Open web-based application to optimise sampling strategies for active surveillance activities at the herd level

A ALBA[1,*], RE MORRISON[1], AN CHEERAN[1], A ROVIRA[2], J ALVAREZ[1], AM PEREZ[1]
[1]Dept. Veterinary Population Medicine. College of Veterinary Medicine.Univ. of Minnesota, St.Paul, MN, USA;
[2]Veterinary Diagnostic Laboratory. Univ. of Minnesota, St. Paul, MN, USA

*aalbacas@umn.edu

Abstract
Design of efficient active surveillance sampling schemes is challenging because optimal surveillance strategies may differ depending on the epidemiological conditions of the farm, including infection status, structure, management, or resources for conducting samplings. We present 'OptisampleTM', an open web-based application designed to optimise farm sampling strategies to substantiate freedom of infection considering also costs of testing. In addition to herd size, expected prevalence, test sensitivity, and desired level of confidence, this tool takes into account the risk of disease introduction at different stages of the production cycle, the structure of the herd, and the procedures used to select the samples over time. We illustrate its functionality through its application for active surveillance of porcine reproductive and respiratory syndrome virus (PRRSv) in hypothetical swine herds under different epidemiological conditions. Diverse sampling schemes for each farm were simulated and the most cost-effective strategy was estimated for each farm. The model shows the importance of considering the epidemiological context and the process of sampling selection to demonstrate freedom from disease. The approach demonstrated for PRRSv may be easily extended to other animal disease surveillance systems using the web-based application available at https://final.shinyapps.io/optisampleTM/.

Keywords: *Surveillance, sampling, open web application; PRRSv*

Introduction
Confidence on freedom from infection at the farm level is often derived from a combination of absence of clinical signs and negative lab results from consecutive samplings conducted in the population (1,2).

Identification of cost-effective sampling strategies is crucial to implement effective and sustainable systems for animal disease surveillance at the farm level. However, there are infinite schemes to conduct samplings and the optimum strategy may differ between farms depending on their infectious status for a given disease, their structure and management, the epidemiological context, and the resources available. Practical methods to identify efficient sampling strategies taking into account the context of the farm and costs are required. Active surveillance routinely conducted in many swine farms to demonstrate freedom from porcine reproductive and respiratory syndrome virus (PRRSv) infection is used here as an example to illustrate this need.

Our aim is to show the use of 'Optisample^{TM}', a flexible and accessible modeling tool for stakeholders and veterinarians to identify appropriate sampling strategies of active surveillance according to the epidemiological context of each herd. We demonstrate its application for active surveillance of PRRSv in different farm types. This example may be easily extended to address similar problems for other species and health conditions.

Materials and methods
'Optisample^{TM}' is an expanded version of the models proposed by Cannon (2002) and Martin (2007 and 2008). Its innovations intend to assess the influence of the sampling selection and compare the efficacy of different sampling schemes to demonstrate freedom from infection at the farm level considering the costs of testing.

Table 1. Summary of the input and output parameters of 'Optisample™'

PARAMETER	NOTATION	DATA TYPE	RANGE
INPUTS			
Demographic and epidemiologic traits of the herd			
Herd size	**N**	**Integer**	**0 - ∞**
Date since when there is information of disease outbreaks occurred in the herd	H_d	Date (yyyy-mm-dd)	No limits
Number of outbreaks occurred until now	n_o	Integer	0–∞
Expected duration (in days) of pathogen persistence in a herd in the event of an outbreak	p_p	Integers (min, max)	1-365
Time elapsed (in years) between two outbreaks occurred in the herd	o_f	Integers (min, max)	1-15
Correlation between successive sampled groups for the pathogen prevalence	ICC_{bt}	Unif (min, max)	0-1
Sampling strategy			
Minimum prevalence to detect	P*	Fixed percentage	0-1
Sample size of consecutive samplings	n_t	Sequence of 12 integers	0–300
Frequency of consecutive testins	freq	Factor of 3 levels	d, wk, mo
Diagnostic test sensitivity	se_{test}	Pert(min, mode , max)	0-1
Price for unit lab test	$Price_{test}$	Numeric value	0–∞
OUTPUTS			
Pr. free of infection after sampling t	$PrFree_t$	min –md – max	0-1
Pr. free of infection over all period assuming homogeneous pathogen distribution and collecting random samples from all herd over time	AUC_s	min –md – max	0-1
Pr. free over all period taking into account heterogeneous pathogen distribution and collecting samples from different animal sub-units over time	AUC_d	min –md – max	0-1
Cost of testing	$Cost_{test}$	Numeric value	0-9999999

Table 2. Inputs and outputs for the proposed scenarios.

