Bayesian outbreak detection in the presence of reporting delays

Maëlle Salmon[*,1], **Dirk Schumacher**[1], **Klaus Stark**[1], and **Michael Höhle**[2]

[1] Department for Infectious Diseases Epidemiology, Robert Koch Institut, Seestrasse 10, 13353 Berlin, Germany
[2] Department of Mathematics, Stockholm University, Kräftriket, 106 91 Stockholm, Sweden

Received 14 July 2014; revised 29 April 2015; accepted 12 May 2015

One use of infectious disease surveillance systems is the statistical aberration detection performed on time series of counts resulting from the aggregation of individual case reports. However, inherent reporting delays in such surveillance systems make the considered time series incomplete, which can be an impediment to the timely detection and thus to the containment of emerging outbreaks. In this work, we synthesize the outbreak detection algorithms of Noufaily et al. (2013) and Manitz and Höhle (2013) while additionally addressing right truncation caused by reporting delays. We do so by considering the resulting time series as an incomplete two-way contingency table which we model using negative binomial regression. Our approach is defined in a Bayesian setting allowing a direct inclusion of all sources of uncertainty in the derivation of whether an observed case count is to be considered an aberration. The proposed algorithm is evaluated both on simulated data and on the time series of *Salmonella* Newport cases in Germany in 2011. Altogether, our method aims at allowing timely aberration detection in the presence of reporting delays and hence underlines the need for statistical modeling to address complications of reporting systems. An implementation of the proposed method is made available in the R package surveillance as the function "bodaDelay".

Keywords: Bayesian inference; Infectious diseases; INLA; Reporting delays; Surveillance.

Additional supporting information including source code to reproduce the results may be found in the online version of this article at the publisher's web-site

1 Introduction

Since 2001 the Protection against Infection Act regulates which infectious diseases are notifiable in Germany so that anonymized information about each diagnosed case is sent from local health authorities to the Robert Koch Institute (RKI) via federal health authorities (Faensen et al., 2006; Krause et al., 2007). The resulting database does not only support the exploration of disease counts for situational awareness and yearly summaries for health statistics, but also the detection of emerging outbreaks by an automated data analysis system (Salmon et al., 2014). However, case reports do not arrive immediately at the RKI because of inherent reporting delays, due to a time lag between disease onset and reporting date as well as to the decentralized structure of the reporting system. This can be a hindrance to the timely detection of outbreaks. For example, Jones et al. (2014) and Noufaily et al. (2015) have analyzed reporting delays in European reporting systems, respectively, for *Salmonella* in France and 12 infections in the United Kingdom. Furthermore, Altmann et al. (2011) and Höhle and an der Heiden (2014) have analyzed reporting delays during the German EHEC O104:H4 outbreak in particular. In this article, we investigate how to improve the real-time automated outbreak detection by adequately addressing reporting delays.

*Corresponding author: e-mail: SalmonM@rki.de. Phone: +49-30-187543802.

The available notification reports can be aggregated over time units (e.g., weeks) and other sublevels (e.g., pathogen subtype or age group) in order to obtain time series of reported incidence counts, which often display characteristics such as seasonality, overdispersion and presence of past outbreaks. Statistical algorithms for aberration detection can fit adequate models to these time series of counts and thus derive a predictive distribution for the current count. If the observed current count lies above a suitable quantile of this distribution, an alarm is flagged, prompting further checks by epidemiologists. A variety of statistical algorithms for aberration detection exist: see , for example, the reviews in Buckeridge et al. (2005) and Unkel et al. (2012). Nearly all of them are of frequentist nature, but recently more Bayesian oriented proposals have emerged (Höhle, 2007; Martínez-Beneito et al., 2008; Conesa et al., 2015; Manitz and Höhle, 2013). One advantage of a Bayesian approach when calculating a predictive distribution is that this distribution takes into account both the uncertainty from estimation and the stochasticity of the model, that is on the one hand the uncertainty in the model parameters which we estimate based on a limited sample, and on the other hand the uncertainty resulting from natural fluctuations of the number of cases. In the present work, we shall build upon these strengths and hence try to move toward Bayesian thinking in the field of outbreak detection, especially because prediction is natural in a Bayesian framework.

The RKI automated system for aberration detection (Salmon et al., 2014), which will be the main motivation of our work, helps uncovering outbreaks of infectious diseases based on the German mandatory reporting data. In this system, aberration detection is performed on a daily basis for the weekly disease counts of hundreds thousands of time series and is based on the state-of-the-art algorithm proposed by Noufaily et al. (2013), which is an improvement of the widely adopted algorithm by Farrington et al. (1996). Nevertheless, the response time of the system when there is an outbreak, a.k.a. the system's timeliness, does not only depend on the implemented algorithm but also on data quality. In Germany, each new case of a notifiable infectious disease has to pass several stages before arriving at the RKI database. This includes onset of disease symptoms in the patient, the patient's visit to the doctor, a diagnostic laboratory analysis and a report of the case to the local health authority followed by its notification to the federal health authority before it is finally notified at the RKI. Moreover, each step within the reporting system requires the fulfilment of certain quality criteria before further transmission. Although technical, legal, and managerial efforts are made in order to reduce these delays, reporting delays remain an issue.

Adjusting time series of public health events for reporting delays became important in biostatistics as part of modeling of the AIDS epidemic (Brookmeyer and Damiano, 1989; Kalbfleisch and Lawless, 1989; Zeger et al., 1989), but applications are also found in noninfectious modeling such as cancer registry data (Midthune et al., 2005) or mortality monitoring (Lin et al., 2008). Moreover, the adjustment for such occurred-but-not-yet-reported events (Lawless, 1994) has strong links to actuarial sciences, where it is part of claims reserve modeling (England and Verrall, 2002; Hess and Schmidt, 2002). Recently, such adjustments have re-emerged in public health settings under the name of *nowcasting* (Donker et al., 2011; Höhle and an der Heiden, 2014). In outbreak detection, we aim at comparing the current number of cases to a threshold computed from past data in order to see if the current number of cases is *unexpectedly high*. Using *nowcasting* would allow us to correct the current number of cases before the comparison to a threshold (Heisterkamp et al., 2006; Gergonne et al., 2011). In the present work we choose an alternative approach: we correct the threshold for reporting delays and hence compare the observed current number of cases to this threshold. This approach allows to have all sources of estimation uncertainties—from the prediction error as well as from the delay correction—on the same side of the comparison rather than to have on the one side a nowcasted number of observed cases with its uncertainty and on the other side a predicted number of cases with its own uncertainty, possibly correlated with each other. Additionally, the use of the actual observed number, rather than of some hypothetical estimate associated with uncertainty, is psychologically advantageous since epidemiologists do not need to make further investigations about the imaginary cases that a nowcasted number of cases automatically entails. Another recent work also follows this strategy of correcting the threshold, and uses a statistical test on a cumulative sum of discrepancies

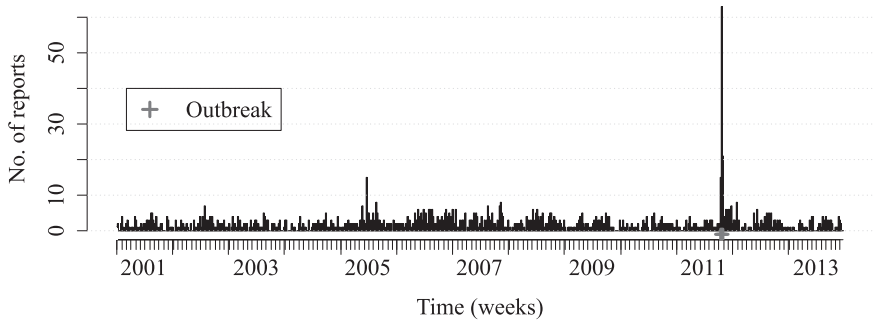

Figure 1 Weekly time series for all 1391 *Salmonella* Newport cases reported in Germany 2002–2013 by onset of disease (as available now in retrospect).

between the observed number of counts and a threshold corrected for reporting delays (Noufaily et al., 2014).

