# SCIENTIFIC REP⚙RTS

**OPEN**

# Value of evidence from syndromic surveillance with cumulative evidence from multiple data streams with delayed reporting

R. Struchen[1,2], F. Vial[1,4] & M. G. Andersson [3]

Delayed reporting of health data may hamper the early detection of infectious diseases in surveillance systems. Furthermore, combining multiple data streams, e.g. aiming at improving a system's sensitivity, can be challenging. In this study, we used a Bayesian framework where the result is presented as the value of evidence, i.e. the likelihood ratio for the evidence under outbreak versus baseline conditions. Based on a historical data set of routinely collected cattle mortality events, we evaluated outbreak detection performance (sensitivity, time to detection, in-control run length) under the Bayesian approach among three scenarios: presence of delayed data reporting, but not accounting for it; presence of delayed data reporting accounted for; and absence of delayed data reporting (i.e. an ideal system). Performance on larger and smaller outbreaks was compared with a classical approach, considering syndromes separately or combined. We found that the Bayesian approach performed better than the classical approach, especially for the smaller outbreaks. Furthermore, the Bayesian approach performed similarly well in the scenario where delayed reporting was accounted for to the scenario where it was absent. We argue that the value of evidence framework may be suitable for surveillance systems with multiple syndromes and delayed reporting of data.

Timeliness is a key measure of any public health surveillance system. System users and decision makers depend on it to take appropriate action based on the urgency and the type of responses required by the situation. As such, it should be assessed regularly[1].

Most classical surveillance algorithms (including Salmon *et al.*[2]; Farrington *et al.*[3]; Noufaily *et al.*[4]) look for peaks with unusually high number of reported syndromic cases within a particular time period (e.g. a week), and generate an alarm if the counts exceed the threshold. Albeit simple, the approach has limitations that may hamper sensitivity and timeliness. Delays in the reporting of syndromic cases may result in counts remaining below a defined threshold until a majority of cases are reported, resulting in a delay or even failure of outbreak detection.

Delays in surveillance data may originate from an intrinsic biological process (e.g. incubation period) or from external processes (e.g. transport delay of the sample to the laboratory)[5]. While delays originating from the former cannot be reduced, analyses may be hindered by delays in case reporting decreasing the overall timeliness and usefulness of the early-warning surveillance systems[6]. Reporting delays depend, among other things, on statutory reporting regulations; on whether an electronic reporting system is in place; on the disease(s) under surveillance (whose identification process may be more or less complex); and on reporting units (e.g. different laboratories)[7]. Time lags between disease onset and notification can be estimated in terms of weeks (mostly as a result of lag between onset and diagnosis) even for notifiable disease reports (e.g. in Korea[8], in the UK[9]). Reporting delays may be monitored to detect trends, for example following an intervention aimed at improving reporting timeliness[10] or following a change in case definition[11]. They may also be modelled in order to better understand the factors leading to increasing time between disease onset and notification to the health systems[5, 12].

[1]Veterinary Public Health Institute, Vetsuisse Faculty, University of Bern, Schwarzenburgstrasse 155, 3003, Bern, Switzerland. [2]Federal Food Safety and Veterinary Office, Schwarzenburgstrasse 155, 3003, Bern, Switzerland. [3]Department of Chemistry, Environment and Feed Hygiene, National Veterinary Institute (SVA), SE-751 89, Uppsala, Sweden. [4]Present address: Epi-Connect, Djupdalsvägen 7, SE-14251, Skogås, Sweden. Correspondence and requests for materials should be addressed to M.G.A. (email: gunnar.andersson@sva.se)

The statistical interest in modelling delays in surveillance data is not new, but has so far mainly focused on the development of methods to obtain valid estimates of recent disease incidence[13]. Indeed, large reporting delays (e.g. as may be seen in cancer registries) may produce downwardly biased incidence trends, especially in the most recent years, when case ascertainment or reporting is subject to delays[14]. In the context of outbreak detection, delays occurring in a short time-window (on a scale of days or a few weeks) are more relevant than data with longer delays as they cannot be acted on promptly[9]. Accounting for reporting delays in outbreak detection algorithms (e.g. in syndromic surveillance SyS), is not trivial. This is, partly, the reason why most surveillance systems use the date of the reception of data, rather than the (often unknown) date of the health event itself. The main drawback of this common approach is the resulting reduction in sensitivity and specificity of the system[15]. In the relatively few systems for which all dates are known, a correction factor can be imputed based on mathematical models adjusting for the under-reporting bias owing to the time lag of the reporting process[16].

Another difficulty may arise when faced with slowly increasing outbreaks. The number of cases reported in each time unit (e.g. each week or each day) may be too small to trigger an alarm. If the baseline is recalculated iteratively as in Noufaily et al.[4], this may also result in outbreak-related cases being incorporated into the baseline. Guard bands leaving a short time lag between the current value under evaluation and the baseline have been used in order to reduce this risk of baseline contamination[17–19].

Finally, many diseases cause more than one syndrome and combining data streams[20] may result in increased sensitivity[21]. It is also desirable to combine the result from surveillance with other information. However, there is no straightforward approach when the algorithm is based on an alarm threshold[22]. Combining syndromic data from multiple data streams with other knowledge may be done within a Bayesian framework where the result is presented in the form of a posterior probability for a disease, or, when the hypotheses in the model are not exhaustive, as the odds for outbreak versus baseline. In Andersson et al.[23] we proposed a framework where the result from SyS is expressed as the value of evidence in favour on outbreak, i.e. the likelihood ratio for the evidence under outbreak versus baseline conditions. This approach was evaluated using three syndromic indicators (nervous syndromes in horses, mortality in both horses and wild birds) for early detection of West Nile virus in France, achieving better performance in a multivariate than univariate system[24]. The most important difference between the value of evidence approach and classical SyS is that the former explicitly incorporates the assumptions about the disease of interest and also refers to these assumptions when the results are presented. By including in the model assumptions the distribution of syndromes under outbreaks and baseline conditions, it is possible to apply change point analysis to estimate the probability that the system is in outbreak or baseline conditions, and the most likely point of transition.

In this study, we show how the empirical Bayes[25] likelihood ratio framework can be applied to perform change point analysis for multiple data streams and estimate the evidence accounting for delayed reporting of syndromes, using routinely collected cattle mortality data as an example.

## Methods

### The reporting system & reporting delay.
Since 2000, it is compulsory for Swiss cattle farmers to report all births and deaths of animals on their holding to the "Tierverkehrsdatenbank" or TVD. Deaths on farm need to be reported within 3 days but the reporting of stillbirths is not compulsory (Animal Health Ordinance (AHO), SR 916.401). It is likely that some farmers report stillbirths as the birth of a calf that was alive and died after a few days[26]. For this reason, we termed perinatal mortality the sum of reported stillbirths and on-farm deaths within the first seven days after birth. We extracted all deaths on farm and perinatal deaths reported to the TVD between 01/01/2009 and 31/12/2011.

