Outbreak Prediction: Aggregating Evidence Through Multivariate Surveillance

Flavie Vial*[1], Wei Wei[2] and Leonhard Held[2]

[1]Vetsuisse Faculty, Veterinary Public Health Institute, Bern, Switzerland; [2]Institute of Social and Preventive Medecine, Zürich, Switzerland

Objective

The question of how to aggregate animal health information derived from multiple data streams that vary in their specificity, scale, and behaviour is not trivial. Our view is that outbreak detection in a multivariate context should be viewed as a probabilistic prediction problem.

Introduction

Production animal health syndromic surveillance (PAHSyS) data are varied: there may be standardized ratios, proportions, counts of adverse events, categorical data and even qualitative 'intelligence' that may need to be aggregated up a hierarchy. PAHSyS provides some unique challenges for event detection. Livestock populations are made up of many subpopulations which are constantly moving around between farms and markets to slaughter. Pathogen expression often varies across production types and rearing-intensity levels. The complexity of animal production systems necessitates monitoring many time series (Figure 1); and makes the investigation of statistical signals imperative and at the same time difficult and resource intensive. Having multivariate surveillance methods that can work across multiple data streams to increase both sensitivity and specificity are much needed.

Methods

Although the benefits of 'model-based' inference and prediction seem to be generally well accepted in numerous scientific disciplines, this does not seem to have found the same resonance in the context of surveillance data. A potentially promising alternative to traditional outbreak detection algorithms is to fit a model to the surveillance data, allowing for historic outbreaks, and to base outbreak detection on the posterior distribution of suitable model parameters or on the predictive distribution of X_{n+1}, the case counts at t_{n+1}. With this approach, X_n is used to fit the model and all available historic information on the disease is taken into account. While both outbreak detection and outbreak prediction have strengths and weaknesses for univariate time series, the benefits of a model-based predictive approach become overwhelming in multivariate surveillance. The challenges of multivariate surveillance, such as different data time lags, different frequencies of observations (e.g. weekly and daily), etc., are most easily met by suitable statistical modelling.

We propose to use Swiss PAHSyS data to develop model-based predictive methods based on the two-component model for counts of disease cases described in (1). Evaluation of the one-step-ahead predictions can be used to assess if the predictions are well calibrated. In contrast, the quality of traditional outbreak detection methods is usually assessed by simulation based on artificially inserted outbreaks.

Results

The concept will be presented since development has not started yet.

Conclusions

While univariate outbreak detection algorithms can be useful in practice, a potentially promising alternative for multivariate PAHSyS is based on model-based predictive methods. These methodological advances will provide a concrete step towards the development of operational multivariate PAHSyS systems which will contribute to disease risk reduction and increased food safety and public health.

Diversity of health data for production animals available to surveillance system developers at the Federal Food Safety and Veterinary Office in Switzerland. Datasets in purple are held by private companies but have been made available.

Keywords

Animal production system; Forecasts; Multivariate health data; Statistical challenges; Syndromic surveillance

References

1. Held L, Hofmann M, Höhle M, Schmid V. A two-component model for counts of infectious diseases. Biostatistics. 2006 Jul 1;7(3):422–37.

*Flavie Vial
E-mail: flavie.vial@vetsuisse.unibe.ch