2016-08

Detection of infectious disease outbreaks from laboratory data with reporting delays

Noufaily, A

http://hdl.handle.net/10026.1/4622

Detection of Infectious Disease Outbreaks From Laboratory Data With Reporting Delays

Angela Noufaily, Paddy Farrington, Paul Garthwaite*

Department of Mathematics and Statistics, The Open University, UK

and

Doyo Gragn Enki

Peninsula Schools of Medicine and Dentistry,Plymouth University, UK

and

Nick Andrews, Andre Charlett

Public Health England, London, UK

October 5, 2015

Abstract

Many statistical surveillance systems for the timely detection of outbreaks of infectious disease operate on laboratory data. Such data typically incur reporting delays between the time at which a specimen is collected for diagnostic purposes, and the time at which the results of the laboratory analysis become available. Statistical surveillance systems currently in use usually make some ad-hoc adjustment for such delays, or use counts by time of report. We propose a new statistical approach that takes account of the delays explicitly, by monitoring the number of specimens identified in the current and past m time units, where m is a tuning parameter. Values expected in the absence of an outbreak are estimated from counts observed in recent years (typically 5 years). We study the method in the context of an outbreak detection system used in the United Kingdom and several other European countries. We propose a suitable test statistic for the null hypothesis that no outbreak is currently occurring. We derive its null variance, incorporating uncertainty about the estimated delay distribution. Simulations and applications to some test data sets suggest the method works well, and can improve performance over ad-hoc methods in current use.

Keywords: Aberrance, Epidemic, Exceedance, Infection, Quasi-Poisson, Surveillance

1

1   Introduction

Laboratory-based statistical surveillance systems for the detection of outbreaks of infectious disease have been in operation since at least the 1990s, and are in routine use in many countries (Farrington et al., 1996; Hulth et al., 2010). In such systems, an observation originates when a symptomatic individual submits a biological sample (typically of blood, saliva, urine or faeces), which we shall call a specimen. The specimen is then sent off for analysis, which can take days before the causative organism is identified and classified. The identified organism is then reported to the surveillance centre. The time elapsed between the date of specimen and the date of report is the reporting delay.

The duration of the reporting delay depends primarily on the time it takes to identify the infectious organism in the biological sample. This varies according to the complexity of the biological investigations required. For example, the median delay for Norovirus is 11 days, whereas for uncommon salmonella serotypes it can be 3 weeks, owing to the detailed serotyping required (Noufaily et al., 2015).

One advantage of obtaining a detailed classification of the causative organism is that cases occurring in an outbreak may be linked. A disadvantage is that a great number of different organism types must be treated separately (Enki et al., 2013). Furthermore, the reporting delay causes a delay in implementing control measures, though there is also merit in not rushing to control what may turn out to be a short-lived problem. As with much else in outbreak detection, the key is to strike an effective balance between the different aspects of a system's performance.

The detection of infectious disease outbreaks is a rapidly evolving field, with new methods continually being developed such as that of Guillou et al. (2014). The many statistical issues and techniques involved have been regularly reviewed (Sonesson and Bock, 2003; Buckeridge et al., 2005; Shmueli and Burkom, 2010; Unkel et al., 2012). None of the methods reviewed deal in any detail with adjustments for reporting delays. In this paper, we focus on the laboratory-based statistical surveillance system used in England, Wales and Northern Ireland (or rUK, rest of the United Kingdom without Scotland) by Public Health England since the early 1990s. This employs a relatively simple Poisson-based regression method (Farrington et al., 1996), which has recently been upgraded (Noufaily et al., 2013),

applied to data from the LabBase database. This system and variants of it is also widely used in Europe (Hulth et al., 2010). It is essentially an elaboration of the Shewhart chart method (Shewhart, 1931). It has proved effective as an adjunct to more traditional methods of surveillance, which rely on the alertness of a network of epidemiologists; see Enki et al. (2013) for examples of outbreaks detected. In this system, as in other laboratory-based systems, analyses are based on counts by date of report. This inevitably causes distortions that may amplify the false positive rate. For example, batching of reports causes spikes in counts by date of report when there may be none by date of specimen. A worthwhile improvement to the system would thus be to switch to analyses based on counts of organisms by date of specimen. This, however, would require an adjustment involving the delay distribution, since the count by current date of specimen is necessarily incomplete in current and recent weeks.

Much of the statistical work on the analysis of reporting delays, and on the estimating disease incidence in the presence of reporting delays, was undertaken in the context of AIDS epidemiology (Brookmeyer and Gail, 1988; Cox and Medley, 1989; Brookmeyer and Damiano, 1989; Brookmeyer and Liao, 1990; Pagano et al., 1994; Green, 1998). New applications and methodological developments nevertheless continue to arise (Midthune et al., 2005; Noufaily et al., 2015). It is convenient to think of the data as arising in two processes: specimen collection with rate $\mu(s)$ and reporting of results with rate $\rho(t)$, linked by the delay distribution, $f(s, u)$ denoting the probability density of a delay $u = t - s$ conditional on a specimen being collected at s. These quantities are linked by the convolution

$$\rho(t) = \int_{-\infty}^{t} \mu(s) f(s, t - s) ds.$$

This relationship, frequently expressed in discrete rather than continuous time, and expanded to accommodate covariates, has been used to model the delay distribution $f(s, u)$ (Brookmeyer and Liao, 1990), or alternatively to infer $\mu(t)$ when the delay distribution is known, using the technique of back-projection (Brookmeyer and Gail, 1988). A common feature of these techniques, however, is that the target parameter is the incidence rate $\mu(s)$, rather than the eventual total but as yet unobserved number of events (in our case, the

3

number of specimens collected) of which $\mu(s)$ is the expectation. Thus, back projection usually involves a degree of smoothing, induced by parametric or step function curves, or non-parametric smoothing (for example, see Rosenberg and Gail (1991) and Becker et al. (1991)). However, when the aim is to detect outbreaks, smoothing must be used sparingly, lest the outbreaks are smoothed out.

Closer to our intended application are the methods of Lawless (1994) for predicting 'occurred but not reported' (OBNR) events, again with application to AIDS surveillance (among others). Translated into our context, his focus is on estimating the current number of specimens collected, based on those reported and knowledge of the delay distribution. This leads to estimators of the form $\hat{N}(s, \infty) = N(s, t)/\hat{F}(t - s)$ where $N(s, t)$ is the number of specimens collected at time s and reported by time $t \geq s$ and $F(u)$ is the cdf of the delay distribution. Related approaches have been proposed by Donker et al. (2011) and Hohle and an der Heiden (2014) to adjust infectious disease outbreak data for reporting delays.

Our application to outbreak detection differs from these in that our aim is not primarily to estimate what the current (but as yet unobserved) total is, but to decide prospectively, on the basis of the current evidence, whether what we observe departs from what one might expect if no outbreak were occurring. To our knowledge, the only attempt to account for reporting delays in outbreak detection is Salmon et al. (2015). This method, based on the algorithm of Noufaily et al. (2013), involves looking back at recent weeks using multiple tests to assess whether an outbreak was occurring. The method thus contains a retrospective element. Our aim is to develop a system that can be used prospectively, based on a single test statistic constructed from recent observations.

To inform the models developed in the present paper, we studied the delay distributions for a dozen pathogens of contrasting characteristics reported in LabBase (Noufaily et al., 2015). We found that the delay distributions have extremely long tails, the hazard of reporting remaining approximately constant in this upper region, though most delays are less than eight weeks. The median delay was typically 2 to 3 weeks. We found evidence that the delay distribution had changed over the 10-year study period for some organisms, in different ways for different organisms. We also found some weak dependence on season and recent frequency. These findings are broadly in keeping with those of other laboratory-

based surveillance systems (Jones et al., 2014).

In Sections 2 to 4, we describe the proposed modelling procedure, the selection of a key tuning parameter m, and estimators for the test statistic and its null variance. In Section 5 we study our proposal by simulations, and in Section 6 we apply it to data from the rUK (Farrington et al., 1996; Noufaily et al., 2013).