The time interval between the date the event occurred and the date it was reported to the TVD is termed reporting delay. It includes the time needed for the farmer to observe the event but may also include data entry errors. Reporting delay in the TVD has been previously assessed with a median of one day for deaths (range: 0–968) and two days for perinatal deaths (range: 0–907)[26]. Over 80% of deaths on farm and over 70% of perinatal deaths were reported within seven days of occurrence (Fig. 1). The focus of this study was on reports with relatively short reporting delays ($\leq$14 days); since the minority of reports with longer delays (9 and 12% of on-farm death and perinatal death reports respectively) are less relevant for timely, "early" outbreak detection. Based on the cumulative probability distribution of the estimated reporting delays (Fig. 1), we used a binomial distribution to calculate for each day the number of cases occurring on day $t$ that were reported on the same day (delay $s=0$), 1 day later ($s=1$), 2 days later ($s=2$), etc. until all cases of day $t$ were reported 14 days later ($s=14$):

$$R_{ts} \sim Bin(n_{ts}, p_{ts})$$

where $n_{ts}$ is the number of deaths occurring on day $t$ if $s=0$ or the number of deaths occurring on day $t$ minus $R_{ts-1}$ if $s>0$, and $p_{ts}$ is defined from the cumulative probability distribution of the reporting delays as the proportion of deaths $c_{ts}$ occurring on day $t$ that were reported on day $s$ if $s=0$ or as $(c_{ts} - c_{ts-1})/(1 - c_{ts-1})$ if $s>0$.

### The value of evidence from syndromic surveillance.
We have previously proposed a tool for evaluating and presenting circumstantial 'evidence' of a disease outbreak from SyS, in which prior information and evidence (E) from the data are explicitly separated[23]. Applying Bayes' theorem, the *a-posteriori* odds ($O_{post}$) define our posterior belief about the disease state of the system given our prior belief and the syndromic evidence:

$$O_{post} = \frac{P(H_1|E)}{P(H_0|E)} = \frac{P(E|H_1)}{P(E|H_0)} * \frac{P(H_1)}{P(H_0)}$$
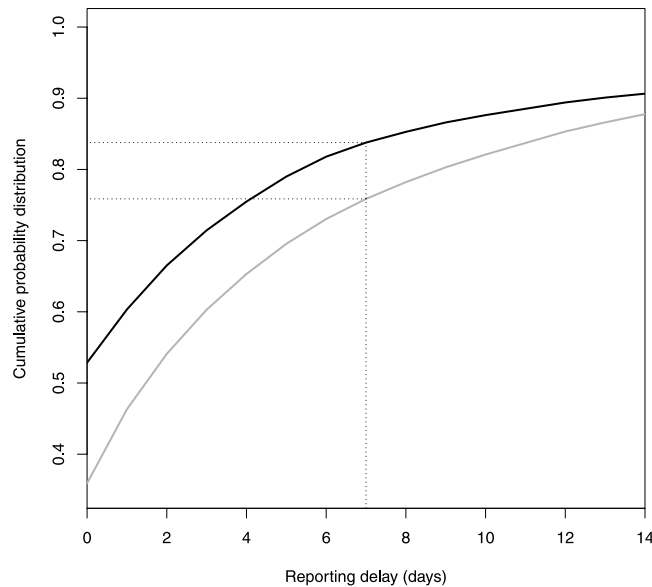
<div align="right">(1)</div>

**Figure 1.** Cumulative probability distribution of the reporting delays for on-farm deaths (black) and perinatal deaths (grey) in the TVD. Over 80% of on-farm deaths and over 70% of perinatal deaths were reported within seven days of occurrence (dotted lines). We focused on reports with relatively short reporting delays ($\leq$14 days) which are the most relevant for timely outbreak detection.

Where:

- $H_1$ is our hypothesis of interest (system is experiencing an outbreak of a specific disease of a group of diseases producing similar syndromes),
- $H_0$ is the "null hypothesis" (the system is operating under baseline, non-outbreak, conditions),
- E is the evidence represented by a set of vectors with reported cases of (a) syndrome(s) (cattle deaths on farm and perinatal deaths in our case).

In reality, the probability of observing a given number of syndrome(s) is not constant throughout an outbreak and the appearance of different syndromes may not be simultaneous. If, for example, syndrome A usually appears before syndrome B, the absence of syndrome A will speak against an outbreak when a peak in syndrome B is observed. However, when a peak of syndrome A is observed, the absence of B does not speak against an outbreak at an early stage.

Our previous model evaluated evidence from only one day or week at a time[23]. In this study, we extend the framework to accumulate evidence over $n$ points in time (30 days in this case).

In order to estimate the evidence in favour of an outbreak, we consider $H_1$ being composed of $n$ sub-hypotheses $H_{11}\dots H_{1n}$, representing an outbreak being in its first 1 to $n$ days. The system may be represented as an $n+1$ state Hidden Markov Model (Fig. 2), where state $S_0$ corresponds to hypothesis $H_0$, no outbreak, and states $S_1$ to $S_n$ corresponds to the sub-hypotheses $H_{11}\dots H_{1n}$. The probability of an outbreak starting, that is a transition from state $S_0$ to $S_1$, is non-constant (i.e. seasonal). We set it so that outbreaks were more likely to occur in summer. Therefore, the prior probabilities $P(S_i)$ of the system being in each state $S_i$ (i$=$0:$n$) also vary.

The probability distributions for the number of reported syndrome cases under a non-outbreak situation and for an outbreak in state $n$, in combination with the prior probability for each state, were used to derive $P(E|H_1)$ and $P(E|H_0)$. $H_0$ is the hypothesis that no outbreak is ongoing (S$=$0) and $H_1$ is the hypothesis that an outbreak is ongoing (S$=$1:$n$).

The posterior probability of each hypothesis and the cumulative probability of an ongoing outbreak were obtained by numerically calculating the marginal probability of evidence, P(E), given the vector of prior probability of introduction. Marginal probability of evidence P(E) at a given time is defined as:

$$P(E) = \sum_{i=0}^{30} P(S_i)P(E|S_i) \tag{2}$$

where $P(S_i)$ is the prior probability of state $S_i$ at time t. Posterior probability of state $S_i$ (post($S_i$)) can subsequently be calculated using the formula

$$post(S_i) = P\frac{(S)P(E|S_i)}{P(E)} \tag{3}$$

The value of evidence (V) in favour of an ongoing outbreak at each time $t$ is defined as the Bayes' factor, that is the ratio between the posterior and prior odds for $H_1$ versus $H_0$:
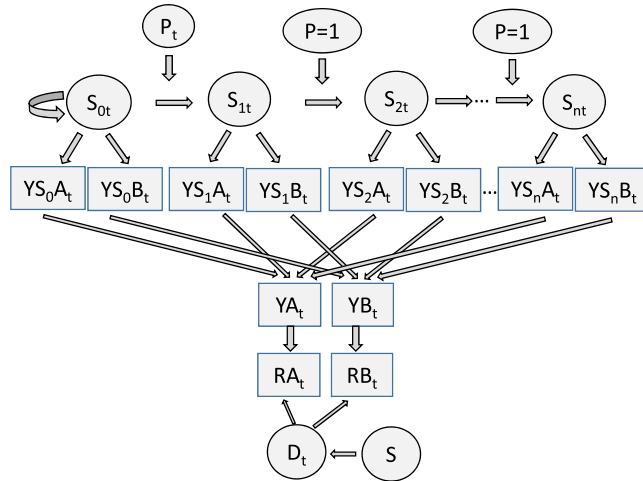
**Figure 2.** Representation of the system as an n + 1 state Hidden Markov Model. $P_t$ is the probability of transition from state $S_0$ to $S_1$ at time t. $S_{it}$ is the probability of the model being in state i at time t. $YS_iA_t/B_t$ is the number of observed cases of syndrome A and B emitted by state i at time t. $D_t$ is the probability that a syndrome observed at time t was reported at the time of observation (s). Finally, $RA_t/B_t$ is the number of observed syndromes that was reported from time t as seen on day s.

