 or a decrease to 0.10 over 52 weeks in two different events. The different statistical monitoring methods were evaluated for each event.

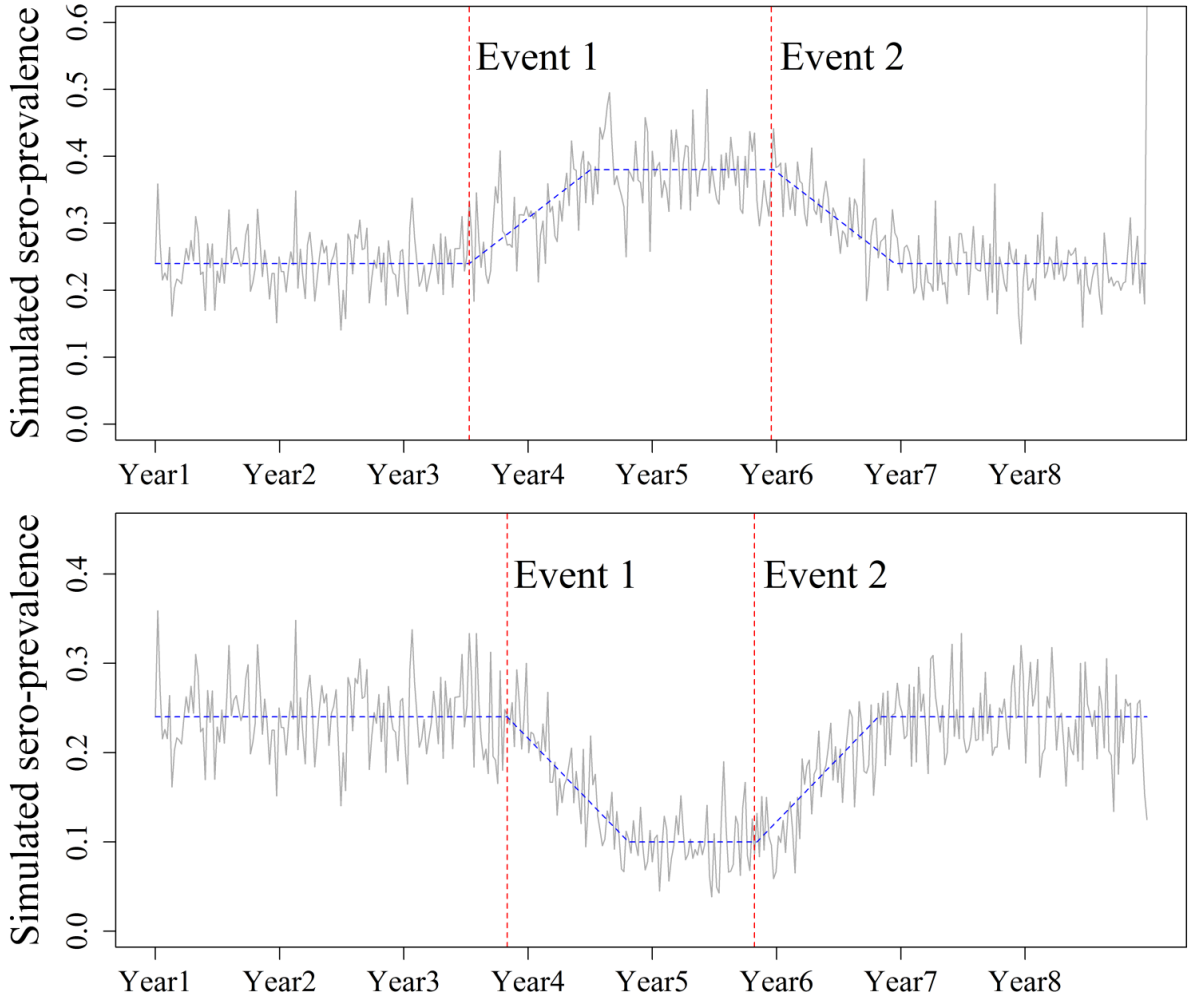


Table 1. Description of the 16 simulated scenarios representing changes in endemic diseases.

		Event 1		Event 2	
	Initial sero-prevalence	Sero-prevalence achieved at the end of the event	Duration of the event (weeks)	Sero-prevalence achieved at the end of the event	Duration of the event (weeks)
Scenario A	0.24	0.33	52	0.24	52
			52		104
			104		52
			104		104
		0.38	52		52
			52		104
			104		52
			104		104
Scenario B	0.24	0.15	52	0.24	52
			52		104
			104		52
			104		104
		0.10	52		52
			52		104
			104		52
			104		104

An initial constant sero-prevalence of 0.24 was simulated over 104 weeks. This was followed by an increase in sero-prevalence to 0.33 or 0.38 (Scenario A) or a decrease to 0.15 and 0.10 (Scenario B) over 52 and 104 weeks (Event 1). Event 1 was followed by a constant level of sero-prevalence, then by a second event (Event 2), corresponding to a decrease (Scenario A) or an increase (Scenario B) to the initial value of 0.24 over 52 and 104 weeks. Different combinations of event durations and changes in the sero-prevalence were tested, resulting in a total of 16 scenarios.

Modeling

A Dynamic Linear Model (DLM) and a Dynamic Generalized Linear Model (DGLM), both with a linear growth component as described previously [11], were used to model the simulated data.

The general objective of state-space models is to estimate an underlying parameter vector from observed data (θ) combined with any prior information available at time 0 (D_0), *i.e.* before an observation is made. The estimated parameter vector is updated each time every time there is a new observation (e.g. of the PRRS sero-prevalence). Specifically, the distribution of θ_t conditional on D_t ($\theta_t|D_t$) is estimated for each time step *t*. These models can be used to estimate a one-step forecast of the mean, allowing for a comparison between observed and forecasted values.

Briefly, the DLM is represented by a set of two equations, defined as the observation equation (Eq. 1) and the system equation (Eq. 2).

$$Y_t = \boldsymbol{F}'\boldsymbol{\theta}_t + v_t,\ v_t \sim N(0, V_t) \quad (1)$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}\boldsymbol{\theta}_{t-1} + w_t,,\ w_t \sim N(0, W_t) \quad (2)$$

where V_t and W_t are referred to as the observational variance and system variance, respectively. In our study, the observational variance was adjusted for the number of submissions in a given week (see Eq. 5 below). The transposed design matrix ($\boldsymbol{F}'$) had the following structure:

$$\boldsymbol{F}' = [1 \quad 0] \quad (3)$$

Eq. 2 describes the evolution of $\boldsymbol{\theta}$ from time *t*-1 to *t*. The system matrix ($\boldsymbol{G}$) for a local linear trend model is given as:

$$\boldsymbol{G} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad (4)$$

The linear trend component enabled us to include a time-varying slope (or local linear trend), allowing the system to adapt to a potential positive or negative trend for each *t*. Assuming that the PRRS sero-prevalence was not auto-correlated over time, the observational variance was defined as:

$$V_t = \frac{\mathrm{Y_{t-1}}\,(1-\mathrm{Y_{t-1}})}{\mathrm{n_t}} \quad (5)$$

where Y_t was the observed sero-prevalence for week *t*, and n_t was the number of herds tested for PRRS that week.