Altogether, our blueprint for a Bayesian outbreak detection algorithm taking reporting delays into account is a synthesis of the outbreak detection algorithms of Noufaily et al. (2013) and Manitz and Höhle (2013) where we extend the regression approach to the so-called reporting triangle, in which information about reporting delays is stored. The structure of the article is as follows. First, we state the problem of reporting delays in routine surveillance data, introduce adequate notation and illustrate it using RKI salmonellosis data in Section 2. Then, Section 3 introduces the methodological framework for aberration detection. The proposed algorithm, including numerical aspects of its implementation, is explained in Section 4 before being tested on simulated data in Section 5 and illustrated on real data in Section 6. Finally, a discussion rounds off the work.

2 Motivational example and notation

Throughout our work, we shall use notification data for *Salmonella enterica* subsp. *enterica* serovar Newport (*Salmonella* Newport) to illustrate matters. Salmonellosis is a bacterial caused gastrointestinal disease, with symptoms such as diarrhea, fever, and vomiting (Sánchez-Vargas et al., 2011). The Newport serovar is rather uncommon in Germany: in 2010 only 0.3% of all 25,307 notified *Salmonella* infections were of this serovar. Nevertheless, during the winter of 2011 there was a strong increase in the number of reported cases as shown in Fig. 1, which was due to an outbreak linked to mung bean sprouts (Bayer et al., 2014). We shall use this outbreak as illustration and motivation throughout our work.

Currently, monitoring at the RKI is performed on time series aggregated by date of report arrival at the local health authority. Nonetheless, as a supplement, there is an interest in monitoring cases based on their symptoms onset date: for *Salmonella* this would typically be the onset of diarrhea. One advantage is that onset dates, despite being self-reported by patients, provide a more precise description of the temporal extension of a possible outbreak. Unlike dates of report they escape noise introduced by delays, that is, due to patients going to the doctor, shipping of samples to laboratories, as well as reporting artifacts such as the late discovery of cases through interviews of other cases. In the RKI *S.* Newport data, onset date is available for about 80% of all cases. Because of the additional delay between symptoms onset and reporting to local health authority, taking reporting delays into account becomes even more important when trying to monitor available time series based on these data. Hence, only a delay adjustment method could justify the additional costs incurred by this supplementary monitoring.

For a case report i let the tuple (t_i^E, t_i^R) denote the time of the *e*vent of interest (e.g., disease onset or receipt of case information at local health authority) and the time the notification of this event

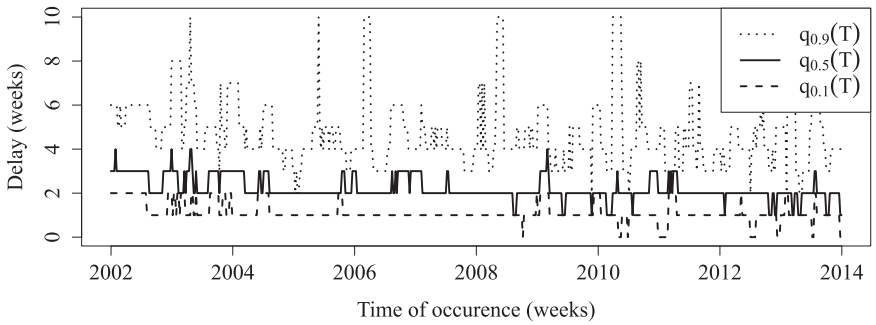

Figure 2 Median reporting delay between symptoms onset and arrival at the RKI (as well as 10% and 90% quantiles) of the smoothed empirical delay distribution as a function of time of occurrence, for all 1391 *Salmonella* Newport cases reported in Germany 2002–2013 (as available now in retrospect).

arrives at the RKI, respectively. Let $d_i = t_i^R - t_i^E$ be the delay between these two events. We place our analysis in a discrete time setting, with units being, for example, days or weeks and assume that the d_i are independent variables drawn from a distribution with probability mass function (PMF) f_d that has support $\{0, 1, 2, \ldots, D\}$, where D is the maximal relevant delay. This can, for example, be the delay after which sufficiently many of the observations have become available or the maximum delay back in time where interventions are still relevant. Reports with a delay larger than D are ignored. For the convenience of the analysis, we shall furthermore assume that the delay distribution is stable over time. This means we ignore increase or decrease of the delays during, for example, past outbreaks. For an outbreak to be detected, the delay can thus be assumed to be the same as for nonoutbreak situations, since no awareness of the outbreak exists yet. Figure 2 contains an illustration of the weekly median as well as the 10% and 90% quantiles of the smoothed empirical delay distribution for *Salmonella* Newport cases in Germany, obtained from tabulating all cases occurring within a moving window of $t-4, \ldots, t+4$ weeks. The distribution can be assumed stable until March 2013 where an amendment to the Protection against Infection Act made it compulsory to notify cases on a daily basis to the RKI.

At fixed observation time $T =$ now which in our context is a specific week, we only observe a case occurred at time t if it was reported with a delay at most equal to $T - t$: the data are thus right-truncated. Following the notation from Lawless (1994), let $n_{t,d}$ be the number of cases of the disease occurred at time t and reported with a delay of d time units. Not all $n_{t,d}$ for $t \in \{0, \ldots, T\}$ and $d \in \{0, \ldots, D\}$ are available at time T. This is illustrated in Fig. 3: each row represents counts of cases occurred at time t, counts for cells located to the right of $\min(T, D-t)$ are not available at time T. For the subsequent analysis, we define three index sets: $A_T = \{(t,d) : 0 \le t \le T, 0 \le d \le D\}$ contains indices for *all* observations $n_{t,d}$ while $O_T = \{(t,d) : 0 \le t \le T, 0 \le d \le T-t\}$ only contains indices for *observed* $n_{t,d}$ at time T and $U_T = A_T \setminus O_T$ only contains indices for *occurred-but-not-yet-reported* $n_{t,d}$ at time T. Note that O_T and U_T, respectively, are indicated as the light gray trapezoid and as the darker gray triangle in Fig. 3. Counts corresponding to those index sets are denoted by $\boldsymbol{n}_{O_T}$ and $\boldsymbol{n}_{U_T}$, respectively. At time T, we only observe events reported by time T so that $N(t,T) = \sum_{d=0}^{\min(D,T-t)} n_{t,d}$.

For situational awareness at time T, we would like to know the row sum $N_t = N(t,\infty) = \sum_{d=0}^{\infty} n_{t,d}$, that is, the total number of cases occurred at time t, where $0 \le t \le T$. One would eventually observe N_t after waiting for an infinite time or for at least a time equal to the maximal possible delay D. Because of right truncation, at time $T < t + D$, $N(t,T)$ is possibly still smaller than N_t as seen in Fig. 3. The practical consequences of this right truncation for the *S.* Newport outbreak associated with sprouts are illustrated in Fig. 4: at week 45 of 2011, the RKI could only observe 15 cases for week 43, that is, $N(\text{W43-2011, W45-2011}) = 15$, but the occurred number of cases that week was 63,