$$V = B = \frac{\sum_{i=1}^{30} P(S_i|E)/P(S_{i=0}|E)}{\sum_{i=1}^{30} P(S_i)/P(S_{i=0})} \tag{4}$$

**Modelling the data.** Seasonal variation in syndromes under baseline conditions was modelled by fitting negative binomial (NB) distributions to each of the two pre-processed[27] mortality time series using dynamic binomial regression based on maximum likelihood (ML) in R with function *glm.nb* (package {MASS[28]}). The fitted models were used for data simulation and generating input parameters in the Bayesian model. First, a set of 200 baseline time series was generated from each of the two fitted NB models. Second, a set of 200 non-specific disease outbreaks of two different magnitudes (further referred to as smaller and larger outbreaks, see Supplementary Fig. S1) resulting in daily excess mortality cases of both syndromes was simulated and added to the simulated baseline counts. Details of the NB models and the baseline and outbreak simulation procedure are provided in the Supplementary Information.

In our previous work[23], the distribution of baseline and outbreak-related syndrome cases was modelled directly using the NB distribution obtained from the regression (a similar method is also used in Salmon *et al.*[2]). The total outbreak distribution, i.e. the probability of observing *n* cases from the sum of baseline distribution and outbreak-related distribution, was calculated by numerical integration. However, the time to compute the distribution is proportional to the square of the maximum number of counts that result in very long computational time as the maximum number of counts increases. In this work, the number of counts at outbreak and baseline conditions is a magnitude higher calling for a faster, approximate solution.

To speed up calculations, the NB distributions were approximated with a (truncated) normal distribution with a standard deviation proportional to the mean.

P(base) ≈ TrunkNormal($\exp_{st}$, $\mathrm{var}_{st}$)

Where

Exp.base$_{si}$ is the expectation value from the NB model for syndrome s on day *i*

$\mathrm{var}_{si} = (\mathrm{exp.base}_{st} \cdot \mathrm{sd.rel}_s)^2$.

sd.rel is the relative standard deviation for each syndrome calculated as:

$$sd.rel_s = \sqrt{\sum_{t=1}^{t\,\max} \frac{(y_{st}/\exp_{st}^{-1})^2}{t\,\max}} \tag{5}$$

The expected distribution of outbreak-related cases of syndrome s on day *i* of an outbreak

P(out) ≈ TrunkNormal($\mathrm{mean}_{si}$, $\mathrm{var}_{si}$)

Where $\mathrm{mean}_{si}$ and $\mathrm{var}_{si}$ would ideally be based on expert judgement based on information on the disease and historical data. In the example, $\mathrm{mean}_{si}$ and $\mathrm{var}_{si}$ were calculated from sets of 1000 simulated outbreaks. For the ideal situation, the parameters were generated from a representative sample of simulated outbreaks, as they appear in the simulations. To investigate the effect of non-perfect expert knowledge, additional distributions were defined by calculating their mean and variance from biased samples of simulated outbreaks which differ in magnitude and progression rate (different parameters k, $\mu$ and $\sigma$, see Supplementary Information) from the outbreaks used in evaluation.
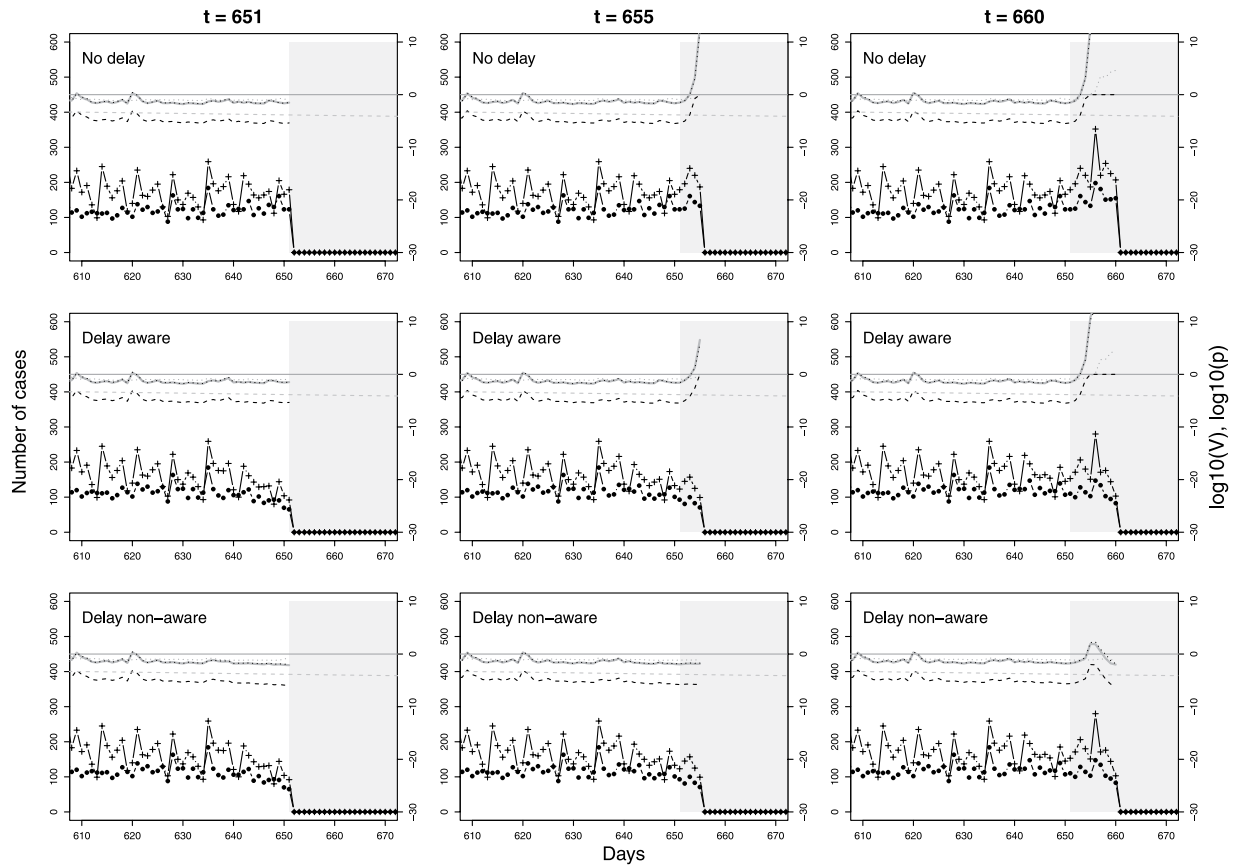
**Figure 3.** Evolution of the value of evidence at three days of observation (t) in comparison for the three scenarios no delay (top row), delay aware (middle row), and delay non-aware (bottom row). For illustration purposes, results for one specific outbreak, starting on day 651, were selected. Number of observed perinatal (black circles) and on-farm deaths (black pluses), value of evidence for both syndromes (solid grey line) and separately (grey and black dotted lines, respectively), prior probability that an outbreak is ongoing (grey dashed line) and posterior probability that an outbreak is ongoing given the evidence (black dashed line). The grey area represents the outbreak interval from days 651 to 678 while the horizontal grey solid line shows a value of evidence equal to 1, i.e. $\log10(V) = 0$.