Unlike the DLM, the DGLM was based on a binomial distribution. The observation equation (Eq. 6) for the DGLM was defined as:

$$p_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t \quad (6)$$

For both DLM and DGLM, the variance-covariance matrix (W_t) describes the evolution of variance and covariance of each parameter for each time step. Rather than estimating W_t, the system variance was modeled using a discount factor (δ), as previously described by [21] and [17].

State-space model initialization and discount factors

Reference analysis was used to estimate the initial parameters $D_0 \sim [m_0, C_0]$ as described by West and Harrison [11].

The discount factors (δ) were defined using the method described by Kristensen [22], and were selected in order to optimize the performance of the model forecasts (i.e. minimizing the normalized forecast errors e_t^{norm}). The DLM and the DGLM models were run for 418 weeks with a constant simulated sero-prevalence of 0.24, using different δ-values ranging from 0.1 up to 1 in increments of 0.01. The δ-value that minimized the sum of the squared normalized forecast errors was chosen for the analysis. For both models, the forecast errors were normalized with respect to the forecast variance Q_t, such that $e_t^{norm} = e_t / \sqrt{Q_t}$.

Monitoring methods

Univariate process control algorithms (UPCA)

Three monitoring methods were used to generate alarms: the Shewhart Control Chart, Tabular Cumulative Sums, and V-Mask [20]. These methods are useful when only small changes are expected in the data [20].

The Shewhart Control Chart and Tabular Cumulative Sums were applied to the normalized forecast errors, whereas the V-Mask was applied to simple cumulative sums of the normalized forecast errors. The first 104 weeks of data were used as a "burn-in"

period for the models and the alarms were generated from the third year onwards (>108 weeks) when the simulated events started.

The fixed upper and lower control limits (*UCL* and *LCL*) required for the Shewhart Control Chart to generate alarms in a given week were calculated based on the following equations [20]:

$$UCL(f)_t = \mu_t + L\,\sigma_t \quad (7)$$

$$LCL(f)_t = \mu_t - L\,\sigma_t \quad (8)$$

where μ_t is the center line ($\mu_t = 0$), L is the selected number of standard deviations and σ_t is the standard deviation of the normalized forecast errors from t>104.

The Tabular Cumulative Sums for week t were calculated as described by Montgomery [20]. This method accumulates derivations from T_0 (target value) that are above the target with one statistic C^+, and below the target with another statistic C^-. The $C^+$and C^- for a given week (t) were calculated as:

$$C_t^+ = max\{0,\ e_t^{norm} - (T_0 + K) + C_{t-1}^+\} \quad (9)$$

$$C_t^- = mix\{0,\ (T_0 - K)\ e_t^{norm} + C_{t-1}^-\} \quad (10)$$

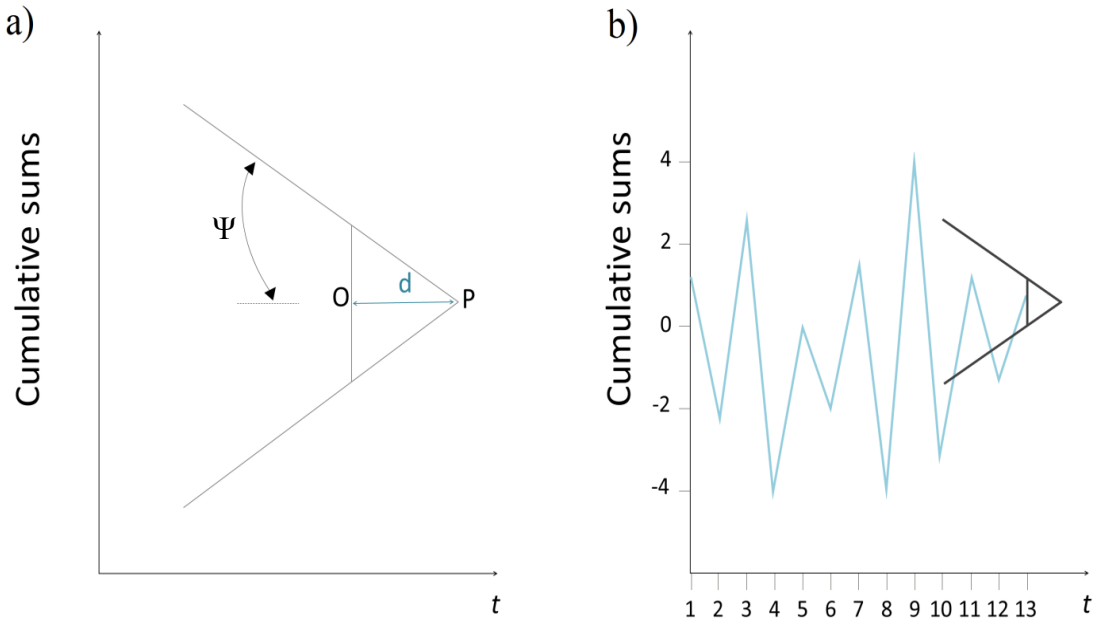
where $T_0 = 0$ and K is the reference value expressed as K=(1*σ_t)/2. Alarms were raised if C_t^+ or C_t^- exceeded a threshold H (expressed in terms of the standard deviation) in a given week t. The starting values of C_0^+ and C_0^- were defined as zero.

The V-Mask was applied to successive values of the cumulative sum of normalized forecast errors, which was calculated as follows [20]:

$$\text{cumulative sum}_{\,t} = \sum_{j=1}^{i} e_t^{norm} \quad (11)$$

The V-Mask is defined by the lead distance d and the angle Ψ, which were equivalent to the cumulative sum as described by Montgomery [20] (Fig 2). The point O of the V-Mask was directly applied to each value of the cumulative sum_t with the line OP parallel to the horizontal axis. The V-Mask was applied to each new point on the cumulative sum chart and the arms extended backwards towards the origin. If all the cumulative sums in previous time steps were within the two arms of the V-Mask, the process was considered to be 'in-control'; if any of the cumulative sums lay outside of the arms, the process was considered 'out-of-control' and an alarm was given. The value of the cumulative sum_t was reset to zero each time an alarm was given.

Fig 2. V-Mask description: (a) Parameters used to define the V-Mask; (b) Illustration of the V-Mask applied to the cumulative sum. The point O is positioned on the cumulative sum for each time *t*, and the line OP defines the lead distance d of the V-mask as expressed using horizontal plotting time steps.



