

# Value of evidence from syndromic surveillance with delayed reporting

Rahel Struchen\*<sup>1,3</sup>, Flavie Vial<sup>2</sup> and Gunnar Andersson<sup>4</sup>

<sup>1</sup>Federal Food Safety and Veterinary Office, Bern, Switzerland; <sup>2</sup>Epi-Connect, Skogas, Sweden; <sup>3</sup>Veterinary Public Health Institute, Bern, Switzerland; <sup>4</sup>National Veterinary Institute, Uppsala, Sweden

## Objective

We apply an empirical Bayesian framework to perform change point analysis on multiple cattle mortality data streams, accounting for delayed reporting of syndromes.

## Introduction

Taking into account reporting delays in surveillance systems is not methodologically trivial. Consequently, most use the date of the reception of data, rather than the (often unknown) date of the health event itself. The main drawback of this approach is the resulting reduction in sensitivity and specificity<sup>1</sup>. Combining syndromic data from multiple data streams (most health events may leave a “signature” in multiple data sources) may be performed in a Bayesian framework where the result is presented in the form of a posterior probability for a disease<sup>2</sup>.

## Methods

We used a historical national database on Swiss cattle mortality to model daily baseline counts of two syndromic time series<sup>3</sup>. Reporting delay was defined as the number of days between reported occurrence and reporting date. The cumulative probability distribution of the estimated reporting delays was used to calculate for each day the proportion of cases that were reported either on the same day or with a delay of 1 to 14 days.

We evaluated outbreak detection performance under three scenarios: (A) delayed data reporting occurs but is not accounted for; (B) delayed data reporting occurs and is accounted for; and (C) absence of delayed data reporting (i.e. an ideal system). Outputs are presented as the value of evidence ( $V$ ) in favour of an ongoing outbreak accumulated over  $n$  points in time (30 days in this case). At each time  $t$ ,  $V$  is defined as the ratio between the posterior and prior odds for  $H_1$  versus  $H_0$ :

[insert equation 1 here]

Using sensitivity, time to detection and in-control run length, performance of the ( $V$ -based) system on large and small non-specific outbreaks was measured.

## Results

The evolution of  $V$  based on the information available on the 1st, 5th and 10th day after the onset of an outbreak can be visualised in Fig. 1. After 5 days,  $V$  shows evidence in favour of an outbreak for both syndromes combined, as well as for on-farm deaths alone, only in the “Delay aware” and “No delay” scenarios. The development of  $V$  for the perinatal deaths alone highlights the importance of considering multiple syndromic data streams for outbreak detection, as it speaks in favour of an outbreak at a later stage than on-farm deaths alone or both syndromes combined.

## Conclusions

Our empirical Bayes approach is an attractive alternative to multivariate CUSUM algorithms offering a logical approach to weighting variables and incorporating additional information such as delayed reporting, and a performance on a comparable level to an ideal (no delay) system. Outbreaks are detected earlier and with only

a marginal loss of specificity compared to a system where reporting delay is present but unaccounted for.

We also found that the accumulation of evidence from several days resulted in a significantly better outbreak detection timeliness, for a given specificity; or a similar timeliness, but higher specificity, compared to an algorithm<sup>4</sup> that only looks for days with unusual high number of counts.

$$V = \frac{\sum_{i=1}^{30} P(S_i | E) / P(S_i = 0 | E)}{\sum_{i=1}^{30} P(S_i) / P(S_i = 0)} \quad (\text{Eq. 1})$$

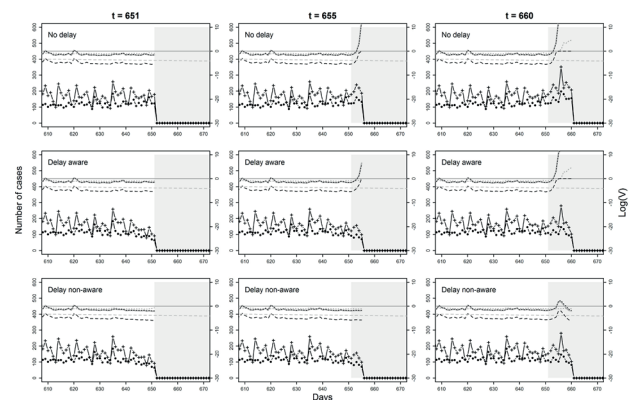


Fig. 1: Evolution of  $V$  over three time points ( $t$ ) for the three scenarios. Outbreak starts at  $t=651$ . Number of observed perinatal (circle) and on-farm deaths (cross),  $V$  for both (solid grey) and individual syndromes (dotted grey and black respectively), prior probability that an outbreak is ongoing (grey dashed) and posterior probability that an outbreak is ongoing given the evidence (black dashed). Horizontal grey solid line shows  $V=1$ .

## Keywords

empirical Bayes; Reporting delay; Multivariate surveillance

## References

1. Farrington C P & Andrews N J. *Monit. Heal. Popul.* 2004. Oxford University Press.
2. Andersson M G, Faverjon C, Vial F, Legrand L & Leblond A. Using bayes' rule to define the value of evidence from syndromic surveillance. *PLoS One.* 2014, 9, e111335.
3. Struchen R, Reist M, Zinsstag J & Vial F. Investigating the potential of reported cattle mortality data in Switzerland for syndromic surveillance. *Prev. Vet. Med.* 2015, 121, 1–7.
4. Salmon M, Schumacher D, Stark K & Höhle M. Bayesian outbreak detection in the presence of reporting delays. *Biom. J.* 2015, 57, 1051–67.

\*Rahel Struchen

E-mail: rahel.struchen@blv.admin.ch