2 Analyses by Date of Specimen

We treat the data as arising in discrete time, with weekly time units (other units can be used). Let n_{id} denote the number of isolates of a given organism for which a specimen is collected in week i and reported in week $i + d$, where d is the delay between specimen collection and report of the identified organism. Let $d_{\max}$ denote the maximum delay considered, beyond which reports are deemed to be of little relevance for outbreak detection. Thus, d belongs to the set $\{0, 1, 2, \ldots, d_{\max}\}$, and we shall ignore all isolates with delay greater than $d_{\max}$.

Let N_{id} denote the number of isolates with specimen taken at time i and reported by week $i + d$, that is, reported in weeks $i, i + 1, \ldots, i + d$. Thus, $N_{id} = \sum_{s=0}^{d} n_{is}$. Also let $N_i = N_{i,d_{\max}}$ be the total number of isolates with specimen taken in week i. Let μ_{id} denote the expectation of N_{id}, if there is no outbreak in week i. Similarly, let μ_i denote the expectation of N_i, if there is no outbreak in week i. Thus,

$$\mu_{id} = E\big\{N_{id}|\text{ No outbreak at } i\big\}, \quad \mu_i = E\big\{N_i|\text{ No outbreak at } i\big\}$$

Let $p_d = P(D = d)$ denote the probability mass function of the delays D and $f_d = P(D \leq d) = p_0 + \cdots + p_d$. Later in the paper, we discuss a specific implementation involving the quasi-Poisson algorithm of Farrington et al. (1996) and recently improved versions of it (Noufaily et al., 2013). Our present focus, however, is how to incorporate delays into the detection mechanism.

Suppose that the current week is t. At this point, our observation on N_t is incomplete: we only observe $n_{t0} = N_{t0}$. If counts are Poisson and the delay distribution is multinomial, then $n_{t0}|N_t \sim \text{Bin}(N_t, p_0)$, and so we could estimate N_t by n_{t0}/p_0. However, the variance of this estimator (assuming that p_0 is known) is $N_t(1 - p_0)/p_0$, and if p_0 is close to 0 as will

often be the case, this will be large. Instead, we shall make use of lagged observations at times $t - m, t - m + 1, \ldots, t$, for some suitable, preferably small integer $m \geq 0$, which we shall call the lag. This will reduce the uncertainty resulting from incomplete observation, at the cost of introducing some bias (and hence some delay in detection) at the start of an outbreak. Discussion of the best choice for the lag m is deferred until the next section.

We propose to use the following test statistic T, computed in week t:

$$T = \sum_{s=t-m}^{t} N_{s,t-s} - \nu_t(m)$$

where

$$\nu_t(m) = \sum_{s=t-m}^{t} \mu_{s,t-s} = \sum_{s=t-m}^{t} \mu_s f_{t-s}.$$

The quantity $\nu_t(m)$, here taken as fixed, is assumed to be estimable independently from the $N_{s,t-s}$. The statistic T is the score test statistic for the null hypothesis that no outbreak has begun in the past m weeks under the Poisson model

$$N_s \sim P(\theta \mu_s); \quad s = t - m, \ldots, t,$$

where $\theta > 0$, $\theta > 1$ denoting the presence and $\theta = 1$ the absence of an outbreak (or more accurately, an aberrance). Thus it will be efficient for detecting an outbreak of this form in its $m + 1$th week; inevitably, it will be less efficient for detecting an outbreak before its $m + 1$th week, whence the need to keep the lag m small.

In order to obtain the variance of T it is necessary to make some distributional assumptions about the counts N_i and the delays. We shall assume that N_i has variance $\phi \mu_i$, as might arise from a quasi-Poisson or gamma-Poisson (and hence negative binomial) model. This assumption is supported by empirical evidence (Enki et al., 2013); the dispersion parameter ϕ is estimated using the Pearson chi-square. We also assume that, conditional on N_i, n_{ij} has variance $\psi N_i p_j (1 - p_j)$, $j = 0, 1, \ldots, d_{\max}$ and that for $j \neq k$ the covariance of n_{ij} and n_{ik} is $-\psi N_i p_j p_k$. Thus, the distribution of $(n_{i,0}, \ldots, n_{i,d_{\max}})$ is overdispersed with respect to the multinomial $M(N_i, p_0, \ldots, p_{d_{\max}})$. With these assumptions, we obtain

$$\text{var}(N_{t,d}) = \psi \mu_t f_d + (\phi - \psi) \mu_t f_d^2,$$

6

and hence $\text{var}(T) = \psi\nu_t(m) + (\phi - \psi)\delta_t(m)$, where

$$\delta_t(m) = \sum_{s=t-m}^{t} \mu_s f_{t-s}^2.$$

Note that if $\psi = \phi$, then $\text{var}(N_{t,d}) = \phi\mu_{t,d}$ and so the partial sums $N_{t,d}$ have the same degree of overdispersion as N_t; if $\psi < \phi$ the overdispersion is less. In practice, the effect of the delays is to increase variability, corresponding to $\psi > \phi$.

Mirroring the assumptions made in Farrington et al. (1996) and Noufaily et al. (2013), it will be convenient to replace the counts and their means by their $\frac{2}{3}$ powers, this representing the transformation of a Poisson variate to symmetry, and to use a normal approximation in order to obtain prediction limits, upon which the outbreak detection threshold will be based. In addition, $\nu_t(m)$ has to be estimated. Thus, we shall use

$$T^* = \Big(\sum_{s=t-m}^{t} N_{s,t-s} \Big)^{\frac{2}{3}} - \hat{\nu}_t(m)^{\frac{2}{3}}.$$

With the $\frac{2}{3}$ transformation, the variance of T^* under the null hypothesis is, approximately,

$$\text{var}(T^*) \simeq \frac{4}{9}\nu_t^{\frac{1}{3}}(m)\Big\{\psi + \frac{(\phi - \psi)\delta_t(m) + \text{var}(\hat{\nu}_t(m))}{\nu_t(m)}\Big\}.$$

This expression is derived using first-order Taylor expansions and mirrors that obtained in Farrington et al. (1996), and indeed reduces to it when $m = 0$ and $p_0 = 1$, that is, when there are no reporting delays. The uncertainty in the $\hat{\mu}_s$ is incorporated in the term $\text{var}(\hat{\nu}_t(m)$, the estimation of which will be described in Section 4. To quantify the magnitude of departure from the null hypothesis, we use the following exceedance score:

$$Z^* = \frac{T*}{z_\alpha\big\{\text{var}(T^*)\big\}^{\frac{1}{2}}},$$

where z_α is the $(1 - \alpha)\%$ quantile of the standard normal density.

Throughout, quantities are starred to indicate that they involve the $\frac{2}{3}$ transformation; in Farrington et al. (1996) the exceedance score was defined without this transformation. If $Z^* > 1$, the organism count exceeds the upper $(1 - \alpha)\%$ prediction limit under the null hypothesis. In this case the organism is classified as currently aberrant, and is investigated further to decide whether an outbreak is occurring.

7

3 Choosing the Lag m

How should the lag m be chosen? Clearly, the longer the median reporting delay, the larger m will need to be to achieve a good power to detect an outbreak lasting at least $m + 1$ weeks. On the other hand, the larger m is, the longer it may take to detect the outbreak, since outbreak counts will initially be combined with non-outbreak counts, thus diluting the impact of the outbreak.