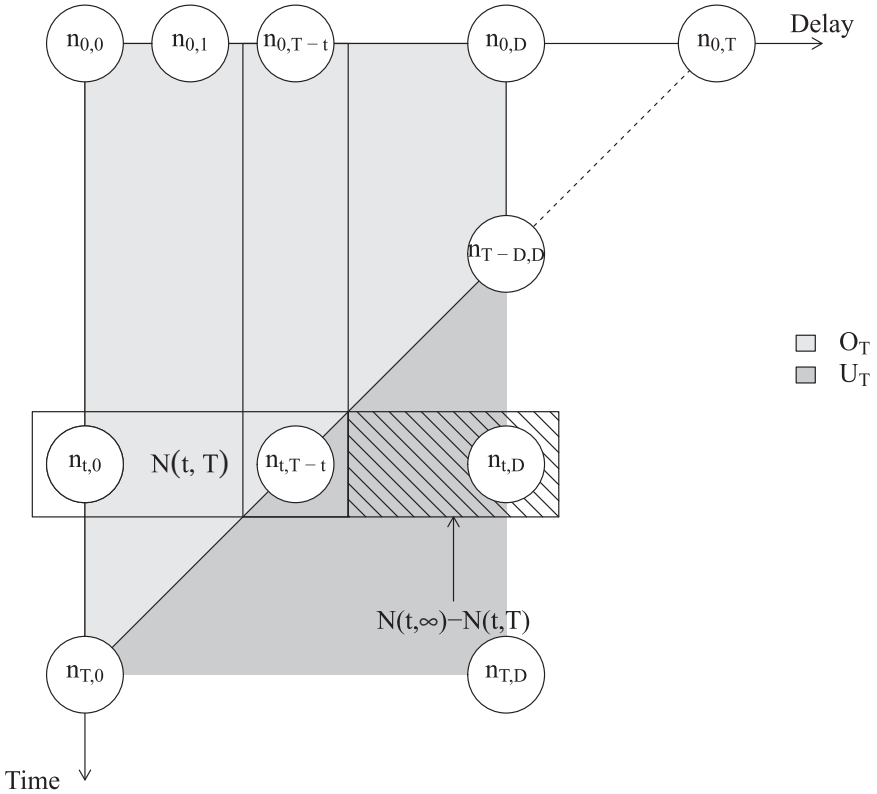

Figure 3 Reporting triangle at time T. Available observations are those in the right-angled trapezoid O_T spanned by $n_{0,0}$, $n_{T,0}$, $n_{T-D,D}$, and $n_{0,D}$. Occurred-but-not-yet-reported events are those in the triangle U_T spanned by $n_{T,1}$, $n_{T-D+1,D}$, and $n_{T,D}$.

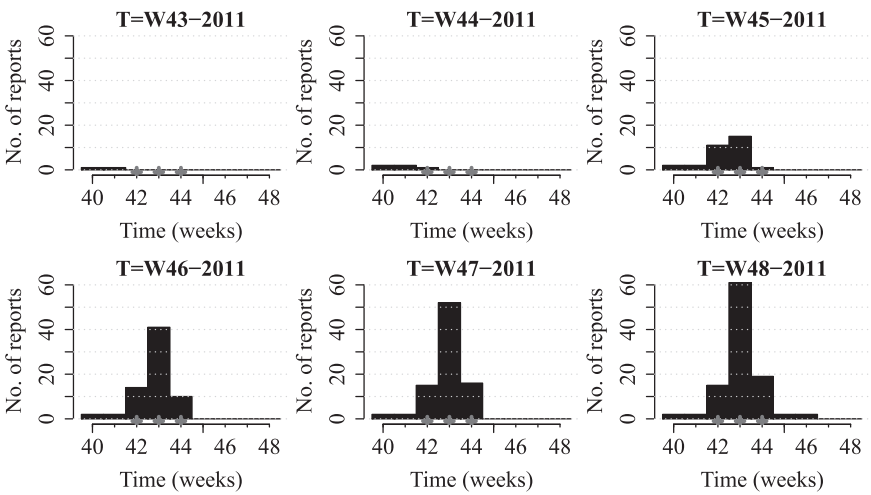

Figure 4 Weekly counts $N(t, T)$ of *Salmonella* Newport cases in the RKI database aggregated by date of disease onset at different observation times T during the outbreak. The six plots show $t =$ W40-48 in 2011 as of $T =$ from W43 in 2011 to W48 in 2011. The gray "+" symbols indicate the three outbreak weeks.

that is, $N_{\text{W43-2011}} = 63$. Such incomplete observations during monitoring are an obstacle to situational awareness and outbreak detection.

3 Methodological framework for outbreak detection

In the context of aberration detection, we want to compare to a threshold the available number of events $N(s, T)$ occurred at time $s \leq T$ and observed at fixed time $T =$ now. Typically, s is the current week, that is, $s = T$. In Section 3.1, we describe the framework for aberration detection in situations where $D = 0$, that is, when there is no reporting delay before Section 3.2 covers the situation when $D > 0$. Section 4 proposes the novel Bayesian algorithm for handling this difficulty.

3.1 Aberration detection without reporting delays

Here, we briefly explain the frequentist algorithm described by Noufaily et al. (2013) for monitoring $s = T$ when $D = 0$. It is based on a quasi-Poisson generalized linear model (GLM) that outputs a predictive distribution for N_s. The GLM is fitted on past counts $N(t, T)$ where t is from the last b years before s except the last 26 weeks prior to s. Mean and variance are given by $\mathrm{E}(N_t) = \mu_t$ and $\mathrm{Var}(N_t) = \phi \cdot \mu_t$, respectively, where $\phi \geq 1$ is ensured. Furthermore,

$$\log(\mu_t) = \beta_0 + \beta_1 \cdot t + \gamma_{c(t)},$$

where β_0 is the intercept, β_1 the coefficient of the linear time trend, $c(t)$ is a periodic function mapping time t to a factor level i and thus $\gamma_i \sim N(0, \lambda_{\gamma_i}^{-1})$, $i = 0, \ldots, S-1$ are the coefficients of a zero order spline with $S = 10$ knots, which can be conveniently represented as a 10-level factor that reflects seasonality. Thus, i indicates in which season or period of the year t belongs to. One of the periods is a $2 \cdot w + 1$-week interval centered around s. The threshold is then defined as the $(1-\alpha)$-quantile Q_s of the predictive distribution of N_s parameterized with plug-in estimates from the fitting of the GLM. If $\hat{\phi} > 1$ one assumes the predictive distribution of N_s to be negative binomial $\mathrm{NB}(\hat{\mu}_s, \hat{\mu}_s/(\hat{\phi}-1))$, otherwise the predictive distribution is taken to be $\mathrm{Po}(\hat{\mu}_s)$. Subsequently, the alarm indicator a_s at time s is defined as $a_s = I(N_s > Q_s)$ which statistically measures whether the current observed count is *unexpectedly high*. Note that the definition of the threshold Q_s in Noufaily et al. (2013) takes the sampling variation of a new observation into account, but disregards the uncertainty of the estimation of $\hat{\mu}_s$ and $\hat{\phi}$ by using plug-in estimates.

3.2 Aberration detection adjusting for reporting delays

When $D > 0$, the right truncation of the data creates two complications. On the one hand, for each observation time T the counts from the previous time units can have increased due to reporting delays and become aberrant so that one should not only monitor the current timepoint $s = T$ but also timepoints before T. For $s < T - D$, we have $N(s, T) = N_s$ so that these counts do not change anymore and hence do not need to be monitored again. Assume therefore, than in the presence of reporting delays, at each observation timepoint T we monitor the set of timepoints $M_T = \{T - D, \ldots, T\}$. If this is repeated for several subsequent observation timepoints then each timepoint gets monitored $D + 1$ times. This is illustrated in Fig. 5.

On the other hand, one needs to adjust the detection algorithm for the incomplete observations because otherwise incomplete counts are compared to complete counts. When $D > 0$ the use of the predictive distribution for N_s *in control* is meaningful only if $s \leq T - D$ so that the observations for s are complete. Otherwise, we make a biased comparison, at the risk of not getting any alarm in the cases where $N(s, T)$ is smaller than the threshold: not because N_s itself is smaller than the threshold,

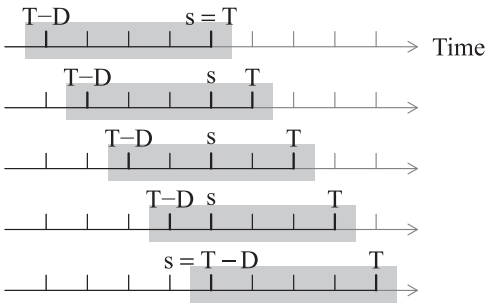

Figure 5 Sets of timepoints M_T monitored for five subsequent observation times T, among which we see a fixed timepoint s. In this illustration, $D = 4$. From top to bottom time goes by so that at the end the timepoint s has been monitored $D + 1 = 5$ times.