Calibration

In order to calibrate the process control algorithms, the generalized DLM and DGLM were applied to 418 weeks of simulated data with a constant sero-prevalence of 0.24. The process control algorithms were calibrated for a false alarm rate of 1% when applied to the weekly e_t^{norm} (excluding the first 104 weeks, which represented the "burn-in" period of both models). The Shewhart Control Chart was calibrated with *L*

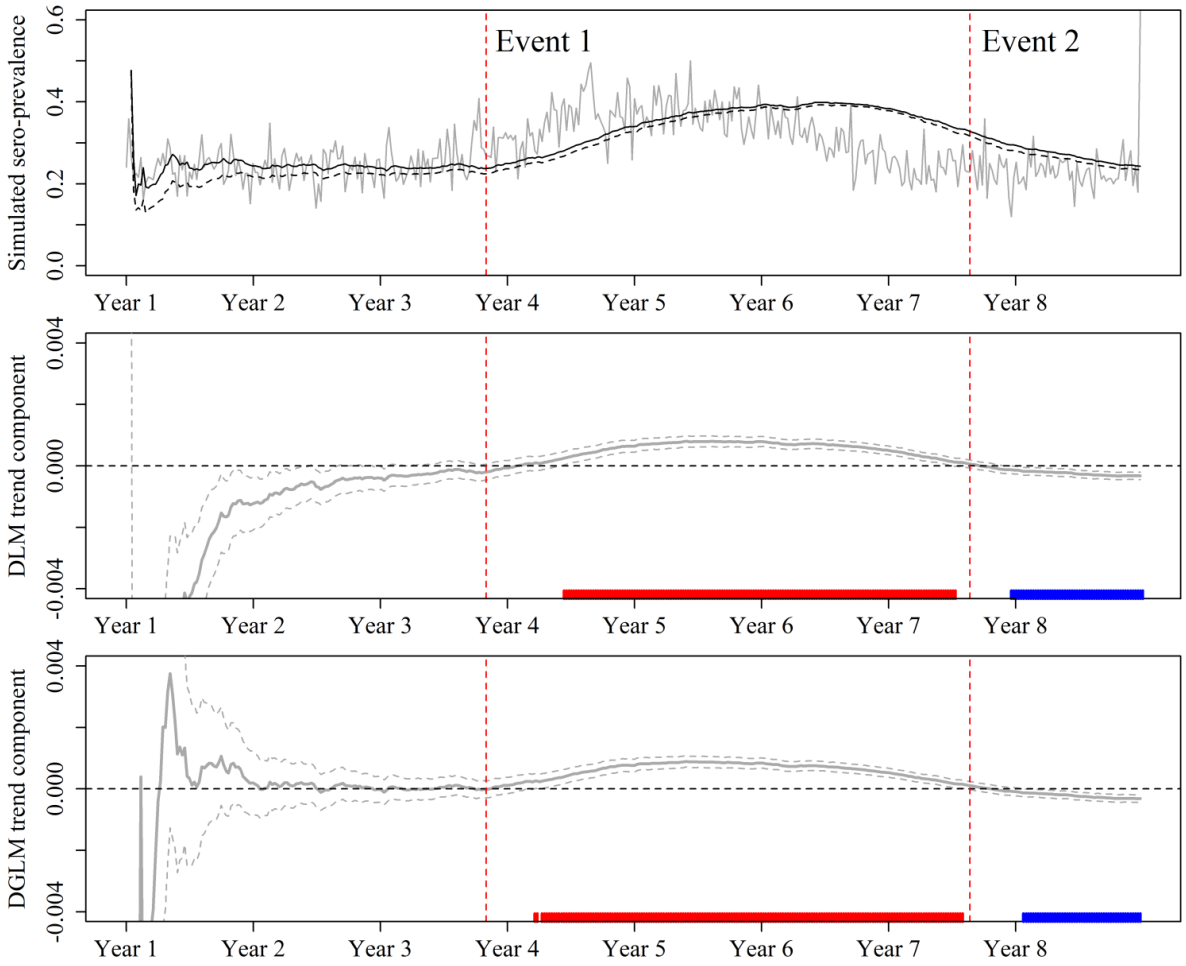
ranging from 1 to 4 standard deviations of the e_t^{norm}, and μ_t was defined as zero. For the Tabular Cumulative Sums, values of *H* ranging from 1 to 4 standard deviations of the e_t^{norm} were tested. This process was simulated 2,000 times for each parameter of the algorithm during calibration, and the median value of the false alarm rate was used as the summary statistic for evaluation.

Montgomery [20] suggested using $\Psi = \tan^{-1}(K)$ and $\mathrm{d} = H/K$ in order for the V-Mask to be comparable to the Tabular Cumulative Sums. For this reason, these values were adopted for the implementation of the V-Mask in this study.

Monitoring the time-series trend

For both the DLM and DGLM, the trend was extracted from the θ vector for each time step *t*. The variance of the trend parameter was calculated from the variance-covariance matrix for the posterior distribution, as previously described [11]. This variance was used to calculate 99% confidence intervals (CI) (Fig 3). Alarms were generated based on the trend when significant differences above and below zero were found according to the 99% CI. In addition, a second method was used to generate alarms when the absolute values of the trend component changed the sign from positive to negative and vice versa (Trend Sign).

Fig 3. The results show the simulated weekly sero-prevalence and the filtered mean obtained from the DLM (black dashed line) and DGLM (solid black line), and the corresponding DLM and DGLM trend component. The rugs indicate where the trend component was significantly above (red) or below (blue) zero.



Performance assessment

The performance was also assessed using the method proposed by Lopes Antunes et al [7]. The cumulative sensitivity (CumSe) was calculated as:

$$CumSe_i = \frac{\sum_{j=1}^{i} x_j}{N_{iter}} \quad (12)$$

where x_j is the number of iterations in which an alarm was given *j* weeks after an event started, and N_{iter} is the total number of iterations. An increase in the sero-prevalence was considered to have been detected if an alarm was generated for each week *i* after the event was started ($i \geq 0$).