Suppose that, at time t, the outbreak is in its $k + 1$th week, where $k \leq m$, with mean count $\theta \mu_s$ for $s = t - k, \ldots, t$, whereas weeks $t - m, \ldots, t - k - 1$ are non-outbreak weeks with mean count μ_s. Then $\sum_{s=t-m}^{t} N_{s,t-s}$ has mean $\nu_t(k, m, \theta)$ where

$$\nu_t(k, m, \theta) = \sum_{s=t-m}^{t-k-1} \mu_s f_{t-s} + \theta \sum_{s=t-k}^{t} \mu_s f_{t-s}.$$

Thus, $\nu_t(k, m, 1) = \nu_t(m)$ for $k \leq m$. For the moment, assume that $\nu_t(m)$ is known, so there is no variability associated with its estimation, and that $\psi = \phi = 1$. The threshold on the $\frac{2}{3}$ power scale for declaring an aberrance at time t is then, approximately,

$$\nu_t(m)^{\frac{2}{3}} + \frac{2}{3} z_\alpha \nu_t(m)^{\frac{1}{6}}.$$

Assume also that $\sum_{s=t-m}^{t} N_{s,t-s}$ has variance $\nu_t(k, m, \theta)$ when an outbreak is occurring, and that the outbreak lasts for at least $m + 1$ weeks. The power for detecting an outbreak of (relative) size θ in its $k + 1$th week at time t is

$$P_t(k+1, m, \theta) = 1 - \Phi\left\{ z_\alpha \left(\frac{\nu_t(m)}{\nu_t(k, m, \theta)} \right)^{\frac{1}{6}} - \frac{3}{2} \frac{\nu_t(k, m, \theta)^{\frac{2}{3}} - \nu_t(m)^{\frac{2}{3}}}{\mu(k, m, \theta)^{\frac{1}{6}}} \right\}.$$

Suppose first that $k = m$. Then $\nu_t(m, m, \theta) = \theta \nu_t(m)$ and so the power is

$$P_t(m+1, m, \theta) = 1 - \Phi\left(\frac{1}{\theta^{\frac{1}{6}}} \left\{ z_\alpha - \frac{3}{2} (\theta^{\frac{2}{3}} - 1) \nu_t(m)^{\frac{1}{2}} \right\} \right).$$

Since $\nu_t(m)$ increases with the lag m, it is clear (if unsurprising) that $P_t(m + 1, m, \theta)$ increases with m when $\theta > 1$. It can also easily be shown that the power increases with θ, as expected. Suppose now that $k = m = 0$. Then $\nu_t(0, 0, 1) = \mu_t p_0$, and so the power drops as p_0 gets smaller. This is why it is advisable, at least when p_0 is small, to use $m > 0$.

In order to separate out the temporality of detection and the probability of detection, we define the detection delay conditionally on detection. Thus, the detection delay is the

8

expected number of weeks until the outbreak is detected, minus 1, conditionally on it having been detected by week $m + 1$. The minus 1 is included so that the detection delay is zero when the outbreak is detected in its first week. We shall assume that the outbreak-free incidence is stationary, and hence drop the subscripts t; for example we replace μ_t by μ. First, let $\pi_{k+1}(m, \theta)$ be the probability of first detecting an outbreak in its $k + 1$th week, $k = 0, \ldots, m$. Thus,

$$\pi_{k+1}(m, \theta) = P_t(k + 1, m, \theta) \prod_{i=1}^{k} \{1 - P_t(i, m, \theta)\}.$$

The overall power to detect the outbreak at any time up to its $m + 1$th week is

$$P(m, \theta) = \sum_{k=0}^{m} \pi_{k+1}(m, \theta),$$

and the detection delay (conditional on detection by outbreak week $m + 1$) is then

$$\tau(m, \theta) = \frac{\sum_{k=0}^{m} k \pi_{k+1}(m, \theta)}{\sum_{k=0}^{m} \pi_{k+1}(m, \theta)}.$$

We illustrate the trade-off between the power to detect the outbreak by its $m + 1$th week and the conditional detection delay, using the above equations with values $\mu = 10$ and $\theta = 2, 3$ and $\mu = 1$, $\theta = 4, 5$, and the delay distribution with probability mass function $(p_0, \ldots, p_4) = (0.15, 0.5, 0.2, 0.1, 0.05)$, which is typical of that observed for many organisms (ignoring the long upper tail). The results are in Table 1.

Table 1: Power and detection delay (τ) for weekly frequencies $\mu = 10$ and $\mu = 1$.

| | $\mu = 10$ | | | | $\mu = 1$ | | | |
| | $\theta = 2$ | | $\theta = 3$ | | $\theta = 4$ | | $\theta = 5$ | |
lag m	Power	τ	Power	τ	Power	τ	Power	τ
0	0.216	0	0.508	0	0.163	0	0.260	0
1	0.708	0.89	0.988	0.82	0.497	0.87	0.717	0.85
2	0.961	1.43	1.000	0.96	0.824	1.53	0.960	1.34
3	0.998	1.70	1.000	1.09	0.968	1.93	0.998	1.57
4	1.000	1.87	1.000	1.21	0.997	2.16	1.000	1.72

As expected, both the power for detecting the outbreak, and the (conditional) detection delays τ increase as the lag m increases. Also as expected, for a given value of m, the power is higher when θ or μ are larger.

These illustrations can be used to get some idea of the performance to be expected from the system. The choice of m ought to ensure that the power to detect an outbreak is acceptable over a useful range of values (μ, θ), with θ reducing as μ increases. The illustrations above suggest that $m = 2$ gives power in excess of 80% in relevant scenarios. We also need to ensure that, for those outbreaks detected, the detection delay is not so long as to render intervention pointless. In these illustrations, the choice $m = 2$ gives a detection delay τ of the same order as the mean reporting delay (1.4 weeks). A detection system with $m = 2$ should therefore have reasonable performance for this and similar delay distributions. The choice of the lag m will be more fully evaluated in simulations, taking into account the uncertainties associated with estimation.

4 Estimators

In order to evaluate the test statistic T^* and its null variance, we need an estimator of ν_t, the expectation of $\sum_{s=t-m}^{t} N_{s,t-s}$ in the absence of an outbreak, and the variance of that estimator. A key requirement is that the method should be robust to changes in the delay distribution over time.

Since reports with delays greater than $d_{\max}$ are rare and irrelevant for outbreak detection, we can assume that the N_i are fully observed at past times $i < t - d_{\max}$. The correct specification of the model for the μ_s is clearly important. We propose to use the quasi-Poisson methods of Farrington et al. (1996) and Noufaily et al. (2013) and regress the baseline counts N_i on time and season at i (assuming that $t - i > d_{\max}$), with seasonal effects represented by Fourier terms rather than constant factors, to obtain estimates of $\hat{\mu}_s$ of μ_s, $s = t - m, \ldots, t$ and of ϕ. The model is as follows:

$$E(N_t) = \mu_t, \qquad \mathrm{var}(N_t) = \phi\mu_t,$$

with

$$\log(\mu_t) = \beta_1 + \beta_2 t + \sum_{r=1}^{K} \left\{ \gamma_{r1} \cos\left(\frac{2\pi t}{52}\right) + \gamma_{r2} \sin\left(\frac{2\pi t}{52}\right) \right\}.$$

10

We shall use $K = 4$ Fourier terms, so the model for μ_t has ten parameters.

To estimate the delay distribution, we assume that it has changed little over the time interval $[t - d_{\text{est}}, t)$ where $d_{\text{est}} > d_{\text{max}}$. This interval is chosen to be short relative to the period over which changes in the delay distribution have been observed empirically (Noufaily et al., 2015). We shall estimate the average delay distribution from data on delays for specimen counts N_s with $s \in [t - d_{\text{est}}, t - d_{\text{max}})$. For a given current week t, set

$$n = \sum_{s=t-d_{\text{est}}}^{t-d_{\text{max}}} N_s$$

and define, for $r \leq d_{\text{max}}$,

$$\hat{p}_r = \frac{1}{n+\epsilon}\Big(\sum_{s=t-d_{\text{est}}}^{t-d_{\text{max}}} N_{sr} + \frac{\epsilon}{d_{\text{max}}+1} \Big), \qquad \hat{f}_r = \hat{p}_0 + \cdots + \hat{p}_r.$$

The ϵ (typically $\epsilon = \frac{1}{2}$) is there to avoid zero estimated delay probabilities.