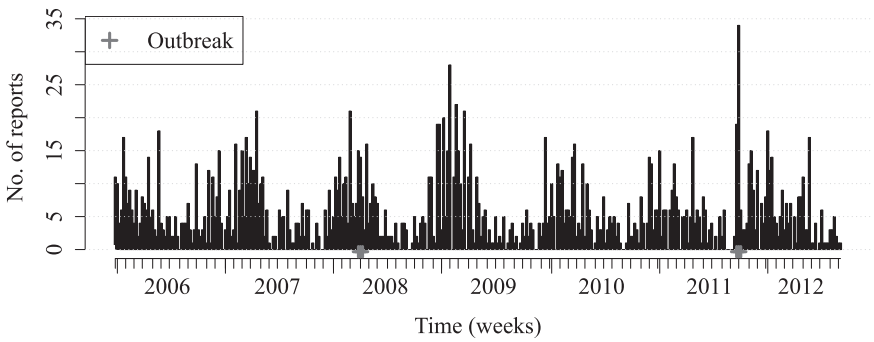

Figure 6 Example of a simulated time series with outbreaks of size coefficient $\sigma = 2$ and $\sigma = 5$ starting at week 13 of 2008 and at week 38 of 2011, respectively.

but simply because of reporting delays. Current algorithms for aberration detection output the same threshold $Q_{s,T}$ no matter how close s is to T. When correcting for right truncation in the threshold calculation we thus hope to be able to spot aberrations for time s at an earlier observation timepoint T. We do this as follows: we aim at inferring a predictive distribution for $N(s, T)$, instead of a predictive distribution for N_s. The actually observed $N(s, T)$ is then compared to a quantile $Q_{s,T}$ from this distribution and an alarm is flagged if this threshold is exceeded, that is, $a_{s,T} = I(N(s, T) > Q_{s,T})$.

4 Bayesian algorithm for delay-adjusted aberration detection

Our model is a synthesis of the modeling presented by Manitz and Höhle (2013) and of the modeling of Noufaily et al. (2013) now extended to account for delays while keeping the treatment in a Bayesian framework. From the HIV/AIDS literature, it is well known that the observations in the reporting triangle can be formulated as an inference problem in an incomplete contingency table (Kalbfleisch and Lawless, 1989; Zeger et al., 1989). The model we define is inspired by the so-called multinomial model for claim counts (Schmidt and Wünsche, 1998), that is, we assume that

$$N_t \sim \mathrm{NB}(\mu_t, \nu), \quad 0 \le t \le T, \tag{1}$$
$$(n_{t,0}, n_{t,1}, \ldots, n_{t,D}) \mid N_t, \boldsymbol{p} \sim \mathrm{M}(N_t, \boldsymbol{p}),$$

with $E(N_t) = \mu_t$ and $\mathrm{Var}(N_t) = \mu_t(1 + \mu_t/\nu)$. Here, $\mathrm{M}(N_t, \boldsymbol{p})$ is the multinomial distribution with size parameter N_t and $\boldsymbol{p} = (p_0, p_1, \ldots, p_D)$ is the PMF of the delay distribution. Then, from Schmidt and Wünsche (1998) we know that:

$$n_{t,d} \sim \mathrm{NB}(\mu_t \cdot p_d, \nu), \quad (t, d) \in O_s. \tag{2}$$

Synthesizing the elements of Noufaily et al. (2013) for the modeling of $\log(\mu_t)$ in the above, we obtain

$$\log(\mu_t) = \beta_0 + \beta_1 \cdot t + \gamma_{c(t)}.$$

We parametrize the delay on $0, \ldots, D$ as $\log(p_d) = \alpha_d$. Altogether, the parameter vector for the linear predictor is $\boldsymbol{\theta} = (\beta_0, \beta_1, \gamma_0, \ldots, \gamma_{S-1}, \alpha_0, \ldots, \alpha_D)'$ and the whole parameter vector is $\boldsymbol{\psi} = (\boldsymbol{\theta}, \nu)'$. Hence, we can write that

$$\log(\mu_t \cdot p_d) = \eta_{t,d} = \boldsymbol{z}_{t,d}'\boldsymbol{\theta}$$

with $\boldsymbol{z}_{t,d}$ the vector of covariates corresponding to the linear predictor $\eta_{t,d}$. As in Manitz and Höhle (2013) we assume the observations $n_{t,d}$ given the model structure (2) are independent, so that the likelihood becomes

$$f(\boldsymbol{n}_{O_s} \mid \boldsymbol{\psi}) = \prod_{(t,d) \in O_s} f(n_{t,d} \mid \boldsymbol{\psi}),$$

where $f(n_{t,d} \mid \boldsymbol{\psi})$ is the PMF of the negative binomial distribution presented in Eq. (2). From this, using Bayes' theorem, we get

$$f(\boldsymbol{\psi} \mid \boldsymbol{n}_{O_s}) \propto f(\boldsymbol{n}_{O_s} \mid \boldsymbol{\psi}) \cdot \pi(\boldsymbol{\psi}).$$

For inferring the predictive posterior distribution, we now propose two methods: a fully Bayesian method using integrated nested Laplace approximations (INLA) (Rue et al., 2009), and a method using an asymptotic normal approximation of the posterior that is much faster and thus better adapted to future routine applications of the algorithm.

4.1 Inference of the predictive posterior distribution with INLA

Priors are chosen as in Manitz and Höhle (2013) where possible. This means, for example, $\beta_i \sim \mathrm{N}(0, \lambda_{\beta_i}^{-1})$, $i = 1, 2$, where the λ_{β_i}'s indicate precision parameters. We choose $\nu \sim \log\mathrm{N}(0, 100)$ for not imposing a too high constraint on overdispersion. For the period effects we use independent priors, that is, $\gamma_i \sim \mathrm{N}(0, \lambda_{\gamma_i}^{-1})$. Furthermore, for each level α_d of the factor variable accounting for delay we use independent priors $\alpha_d \sim \mathrm{N}(0, \lambda_{\alpha_d}^{-1})$. For ensuring identifiability, we add two corner constraints for the blocks of parameters for season and delay: $\alpha_0 = 0$ and $\gamma_0 = 0$. For all fixed effects, we used the default value for the precision, 0.001. Finally, we assume independence of the priors so that the prior $\pi(\boldsymbol{\psi})$ is the product of the parameters priors. Inference for the posterior is then performed with INLA as described in Rue et al. (2009, 2015).