	INPUTS	
	Farm A	Farm B
Infection status	Sow Herd Negative (IV)	Commercial pig herd - Positive stable undergoing elimination (II-B)
Herd size (N)	3000	3000
Date since the occurrence is known (Hd)	(~Date 5 years ago)	Initial date of the outbreak
No. of outbreaks (no)	0	1
Pathogen persistence (pp)	Unif (147, 231)	Unif (147–231)
Frequency of outbreaks (of)	5-6	3-4
Initial probability of infection (PrIt=0)	~ Beta (α=7, β=55)	~0.90-0.99
Probability of infection between consecutive samplings (PrIbt)	Pert (0.07, 0.09, 0.13)	Pert (0.10, 0.15, 0.21)
Minimum prevalence to detect (PI)	.05	.05
Test sensitivity(setest)	Pert (.97, .98, .99)	Pert (.97, .98, .99)
Correlation between sampled groups (ICCbt)	Unif (.5,.7)	Unif (.5,.7)
Pricetest	5	10

SAMPLING STRATEGIES

Scheme	I	II	IIIa	I	II	IIIB
Total number of samples	360	600	300	360	600	475
	30 samples monthly	50 samples onthly	50 samples bimonthly	30 samples monthly	50 samples monthly	90 samples at t=1 and 35 monthly

OUTPUTS Area under the curve (AUC) expressed as min-md-max)

AUCs	.96 - .97 -.98	.98-.99-.99	.91-.93-.95	.78-.82-.84	.92-.93-.94	.93-.95-.96
AUCd	.76 - .78 -.8	.92-.93-.93	.61-.63-.66	.52-.58-.61	.85-.86-.87	.76-.79-.81
Costtest	1800	3000	1500	3600	6000	

Plots in Figures 1 and 2 show the Pr_{Free} computed by month of the cheapest and most effective scheme for farms A and B. The modeling process comprises different steps.

1. To estimate the probability that the farm is infected before conducting any sampling ($PrI_{t=0}$) the user sets: i) date since when there is information of the disease occurrence in the herd (H_d), ii) number of outbreaks that occurred between H_d and the current date (n_o), and iii) expected duration of pathogen persistence (in days) in the herd in the event of outbreak (p_p) expressed as a range. The $PrI_{t=0}$ is expressed as a Beta distribution with parameters α and β [9] automatically derived from the input values of H_d, n_o and p_p:

$$\mathbf{PrI_{t=0}} = \boldsymbol{Beta}(\propto = \boldsymbol{s+1}, \boldsymbol{\beta = n - s + 1})$$

where "*s*" represents the period of time in which the pathogen may persist in the herd obtained from multiplying n_o by p_p; and "n" corresponds to the period of time with available information of the disease occurrence computed as the difference between the current date and H_d.

2. At time t=1 a first sampling is conducted on a number of animals (n_1) using a given diagnostic test. The probability of detecting at least one infected animal if the herd is infected ($Se_{t=1}$) is estimated considering: a minimum proportion of infected animals within the herd that we would expect if the disease was present (P*), the size of the herd (N), the number of sampled animals at t=1 (n_1) and the sensitivity of the diagnostic test (se_{test}). The value of P* is included as a fixed value that ranges between 0 and 1 and is set by the user based on the market-requirements or accreditation purposes. The se_{test} may be determined based on the information provided by the veterinary diagnostic lab or available references expressed as a Pert distribution with possible values ranging between 0 and 1.The $Se_{t=1}$ is calculated using a hypergeometric approximation expressed as:

$$\boldsymbol{Se_{t=1}} = \mathbf{1} - \left(\mathbf{1} - \boldsymbol{se_{test}}\, \boldsymbol{x} \frac{\boldsymbol{n_1}}{\boldsymbol{N}}\right)^{\boldsymbol{NxP^*}}$$