As explained in section 2, we use a quasi-multinomial model for the delays. Thus, the variance-covariance matrix of the vector $\hat{p}^T = (\hat{p}_m, \hat{p}_{m-1}, \ldots, \hat{p}_0)$ is the $(m+1) \times (m+1)$ matrix ψW, say, where

$$W_{ii} = \frac{p_i(1-p_i)}{n}, \qquad W_{ij} = -\frac{p_i p_j}{n} \quad \text{if } i \neq j$$

and ψ is the dispersion parameter described in the previous section. We use an estimator evaluated at the current time, conditional on $(N_{t-m,m}, \ldots, N_{t-1,1})$. We evaluate the Pearson chi-squared goodness of fit statistic:

$$X^2 = \sum_{k=1}^{m} \sum_{j=0}^{k} \frac{(n_{t-k,j} - N_{t-k,k} p_j / f_k)^2}{N_{t-k,k} p_j / f_k}$$

with $df = \frac{1}{2}m(m+1)$ degrees of freedom. If for some k, $1 \leq k \leq m$, $N_{t-k,k} = 0$, then the corresponding term is omitted from the calculation of X^2 and the degrees of freedom are reduced by k. Then set

$$\psi = \max\Big(\frac{X^2}{df}, \phi\Big).$$

Thus $\psi \geq \phi$ and any overdispersion associated with the delay distribution tends to increase the variability of the counts, as observed in practice. When the data are very sparse, we set $\psi = \phi$. This estimator allows for fluctuations in the delay distribution which may

11

spuriously inflate T^*: in such cases ψ will be large. The purpose of allowing for such fluctuations is to increase the robustness of the system to short-term changes in the delay distribution.

We take $\hat{p}_r$ and $\hat{f}_r$ as the estimators of p_r and f_r. This yields an estimator of ν_t:

$$\hat{\nu}_t = \sum_{s=t-m}^{t} \hat{\mu}_s \hat{f}_{t-s}.$$

It remains to find $\text{var}(\hat{\nu}_t)$. It is easier to work with the p_r rather than the f_r. We can rewrite

$$\nu_t = \sum_{s=1}^{m+1} \gamma_s p_{m+1-s} = \gamma^T p$$

where $\gamma^T = (\mu_{t-m}, \mu_{t-m} + \mu_{t-m+1}, \ldots, \mu_{t-m} + \cdots + \mu_t)$. Also, we can write

$$f^T \text{diag}(\mu) = p^T M$$

where M is the matrix with

$$M_{ij} = \mu_{t-m+j-1} \text{ if } i \geq j, \qquad 0 \quad \text{otherwise.}$$

So the first column of M consists of μ_{t-m}, the second column of a zero followed by μ_{t-m+1}, and the last column is all zeroes except for μ_t in the last position.

We now obtain $\text{var}(\hat{\nu}_t)$. Suppose that the regression of the baseline counts on time and season at time t involves q covariates $x_{t,1}, \ldots, x_{t,q}$ and let X denote the $(m+1) \times q$ design matrix at times $t-m, \ldots, t$. Let $f^T = (f_m, \ldots, f_0)$, and let $\text{diag}(\mu)$ denote the diagonal matrix with diagonal elements $\mu_{t-m}, \ldots, \mu_t$. Let V denote the variance-covariance matrix of the q parameters estimates, and suppose that the model involves the log link

$$\log(\mu_s) = \sum_{j=1}^{q} \beta_j x_{s,j}.$$

Let

$$A = (MX)V(MX)^T.$$

Then we have

$$\text{var}(\hat{\nu}_t | \hat{p}) \simeq \hat{p}^T A \hat{p}.$$

Thus, unconditionally,

$$
\begin{aligned}
\mathrm{var}(\hat{\nu}_t) &= \mathrm{var}\Big(E(\hat{\nu}_t|\hat{p})\Big) + E\Big(\mathrm{var}(\hat{\nu}_t|\hat{p})\Big) \\
&= \mathrm{var}\Big(E(\hat{\gamma}^T\hat{p}|\hat{p})\Big) + E\Big(\mathrm{var}(\hat{\nu}_t|\hat{p})\Big) \\
&\simeq \mathrm{var}\Big(\gamma^T\hat{p}\Big) + E\Big(\hat{p}^T A\hat{p}\Big) \\
&= \psi\gamma^T W\gamma + \psi\,\mathrm{tr}(WA) + p^T Ap
\end{aligned}
$$

where in the last line we have used the identity $E(x^T Ax) = \mathrm{tr}(VA) + \mu^T A\mu$ for a random vector x with mean μ and covariance matrix V. The last term $p^T Ap$ is $\mathrm{var}(\hat{\nu}_t|p)$; the other terms represent the inflation of the variance resulting from estimating the delay distribution. The variance estimator is obtained by substituting estimated quantities throughout the expression for $\mathrm{var}(\hat{\nu}_t)$. Likewise we obtain $\mathrm{var}(T^*)$.

5 Simulations

In this section, we check by simulations the proposed methods for incorporating reporting delays in outbreak detection, for a range of settings relevant to the rUK LabBase data.

5.1 Generating the data

We generated data over a total 322 weeks, where weeks 1 to 310 are used as baseline weeks, and weeks 311 to 322 are 'current' weeks, which may include outbreaks; the detection algorithm will be applied to these current weeks iteratively. We set the maximum delay to be $d_{max} = 25$ weeks and ignored all specimens reported after a delay of 25 weeks. A delay of 0 weeks means that the organism is reported in the same week that the specimen is taken.

We based the delay distribution on a discretized version of the Weibull density

$$
f(d) = \frac{k}{\eta} d^{k-1} e^{-d^k/\eta},
$$

where the delay $d > 0$, $\eta > 0$ is a scale parameter and $k > 0$ is a shape parameter. The mode is then $(\eta(k-1)/k)^{(1/k)}$. The distribution function is $F(d) = 1 - \exp(-d^k/\eta)$ and the hazard function is $h(d) = kd^{(k-1)}/\eta$. For some simulations, we obtained a time-varying

13

hazard by replacing η by $\eta_t = \eta \exp(-\alpha(t-322))$, where t is the week of specimen collection, and α governs the change over time. The relative hazard over a 1-year period (with $k = 2$) is $\delta = \exp(52\alpha)$; expressed in terms of δ, we have $\alpha = \log(\delta)/52$. The year-on ratio of modes is $\delta^{-\frac{1}{2}}$.

We discretized this distribution by setting

$$p_{ij} = (F_i(j+1) - F_i(j))/F_i(26), \text{ for } j = 0, 1, ..., 25 \text{ and } i = 1, 2, ..., 322,$$

where p_{ij} is the probability of a delay j for a specimen taken in week i, and $F_i(d)$ is the cdf of the Weibull with parameters η_i and k.

In a first set of simulations, we assumed there was no seasonality. The total number N_i of specimens collected in week $i = 1 \ldots 310$ (with delay up to 25 weeks) was simulated as $N_i \sim \text{Poisson}(\mu)$ where μ is a pre-specified incidence parameter. For the 12 'current' weeks $i = 311, ..., 322$, we simulated $N_i \sim \text{Poisson}(\mu + b\mu^{1/2})$ where b controls the size of the outbreak (if there is no outbreak, $b = 0$). In addition, we undertook simulations with seasonality in which $N_i \sim \text{Poisson}(\mu_i)$ in weeks $i = 1, \ldots, 310$ with

$$\mu_i = \exp\left\{\theta_1 + \theta_2 \sin\left(\frac{2\pi i}{52} + \theta_3\right)\right\}.$$

The parameters θ_1 and θ_2 were chosen so as to obtain pre-specified maxima and minima of the μ_i. The phase θ_3 was chosen so as to start the outbreak in week 311 at different points of the seasonal cycle. For the 12 current weeks $i = 311, \ldots, 322$, we simulated $N_i \sim \text{Poisson}(\mu_i + b\mu_i^{1/2})$.

Finally, we generated n_{id}, the number of specimens taken in week i and reported in week $i + d$ from the multinomial distribution with index N_i and probabilities $p_{i0},...,p_{i25}$.

5.2 Applying the algorithm

The first step is to estimate the delay distribution from the simulated data in 'recent' weeks (only recent weeks are used to reduce the impact of past variation in the delay distribution). Thus, we use data from some past week S to week 285. We do not use data after week 285 as these are incomplete (we don't have the full delay distribution for these). To ensure

14

that sufficient data are available to estimate the delay distribution, we set $S = \min\{259,$ the latest week s such that the sum of the N_i for i between s and 285 is at least 100$\}$. We then calculate

$$\hat{p}_j = \Big(\sum_S^{285} n_{ij} + \frac{\epsilon}{26} \Big) / \Big(\sum_S^{285} N_i + \epsilon \Big)$$

with $\epsilon = \frac{1}{2}$ and set $\hat{f}_j = \hat{p}_0 + \hat{p}_1 + ... + \hat{p}_j$ (with $\hat{f}_0 = \hat{p}_0$).