4.2 Asymptotic inference of the predictive posterior distribution

In this method, we first fit a negative binomial GLM type regression model with log-link to the data in O_s according to Eq. (2), providing a frequentist estimation of the parameters. This results in the estimators $(\hat{\eta}_{t,d}, \hat{\nu})'$. We then use the asymptotic normal distribution of the parameters for

approximating the posterior $f(\boldsymbol{\psi} \mid \boldsymbol{n}_{O_s})$. As explained in, for example, Held and Sabanés Bové (2014), the posterior distribution is under suitable regularity conditions asymptotically Gaussian with mean equal to the maximum-likelihood estimator and covariance equal to the inverse observed Fisher information matrix, no matter the priori:

$$\boldsymbol{\psi} \mid \boldsymbol{n}_{O_s} \overset{a}{\sim} \mathrm{N}(\hat{\boldsymbol{\psi}}, I(\hat{\boldsymbol{\psi}})^{-1}).$$

Moreover, due to the estimation method of the negative binomial model, we assume that $\eta_{t,d}$ and ν are information-orthogonal. With these assumptions, $\eta_{t,d}$ is the linear combination of Gaussian variables so that we can write that

$$\eta_{t,d} \overset{a}{\sim} \mathrm{N}\left(\boldsymbol{z}'_{t,d}\hat{\boldsymbol{\theta}}\,,\ \boldsymbol{z}'_{t,d} I(\hat{\boldsymbol{\theta}})^{-1} \boldsymbol{z}_{t,d}\right),$$

where $I(\hat{\boldsymbol{\theta}})$ is the block related to $\hat{\boldsymbol{\theta}}$ in $I(\hat{\boldsymbol{\psi}})$. Note that with $b = 4$ years data, $S = 10$, and $D = 10$, we have over $150 \cdot (D+1)$ observations for inference, that is to say there are over 1600 elements in O_s for the 24 parameters to be estimated. We hence expect the difference between the asymptotic method and the INLA method to be small.

4.3 Predictive posterior distribution integration and threshold calculation

One can directly compute the posterior distribution of $N(t, T)$ from the inferred distribution of the parameters because

$$N(t, T) \mid N_t, \boldsymbol{p} \sim \mathrm{Bin}\left(N_t, \sum_{d=0}^{\min(T-t,D)} p_d\right).$$

Therefore, as N_t is a negative binomial variable, we obtain (Schmidt and Wünsche, 1998)

$$N(t, T) \sim \mathrm{NB}\left(\sum_{d=0}^{\min(T-t,D)} p_d \mu_t, \nu\right).$$

The joint posterior distribution of the current observations and of the parameter $\boldsymbol{\theta}_T$ can then be computed. For getting the predictive posterior distribution for the $n_{t,d}$, we have to integrate with respect to $\boldsymbol{\psi}$,

$$\begin{aligned} f(N(s, T) \mid \boldsymbol{n}_{O_s}) &= \int f(N(s, T), \boldsymbol{\psi} \mid \boldsymbol{n}_{O_s}) d\boldsymbol{\psi} = \\ &= \int f(N(s, T) \mid \boldsymbol{\psi}) f(\boldsymbol{\psi} \mid \boldsymbol{n}_{O_s}) d\boldsymbol{\psi}. \end{aligned}$$

The threshold $Q_{s,T}$ is defined as the $(1-\alpha)$-quantile from this distribution. Then the alarm indicator is $a_{s,T} = I(N(s, T) > Q_{s,T})$. This comparison encompasses both estimation and observation uncertainties. Independent of whether the INLA or asymptotic method was used, we shall proceed as in Manitz and Höhle (2013) and use Monte Carlo simulation to obtain the desired quantile of the posterior predictive, as explained in Algorithm 1.

Algorithm 1. Algorithm for computing the threshold in the Bayesian setting.

Data $\boldsymbol{n}_{O_s}$
Result $\hat{Q}_{s,T}$
Determine $f(\boldsymbol{\psi} \mid \boldsymbol{n}_{O_s})$ using either the INLA method or the asymptotic method.
for $1 \leq r \leq R$ **do**
  Sample $\boldsymbol{\psi}^{(r)} \sim f(\boldsymbol{\psi} \mid \boldsymbol{n}_{O_s})$.
  Sample the $N(s, T)^{(r)} \sim f(N(s, T) \mid \boldsymbol{\psi}^{(r)})$.
end
Determine $\hat{Q}_{s,T}$ as the empirical $(1-\alpha)$-quantile of the responses $N(s, T)^{(r)}$, $r \in \{1, \ldots, R\}$.

5 Simulation study

This section explores the performance of the proposed delay-adjusting algorithm by first looking on simulated data. Section 6 contains an analysis on the *S.* Newport surveillance time series.

5.1 Simulated data

We used simulated time series built as follows to evaluate the algorithm: Baseline counts B_t are drawn from a negative binomial distribution with size parameter ν and mean μ_t for each timepoint. The structure for the mean is the same as in Noufaily et al. (2013), that is,

$$\log(\mu_t) = \beta_0 + \beta_1 t + \sum_{j=1}^{m} \left\{ \gamma_{2m-1} \cos\left(\frac{2\pi jt}{52}\right) + \gamma_{2m} \sin\left(\frac{2\pi jt}{52}\right) \right\}.$$

In this expression β_0 is the baseline frequency of reports, β_1 is the time trend, and if $m > 0$ then the γ's specify the annual ($m = 1$) and biannual ($m = 2$) seasonality with $\gamma_1 = \gamma_3$ and $\gamma_2 = \gamma_4$. We used the 42 scenarios of Noufaily et al. (2013) but had to change their parameterization of the variance which was $\mathrm{Var}(B_t) = \phi\mu_t$, that is, the variance structure adapted to a quasi-Poisson regression model. In order to keep the variety of the scenarios and to have comparable scenarios, we chose the values for ν so that the variance obtained for $\mu^* = \exp(\beta_0)$ would be the same in both cases. This gives, for each scenario, $\nu = \mu^*/(\phi - 1)$.

Outbreaks were added as in Noufaily et al. (2013). Their size is fixed by the parameter σ, the outbreak *size coefficient*, so that the total number of cases of the outbreak starting at t is $\mathrm{Po}(\sigma \cdot \mathrm{Var}(B_t))$. The outbreak cases are distributed in time according to a discretized lognormal distribution with mean 0 and standard deviation 0.5. At each timepoint, we have $N_t = B_t + K_t$ with K_t the outbreak count which is possibly zero. An example of a simulated time series is shown in Fig. 6.

We also simulated the reporting of these cases with the delay distributed according to the empirical distribution of reporting delays of *Salmonella* cases to the RKI in 2011 and with maximal delay set to $D = 10$ weeks. Cases with delays longer than 10 are ignored. The probability associated with each possible value of the delay from 0 to 10 is given by $\boldsymbol{p} = (0.035, 0.369, 0.357, 0.139, 0.049, 0.020, 0.01, 0.005, 0.004, 0.002, 0.009)'$. Using this delay distribution we are able to generate the reporting triangles and hence the partial time series at a given timepoint T, that is, $N(t, T)$ instead of N_t.

5.2 Specification of the algorithms

We compare the algorithm of Noufaily et al. (2013) and the algorithm proposed in Section 4 with and without delay correction, that is, using $D = 10$ and $D = 0$. In all cases we use $S = 10$ periods for the seasonal effect, $b = 4$, and $w = 3$. We generate time series of 350 weeks. Since we wish to study

an algorithm correcting for reporting delays, we need to monitor each timepoint at different dates as shown in Fig. 5: for each time series monitored, each timepoint is monitored $D + 1 = 11$ times. This means that our simulation study involves many more computations than studies ignoring reporting delays and only monitoring complete counts. For methods that do not correct for reporting delays, all upperbounds for a fixed s are equal — although the proposed method with $D = 0$ can produce slightly different thresholds because of sampling uncertainty. Thus, when using a nonreporting delay adjusting method for monitoring, if $a_{s,T} = 1$ then $a_{s,T+1} = 1$, that is, alarms do not disappear.

5.3 Evaluation measurements

We used the same evaluation measurements as in Noufaily et al. (2013): the probability of detection (POD) and the false positive rate (FPR). The FPR is the number of alarms in weeks without outbreaks divided by the number of weeks in the monitored period. The POD is estimated as the number of detected outbreaks, that is to say when an alarm corresponds to one of the outbreak weeks, divided by the number of outbreaks. Since we only include one outbreak in the monitored period of time, each simulation gives a value of either 1 or 0 for a given algorithm, whose mean among several simulations gives the POD. Note that the POD and the FPR do not measure timeliness. Therefore, for each tested algorithm, we calculate the POD and the FPR for each possible value of $T - s \in \{0, \ldots, D\}$. In methods that do not correct for reporting delays, the threshold is only valid for complete observations so that we expect the FPR to be controlled only for the $N(s, T)$ with $s = T - D$. Moreover, as the methods that do not correct for reporting delays make the same guess for every $N(s, T)$ no matter how close s is to T, they are also less likely to detect an outbreak for small $T - s$. Thus, we expect our algorithm to have a fairly constant FPR over different values of $T - s$ and to detect outbreaks for smaller $T - s$ in comparison with methods that do not correct for reporting delays.