With this approximation, we computed the mean and variance of the total outbreak distribution (out.tot) for each possible combination of syndrome (s), time (t) and stage of the disease (i) analytically.

$P(out.tot_{sti}) = TrunkNormal(exp.tot_{sti}, var.tot_{sti})$

where

$exp.tot_{sti} = exp.base_{st} + exp.out_{si}$

$var.tot_{sti} = var.base_{st} + var.out_{si}$

However, the normal approximation results in erroneous probabilities at very low counts, since it implicitly assumes the possibility of negative number of outbreak-related syndromic cases. This will lead to artificially high values of evidence in favour of an outbreak on days when an extremely small number of counts are reported. This primarily happens when a reporting delay is present but not accounted for. To handle this, we introduced a heuristic 2-steps filtering algorithm:

- For each day, the algorithm finds the number of reported cases ($nmin_{ts}$) that returns the minimum value of V (i.e. provides the strongest support against an ongoing outbreak);
- If the number of observed and reported cases ($nobs_{ts}$) is smaller than $nmin_{ts}$, it substitutes $obs_{ts}$ with $nmin_{ts}$.

This approximation effectively means that the cumulative probability for the left tail of the distribution of outbreak-related cases corresponding to negative counts is added to the probability of zero outbreak-related cases. The substitutions taking place were logged to estimate how frequent substitutions were and to confirm that it does not significantly impact the performance.

We compared the evolution of V for $H_0$ and $H_1$ in both syndromic time series before, during and after a set of simulated outbreak under three reporting scenarios:
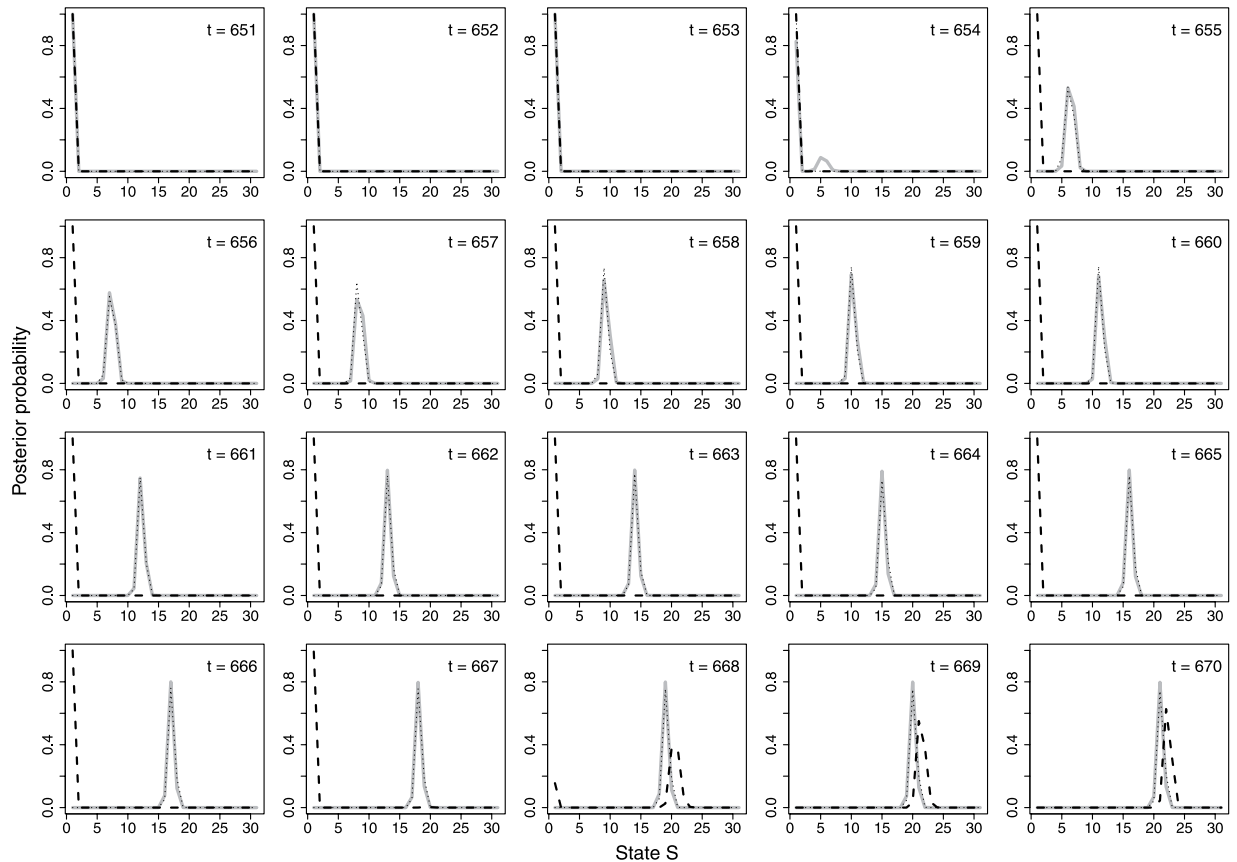
**Figure 4.** Posterior probability of being in state S (0–30) at a given day of observation (t) for the three scenarios no delay (grey solid line), delay aware (black dotted line) and delay non-aware (black dashed line). Twenty days of observation, starting on the first day of outbreak, were chosen to illustrate the transition of the system from baseline to outbreak condition.

- "Delay non-aware" scenario: the number of deaths occurring on day $t$ is equal to the number of deaths reported on day $t$ (i.e. regardless of when they truly occurred).
- "Delay aware" scenario: the number of deaths occurring on day $t$ is estimated based on the number of deaths reported on day $t$ and the probability distribution of the reporting delay.
- "No delay" scenario: all deaths are reported on the same day they occur and the number of deaths occurring on day $t$ is equal to the number of deaths reported on day $t$.

**Evaluation of the framework.** The performance of the system with regard to outbreak detection was compared among the three reporting scenarios using (1) the empirical Bayes approach introduced in this paper and (2) a classical process control approach adapted for reporting delay, as suggested by Salmon *et al.*[2], and modified to handle multivariate data. For both methods, sensitivity and timeliness, measured as average time to detection, were calculated as a function of in-control average run-length (ARL), in practice by changing the threshold for the statistics (Log(V) or z-score) in small steps. In approach 1, the value of evidence, Log(V) of both syndromes for a particular day was compared to all potential thresholds between 0 (evidence neither in favour of or against outbreak) and 15 (evidence extremely strong in favour of outbreak), with $\Delta \text{Log(V)} = 0.125$. An alarm was recorded if Log(V) exceeded these thresholds. In approach 2, the cumulative probability of the system producing the observed number of counts given the baseline condition was transformed to a z-score. To incorporate changes due to cases reported with delay, the z-scores of the last $d$ days were considered, where $d$ corresponds to the maximum reporting delay. For each day of observation, an alarm was recorded if the largest of these values exceeded a set of thresholds defined between 0 and 10, with $\Delta z = 0.05$.