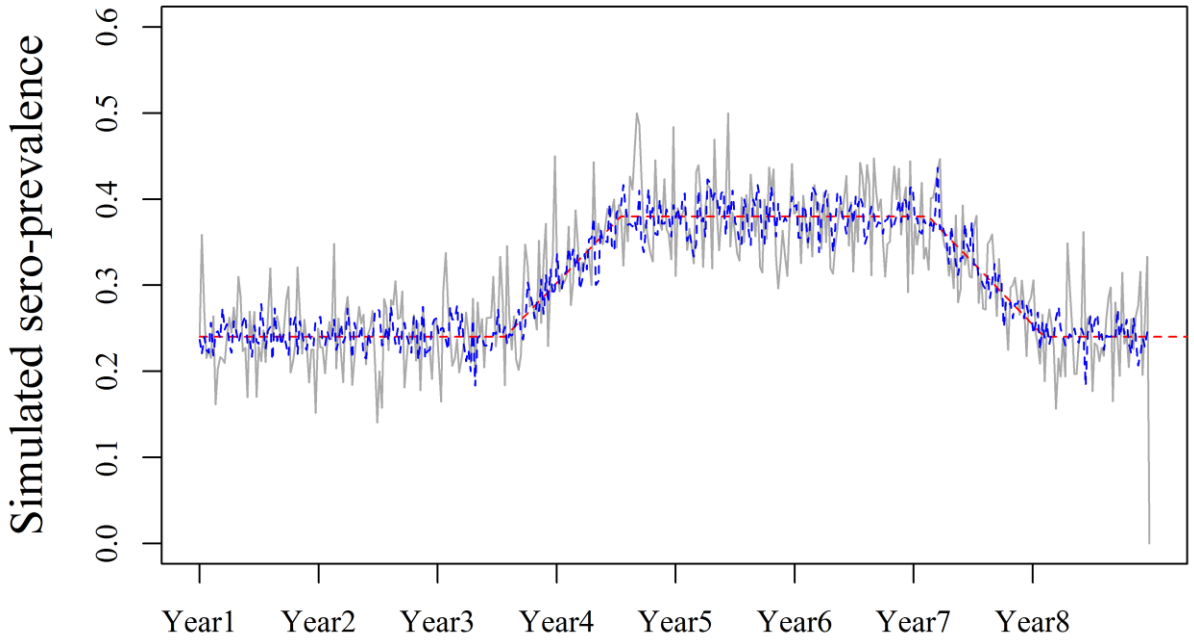
Convergence

A total of 10,000 iterations were simulated, with an initially constant sero-prevalence of 0.24 followed by a steady decrease to 0.15 over a period of 52 weeks. The decrease was randomly started between weeks 104 and 156. The number of iterations required to reach a stable detection time was determined visually using a plot of the variance in time to generate an alarm. This was done for each of the five statistical monitoring methods based on both types of models after the event was started with a stepwise increase of 100 iterations. Stable variance was observed after 2,000 iterations, therefore all simulated scenarios were run using this number of iterations.

Assessing the impact of noise in the data on the performance of detection methods

In order to assess the impact of noise in the data, the simulation study was repeated with *n* fixed at 600 herds tested per week. This value corresponds to a five-fold increase in the average number of Danish swine herds tested for PRRSV per week between 2007 and 2014, and it reduced the variation in the baseline (Fig 4).

Fig 4. Simulated sero-prevalence representing endemic disease monitoring. The weekly sero-prevalence was simulated using a binomial distribution based on the Danish herds tested for PRRSV during the corresponding week (grey line), and with five times the average number of Danish herds (n=600) tested for PRRSV (blue line). The red straight lines indicate the actual values of the simulated sero-prevalence.



Results

Parameters used for calibration

The selected values used to define a 1% false alarm rate for the UPCA based on the DLM model corresponded to L=2.6 for the Shewhart Control Chart, H=6 and K=6 for the Tabular Cumulative Sums, and a distance of 2 units for the V-Mask. For the

DGLM model, the values corresponded to L=2.5, H=16, K=5, and a distance of 3.2 units. These parameters were recalibrated to maintain a 1% false alarm rate when the number of herds tested per week was increased to 600, in order to simulate the baseline. The DLM model used parameters of L=2.3 for the Shewhart Control Chart, H=1.8 and K=1 for the Tabular Cumulative Sums and a distance of 1.8 units for the V-Mask for a constant number of herds tested. For the DGLM model, these parameters were defined as L=2.2, H=11, K=6 and a distance equal to 1.07 units.

A discount factor δ=0.99 was used to define the system variance for the DLM and the DGLM.

Statistical monitoring methods based on the DLM

The number of weeks needed to identify 50% of all iterations simulated (CumSe=50%) for each event is given in Table 2. A CumSe=50% was achieved most rapidly by the Trend Sign, followed by the V-Mask for Event 1 of Scenario A based on the DLM. For Event 2, the fastest CumSe=50% was achieved using the V-Mask and Shewhart Control Chart. Using the Trend Sign to monitor the changes, we noted an increase in the number of weeks needed to achieve CumSe=50% when comparing Event 1 and Event 2. As an example: for Event 1, 37 weeks were required to detect an increase in sero-prevalence from 0.24 to 0.38 over a period of 104 weeks based on 99% CI, and 2 weeks were required for the same increase and time period based on the Trend Sign. The same CumSe was achieved 74 and 59 weeks after the start Event 2 for the 99% CI and the Trend Sign, respectively. Furthermore, the Tabular Cumulative Sums detected

changes in Event 1 of Scenario A more quickly than Event 2, with the exception of scenarios where changes occurred over 104 weeks. The main differences found when comparing scenarios A and B (Table 2) were: the Tabular Cumulative Sums was able to achieve a CumSe=50% more quickly Event 2 of Scenario B than Scenario A; the Shewhart Control Chart achieved CumSe=50% faster during Event 1 of Scenario B, and this value could not be achieved for Event 2 (expressed as NA in Table 2); the V-Mask quickly detected changes in Event 2 for Scenario B. Moreover, the 99% CI and the Trend Sign had similar results in both scenarios.

Table 2. Number of weeks needed to achieve a CumSe=50% for the different statistical monitoring methods based on the DLM model.

		Event 1						Event 2					
	Sero-prevalence achieved	Duration (weeks)	Shewhart Control Chart[1]	Tabular Cumulative Sums[1]	V-Mask[2]	99% CI[3]	Trend sign[3]	Duration (weeks)	Shewhart Control Chart	Tabular Cumulative Sums	V-Mask	99% CI	Trend sign
Scenario A	0.33	52	119	27	18	34	2	52	31	52	18	121	93
	0.33	52	123	27	17	34	3	104	13	32	9	93	68
	0.33	104	146	44	20	50	1	52	30	52	18	113	89
	0.33	104	131	49	19	48	2	104	13	33	10	82	58
	0.38	52	121	19	13	27	2	104	26	43	16	109	93
	0.38	52	123	19	13	22	1	52	6	17	6	84	69
	0.38	104	158	33	18	39	1	104	25	43	16	103	89
	0.38	104	144	38	18	37	2	104	6	18	6	74	59
Scenario B	0.15	52	25	42	14	30	0	52	193	23	17	111	90
	0.15	52	25	42	13	30	0	104	NA	1	6	83	62
	0.15	104	35	70	18	46	0	52	NA	23	17	106	88
	0.15	104	39	75	18	43	0	104	NA	1	8	73	52
	0.10	52	19	29	10	23	0	52	NA	8	11	99	89
	0.10	52	19	28	10	23	0	104	NA	0	2	70	57
	0.10	104	28	51	14	35	0	52	NA	7	10	98	88
	0.10	104	32	58	16	33	0	104	NA	0	2	62	49

NA indicates that a CumSe=50% was not achieved by the monitoring method.

[1] Statistical monitoring methods applied to normalized forecast errors.

[2] Statistical monitoring methods applied to the simple cumulative sum of normalized forecast errors.

[3] Statistical monitoring methods applied to the trend component.

Table 3 shows the $CumSe_{52}$ (CumSe achieved 52 weeks after the event started) for the different statistical monitoring methods based on the DLM, indicating the likelihood of detecting the simulated events in the baseline for each method. For Scenario A, higher $CumSe_{52}$ was achieved by the trend-based methods (99% CI and Trend Sign) and the V-Mask for Event 1. For Event 2, the Shewhart Control Chart and the V-Mask had higher $CumSe_{52}$, and the trend-based methods were the worst performing ($CumSe_{52} \leq 0.3$). When comparing scenarios A and B, the major differences were seen for the Shewhart Control Chart, corresponding to a better performance (higher $CumSe_{52}$) for Event 1and a poorer performance for Event 2 of Scenario B. The other statistical monitoring methods presented similar results in both scenarios.

Table 3. CumSe achieved 52 weeks after the events were started for the different statistical monitoring methods based on the DLM model.