We also introduce a measure of timeliness: the reaction time (RT) needed for the algorithm to produce an alarm for an outbreak after its beginning if the outbreak was detected. In other words, if there is an alarm for the outbreak starting at t_{out} and spanning over l_{out} time units, $\text{RT}(t_{\text{out}}, l_{\text{out}})$ is the difference between t_{out} and the smallest observation time T at which we get an alarm for any $s \in \{t_{\text{out}}, \ldots, t_{\text{out}} + l_{\text{out}}\}$. If the outbreak is not detected at all, we set $\text{RT}(t_{\text{out}}, l_{\text{out}})$ to NA.

5.4 Results

We start by comparing the results of the two methods of inference as regards estimation of the posterior $\boldsymbol{\psi}$. In Fig. 7, we show two contourplot examples for two randomly selected simulation scenarios. Ideally one should use the INLA method for performing full Bayesian inference, but as our goal was the creation of an algorithm also suitable for routine use, the asymptotic method being 15 times faster made the computation time of the INLA approach intractable for the simulation study. In Section 5, we also compare the alarm threshold of the two methods for the *S.* Newport data.

In a first study of the efficiency of the algorithm, we explore the specificity of the new algorithm and thus only simulate time series with no outbreak. We generate 10 time series for each of the 42 scenarios and monitor each of them over one year, that is, 52 weeks. Our goal with this study is to explore the FPR of our method compared to the same algorithm without reporting delays correction and compared to the established Noufaily algorithm. We performed the exploration of the FPR of our algorithm with $\alpha = 0.05$. As the algorithm proposed by Noufaily et al. (2013) uses a $(1 - \alpha/2)$-quantile, we used $\alpha = 0.1$ for this algorithm.

The results are shown in Fig. 8 with the mean of the FPR obtained over all 420 time series, and the standard deviation of the mean of FPR over the 42 scenarios, for each possible value of $T - s$. In Fig. 8, one sees that the FPR does vary with the scenario. This may be due to the fact that how well the parameters are estimated depends on the parameters values (Lloyd-Smith, 2007). As expected, the new method adjusting for delay ($D = 10$) has a higher FPR for small values of $T - s$. The FPR

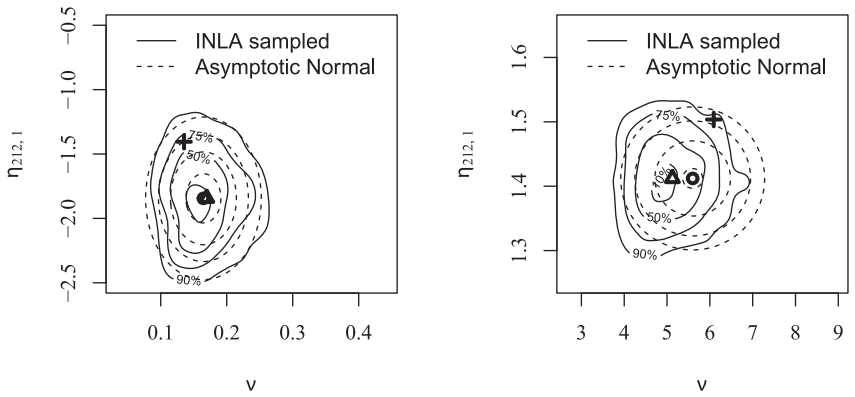

Figure 7 Contourplots of the sampled posterior for $(\eta_{t,d}, \nu)'$ obtained from the negative binomial model for the 212th timepoint of a time series simulated with the scenarios (a) 12 and (b) 25 of the simulation study and $d = 1$, with data available as of the 350th timepoint of the time series. True values of the parameters are represented with a cross, while the mean values of the posterior distributions of the parameters are represented with a triangle and a circle, respectively, for the INLA method and the asymptotic method.

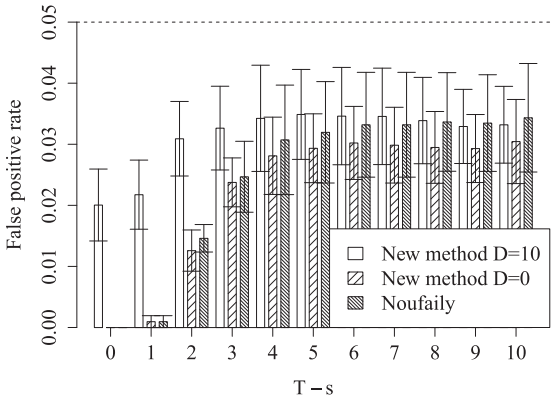

Figure 8 FPR for the 11 possible values of $T - s \in \{0, \ldots, D\}$ for the three algorithms tested and using $\alpha = 0.05$. The horizontal bars represent the mean of the FPR $\pm$ its standard deviation over the 42 scenarios.

of the two other methods is smaller for these small values of $T - s$ because the threshold of the two other methods, defined for complete counts, is nearly always too high for having false alarms on very incomplete counts. For all three methods, the average FPR is never higher than the nominal level.

In a second series of simulations, we investigated how the three methods compare at detecting outbreaks. For this, we again use $\alpha = 0.05$. We generated 420 time series, 10 per scenario, and added outbreaks as in Noufaily et al. (2013) but with only three outbreaks in the baseline of size coefficient $\sigma \in \{2, 3, 5, 10\}$, and an outbreak in the monitored weeks with size coefficient $\sigma \in \{1, \ldots, 10\}$. The starting dates of the outbreaks were randomly drawn. We only monitored the weeks where the outbreak was and record whether an alarm was produced by the algorithm and if so for which s and which T.

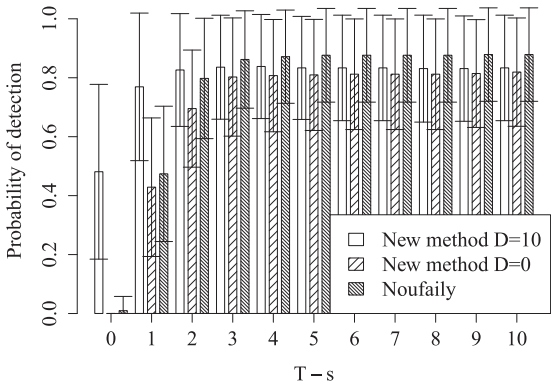

Figure 9 POD for the 11 possible values of $T - s \in \{0, \ldots, D\}$ for the three algorithms tested and using $\alpha = 0.05$. The horizontal bars represent the mean of the POD $\pm$ its standard deviation over the 42 scenarios.

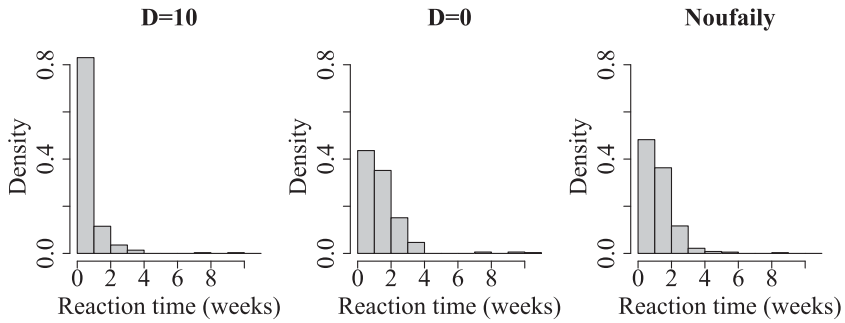

Figure 10 Histogram of the reaction time RT for the three compared methods measured on 420 time series with simulated outbreaks. The *RT* can only be calculated if the outbreak was indeed detected. The medians for the three compared methods are, respectively, 1, 2 and 2 weeks.