Since the counts from the two syndromes in the example may be considered independent conditional on the season and outbreak status the combined z-scores for the two syndromes together was defined as:

$$z_{tot} = \sqrt{z_{A*}^2 + z_{B*}^2} \tag{6}$$

Where

$z_A$ is z-score for perinatal deaths
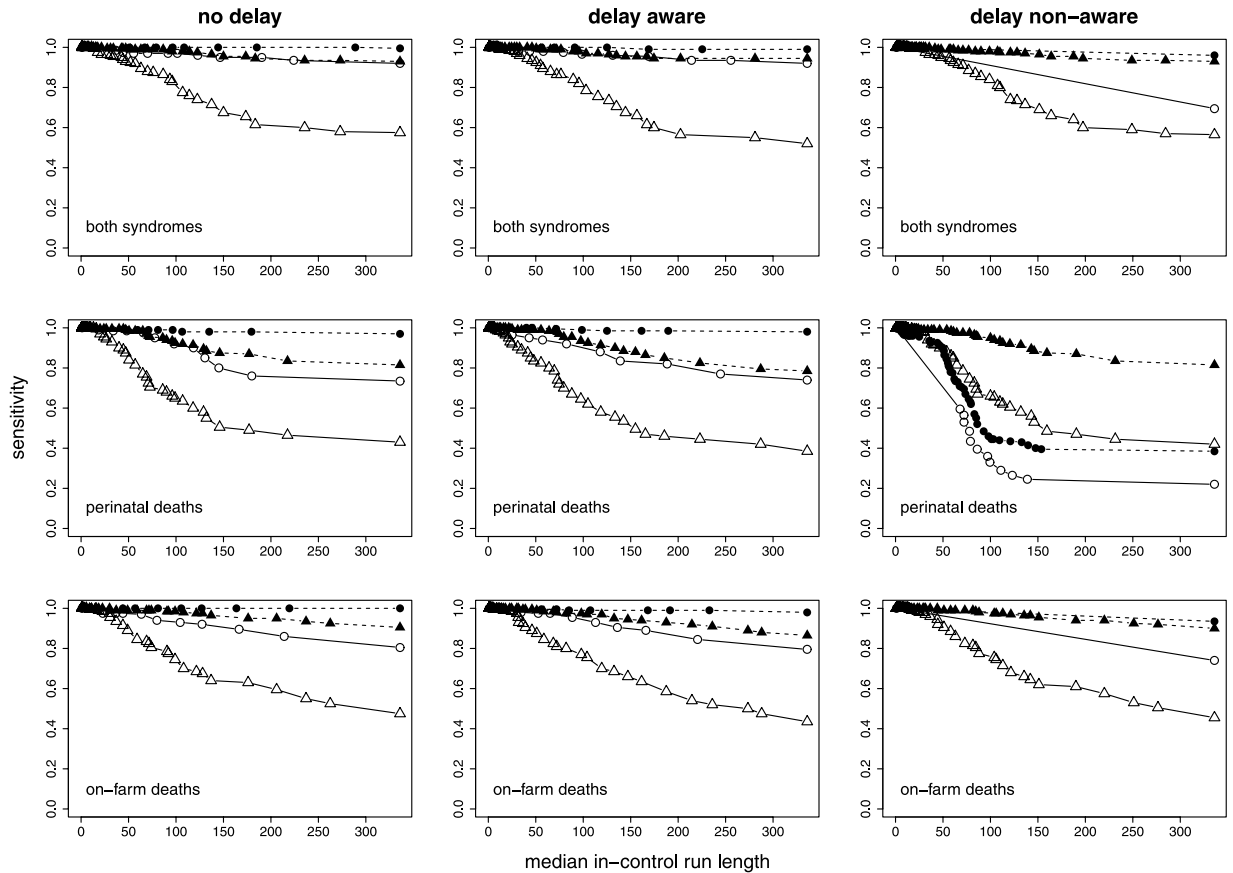$z_A$ is z-score for on-farm deaths

**Figure 5.** Sensitivity and median in-control run length for a range of alarm thresholds based on the value of evidence (circles) and the z-score (triangles), summarised for the smaller (empty symbols with solid lines) and larger (filled symbols with dashed lines) outbreak type. Values in the top right corner of the graph tend towards an ideal system with sensitivity = 1 (i.e. all outbreak signals are detected) and run length = 336 days (i.e. no false alarm occurs within 1 year).

$$z_{A*} = \max(z_A, 0)$$
$$z_{B*} = \max(z_B, 0)$$

The max() function ensures that a negative value of $z_A$ or $z_B$ will not result in a high $z_{tot}$ thereby triggering an alarm.

Sensitivity was defined as the proportion of detected outbreaks among the total number of simulated outbreaks. An outbreak was considered as detected if an alarm was generated on at least one outbreak day. Time to detection was estimated for the outbreak period (i.e. outbreak start plus $n$ days) and defined as the first day of outbreak detection (i.e. when an alarm was raised). In-control run length was estimated for the non-outbreak period (i.e. days $n + 1$ to 365) and defined as the first day with a false alarm. In case there was no alarm during the defined period, time to detection and in-control run length were set to 31 and 336 days, respectively.

## Results

The evolution of V based on the information available on the 1st, 5th and 10th day after the onset of a selected outbreak for each of the three reporting scenarios is presented in Fig. 3. For a larger outbreak starting on day 651. As a result of delayed reporting, daily counts of observed perinatal and on-farm deaths are lower for the most recent days in the two scenarios with reporting delay (middle and bottom row) compared to the scenario without reporting delay (top row). While V estimated for day 651 (left column) speaks against an outbreak for all scenarios, estimates for day 655 (middle column) show evidence in favour of an outbreak for both syndromes combined as well as for the on-farm deaths alone for the scenarios "Delay aware" and "No delay". 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 (Fig. 3, right column, day 660) than on-farm deaths alone or both syndromes combined. The change in corresponding posterior probabilities of each state $S_i$ for different days of observation is illustrated in Fig. 4 for all scenarios.