		Event 1						Event 2					
	Sero-prevalence achieved	Duration (weeks)	Shewhart Control Chart[1]	Tabular Cumulative Sums[1]	V-Mask[2]	99% CI[3]	Trend sign[3]	Duration (weeks)	Shewhart Control Chart	Tabular Cumulative Sums	V-Mask	99% CI	Trend sign
Scenario A	0.33	52	0.13	0.95	0.89	1.00	1.00	52	0.83	0.51	0.92	0.00	0.00
	0.33	52	0.11	0.94	0.87	1.00	1.00	104	0.98	0.70	0.96	0.00	0.12
	0.33	104	0.22	0.61	0.80	0.60	1.00	52	0.81	0.51	0.91	0.00	0.00
	0.33	104	0.24	0.52	0.81	0.65	1.00	104	0.98	0.68	0.96	0.01	0.37
	0.38	52	0.08	1.00	0.96	1.00	1.00	52	0.91	0.66	0.96	0.00	0.00
	0.38	52	0.07	1.00	0.96	1.00	1.00	104	1.00	0.93	0.99	0.00	0.08
	0.38	104	0.15	0.81	0.84	0.98	1.00	52	0.92	0.68	0.96	0.00	0.00
	0.38	104	0.18	0.70	0.82	0.97	1.00	104	1.00	0.91	1.00	0.04	0.33
Scenario B	0.15	52	0.94	0.66	0.95	1.00	1.00	52	0.14	0.93	0.90	0.00	0.00
	0.15	52	0.94	0.66	0.97	1.00	1.00	104	0.05	1.00	0.94	0.00	0.26
	0.15	104	0.71	0.37	0.86	0.74	1.00	52	0.16	0.93	0.91	0.00	0.00
	0.15	104	0.63	0.36	0.84	0.77	1.00	104	0.07	1.00	0.94	0.04	0.50
	0.10	52	1.00	0.93	0.99	1.00	1.00	52	0.08	1.00	0.96	0.01	0.01
	0.10	52	1.00	0.93	0.99	1.00	1.00	104	0.04	1.00	0.99	0.06	0.37
	0.10	104	0.84	0.52	0.94	1.00	1.00	52	0.17	1.00	0.97	0.00	0.00
	0.10	104	0.77	0.46	0.89	0.99	1.00	104	0.05	1.00	0.98	0.27	0.58

[1]Statistical monitoring methods applied to normalized forecast errors.
[2] Statistical monitoring methods applied to the simple cumulative sum of normalized forecast errors.
[3] Statistical monitoring methods applied to the trend component.

Comparing the results from both models

Results revealed that the statistical monitoring methods required more time to achieve CumSe=50% when applied to DGLM (Table 4) compared to DLM (Table 2), with the exception of monitoring the Trend Sign in Event 1 (Scenario A) and the V-Mask in Event 1 (Scenario B). In these cases, CumSe=50% was achieved at least twice as quickly for the DLM.

The trend-based methods produced identical results based on the DGLM (Table 5) and the DLM (Table 3). In general, these methods achieved the highest $CumSe_{52}$ based on the DLM for all simulated scenarios.

Table 4. Number of weeks needed to achieve a CumSe=50% for the different statistical monitoring methods based on the DGLM model.

		Event 1						Event 2					
	Sero-prevalence	Duration (weeks)	Shewhart Control Chart[1]	Tabular Cumulative	V-Mask[2]	99% CI[3]	Trend sign[3]	Duration (weeks)	Shewhart Control	Tabular Cumulative	V-Mask	99% CI	Trend sign
Scenario A	0.33	52	123	38	5	33	0	52	31	95	73	124	95
	0.33	52	127	37	6	35	0	104	11	72	71	96	69
	0.33	104	159	70	6	49	0	52	24	67	71	119	91
	0.33	104	157	214	3	47	0	104	12	81	68	83	58
	0.38	52	118	26	5	25	0	52	23	64	43	113	96
	0.38	52	120	25	5	21	0	104	5	38	88	86	71
	0.38	104	157	48	5	38	0	52	19	53	82	108	93
	0.38	104	152	95	3	36	0	104	5	41	77	75	59
Scenario B	0.15	52	52	128	5	32	1	52	162	27	9	101	82
	0.15	52	52	131	5	33	0	104	129	2	8	78	56
	0.15	104	93	172	5	48	0	52	141	24	8	97	80
	0.15	104	290	171	3	46	1	104	127	2	7	68	47
	0.10	52	36	117	5	25	1	52	164	17	10	84	76
	0.10	52	36	118	5	25	0	104	135	0	9	61	48
	0.10	104	65	161	5	37	0	52	153	13	8	82	74
	0.10	104	NA	154	3	34	0	104	133	0	7	52	39

NA indicates that a CumSe=50% was not achieved by the monitoring method.

[1]Statistical monitoring methods applied to normalized forecast errors.

[2] Statistical monitoring methods applied to the simple cumulative sum of normalized forecast errors.

[3] Statistical monitoring methods applied to the trend component.

Table 5. CumSe achieved 52 weeks after the events were started for the different statistical monitoring methods based on the DGLM model.