Figure 10 shows the RT for the three methods. One sees that outbreaks are detected earlier with the proposed algorithm with $D = 10$ than with the other methods. The associated POD is represented in Fig. 9. With the delay distribution we use, after three weeks the new method with $D = 10$ offers no advantage to the algorithm developed by Noufaily et al. (2013). Note that with the chosen delay distribution, after three weeks about 76% of the cases have already been reported.

6 Application to *S.* Newport

As an illustration of how the proposed algorithms operate, we applied them to the time series of *Salmonella* Newport cases shown in Section 2. We used data available as of $T =$ W43-2011, W44-2011, W45-2011 (shown in Fig. 11), with cases aggregated by date of disease onset as in Fig. 4. We present the results for the method of Noufaily et al. (2013) and for the proposed algorithm with $D = 10$ for the two inference methods. When using the method of Noufaily et al. (2013), the threshold is not adjusted for delay and does not change depending on the observation time: for each observation time T we get a new threshold value for $s = T$ while the threshold values for monitored timepoints $s < T$ do not change. In contrast, the threshold computed by delay-adjusting algorithms

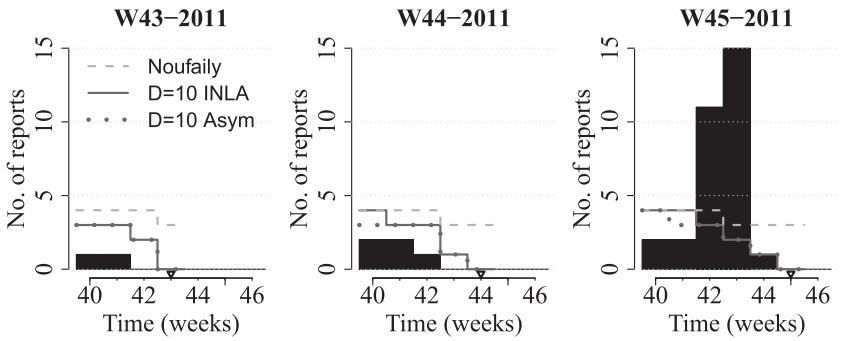

Figure 11 Weekly counts $N(s, T)$ of *Salmonella* Newport cases in the RKI database aggregated by date of disease onset at different observation times $T =$ W43-W45 2011 during the 2011 outbreak and for $s =$ W40-46 2011. The three plots correspond to different values of T whose position is indicated on the x-axis by a black triangle. We present the results for the Noufaily method and for the proposed method with $D = 10$ using both inference methods.

($D = 10$) can change for all timepoints $s \leq T$: in Fig. 11 one sees for instance that when $T =$ W43-2011 the threshold for week $s =$ W42-2011 is 2 with both inference methods, and when $T =$ W44-2011 it is 3 with both inference methods. In other words, the threshold was adapted to the number of cases we could have expected to see for this week ($s =$ W42-2011) *by then* ($T =$ W43-2011, $T =$ W44-2011). Adjusting the threshold for right truncation could make the alarm being sound sooner: a one week earlier detection could already make a difference for an outbreak.

A slightly disappointing result is that for this specific outbreak, the new methods adjusting for delay do not provide an earlier alarm: for all three methods considered the first time that the number of observed cases is higher than the threshold is week 45. The number of cases reported during week 45 was so high that both N(W42-2011, W45-2011) and N(W43-2011, W45-2011) were above the threshold computed by all investigated methods, whereas N(W42-2011, W44-2011) and N(W43-2011, W44-2011) were not above the threshold computed by any of the investigated methods—even the ones adjusting for delay. Back in 2011, the outbreak was manually detected by the National Reference Centre for *Salmonella* because of a cluster of *S.* Newport isolates from a clinic in Northern Germany, not by automatic outbreak detection. These algorithms were not as thoroughly implemented then as compared to now. Hence, the example shows more the virtue of automatic detection and not so much the added value of adjustment for reporting delays. Furthermore, the peak was huge so that despite reporting delays the number of cases was quickly high enough to be detected by any algorithm. However, with another pattern of cases diagnosis and reporting for an outbreak, delay adjustment could have made the difference for getting an earlier alarm. The present example thus shows that the usefulness of the method depends on the pattern of cases reporting, on the form of the outbreak but also on the baseline counts.

As regards the comparison of the two inference methods for $D = 10$, the two thresholds computed are very similar. On our computer the inference using INLA needed 170 s for providing the results for $T =$ W45-2011, whereas the other method only needed 11 s, which is about 15 times faster. This makes the implementation of the algorithm based on the asymptotic approximation more realistic for practical routine use.

7 Discussion

In this article, we introduced a novel regression-based statistical algorithm for aberration detection in public health surveillance data if reporting delays are present. For doing so, our method corrects the

alarm threshold, rather than the observed counts, which we think is a better way of communicating the delay adjustment. Moreover, we developed our method in a Bayesian framework, so that the derivation of the decision-supporting threshold encompasses all sources of uncertainty. Finally, the proposed algorithm was implemented in the R package "surveillance" (Höhle, 2007; Salmon et al., 2015) as the function `bodaDelay`.

In our simulation study, the suggested method does detect outbreaks earlier without producing more false alarms than the nominal level. However, further work remains to be done regarding the application of our algorithm in practice, including tests at a public health institute. If adjusting aberration detection for reporting delays in routine use, for each time unit of aggregation (week, day) to be monitored one could get many alarms in a row. This can be quite cumbersome: each time unit is monitored until no new case is reported for this week, that is, $D + 1$ times. We reckon that one must develop an efficient routine so that the alarms that have already been marked once do not lead to user fatigue, while at the same time accounting for the possible increase of evidence for an aberration. Lastly, a possible development could be a correction for delay at the daily level, but this would make the workload even higher, and a solution for dealing with day-of-the-week effects remains to be found.

The use of a Bayesian framework offered a very flexible framework since the right truncation can be viewed as a missing data problem and there is no fundamental distinction between parameters and data in Bayesian settings. This would also allow for the integration of more flexible delay distribution models in case this is needed. In the future, we aim at exploring two features of such models: how feasible and how much an improvement it is to integrate time-dependent delay distributions, and how the maximal delay chosen and the number of classes in the multinomial distribution of delay impact the results. Furthermore, we currently do not try to online identify aberrant past counts, which is a feature one could later add for increasing the POD. In Manitz and Höhle (2013) this was handled by having an outbreak indicator in the model, for which one has to either know which time units contained outbreaks, or to spot outliers. Since one rarely knows all past outbreaks, one could spot outliers based on the predictive mid-p value instead. This measure was recommended for count data in Held et al. (2010).

It is worth pointing out that reporting delays do not only include cases being notified later but also information such as pathogen subtype to be informed later than other case details. We think that including such features of case reports, possibly by the use of multistate models, would be yet another step in bringing a more statistical modeling perspective into the data transmission mechanisms underlying surveillance systems.

Acknowledgments The work of M.S. was financed by a PhD grant of the RKI. We thank S. Behnke, M. Faber, B. Rosner, and H. Wilking for valuable discussions about reporting delays and aberration detection at the RKI. Furthermore, we are grateful to Håvard Rue for his comments and suggestions on the use of INLA.

Conflict of interest
The authors have declared no conflict of interest.

References

Altmann, M., Wadl, M., Altmann, D., Benzler, J., Eckmanns, T., Krause, G., Spode, A. and an der Heiden, M. (2011). Timeliness of surveillance during outbreak of shiga toxin-producing *Escherichia coli* infection, Germany, 2011. *Emerging Infectious Diseases* **17**, 1906–1909.