With a more restrictive alarm threshold (i.e. when only large values of V and z-scores raised an alarm), sensitivity and median time to detection generally impaired whereas in-control ARL improved (i.e. decreased). This trade-off between in-control run length and sensitivity or time to outbreak detection, respectively, is illustrated in Figs 5 and 6. Under the V-based approach, the performance of the system was comparable for the no delay and the delay aware scenarios. Under the z-based approach, however, the system's performance in the delay aware
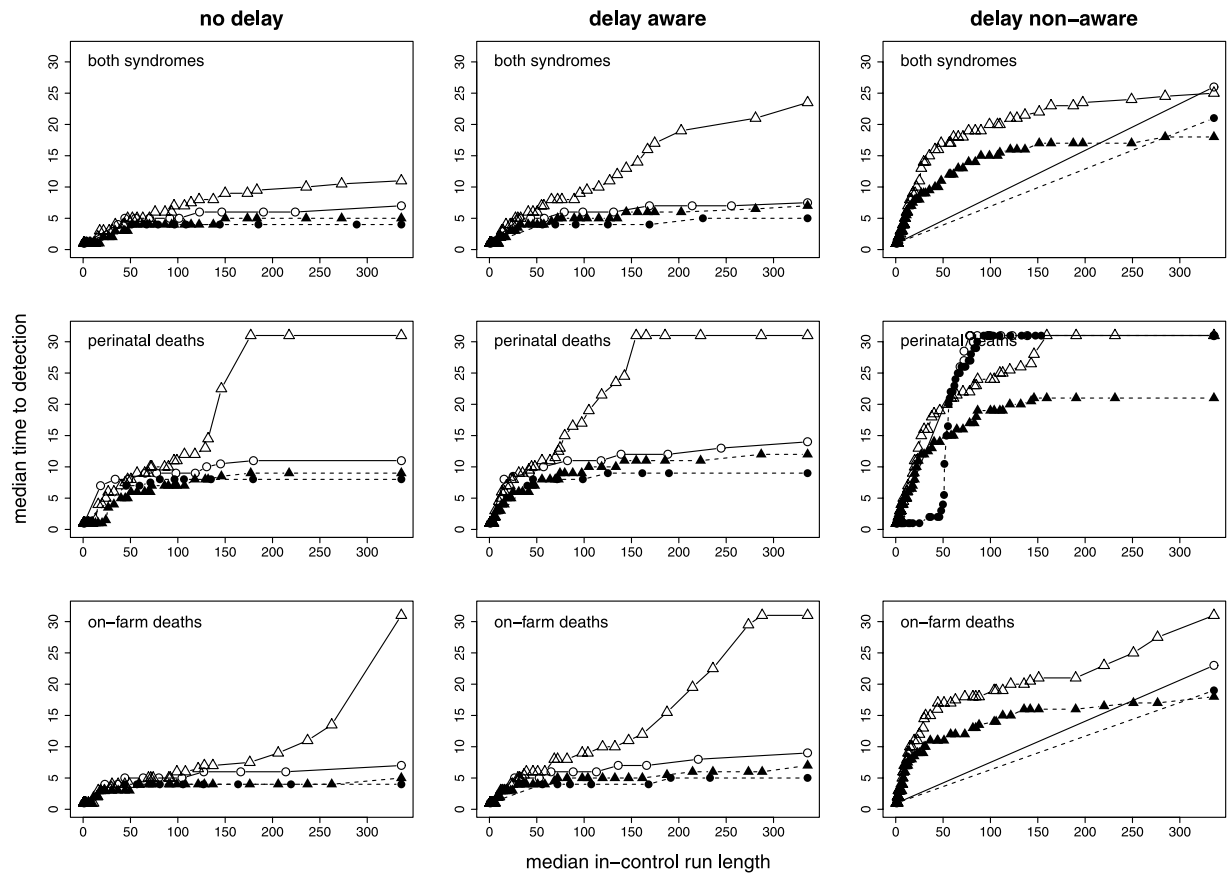
**Figure 6.** Median time to detection and in-control run length for a range of alarm thresholds based on the value of evidence (circles) and the z-score (triangles), summarised for the smaller (empty symbols with solid lines) and larger (filled symbols with dashed lines) outbreak type. Values in the bottom right corner of the graph tend towards an ideal system with a time to detection = 0 (i.e. an outbreak is detected on the day it starts) and run length = 336 days (i.e. no false alarm occurs within 1 year).

| | smaller outbreak | | | larger outbreak | | |
|---|---|---|---|---|---|---|
| | threshold | sensitivity | time to detection | threshold | sensitivity | time to detection |
| **V** | | | | | | |
| no delay | 1.250 | 0.920 | 7 | 1 | 0.995 | 4 |
| delay aware | 1.125 | 0.920 | 7.5 | 0.875 | 0.990 | 5 |
| **z** | | | | | | |
| no delay | 3.650 | 0.575 | 11 | 3.650 | 0.930 | 5 |
| delay aware | 3.900 | 0.520 | 23.5 | 3.900 | 0.945 | 7 |

**Table 1.** Performance of the system under the V- and z-based approach, given the lowest alarm threshold resulting in a maximum in-control run length of 336 days.

scenario considerably decreased with regard to time to detection of the smaller outbreak type. The alarm thresholds resulting in a maximum in-control run length of 336 days (i.e. no false alarm occurred within 1 year) were less restrictive for the delay aware than the no delay scenario under the V-based approach, in contrast to the z-based approach with more restrictive thresholds (Table 1).

Using non-perfect expert knowledge, i.e. biased distributions of expected outbreak-related mortality cases, mainly affected the ability to detect smaller outbreaks as shown in Fig. 7. Erroneously expecting large, steeply increasing outbreaks (black circles) decreased sensitivity of the smaller outbreak type in comparison to results based on perfect expert knowledge (white circles). In contrast, sensitivity of the larger outbreak type was comparable for all (ideal and biased) distributions. With regard to time to detection, outbreak detection performance was not substantially affected by the choice of expected outbreak distributions (Fig. 8).
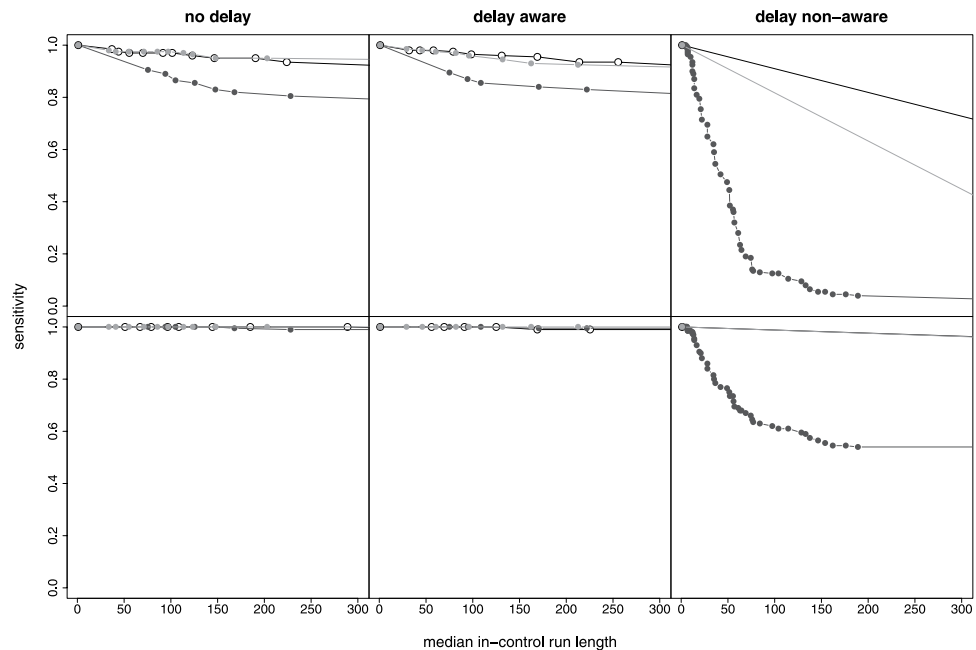
**Figure 7.** Influence of the choice of expert knowledge on the performance (sensitivity, in-control ARL) of detecting smaller (top row) and larger (bottom row) outbreak types under the V-based approach. Perfect expert knowledge was based on an ideal distribution of expected outbreak-related mortality cases (empty circles) for the given outbreak type. Non-perfect expert knowledge was simulated with expected outbreak distribution being biased either towards larger, steeply increasing outbreaks (black circles) or smaller, slowly increasing outbreaks (grey circles). Results are shown for combined mortality syndromes only.

## Discussion

Our results indicate that disease outbreaks can be detected much earlier and with only a marginal loss of specificity when incorporating delayed reporting compared to a system where reporting delay is present but not considered. Furthermore, the performance is on a comparable level to an ideal system in which no reporting delay is present, i.e. all cases are reported on the day they occur.