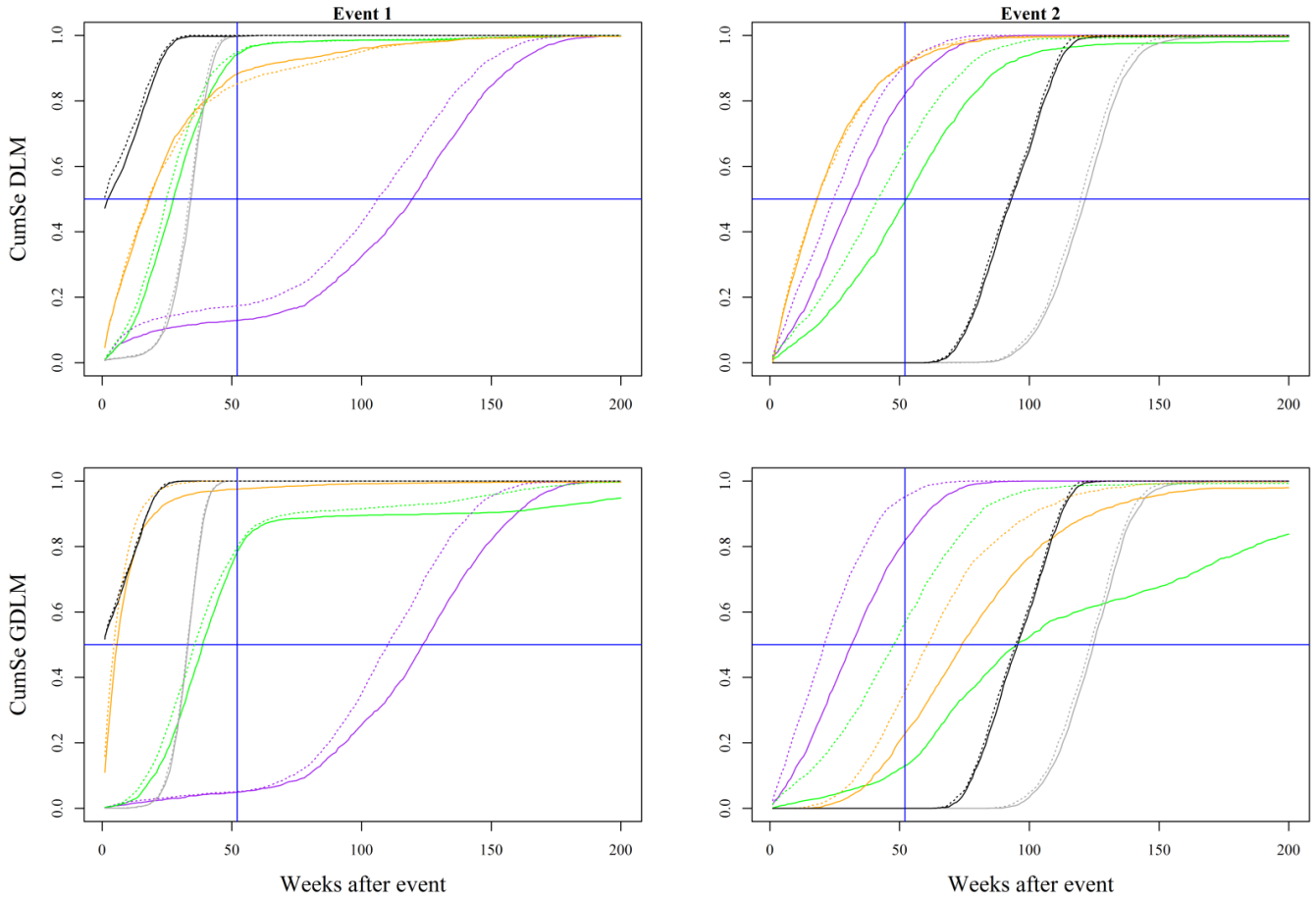
		Event 1						Event 2					
	Sero-prevalence achieved	Duration (weeks)	Shewhart Control Chart[1]	Tabular Cumulative Sums[1]	V-Mask[2]	99% CI[3]	Trend sign[3]	Duration (weeks)	Shewhart Control Chart	Tabular Cumulative Sums	V-Mask	99% CI	Trend sign
Scenario A	0.33	52	0.05	0.80	0.98	1.00	1.00	52	0.83	0.13	0.24	0.00	0.00
	0.33	52	0.04	0.78	0.98	1.00	1.00	104	0.99	0.39	0.22	0.00	0.09
	0.33	104	0.08	0.33	0.98	0.65	1.00	52	0.88	0.36	0.23	0.00	0.00
	0.33	104	0.07	0.12	1.00	0.70	1.00	104	0.99	0.37	0.26	0.00	0.36
	0.38	52	0.03	1.00	0.98	1.00	1.00	52	0.94	0.30	0.08	0.00	0.00
	0.38	52	0.03	0.99	0.98	1.00	1.00	104	1.00	0.63	0.09	0.00	0.05
	0.38	104	0.06	0.58	0.98	0.99	1.00	52	0.97	0.50	0.15	0.00	0.00
	0.38	104	0.05	0.22	1.00	1.00	1.00	104	1.00	0.60	0.15	0.02	0.32
Scenario B	0.15	52	0.51	0.03	0.98	1.00	1.00	52	0.03	0.90	0.98	0.00	0.00
	0.15	52	0.50	0.04	0.99	1.00	1.00	104	0.13	1.00	0.89	0.01	0.41
	0.15	104	0.26	0.03	0.99	0.69	1.00	52	0.20	0.94	0.96	0.00	0.00
	0.15	104	0.14	0.04	1.00	0.77	1.00	104	0.12	1.00	0.91	0.13	0.67
	0.10	52	0.82	0.04	0.98	1.00	1.00	52	0.01	0.99	0.98	0.00	0.01
	0.10	52	0.81	0.04	0.99	1.00	1.00	104	0.07	1.00	0.78	0.26	0.63
	0.10	104	0.38	0.03	0.99	1.00	1.00	52	0.11	0.99	0.96	0.00	0.02
	0.10	104	0.16	0.04	1.00	1.00	1.00	104	0.07	1.00	0.81	0.51	0.83

[1]Statistical monitoring methods applied to normalized forecast errors.
[2]Statistical monitoring methods applied to the simple cumulative sum of normalized forecast errors.
[3] Statistical monitoring methods applied to the trend component.

Impact of noise on the different detection methods

Reducing noise in the data (by increasing the sample size to 600 herds tested per week) resulted in higher CumSe for the statistical monitoring methods (Fig 5). The time required to achieve a CumSe=1 was reduced by a factor ≥2 for the Shewhart Control Chart and Tabular Cumulative Sums. Similar results were found for the remaining 15 simulated scenarios (including Scenario B). The time required to achieve CumSe=50% was reduced by 117 weeks for the Shewhart Control Chart for Event 1 of Scenario A, with an increase in sero-prevalence from 0.24 to 0.33 over 52 weeks based on the DLM. The Tabular Cumulative Sums achieved similar CumSe 8 weeks earlier based on the DLM than when based on the DGLM. The impact of baseline noise in the V-Mask and both trend-based methods had similar results, with only small differences (up to 2 weeks) in the time required to achieve CumSe=50%.

Fig 5. The impact of baseline variation on the cumulative sensitivity (CumSe) of the algorithms. The results are shown for Scenario A, corresponding to an increase in sero-prevalence from 0.24 to 0.33 over 52 weeks (Event 1), followed by a decrease from 0.33 to 0.24 over 52 weeks (Event 2). The CumSe of the Shewhart Control Chart (purple), Tabular Cumulative Sums (green), V-Mask (orange), 99% CI (grey) and Trend Sign (black) are shown based on the actual number of herds tested for PRRSV (straight lines) and on a fixed number (*n*=600) of herds tested per week (dashed lines). The horizontal and vertical blue lines represent a CumSe=50% and the CumSe achieved 52 weeks after the start of the event, respectively.



Discussion

We investigated the performance of different methods for detecting changes in endemic disease (sero-) prevalence. The study included: 1) univariate process control methods applied to residuals, and 2) monitoring changes in the trend component of the time series based on CI and absolute values. The Shewhart Control Chart detected increases in sero-prevalence better than decreases for both scenarios, whereas the opposite was observed for the Tabular Cumulative Sums. The trend-based methods were effective when detecting Event 1, but their performance was inferior when adapting to several consecutive events. The V-Mask seemed to be the method with the most consistent performance seemed to be. Additionally, the impact of noise in the baseline was more profound for the Shewhart Control Chart and Tabular Cumulative Sums, and lower for the V-Mask and the trend-based methods.

Study design

This study was conducted based on sero-prevalence data from the Danish PRRS monitoring program. The different simulated scenarios were chosen to represent potential changes in sero-prevalence in the context of disease control programs, and were based on Danish pig production, where almost 40% of herds must follow rules concerning biosecurity, health control and transportation [9].