3. If all the samples of t=1tested negative, the model estimates the probability of the farm being free ($PrFree_{t=1}$) using a Bayesian inference approach that considers the $PrI_{t=0}$ and the $Se_{t=1}$ as follows:

$$\mathbf{PrFree_{t=1}} = \frac{\mathbf{1} - \boldsymbol{PrI_{t=0}}}{\mathbf{1} - \boldsymbol{Se_{t=1}\, x\, PrI_{t=0}}}$$

4. Once the $PrFree_{t=1}$ is computed the model estimates the probability of being infected immediately after the first sampling ($PrI_{t=1}$):

$$\mathbf{PrI_{t=1}} = \mathbf{1} - \mathbf{PrFree_{t=1}}$$

5. However, it also exists the probability of pathogen incursion between consecutive samplings (PrI_{bt}).The value of PrI_{bt} mainly depends on trade movements, biosecurity measures, proximity to other infected farms and environmental viability and is highly variable and uncertain. Here, this value is assumed as constant over time and to facilitate its definition the user can use historical data of the frequency of outbreaks occurred in the herd. This information is incorporated as:

$$\mathbf{PrI_{bt}} = \boldsymbol{Pert}\left(\boldsymbol{min} = \frac{\boldsymbol{p_{p(min)}}}{\boldsymbol{o_{f(\min)}}}, \boldsymbol{mode} = \boldsymbol{mean}\left(\frac{\boldsymbol{p_{p(min)}}}{\boldsymbol{o_{f(\min)}}}, \frac{\boldsymbol{p_{p(max)}}}{\boldsymbol{o_{f(\max)}}}\right), \boldsymbol{max} = \frac{\boldsymbol{p_{p(max)}}}{\boldsymbol{o_{f(\max)}}}\right)$$

where the $p_{p(min)}$ and $p_{p(max)}$ corresponds to the minimum and maximum time of pathogen persistence and $o_{f(min)}$ and $o_{f(max)}$ corresponds to the minimum and maximum frequency of outbreaks occurrence.

6. Using the PrI_{bt} and the $PrI_{t=1}$ the model computes the overall probability that the herd is infected before the second sampling ($PrItot_{t=1}$) so that:

$$\boldsymbol{PrItot_{t=1}} = \boldsymbol{PrI_{bt}} + \mathbf{PrI_{t=1}} - \mathbf{PrI_{bt}}\ \mathbf{x}\ \mathbf{PrI_{t=1}}$$

7. For each sampling for t=1 to 12 the model develops an analogous process to the previous calculations (steps 2-6). Recursively the model computes the probability of being infected for t=2,...,12 ($PrItot_t$), Se_t and the $PrFree_t$ given all samples are negative.

8. The previous calculations assume that consecutive samplings are conducted on a representative sample of the herd and the pathogen distribution is homogeneous. However, the sampling can be performed in different animal groups or sub-units over time to facilitate the logistics of collection. Here, to assess the influence of selecting different sub-units over time, the model includes a parameter that represents the degree of correlation between successive sampled groups for the prevalence of infection. This parameter is named intraclass correlation coefficient (ICC_{bt}). 'Optisample™' provides two different outputs based on this parameter. The value of ICC_{bt} is defined as a continuous uniform distribution that can take values between 0 and 1 according to the structure and management of the herd. In these cases the PrIt will be computed as:

$$\boldsymbol{PrI_{t=1}} = \mathbf{1} - \boldsymbol{PFree_{t=1}X\ ICC_{bt}}$$

9. The previous steps estimate the $PrFree_t$ after each sampling t. To get the overall probability of being free of infection the area under the curve (AUC) is computed over all the periods. Here the AUC is an integrated metric of the confidence of disease freedom for all the periods. Its computation uses the sum of consecutives values of $PrFree_t$ obtained from all consecutive samplings based on the trapezoidal rule as follows:

$$\mathbf{AUC} = \Delta\boldsymbol{t}(\ \boldsymbol{〚PFree\_(t=1)/2 + PFree〛\_(t=2)} + \cdots + \boldsymbol{PFree\_(t=12)/2})$$

The AUC value ranges between 0 and 1 and indicates the probability that the farm is free from the infection throughout the assessed period, being 1 if PrFree=100%, and 0 if PrFree=0%. Depending on the distribution of the pathogen and the strategy of sampling, AUC is denoted as AUC_s (for homogeneous pathogen distribution and random sampling over time) or AUC_d (for heterogeneous pathogen distribution and sampled sub-units varying over time).