Next, we select the lag m; in fact, we will only actually need $\hat{p}_j$ for $j = 0, 1, .., m$. We compute the response in the 'current' week (which iterates from week 311 to week 322) as the total number of organisms with specimen date in the current week and the m weeks preceding it. For example, if the current week is week 311 and the lag is $m = 2$, the response is $y_{311} = n_{311,0} + (n_{310,1} + n_{310,0}) + (n_{309,2} + n_{309,1} + n_{309,0})$. If for the 'current' week i, $y_i = 0$, we conclude straight away that there is no outbreak in week i, and that the current value does not exceed the threshold; there is no need to complete the other steps. If $y_i > 0$, we proceed as follows.

First, we estimate the expected counts (by specimen date) for current week and the m previous weeks. This is done by applying the detection algorithm of Noufaily et al. (2013) just once for the current week; the same run is used to obtain the expected value in the m previous weeks, by adjusting the trend and season covariates. The expected value of the response variable is then obtained by combining the expected counts with the estimated delay distribution.

For example, if the current week is week 311 and $m = 2$, application of the detection algorithm gives the expected counts $\hat{\mu}_{311}$, $\hat{\mu}_{310}$ and $\hat{\mu}_{309}$. The expected value of the response y_{311} is then $\hat{\nu}_{311} = \hat{\mu}_{311}\hat{f}_0 + \hat{\mu}_{310}\hat{f}_1 + \hat{\mu}_{309}\hat{f}_2$. The test statistic is $T^* = y_{311}^{2/3} - \hat{\nu}_{311}^{2/3}$ and its variance is obtained as described in Section 5, using the design matrix X, the dispersion parameter ϕ, the matrix of estimated expected counts M, and the covariance matrix V of the estimated regression parameters obtained from the detection algorithm. In the absence of systematic week-to-week variation, we set $\psi = \phi$. Thus for week 311, we have

$$\mathrm{var}(T^*) \simeq \frac{4}{9}\hat{\nu}_{311}^{\frac{1}{3}}\Big\{ \phi + \frac{\mathrm{var}(\hat{\nu}_{311})}{\hat{\nu}_{311}} \Big\}.$$

The upper threshold U for T^* is then obtained, under the null hypothesis that there is no outbreak, where T^* is approximately normal with mean 0 and variance $\mathrm{var}(T^*)$. In these

simulations we use the 0.99 quantile as the upper threshold in order to keep the number of runs to an acceptable level, unless specified otherwise; in practice a higher threshold may be advisable, especially if many different organisms are investigated, as is the case in LabBase. Thus, in current week 311, the upper threshold is $U = \hat{\nu}_{311}^{2/3} + z_{0.99}\sqrt{\mathrm{var}(T^*)}$. An aberrance is flagged if

$$X = \frac{T^*}{U - \hat{\nu}_{311}^{2/3}} > 1.$$

This procedure is iterated for 'current' weeks 311 to 322.

5.3 Choice of simulation scenarios

The parameter values used to define the simulation scenarios are based on the features of the LabBase data (Enki et al., 2013; Noufaily et al., 2015). For the simulations without seasonality, we considered a wide range of weekly incidences: $\mu = 100, 10, 1, 0.1$, for organisms with short ($k = 2$ and $\eta = 2$) and long ($k = 2$ and $\eta = 8$) reporting delays, assumed to be stationary (so $\delta = 1$). For the simulations with seasonality, we investigated two sets of scenarios: in the first, $\min\{\mu_i\} = 10$ and $\max\{\mu_i\} = 20$, and in the second, $\min\{\mu_i\} = 100$ and $\max\{\mu_i\} = 200$. Guided by the theoretical investigation of Section 3, we used lags of $m = 2$ and $m = 4$. The short reporting delays are typical of organisms such as Norovirus; the longer delays are typical of organisms requiring additional serotyping, such as *Salmonella enterica* serovar abony. We considered situations with no outbreaks ($b = 0$) and outbreaks of different relative sizes ($b = 1, 2, 3, 4$). For the simulations with seasonality, the outbreak was started either at a trough or at a peak. The algorithm was run for the 12 'current' weeks, with 1000 runs for each scenario.

For each scenario, we obtained (a) the proportion of runs for which an aberrance was flagged in that 'current' week (when $b = 0$, this is the weekly false positive rate; when $b > 0$ it is the power); (b) the cumulative proportion of runs for which an aberrance was detected between week 311 and the 'current week' inclusive (these are cumulative false positive rates when $b = 0$ and cumulative powers when $b > 0$); and in the presence of an outbreak ($b > 0$) (c) the mean number of weeks to detection (between 1 and 12 weeks) for outbreaks detected by or including week 322.

We also investigated the impact of moderate variation over time of the delay distribution, as quantified by the year-on relative hazard ($\delta = 0.85(0.05)1.15$), on the weekly false positive rate, that is, the probability of detecting an aberrance when there is no outbreak (that is, when $b = 0$). In these scenarios, we used the variance inflation factor ψ described in section 4 and the 0.995 quantile.

5.4 Simulation results

The results of the simulations with no seasonality are presented in detail for $\mu = 10$ and $\mu = 1$; the detailed results for $\mu = 100$ and $\mu = 0.1$ are available in the Supplementary Materials. Figure 1 shows the false positive rates and powers for $\mu = 1$ and $\mu = 10$, for short reporting delays; Figure 2 shows the corresponding plots for long reporting delays. The power rises rapidly to a plateau in week $m + 1$. Figures 1 and 2 show that good power is achieved for both values of m for short delays with $b \geq 3$, but the higher value of m incurs a cost in speed of detection. Figure 2 shows that, for long delays, too low a value of m results in low power. With these delay distributions, good results are obtained with $m = 2$ for short delays, and with $m = 4$ for long delays. The power curves for $\mu = 100$ are steeper than for $\mu = 10$; those for $\mu = 0.1$ are flatter than for $\mu = 1$ (see Supplementary Materials). The cumulative power of detection by 12 weeks is typically in excess of 99% for outbreaks with $b \geq 2$ in all scenarios, with conditional times to detection typically under 4 weeks (details in Supplementary Materials).
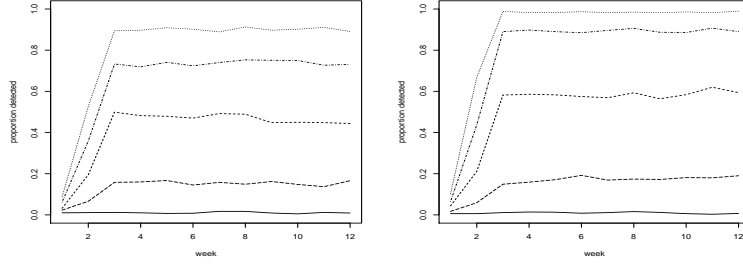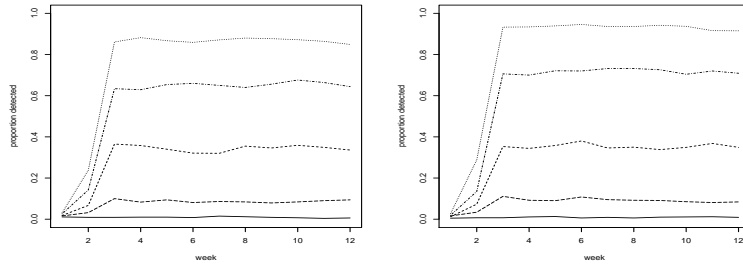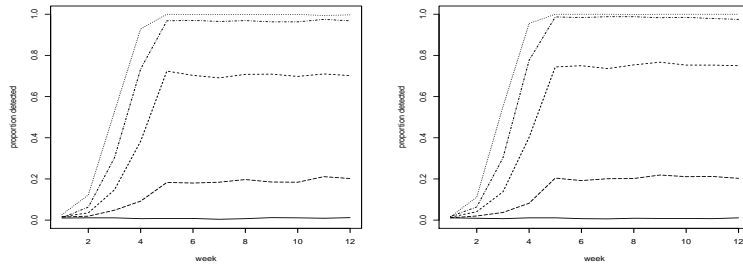
(a) lag $m = 2$
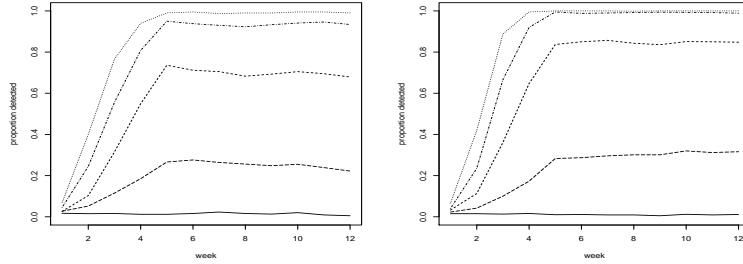
(b) lag $m = 4$