The approach used to simulate sero-prevalence was based on a binomial distribution defined by n and p. Both parameters have an effect on the variance of the binomial distribution, as higher values of p (up to 0.5) result in greater variance in the data obtained in each trial for a constant n, and lower values of p reduce the variance [23]. Event 1 of Scenario A and Event 2 of Scenario B represented an increase in sero-prevalence (p), resulting in greater variance of the data, which might have affected the detection rates presented in this study. However, higher values of n for the same value of p also have an

impact on the variance of the simulated data, which facilitates the reduction of noise in the simulated time-series by defining *n* as five times the average number of herds tested.

A predefined false alarm rate of 1% was used for standardization, and to enable comparison between the different statistical monitoring methods. The value of 1% was chosen as a compromise between false alarms and maintaining confidence in the system.

Results of the performance evaluation

Event 1 was started after 104 weeks in order to guarantee that the "burn-in" period of the model was sufficient for representative inferences to be made. From a practical point of view, false alarms can be generated, and true alarms can be masked thus reducing the sensitivity of the system for monitoring changes during this period.

As anticipated, larger changes in sero-prevalence were indicated earlier. These results are consistent with the expected performance of control charts [20].

The simulations showed that the Shewhart Control Chart was faster than the Tabular Cumulative Sums for detecting decreases in sero-prevalence. Conversely, the Tabular Cumulative Sums was faster at detecting increases. According to Montgomery [20], the Tabular Cumulative Sums is the recommended method for detecting gradual changes. However, the same author also mentioned that the Shewhart Control Chart might detect decreases earlier than the Tabular Cumulative Sum, as verified in this study. In addition, the variance in the simulated time-series was higher (due to a higher *p*) during Event 2 for Scenario B, which might explain the superior performance of the Tabular Cumulative Sums. Furthermore, the results for the trend component showed that both models needed time to adapt to Event 2 of both scenarios. It is possible that the models are forced to adapt to three

consecutive stages of the sero-prevalence ("constant-event-constant") prior to Event 2. This occurred because the system variance (modeled using a discount factor) was optimized for a constant level, resulting in slower model-trend changes for Event 2. As a consequence, the normalized forecast errors were higher and the Tabular Cumulative Sums generated alarms earlier, and as a result CumSe=50% was achieved more quickly. The same argument can also be used to explain why the V-Mask attained a faster CumSe=50% in Event 2 of Scenario B.

The V-Mask showed the most consistent results among the univariate methods in relation to the number of weeks required to achieve a CumSe=50%. This can be explained by the greater flexibility of the V-Mask method compared to other univariate process control methods based on pre-defined control limits.

Regarding the trend-based methods, the Trend Sign was quicker at detecting changes than the 99% CI. However, it is possible that the instantaneous detection of Event 1 for both scenarios based on the Trend Sign might occur due to the variation (above and below zero) of the trend component. In this case, changes in the sign (from positive to negative and vice versa) might occur by chance.

Impact of noise in the baseline

Decreasing the noise in the time-series resulted in higher CumSe for the Shewhart Control Chart and Tabular Cumulative Sums, whereas no important changes were found for the V-Mask or the trend-based methods. This shows the impact of variation in the time series and the importance of choosing the correct monitoring method. When the Shewhart Control Chart and Tabular Cumulative Sums were used, alarms were generated according to the intensity of noise in the data, regardless of whether they were applied to forecast errors or directly to the data. The superior performance of the

Shewhart Control Chart may be due to the upper and lower control limits being defined based on data with less variation. Despite recalibrating to a 1% false alarm rate, the applied control limits were defined based on lower standard deviations, which contributed to the alarms being generated earlier. One possible explanation for the superior performance of the Tabular Cumulative Sums is that the noise in the simulated data was greater during the increase in sero-prevalence, thus increasing the chances of alarms being generated. There has also been previous reference to the impact of noise in the data on the Tabular Cumulative Sums [1,7].

Decomposing the time-series also offers a way to monitor the underlying trend usually masked by random noise in the data. Monitoring the trend component based on CI or target values provides a more stable pattern compared to monitoring the forecast errors.

Perspectives

Choosing the correct methods for the prediction and determination of anomalies is critical for their effective detection [24]. Over the last decade, research has focused on the detection of (re-)emerging disease outbreaks [1–3]. Nevertheless, it is also important to follow up on implemented strategies in order to reduce and/or eliminate specific endemic diseases [6], and control and eradication programs play an important role within this context [25].

In this study, we showed that there is no robust method for all scenarios. Similar conclusions were drawn in previous studies on syndromic surveillance for (re)-emerging diseases [1,2,26,27], where the authors concluded that no single method was suitable for use with all outbreak signals. A surveillance system should be able to detect a variety of outbreaks with different characteristics [28,29]. This is important when the outbreak signature is unknown. The same challenges are

extrapolated to the context of endemic diseases and eradication programs for monitoring changes in (sero-)prevalence.

The efficiency with which changes in prevalence were monitored varied among the different methods. Choosing one specific monitoring method is therefore challenging, and the objectives of the monitoring program and the performance of the statistical monitoring methods in different time patterns should be taken into account [30]. Furthermore, it is important to consider the objectives of the control program, the nature of the disease, political and economic factors, and the infrastructure of the country in which it will be implemented [31].

In this study, state-space models were used to monitor endemic disease and control programs using two distinctive monitoring approaches for the time-series components. The principles can also be applied to general modeling, and the monitoring and surveillance of (re-)emerging diseases in human and veterinary sciences. The need to monitor declining changes in the context of veterinary syndromic surveillance has previously been discussed [32]. This author referenced the importance of monitoring decreases in the number of submissions (such as a decrease in the compliance of farms with passive disease surveillance) and the need for detection and action in the context of active surveillance.

Conclusions

Surveillance and monitoring systems are critical for the timely and effective control of infectious diseases. The different statistical monitoring methods used in this study performed differently in monitoring changes in disease sero-prevalence. In this context, choosing a single method is challenging, and the objectives of the monitoring program as well as the performance of the

591 statistical monitoring methods in different time patterns should be taken into account. Furthermore,
592 noise in the simulated baseline had an impact on the Shewhart Control Chart and the Tabular
593 Cumulative Sums, whereas no substantial changes were found for the trend-based methods. Using the
594 V-Mask or monitoring the trend component provided a consistent approach to monitoring changes in
595 disease sero-prevalence.
596

597 Acknowledgments

598 The authors would like to thank the Pig Research Centre – SEGES for providing part of the data
599 used in this study, and the Danish Food and Agriculture Administration for funding the project.