10. Finally the model computes the cost of testing (Cost). The model sums all the samples tested over time and multiplies this value by a given cost of each individual test ($Price_{test}$) provided by the user.

    To show its functionality, the probability of PRRSv freedom was assessed in two hypothetical swine farms with disparate contexts and different sampling schemes.

    Farm A was a multiplier herd in a zone of very low incidence, with PRRSv negative status, without any pigs recently introduced and good biosecurity. The samples were tested using a commercial PRRSv antibody ELISA kit (6,7). Farm B was a commercial herd in a zone of medium incidence classified as positive stable undergoing elimination. The samples were tested using a PRRSv RT-qPCR (19).

Results

The basic traits of each farm, strategy of sampling and probabilities of being free from PRRSv and costs of testing over one year using the previous sampling schemes in farms A and B are shown in Table 2.

Figure 1. Pr of being free of PRRSv in Farm A resulting 30 negative tests by month.

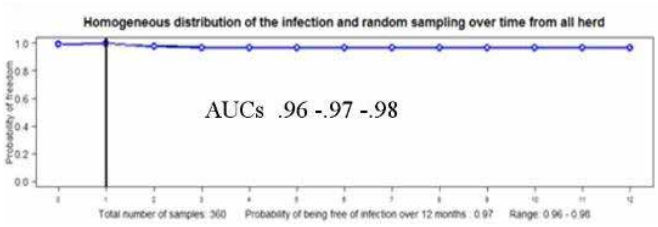
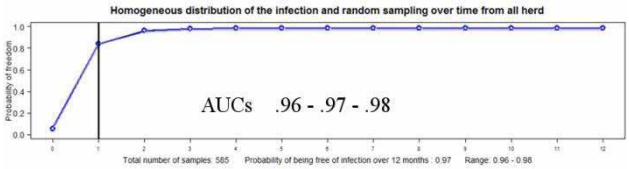


Figure 2. Pr of being free of PRRSv in Farm B resulting 90 negative tests initially and 35 negative tests by month.



Discussion

'Optisample™'shows that a single strategy to demonstrate disease freedom may not serve equally well in different epidemiological contexts. The suitable protocol differs depending on the initial infection status of the herd, the risks of introduction and the immediate aims to achieve.

The model illustrates the importance of checking the sanitary status of animals at the arrival to maximise the likelihood that introduced animals are not infected. The confidence in being free from PRRSv diminished over the entire period when the risk of being infected at arrival or between samples increased. The results show that to demonstrate freedom from infection when the risk of being initially infected is high, it is necessary to substantially increase the sample size during the first samplings.

The model also evidences that when the risk of disease introduction between consecutive samplings is low, the previous negative outputs provide cumulative information to substantiate that the herd is free from infection.

'Optisample™' also puts in evidence the importance of assessing the sample selection process. A potential extension to improve the accuracy of outputs in future versions would be the inclusion of updated values of prevalence, herd size, and risk of incursion between consecutive samplings, taking into account available continuous information of each context.

This work may not only enhance the design of active surveillance for PRRSv at farm level, but can also be easily extended to other surveillance contexts for a variety of species and animal diseases. This application contributes to assess and explain to veterinarians and stakeholders the importance of the main factors affecting the probability of disease freedom at the farm level.

Acknowledgements

We thank E Sargeant, T Martin and U Muellner. This study was funded by grants of the Univ. of Minnesota MnDrive program and the US Center of Excellence in Emerging and Zoonotic Animal Diseases (CEEZAD/KBA).

References

11. **Stark KD, Hasler B.** *Prev Vet Med.* 122(1), 229-234, 2015
12. **Holtkamp DJ *et al*.** *Prev Vet Med.* 96(3), 186-193, 2010
13. **Cannon RM.** *Prev Vet Med.* 52(3), 227-24, 2002
14. **Martin PAJ.** *Prev Vet Med.* 84(3), 291-309, 2008
15. **Martin PAJ *et al.*** *Prev Vet Med.* 79(2), 98-115, 2007
16. **Collins J *et al*.** SHP: AASP (USA). 1996
17. **Seo BJ *et al.*** *J Vet Med Sci.* 78(1), 133-138, 2016
18. **Rovira A *et al.*** *J Vet Diagn Invest.* 19(5), 502-509, 2007