Figure 1: False positive rates (solid lines) and power curves for outbreaks of sizes $b = 1$ (long dashes), 2 (short dashes), 3 (dots and dashes), 4 (dots), with $\mu = 1$ (left) and 10 (right) and short delays.

Table 2 shows the false positive rate for different values of μ when $m = 2$ with short delays. For these calculations, the adjustment involving the computation of ψ was incorporated into the model. The value $\delta = 1$ corresponds to a stationary delay distribution. The false positive rates vary between 0.4% and 0.9%, in line with the nominal value of 0.5%. When $\delta > 1$, corresponding to shortening delays, the false positive rate increases; when $\delta < 1$, corresponding to lengthening delays, the false positive rate decreases. The effect is only apparent when $\mu = 100$. Thus, results are not unduly affected by changing delay distributions unless the weekly incidence is very high.

The results of the simulations with seasonality for outbreaks starting at a peak, for long delays, are shown in Figure 3. The results are similar to those without seasonality. To keep the false positive rate at a constant level, it is necessary to represent seasonality with Fourier terms, rather than piecewise constant terms, in the model for the μ. The results

18
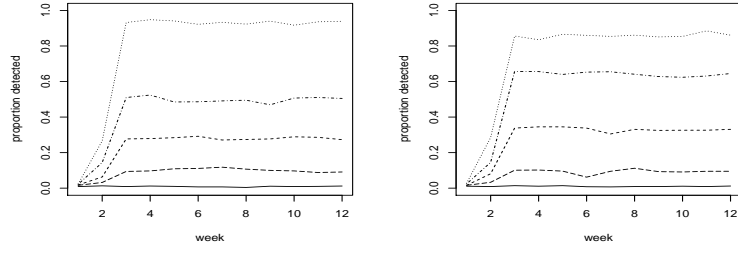
(a) lag $m = 2$
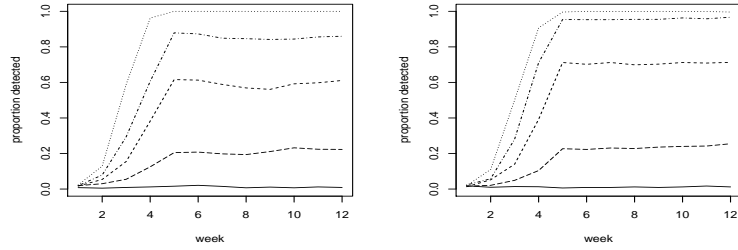
(b) lag $m = 4$

Figure 2: False positive rates (solid lines) and power curves for outbreaks of sizes $b = 1$ (long dashes), 2 (short dashes), 3 (dots and dashes), 4 (dots), with $\mu = 1$ (left) and 10 (right) and long delays.

Table 2: False positive rates for annual relative hazards $\delta = 0.85(0.05)1.15$ and incidence μ, for $m = 2$ and short delays.

	relative hazard δ						
μ	0.85	0.9	0.95	1	1.05	1.10	1.15
100	0.001	0.001	0.002	0.005	0.007	0.012	0.018
10	0.003	0.002	0.004	0.004	0.005	0.003	0.005
1	0.004	0.006	0.006	0.004	0.008	0.006	0.005
0.1	0.003	0.004	0.006	0.009	0.010	0.005	0.009

19

(a) lag $m = 2$

(b) lag $m = 4$

Figure 3: With seasonality: false positive rates (solid lines) and power curves for outbreaks of sizes $b = 1$ (long dashes), 2 (short dashes), 3 (dots and dashes), 4 (dots); outbreaks begin at a peak. μ varying seasonally from 10 to 20 (left) and from 100 to 200 (right), with long delays.

for outbreaks starting at a trough are similar to those for outbreaks starting at a peak; see the Supplementary Materials.

We also undertook simulations with seasonality and short delays, and outbreaks starting either at a peak or at a trough. The results are very similar to those obtained without seasonality; see Supplementary Materials.

6   Implementation for the rUK Data

We applied the new algorithm to the 12 organisms previously studied by Noufaily et al. (2015). The values of m were chosen as the smallest values such that $p_r < 0.10$ for $r > m$. Thus, for the five rare salmonellas (serovars abony, braenderup, brandenburg, infantis and senftenberg) with relatively long delay distributions we used $m = 4$. For the common salmonellas (serovars enteritidis PT 24 and tythimurium DT 104) and *Acinetobacter baumanii* the delays are a little shorter, so we used $m = 3$. For the remaining organisms (*Campylobacter jejuni*, *Chlamydia* sp, *Giardia lamblia* and Norovirus) the delays are shorter still and so $m = 2$.

We used LabBase data by date of report and date of specimen from 2004 to 2011; the data are described in Noufaily et al. (2015) and Enki et al. (2013). We dropped the last 6 months, for which the data by date of specimen are truncated owing to the reporting delays and hence incomplete, and sought to detect aberrations in the last 52 weeks of the remaining data, using the counts in earlier weeks as baselines. First, we ran the standard algorithm (Noufaily et al., 2013) as it is currently used, on the weekly counts by date of report. Second, we ran the same algorithm retrospectively on the weekly counts by week of specimen with reporting delays $d \leq 25$ weeks. Since full information is available (by virtue of the fact that the last 6 months' data are excluded), this analysis is unaffected by reporting delays. Note that this analysis cannot be undertaken prospectively for outbreak detection: our purpose is to obtain a 'gold standard' analysis unaffected by reporting delays. Finally, we ran the new algorithm as if the data were accruing prospectively, and thus subject to reporting delays.

Table 3 shows the number of weeks (out of the 52 investigated) which were flagged by each of the three algorithms. The weeks flagged differ: the standard algorithm (whether

21

run by date of report, or retrospectively by date of specimen) flags individual 1-week spikes, whereas the new algorithm smooths these by taking $m + 1$-week totals and is therefore less likely to detect them. On the other hand, the new algorithm is more likely to detect longer runs of moderately elevated counts.

Table 3: Numbers of weeks flagged as aberrant in 52 weeks' data.

Organism name	Old system run prospectively by report week	Old system run retrospectively by specimen week	New system run prospectively by specimen week
Acinetobacter baumanii	0	0	1
Campylobacter jejeuni	0	14	6
Chlamydia sp	0	0	0
Giardia lamblia	0	1	4
Norovirus	1	0	0
Salmonella abony	2	1	1
Salmonella braenderup	3	4	10
Salmonella brandenburg	1	0	3
Salmonella enteritidis PT21	0	0	0
Salmonella infantis	2	1	1
Salmonella senftenberg	2	0	0
Salmonella typhimurium DT104	1	1	2

Figures 4, 5 and 6 show the time series for three organisms. The top graph shows the time series by date of report with the values flagged prospectively using the standard system as it is currently operated. The middle graph shows the time series N_t (by week of specimen) with the values flagged retrospectively using the standard system. The lower graph shows the time series of $m + 1$-point moving totals, that is the values for each week t of $N_{t-m} + \cdots + N_t$, along with the values flagged prospectively using the new system.