600
601

602 Bibliography

603 1. Dórea FC, McEwen BJ, McNab WB, Sanchez J, Revie CW. Syndromic Surveillance Using
604 Veterinary Laboratory Data: Algorithm Combination and Customization of Alerts. Pappalardo F,
605 editor. PLoS One. Public Library of Science; 2013;8: e82183. doi:10.1371/journal.pone.0082183
606 2. Jackson ML, Baer A, Painter I, Duchin J. A simulation study comparing aberration detection
607 algorithms for syndromic surveillance. BMC Med Inform Decis Mak. 2007;7: 6.
608 doi:10.1186/1472-6947-7-6
609 3. Buckeridge DL, Switzer P, Owens D, Siegrist D, Pavlin J, Musen M. An evaluation model for
610 syndromic surveillance: assessing the performance of a temporal algorithm. MMWR Morb
611 Mortal Wkly Rep. 2005;54 Suppl: 109–115. doi:su5401a18 [pii]

4. Rodríguez-Prieto V, Vicente-Rubiano M, Sánchez-Matamoros A, Rubio-Guerri C, Melero M, Martínez-López B, et al. Systematic review of surveillance systems and methods for early detection of exotic, new and re-emerging diseases in animal populations. Epidemiol Infect. 2014; 1–25. doi:10.1017/S095026881400212X

5. Carslake D, Grant W, Green LE, Cave J, Greaves J, Keeling M, et al. Endemic cattle diseases: comparative epidemiology and governance. Philos Trans R Soc Lond B Biol Sci. 2011;366: 1975–86. doi:10.1098/rstb.2010.0396

6. Doherr MG, Audigé L. Monitoring and surveillance for rare health-related events: a review from the veterinary perspective. Philos Trans R Soc Lond B Biol Sci. 2001;356: 1097–1106. doi:10.1098/rstb.2001.0898

7. Lopes Antunes AC, Dórea F, Halasa T, Toft N. Monitoring endemic livestock diseases using laboratory diagnostic data: A simulation study to evaluate the performance of univariate process monitoring control algorithms. Prev Vet Med. Elsevier B.V.; 2016;127: 15–20. doi:10.1016/j.prevetmed.2016.03.005

8. Kvisgaard L, Hjulsager C, Rathkjen P, Breum S, Trebbien R, Larsen LE. PRRSV outbreak with high mortality in northen part of Denmark. EuroPRRS2011-"Understanding and combating PRRS in Europe." Novi Sad, Serbia; 2011. p. 60.

9. SPF-DANMARK [Internet]. 2015 [cited 12 May 2015]. Available: https://www.spf.dk/

10. Dórea FC, McEwen BJ, Mcnab WB, Revie CW, Sanchez J. Syndromic surveillance using veterinary laboratory data: data pre-processing and algorithm performance evaluation. J R Soc Interface. 2013;10: 20130114. doi:10.1098/rsif.2013.0114

11. West M, Harrison J. Bayesian Forecasting and Dynamic Models. 2nd Ed. New York, USA: Springer; 1997.

12. West M. Time Series Decomposition. Biometrika. 1997;84: 489–494.
13. Cao P-H, Wang X, Fang S-S, Cheng X-W, Chan K-P, Wang X-L, et al. Forecasting influenza epidemics from multi-stream surveillance data in a subtropical city of China. PLoS One. Public Library of Science; 2014;9: e92945. doi:10.1371/journal.pone.0092945
14. Smith a F, West M. Monitoring renal transplants: an application of the multiprocess Kalman filter. Biometrics. 1983;39: 867–878.
15. Cowling BJ, Wong IOL, Ho L-M, Riley S, Leung GM. Methods for monitoring influenza surveillance data. Int J Epidemiol. 2006;35: 1314–1321. doi:10.1093/ije/dyl162
16. Thysen I. Monitoring Bulk Tank Somatic Cell Counts by a Multi-Process Kalman Filter. Acta Agric Scand. 1993; 58–64.
17. Madsen TN, Kristensen AR. A model for monitoring the condition of young pigs by their drinking behaviour. Comput Electron Agric. 2005;48: 138–154. doi:10.1016/j.compag.2005.02.014
18. Ostersen T, Cornou C, Kristensen a. R. Detecting oestrus by monitoring sows' visits to a boar. Comput Electron Agric. Elsevier B.V.; 2010;74: 51–58. doi:10.1016/j.compag.2010.06.003
19. Jensen DB, Cornou C, Toft N, Kristensen AR. A multi-dimensional dynamic linear model for monitoring slaughter pig production. 7th European Conference on Precision Livestock Farming. Milan; 2015. pp. 503–512. doi:10.1017/CBO9781107415324.004
20. Montgomery D. Introduction to Statistical Quality Control. 6th Ed. Arizona State University: John Wiley & Sons, Inc.; 2009.
21. Cornou C, Vinther J, Kristensen AR. Automatic detection of oestrus and health disorders using data from electronic sow feeders. Livest Sci. Elsevier B.V.; 2008;118: 262–271. doi:10.1016/j.livsci.2008.02.004

22. Kristensen AR, Jørgensen E, Toft N. “Advanced” topics from statistics. Herd management science II Advanced topics. Copenhagen: University of Copenhagen, Faculty of life sciences, Department of large animal sciences.; 2010. pp. 331–348.
23. Reddy TA. Applied Data Analysis and Modeling for Energy Engineers and Scientists. Springer Science & Business Media; 2011.
24. Elbert Y, Burkom HS. Development and evaluation of a data-adaptive alerting algorithm for univariate temporal biosurveillance data. Stat Med. 2009;28: 3226–3248. doi:10.1002/sim
25. Salman MD. Surveillance and Monitoring Systems for Animal Health Programs and Disease Surveys. Animal Disease Surveillance and Survey Systems Methods and Aplications. 1st Editio. Iowa: Blackwell Publishing; 2003. pp. 3–13.
26. Yahav I, Shmueli G. Algorithm Combination for Improved Performance in Biosurveillance Systems. In: Zeng D, Gotham I, Komatsu K, Lynch C, Thurmond M, Madigan D, et al., editors. Intelligence and Security Informatics: Biosurveillance. Springer; 2007. pp. 91–102. doi:10.1007/3-540-68339-9_34
27. Dupuy C, Morignat E, Dorea F, Ducrot C, Calavas D, Gay E. Pilot simulation study using meat inspection data for syndromic surveillance: use of whole carcass condemnation of adult cattle to assess the performance of several algorithms for outbreak detection. Epidemiol Infect. Cambridge University Press; 2015;143: 2559–2569. doi:10.1017/S0950268814003495
28. Reis BY, Pagano M, Mandl KD. Using temporal context to improve biosurveillance. Proc Natl Acad Sci U S A. 2003;100: 1961–5. doi:10.1073/pnas.0335026100
29. Lombardo JS, Burkom H, Pavlin J. ESSENCE II and the framework for evaluating syndromic surveillance systems. MMWR Morb Mortal Wkly Rep. 2004;53 Suppl: 159–165. doi:Article
30. Wagner MM, Moore AW, Aryel RM. Handbook of Biosurveillance. Handbook of

Biosurveillance. Elsevier; 2006. pp. 217–234. doi:10.1016/B978-012369378-5/50000-4

31. Christensen J. Application of Surveillance and Monitoring Systems in Disease Control Programs. In: Salman M, editor. Animal Disease Surveillance and Survey Systems Methods and Aplications. 1 edition. Blackwell Publishing; 2003. pp. 15–34.

32. Dórea FC, Lindberg A, McEwen BJ, Revie CW, Sanchez J. Syndromic surveillance using laboratory test requests: A practical guide informed by experience with two systems. Prev Vet Med. 2014;116: 313–24. doi:10.1016/j.prevetmed.2014.04.001

Ana Carolina Lopes Antunes
PhD thesis 2016

DTU Vet
National Veterinary Institute
Technical University of Denmark

Bülowsvej 27
DK-1870 Frederiksberg C
www.vet.dtu.dk