Bayer, C., Bernard, H., Prager, R., Rabsch, W., Hiller, P., Malorny, B., Pfefferkorn, B., Frank, C., de Jong, A., Friesema, I., Stark, B. and Rosner, B. M. (2014). An outbreak of *Salmonella* Newport associated with mung bean sprouts in Germany and the Netherlands, October to November 2011. *Euro Surveillance* **19**, 1–9.

Brookmeyer, R. and Damiano, A. (1989). Statistical methods for short-term projections of AIDS incidence. *Statistics in Medicine* **8**, 23–34.

Buckeridge, D. L., Burkom, H., Campbell, M., Hogan, W. R. and Moore, A. W. (2005). Algorithms for rapid outbreak detection: a research synthesis. *Journal of Biomedical Informatics* **38**, 99–113.

Conesa, D., Martínez-Beneito, M., Amorós, R. and López-Quılez, A. (2015). Bayesian hierarchical Poisson models with a hidden Markov structure for the detection of influenza epidemic outbreaks. *Statistical Methods in Medical Research* **24**, 206–223.

Donker, T., Boven, M., Ballegooijen, W., Klooster, T., Wielders, C. and Wallinga, J. (2011). Nowcasting pandemic influenza A/H1N1 2009 hospitalizations in the Netherlands. *European Journal of Epidemiology* **26**, 195–201.

England, P. and Verrall, R. (2002, 8). Stochastic claims reserving in general insurance. *British Actuarial Journal* **8**, 443–518.

Faensen, D., Claus, H., Benzler, J., Ammon, A., Pfoch, T., Breuer, T. and Krause, G. (2006). SurvNet@RKI—a multistate electronic reporting system for communicable diseases. *Euro Surveillance* **11**, 100–103.

Farrington, C., Andrews, N., Beale, A. and Catchpole, M. (1996). A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **159**, 547–563.

Gergonne, B., Mazick, A., O'Donnell, J., Oza, A., Cox, B., Wuillaume, F., Kaufman, Z., Virtanen, M., Green, H., Hardelid, P., Andrews, N., Pebody, R., Holmberg, M., Detsis, M., Danis, C., Uphoff, H., Josseran, L., Fouillet, A., Nunes, B., Nogueira, P., Junker, C., van Asten, L., van Klooster, T., Fernando, S., Flores, V. M., Tomsic, S., Spiteri, G., Nielsen, J. and Mølbak, K. (2011). Work package 7 report: a European algorithm for a common monitoring of mortality across Europe. Technical report, Statens Serum Institut, Copenhagen.

Heisterkamp, S. H., Dekkers, A. L. M. and Heijne, J. C. M. (2006). Automated detection of infectious disease outbreaks: hierarchical time series models. *Statistics in Medicine* **25**, 4179–4196.

Held, L. and Sabanés Bové, D. (2014). *Applied Statistical Inference*. Springer, Berlin/Heidelberg, DE.

Held, L., Schrödle, B. and Rue, H. (2010). *Posterior and Cross-validatory Predictive Checks: A Comparison of MCMC and INLA*. Statistical Modelling and Regression Structures. Physica-Verlag, Heidelberg, pp. 91–110.

Hess, K. T. and Schmidt, K. D. (2002). A comparison of models for the chain-ladder method. *Insurance: Mathematics and Economics* **31**, 351–364.

Höhle, M. (2007). surveillance: an R package for the monitoring of infectious diseases. *Computational Statistics* **22**, 571–582.

Höhle, M. and an der Heiden, M. (2014). Bayesian nowcasting during the STEC O104:H4 outbreak in Germany, 2011. *Biometrics* **70**, 993–1002.

Jones, G., Le Hello, S., Jourdan-da Silva, N., Vaillant, V., de Valk, H., Weill, F.-X., and Le Strat, Y. (2014). The French human Salmonella surveillance system: evaluation of timeliness of laboratory reporting and factors associated with delays, 2007 to 2011. *Euro Surveillance* **19**, 1–10.

Kalbfleisch, J. D. and Lawless, J. F. (1989). Inference based on retrospective ascertaintment: an analysis of the data on tranfusion related AIDS. *Journal of the American Statistical Association* **84**, 360–372.

Krause, G., Altmann, D., Faensen, D., Porten, K., Benzler, J., Pfoch, T., Ammon, A., Kramer, M., and Claus, H. (2007). SurvNet electronic surveillance system for infectious disease outbreaks, Germany. *Emerging Infectious Diseases* **10**, 1548–1555.

Lawless, J. (1994). Adjustments for reporting delays and the prediction of occurred but not reported events. *Canadian Journal of Statistics* **22**, 15–31.

Lin, H., Yip, P. S. F., and Huggins, R. M. (2008). A double-nonparametric procedure for estimating the number of delay-reported cases. *Statistics in Medicine* **27**, 3325–3339.

Lloyd-Smith, J. O. (2007). Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. *PloSone* **2**, e180.

Manitz, J. and Höhle, M. (2013). Bayesian outbreak detection algorithm for monitoring reported cases of campylobacteriosis in Germany. *Biometrical Journal* **55**, 509–526.

Martínez-Beneito, M. A., Conesa, D., López-Quílez, A. and López-Maside, A. (2008). Bayesian Markov switching models for the early detection of influenza epidemics. *Statistics in Medicine* **27**, 4455–4468.

Midthune, D. N., Fay, M. P., Clegg, L. X., and Feuer, E. J. (2005). Modeling reporting delays and reporting corrections in cancer registry data. *Journal of the American Statistical Association* **100**, 61–70.

Noufaily, A., Enki, D. G., Farrington, P., Garthwaite, P., Andrews, N. and Charlett, A. (2013). An improved algorithm for outbreak detection in multiple surveillance systems. *Statistics in Medicine* **32**, 1206–1222.

Noufaily, A., Farrington, P., Garthwaite, P., Enki, D. G., Andrews, N. and Charlett, A. (2014). Detection of infectious disease outbreaks from laboratory data with reporting delays. *The Open University technical reports*, Milton Keynes.

Noufaily, A., Ghebremichael-Weldeselassie, Y., Enki, D. G., Garthwaite, P., Andrews, N., Charlett, A. and Farrington, P. (2015). Modelling reporting delays for outbreak detection in infectious disease data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **178**, 205–222.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**, 319–392.

Rue, H., Martino, S., Lindgren, F., Simpson, D., Riebler, A., and Krainski, E. T. (2015). *INLA: Functions Which Allow to Perform Full Bayesian Analysis of Latent Gaussian Models Using Integrated Nested Laplace Approximaxion*. R package version 0.0-1420281647.

Salmon, M., Schumacher, D., Frank, C., Claus, H. and Höhle, M. (2014). A system for automated outbreak detection of communicable diseases in Germany. In revision.

Salmon, M., Schumacher, D. and Höhle, M. (2015). Monitoring count time series in R: aberration detection in public health surveillance. *Journal of Statistical Software*. Conditionally accepted.

Sánchez-Vargas, F. M., Abu-El-Haija, M. A. and Gómez-Duarte, O. G. (2011). Salmonella infections: an update on epidemiology, management, and prevention. *Travel Medicine and Infectious Disease* **9**, 263–277.

Schmidt, K. D. and Wünsche, A. (1998). Chain ladder, marginal sum and maximum likelihood estimation. *Blätter der Deutschen Gesellschaft für Versicherungs und Finanz Mathematik* **23**, 267–277.

Unkel, S., Farrington, C. P., Garthwaite, P. H., Robertson, C. and Andrews, N. (2012). Statistical methods for the prospective detection of infectious disease outbreaks: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **175**, 49–82.

Zeger, S. L., See, L.-C. and Diggle, P. J. (1989). Statistical methods for monitoring the AIDS epidemic. *Statistics in Medicine* **8**, 3–21.