Figure 4 shows the current system by date of report failing to detect a rise in *Campylobacter jejuni*, which is unusual relative to the recent past. The rise is detected both by the system run retrospectively and by the new system; there is evidence of a slight delay in detection using the new system. The failure to detect anything using the current system

is due to the higher variability in the counts by date of report: for example the high peaks close to week 200 are entirely artefactual. Figure 5 shows the current system detecting a peak close to a trough in seasonal Norovirus activity. The peak is not present in the series by week of specimen: it results from batching of reports, and so is spurious. Neither the standard system run retrospectively nor the new prospective system detect anything as there is no genuine outbreak to detect. In Figure 6 all three systems repeatedly flag a large peak in *Salmonella braenderup* incidence. The new system also detects a smaller peak before the main one. Figures for the other nine organisms are in the Supplementary Materials.
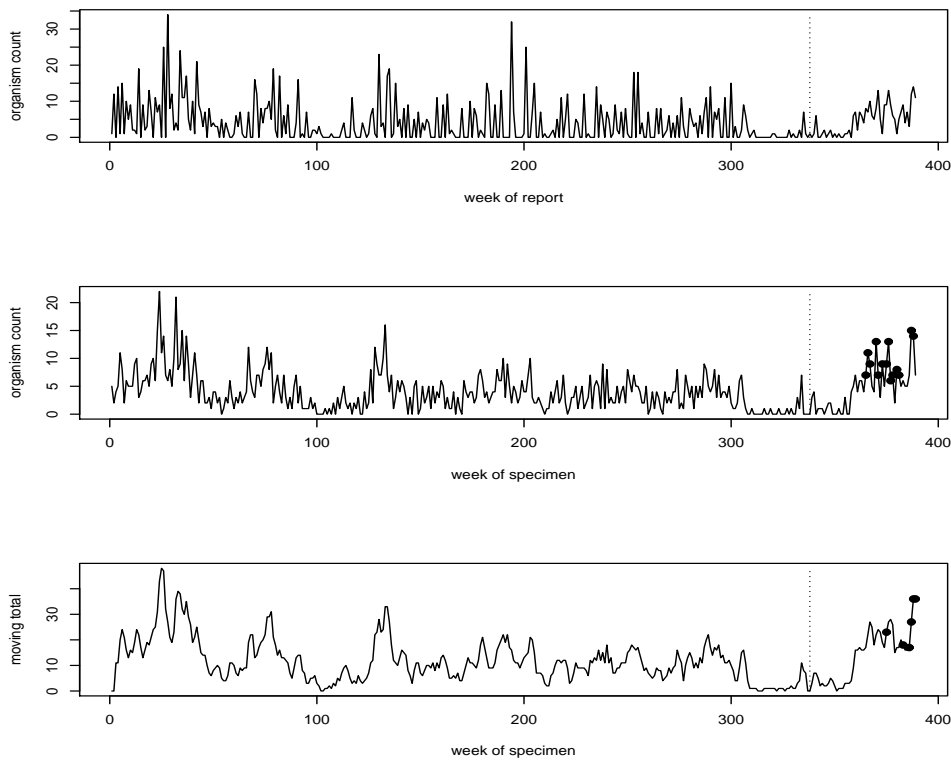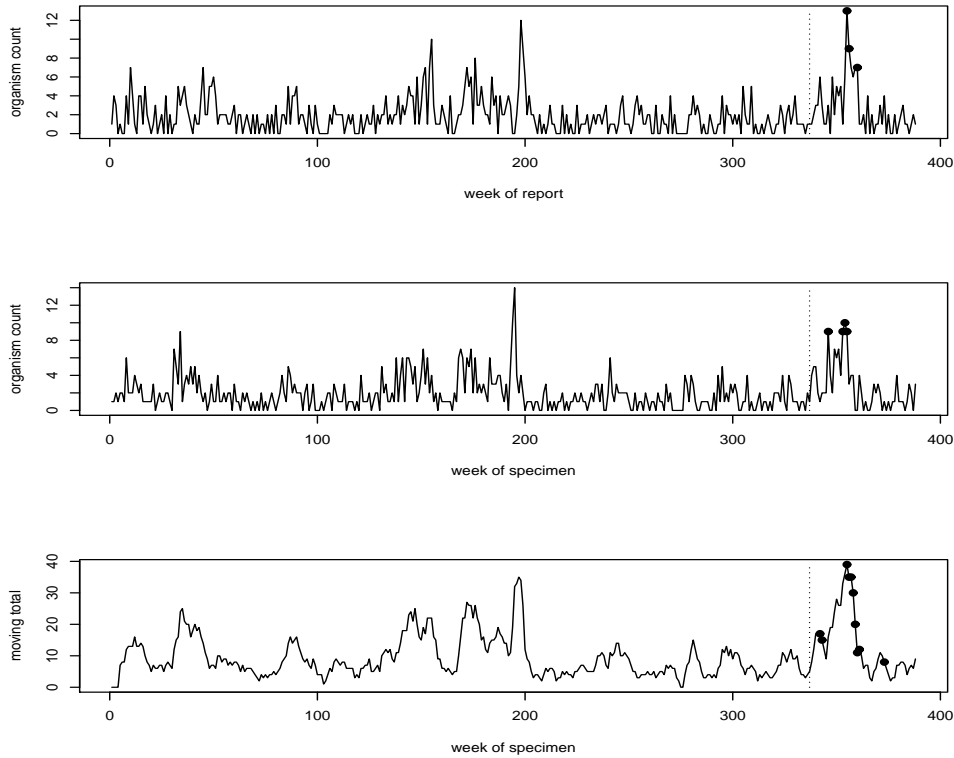

Figure 4: *Campylobacter jejuni*: Time series of counts (full lines) and weeks flagged (dots). Top: standard algorithm applied prospectively by week of report; middle: standard algorithm applied retrospectively by week of specimen (gold standard); bottom: new algorithm applied prospectively, line represents 3-point moving totals (see text). The algorithms were applied to the 52 weeks to the right of the vertical dotted line.

23
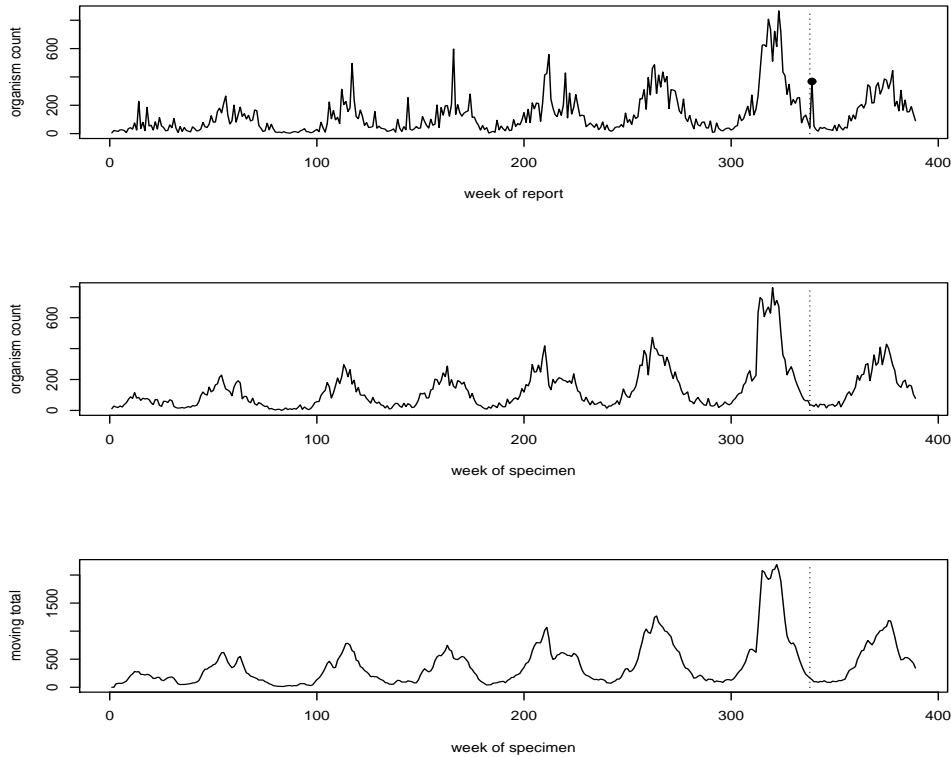
Figure 5: Norovirus: Time series of counts (full lines) and weeks flagged (dots). Top: standard algorithm applied prospectively by week of report; middle: standard algorithm applied retrospectively by week of specimen (gold standard); bottom: new algorithm applied prospectively, line represents 3-point moving totals (see text). The algorithms were applied to the 52 weeks to the right of the vertical dotted line.