Moreover, the accumulation of evidence from several days resulted in a significantly better outbreak detection timeliness, at a given specificity; or a similar timeliness, but higher specificity, compared to an algorithm that, analogous to Salmon et al.[2], only looks for days with unusual high number of counts. We expect that this pattern will be even more pronounced for outbreaks that initially produce a moderate number of outbreak-related counts. If, for example, we expect a rather flat (i.e. slowly propagating) outbreak, the threshold based on $P(E|H_0)$ may need to be set at a low level, resulting in many false alarms. On the other hand, a higher threshold could only be used for spike-like (i.e. rapidly propagating) outbreaks. The method based on the value of evidence might be more appropriate for timely detection of flat outbreaks since it is computed by considering the evidence of the past d days. In the example, we used a threshold for V as the alarm trigger, allowing for a comparison with the z-based approach. As discussed in Andersson et al.[23], it is also possible to set a threshold for the posterior probability and use information about the expected seasonality of the disease of interest to further optimize the algorithm.

The effect of delayed reporting in SyS was assessed in a Bayesian framework by Salmon et al.[2] using a similar approach to ours. However, while the authors used a full Bayes approach[20] to model the baseline distribution and impact on delayed reporting, the SyS model is basically a Shewhart plot for which, in each time step, the count is compared with a threshold based on quantiles of the baseline distribution. In comparison, our work is an empirical Bayes approach[25] where parameters for baseline and outbreak distributions and delayed reporting are estimated separately using classical methods. This information is then combined to estimate the posterior probability for an ongoing outbreak or the odds between outbreak and baseline. As proposed in Andersson et al.[23], we separate the prior probability and the strength of the signal from SyS by estimating the value of evidence, i.e. the odds ratio between prior and posterior odds which is a proxy for the likelihood ratio (LR) for the observed evidence under the competing hypothesis ($H_1$ ongoing outbreak and $H_0$ baseline conditions).

Whereas neither approach can be claimed to be generally superior, they fulfil different niches. The approach of Salmon et al.[2] is intended for automatic analysis of a large number of data streams with little human intervention, in which case it may be impractical to formulate assumptions for each possible disease. The disadvantage is an output that may not always be very informative. On the contrary, our approach is designed to provide comprehensive decision support when SyS is implemented for surveillance of a moderate number of specific diseases or classes thereof. By building in empirical knowledge from previous outbreaks in the form of probability distributions, we can present the user with a clearer picture of the possible significance of a peak. In veterinary SyS, the number of data streams is generally small, compared to human SyS. In this situation, it may no longer be optimal
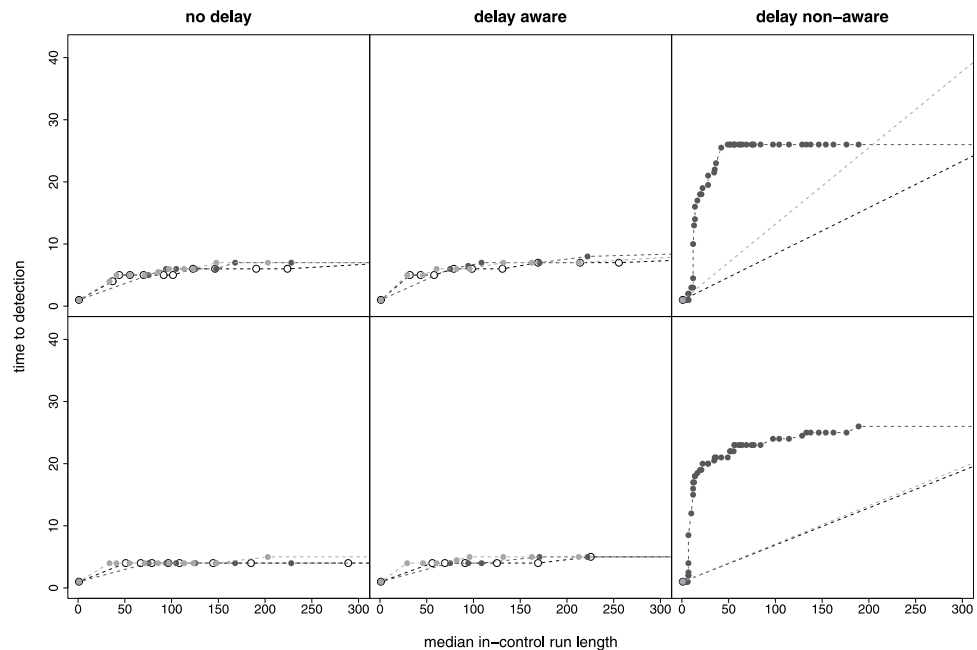
**Figure 8.** Influence of the choice of expert knowledge on the performance (median time to detection, in-control ARL) of detecting smaller (top row) and larger (bottom row) outbreak types under the V-based approach. Perfect expert knowledge was based on an ideal distribution of expected outbreak-related mortality cases (empty circles) for the given outbreak type. Non-perfect expert knowledge was simulated with expected outbreak distribution being biased either towards larger, steeply increasing outbreaks (black circles) or smaller, slowly increasing outbreaks (grey circles). Results are shown for combined mortality syndromes only.

to focus on algorithms that can be easily automatized. Rather, we argue, that one should incorporate as much knowledge as possible in the system to make maximum use of the data that is available.

Several approaches for constructing control charts based on cumulative sum control chart (CUSUM) or exponentially weighted moving average (EWMA) for multivariate data have been proposed, as reviewed by Frisén[29]. We do not doubt that, for example, multivariate CUSUM could be used to construct an algorithm with similar performance to ours. However, as discussed by Frisén[29], there is not a general best method applicable to all situations. Factors such as the relative importance of variables and the expected order of change points for two variables following an event would still have to be accounted for[29]. The Bayesian approach offers a straightforward and transparent way of incorporating such information and generates a model that not only triggers an alarm but may also explain why the alarm was triggered and how strong the evidence is in favour of an outbreak.

At first glance, incorporating a prior distribution for the expected number of outbreak related counts may be deemed more subjective than a method that solely relies on how unusual the peak is. However, even in such cases, the definition of the alarm threshold and estimation of the sensitivity is typically performed by simulating outbreaks. Thus, the Shewhart method also makes use of implicit assumption about the size and shape of the outbreak distribution. In practice, the decision whether the detection of an unusual peak should be considered sufficient to trigger an alarm depends also on what we expect to see in case of an outbreak. We argue that explicitly defining the outbreak distribution makes the reasoning of the system more transparent. As indicated in Figs 7 and 8, the approach works also when there is a discrepancy between the actual and expected outbreak distribution although, as expected, specifying that the expected outbreaks are large and steep, will reduce the sensitivity to detect small flat outbreaks.

Although in this work, we exemplify with baseline and outbreak distributions modelled by truncated normal distributions, it is not a limitation of the framework. In practice, each syndrome would be modelled by a distribution that is appropriate for its data structure. For example, low count data with and without clustering may be modelled by Poisson and NB distributions respectively as in ref. 23. Bayesian methods for fitting time series data have been described elsewhere[2, 30] and may be combined with the framework presented here by integrating over the posterior distribution of the model parameters when calculating $p(E|H_0)$. However, using the full posterior distribution rather than point estimates as input would result in increased complexity and computational time which may be a hurdle for implementation in operative surveillance. We believe that the relatively simple approach presented here may be a competitive alternative to methods like multivariate CUSUM.

