The use of scenario-tree modelling using multiple complex data sources to demonstrate Danish freedom from classical swine fever.

Cameron, A.R.*[1], Martin, P.A.J.[2], Greiner, M.[3], and Barfod, K[4].
[1] AusVet Animal Health Services, 140 Falls Road Wentworth Falls, NSW, 2782, Australia, [2] Department of Agriculture – Western Australia, North Boyanup Road, Bunbury, Western Australia, 6230, Australia, [3] International EpiLab, Danish Veterinary Institute, Bülowsvej 27, DK-1790 Copenhagen V, Denmark, [4] Danske Slagterier, Axeltorv 3, 1609 Copenhagen V, Denmark.

Summary

A methodology is presented for demonstrating freedom from disease, based on scenario trees and stochastic modelling to calculate the sensitivity of a surveillance system using non-random data sources, and the combination of multiple data sources. This method was used to demonstrate Denmark's freedom from classical swine fever. Three different components of the surveillance system for CSF were examined and were found to have estimated monthly sensitivities of between 24% and 96%. The sensitivity of abattoir sero-surveillance was found to be 3.7 times better than a hypothetical equivalent random population survey. The systems had a combined annual sensitivity of close to 100%.

Introduction

Under the World Trade Organisation's Agreement on Sanitary and Phytosanitary Measures, barriers to trade based on animal health status must be substantiated by scientifically valid evidence. To facilitate export of animals or animal products, or to prevent their import from an infected country or zone, a country or zone should be able to demonstrate that particular diseases or agents of interest are not present. In the past, two approaches have been used to provide the evidence. The first is the use of a structured representative sample survey of the relevant population. The data are used to estimate the probability that the results of the survey would be achieved if disease were present at a specified level (the threshold prevalence). If the probability is less than an agreed threshold, the country is considered free.

The second approach is based on a qualitative assessment of multiple sources of evidence, usually by a panel of experts. These may include non-random data sources such as laboratory records, abattoir sampling, notifiable disease reporting system results, etc, as well as the results of the type of structured surveys referred to above. Unlike structured surveys alone, which are able to provide an objective, quantitative measure of probability, qualitative assessments tend to be more subjective, and generally provide only a dichotomous outcome – free from infection or not.

This paper describes research which aims to combine both of these approaches, by enabling multiple sources of surveillance data (both random and non-random) to be combined into a quantitative probability estimate to support claims of freedom from disease or infection.

Objectives

The objective of the research was to use an analytical approach based on the use of scenario tree models to demonstrate that Denmark is free from Classical Swine Fever (CSF). A secondary objective was to use the same methodology to evaluate the relative efficiency of different approaches to surveillance for CSF in Denmark.

Materials and Methods

The two major components to the methodology are 1) a stochastic model based on a scenario tree to estimate the sensitivity of non-random surveillance components, and 2) the combination of quantitative estimates of sensitivity of different components of a surveillance system into a single sensitivity estimate for the surveillance system.

The scenario tree is a branching probability tree which describes the steps required for a single infected unit (eg animal) to be detected as infected by the surveillance system. The steps included in the tree are designed to capture all the potential biases in the system. Each branch of the tree is associated with a probability, which is conditional on all the previous probabilities preceding it in the tree. A set of criteria for the inclusion and ordering of nodes in the tree was established.

Probabilities for each branch in the tree were either estimated from available data, or, where no suitable data was available, estimated using expert opinion. Probabilities were expressed in terms of distributions, in order to capture the variability and uncertainty in the estimates. Key probabilities used in the tree were the threshold prevalence values at different levels (eg herd- and animal-level threshold prevalences). These were selected based on a combination of international standards and the biology of the disease, taking into account differential risk of infection in different sub-populations. In essence, the threshold prevalence values determine the detectable level of disease below which a country is considered to be free.

Once constructed, the tree was used as the basis for a stochastic model. The model simulated the results of the surveillance component being applied to n units (animals), where n is the actual number of records from the surveillance component available for analysis. This simulation was iterated 1000 times to generate a distribution of the proportion of times that one or more animals would be detected as diseased, based on the assumption that the population is diseased at the levels described by the threshold prevalences. This is equivalent to the sensitivity of the surveillance component (the probability of detecting disease if the population is actually diseased at the level of the threshold prevalence).

The performance of the surveillance component was evaluated using the sensitivity ratio. The sensitivity of the surveillance component was first calculated as described above (Se_{Actual}). The process was then repeated after adjusting the scenario tree branch probabilities to remove differential selection probabilities and thereby simulate representative sampling.. This simulated a hypothetical representative surveillance component using the same sample size as the actual surveillance component. The result was an estimate of the sensitivity of an 'ideal', representative surveillance component ($Se_{Representative}$). The sensitivity ratio is the ratio of these two estimates (Se_{Actual} / $Se_{Representative}$) and measures the relative sensitivity of the actual component, compared to a hypothetical representative component. Well-targeted sampling of at-risk populations results in a sensitivity ratio of greater than 1, while

biased sampling of populations with a lower probability of disease results in a sensitivity ratio of less than 1.

Three surveillance components for CSF were analysed: abattoir sero-surveillance for CSF (based on a program of collection of blood samples from targeted groups in abattoirs), slaughterhouse meat inspections (pre and post slaughter inspections), and clinical surveillance (based on records of farm veterinary visits). Data were derived from a number of sources including the Danish Central Husbandry Register, slaughter records, laboratory records, the VetStat drug prescription database (used as a proxy for veterinary farm visits), and expert opinion from Danish veterinary personnel.

The combined sensitivity of multiple surveillance components was calculated using the following formula (Cannon 2002):

$$Se_{combined} = 1 - \prod_{j=1}^{J} (1 - Se_{system\ j})$$

Results

When reduced to a common period of one month's surveillance data, the sensitivity of the sero-surveillance component was estimated as 26.37% with a 5th to 95th percentile range of 23.44% to 27.87%. The sensitivity for the meat inspection component was 67.80% (39.46% to 90.34%) and for the clinical surveillance component was 93.80% (90.77% to 96.43%). The sensitivity ratio for the sero-surveillance component was 3.73, for the meat inspection component was 0.998 and for the clinical surveillance component was 0.991.

The combined sensitivity of the three surveillance components was calculated providing a monthly sensitivity of 98.53%. If surveillance data over the period of one year were considered, the sensitivity would increase to close to 100% ($1 - (1 \times 10^{-22})$).

Discussion

The estimated system sensitivity (or equivalently, confidence in the surveillance system) was strongly influenced by the threshold prevalence assumptions (sensitivity analysis data not shown). The high level of sensitivity of the clinical surveillance component is partly due to the large number of units under surveillance. However, an informal approach was used to obtaining expert opinion for some parameters in this model, and the resultant probabilities may be somewhat optimistic. The sensitivity ratio estimates indicate that the sero-surveillance component was very well targeted and much more efficient than simple random population sampling. The other two components were essentially equivalent to representative population sampling.

The combined sensitivity of the three components is likely to be over-estimated. This is because the lack of independence between the components has not been taken into account. Further research is needed to address this issue, as well as accounting for changes in the value or weighting of data collected at different times.

References

Cannon RM (2002). Demonstrating disease freedom – combining confidence levels. Prev Vet Med 52; 27-249.