## References

1. Jajosky, R. A. & Groseclose, S. L. Evaluation of reporting timeliness of public health surveillance systems for infectious diseases. *BMC Public Health* **4**, 29, doi:10.1186/1471-2458-4-29 (2004).
2. Salmon, M., Schumacher, D., Stark, K. & Höhle, M. Bayesian outbreak detection in the presence of reporting delays. *Biom. J.* **57**, 1051–67, doi:10.1002/bimj.201400159 (2015).

3. Farrington, C. P., Andrews, N. J., Beale, A. D. & Catchpole, M. A. A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease. *J. R. Stat. Soc. Ser. A Statistics Soc.* **159**, 547–563, doi:10.2307/2983331 (1996).

4. Noufaily, A. *et al.* An improved algorithm for outbreak detection in multiple surveillance systems. *Stat. Med.* **32**, 1206–22, doi:10.1002/sim.v32.7 (2013).

5. Jones, G. *et al.* The French human salmonella surveillance system: Evaluation of timeliness of laboratory reporting and factors associated with delays, 2007 to 2011. *Eurosurveillance* **19**, 1–10, doi:10.2807/1560-7917.ES2014.19.1.20664 (2014).

6. Jefferson, H. *et al.* Evaluation of a syndromic surveillance for the early detection of outbreaks among military personnel in a tropical country. *J. Public Health (Oxf).* **30**, 375–83, doi:10.1093/pubmed/fdn026 (2008).

7. Freeman, R. *et al.* Evaluation of a national microbiological surveillance system to inform automated outbreak detection. *J. Infect.* **67**, 378–84, doi:10.1016/j.jinf.2013.07.021 (2013).

8. Yoo, H.-S. *et al.* Timeliness of national notifiable diseases surveillance system in Korea: a cross-sectional study. *BMC Public Health* **9**, 93, doi:10.1186/1471-2458-9-93 (2009).

9. Noufaily, A., Ghebremichael-weldeselassie, Y., Enki, D. G. & Garthwaite, P. Modelling reporting delays for outbreak detection in infectious disease data. *J. R. Stat. Soc. A* **178**, 205–222, doi:10.1111/rssa.2014.178.issue-1 (2015).

10. Silin, M., Laraque, F., Munsiff, S. S., Crossa, A. & Harris, T. G. The Impact of Monitoring Tuberculosis Reporting Delays in New York City. *J. PUBLIC Heal. Manag. Pract.* **16**, E9–E17, doi:10.1097/PHH.0b013e3181c87ae5 (2010).

11. Tabnak, F., Muller, H., Wang, J., Chiou, J. & Sun, R. A change-point model for reporting delays under change of AIDS case definition. *Eur. J. Epidemiol.* **16**, 1135–1141, doi:10.1023/A:1010955827954 (2000).

12. Midthune, D. N., Fay, M. P., Clegg, L. X. & Feuer, E. J. Modeling Reporting Delays and Reporting Corrections in Cancer Registry Data. *J. Am. Stat. Assoc.* **100**, 61–70, doi:10.1198/016214504000001899 (2005).

13. Lawless, J. F. Adjustments for reporting delays and the prediction of occurred but not reported events. *Can. J. Stat.* **22**, 15–31, doi:10.2307/3315826.n1 (1994).

14. Clegg, L. X. Impact of Reporting Delay and Reporting Error on Cancer Incidence Rates and Trends. *CancerSpectrum Knowl. Environ.* **94**, 1537–1545 (2002).

15. Farrington, C. P. & Andrews, N. J. In *Monit. Heal. Popul.* (Brookmeyer, R. & Stroup, D. E.) (Oxford University Press, 2004).

16. Lui, K. J. & Rudy, R. K. An application of a mathematical model to adjust for time lag in case reporting. *Stat. Med.* **8**, 259–62; discussion 279–81 (1989).

17. Burkom, H. S., Elbert, Y., Feldman, A. & Lin, J. Role of data aggregation in biosurveillance detection strategies with applications from ESSENCE. *MMWR. Morb. Mortal. Wkly. Rep.* **53** (Suppl), 67–73 (2004).

18. Dórea, F. C., McEwen, B. J., McNab, W. B., Revie, C. W. & Sanchez, J. Syndromic surveillance using veterinary laboratory data: data pre-processing and algorithm performance evaluation. *J. R. Soc. Interface* **10**, 20130114–20130114, doi:10.1098/rsif.2013.0114 (2013).

19. Buckeridge, D. L. *et al.* Predicting outbreak detection in public health surveillance: quantitative analysis to enable evidence-based method selection. AMIA Annu. Symp. Proc. 76–80 at http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2656053&tool=pmcentrez&rendertype=abstract (2008).

20. Sonesson, C. & Frisén, M. *In Spat. Syndr. Surveill. Public Heal.* (Lawson, A. B. & Kleinman, K.) 153–166 doi:10.1002/0470092505 (John Wiley & Sons, Ltd, 2005).

21. Dubrawski, A. *In Infect. Dis. Informatics Biosurveillance* (Castillo-Chavez, C., Chen, H., Lober, W. B., Thurmond, M. & Zeng, D.) 27, 145–171 (Springer US, 2011).

22. Vial, F., Wei, W. & Held, L. Methodological challenges to multivariate syndromic surveillance: a case study using Swiss animal health data. *BMC Vet. Res.* **12**, 288, doi:10.1186/s12917-016-0914-2 (2016).

23. 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* **9**, e111335, doi:10.1371/journal.pone.0111335 (2014).

24. Faverjon, C. *et al.* Evaluation of a Multivariate Syndromic Surveillance System for West Nile Virus. *Vector borne zoonotic Dis.* **16**, 382–390, doi:10.1089/vbz.2015.1883 (2016).

25. Lawson, A. B. In *Spat. Syndr. Surveill. Public Heal.* (Lawson, A. B. & Kleinman, K.) 53–76, doi: 10.1002/0470092505.ch4 (John Wiley & Sons, Ltd, 2005).

26. Struchen, R., Reist, M., Zinsstag, J. & Vial, F. Investigating the potential of reported cattle mortality data in Switzerland for syndromic surveillance. *Prev. Vet. Med.* **121**, 1–7, doi:10.1016/j.prevetmed.2015.04.012 (2015).

27. Dórea, F. C. *et al.* Retrospective time series analysis of veterinary laboratory data: Preparing a historical baseline for cluster detection in syndromic surveillance. *Prev. Vet. Med.* **109**, 219–227, doi:10.1016/j.prevetmed.2012.10.010 (2012).

28. Venables, W. N. & Ripley, B. D. Statistics and Computing, Modern Applied Statistics with S. at http://www.stats.ox.ac.uk/pub/MASS4, doi:10.1007/978-0-387-21706-2 (2002).

29. Frisén, M. On multivariate control charts. *Produção* **21**, 235–241, doi:10.1590/S0103-65132011005000010 (2011).

30. Jung, R. C., Kukuk, M. & Liesenfeld, R. Time series of count data: modeling, estimation and diagnostics. *Comput. Stat. Data Anal.* **51**, 2350–2364, doi:10.1016/j.csda.2006.08.001 (2006).

## Acknowledgements

## Author Contributions

F.V. and G.A. conceived the research. F.V. and G.A. developed the model framework. G.A. and R.S. conceived the final model and performed the analyses. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-01259-5

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.