7   Discussion

We have sought to adapt a widely used regression-based outbreak detection algorithm to incorporate the effect of reporting delays, which are an unavoidable feature of laboratory-based reporting systems. Our solution is to base detections not just on the current count, but also on the most recent m counts, where m is a small value related to the spread of the reporting delay distribution.

Figure 6: *Salmonella braenderup*: Time series of counts (full lines) and weeks flagged (dots). Top: standard algorithm applied prospectively by week of report; middle: standard algorithm applied retrospectively by week of specimen (gold standard); bottom: new algorithm applied prospectively, line represents 5-point moving totals (see text). The algorithms were applied to the 52 weeks to the right of the vertical dotted line.

The value of m should be chosen in relation to the delay distribution, but should be kept as low as possible to keep the conditional detection delay to a minimum. In our applications, we have found that values of m between 2 and 4 are suitable, based on an ad-hoc rule such that $p_r < 0.10$ when $r > m$; the estimated values of f_m were typically in excess of 80%. A more formal method for choosing m might be desirable, though ultimately this choice involves a trade-off between power and detection delay. This trade-off could also depend on the severity of the infection and the costs of delaying detection, which would most likely vary between infections.

25

Using several weeks' data in this way has the additional benefit of making it less likely to flag 1-week spikes in reports which, whether genuine or not, are of little practical public health interest since there is no opportunity to implement any control measures. The new system, on the other hand, can better detect longer runs of moderately elevated counts. The new system may need a different choice of threshold to accommodate these differences, so as to keep the number of detections to a manageable level. Further experimentation with the number of Fourier terms used to model seasonality (if these are used) may also be advisable.

A key problem in devising the proposed algorithm was how to handle changes in the delay distribution. This is important because the system may flag a report either because the as yet unobserved total number of specimens taken in that week is high, or because the delays are shorter than expected.We dealt with this problem in two ways. First, the delay distribution was estimated from relatively recent data (typically data from the past year). Second, we incorporated an adjustment to account for overdispersion of the currently observed delays relative to the estimated distribution, represented by our ψ parameter. An alternative, but more complicated, approach would be to model changes in the delay distribution explicitly, for example using the method of Brookmeyer and Liao (1990). Our preference for a simple yet robust system was guided by the need to automate the processing of hundreds of different organisms each week.

SUPPLEMENTARY MATERIAL

Supplementary graphs of false positive rate and power, for $\mu = 100$ and $\mu = 0.1$
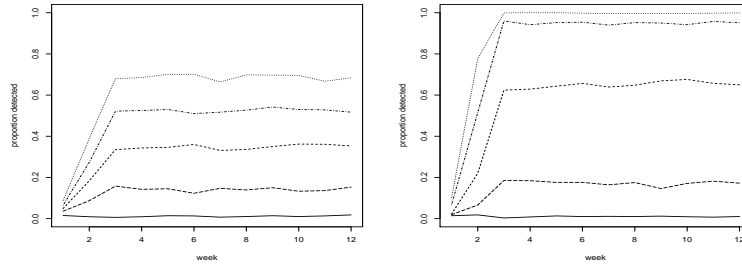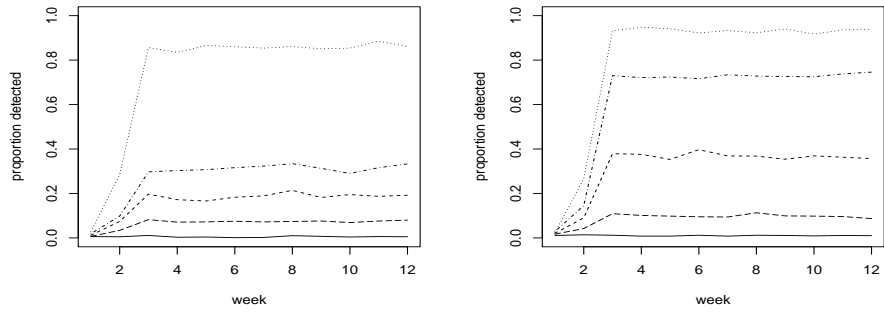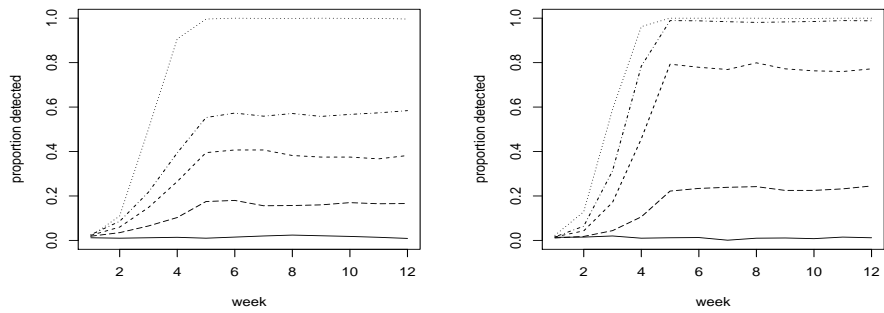

(a) lag $m = 2$
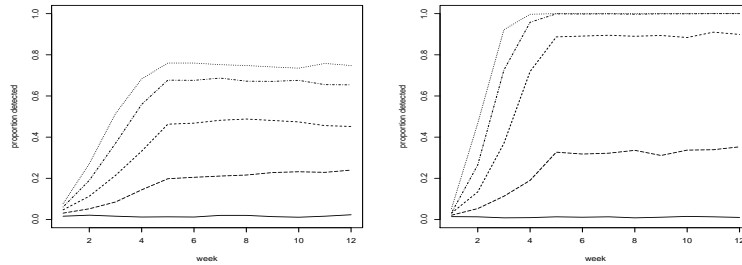
(b) lag $m = 4$

Figure 7: False positive rates (solid lines) and power curves for outbreaks of sizes $b = 1$ (long dashes), 2 (short dashes), 3 (dots and dashes), 4 (dots), with $\mu = 0.1$ (left) and 100 (right) and short delays.

(a) lag $m = 2$

(b) lag $m = 4$

Figure 8: False positive rates (solid lines) and power curves for outbreaks of sizes $b = 1$ (long dashes), 2 (short dashes), 3 (dots and dashes), 4 (dots), with $\mu = 0.1$ (left) and 100 (right) and long delays.

## Cumulative power at 12 weeks and conditional detection delay

Table 4: Cumulative power at 12 weeks and conditional mean time to detection for $\mu = 100$, by outbreak size b, lag m and delay type.

lag m	outbreak size b	short delays		long delays	
		cumulative power	detection delay	cumulative power	detection delay
2	1	0.71	3.77	0.56	4.22
	2	0.99	1.59	0.95	2.77
	3	1	0.47	1	1.30
	4	1	0.12	1	0.80
4	1	0.78	4.17	0.68	4.85
	2	1	1.89	0.99	2.68
	3	1	1.01	1	1.82
	4	1	0.56	1	1.29

Table 5: Cumulative power at 12 weeks and conditional detection delay for $\mu = 10$, by outbreak size b, lag m and delay type.

lag m	outbreak size b	short delays		long delays	
		cumulative power	detection delay	cumulative power	detection delay
2	1	0.72	4.08	0.52	4.23
	2	0.99	1.85	0.93	2.99
	3	1	0.65	1	1.56
	4	1	0.24	1	0.90
4	1	0.73	4.17	0.68	4.69
	2	1	2.12	0.99	3.01
	3	1	1.14	1	1.94
	4	1	0.62	1	1.45

Table 6: Cumulative power at 12 weeks and conditional mean time to detection for $\mu = 1$, by outbreak size b, lag m and delay type.

lag m	outbreak size b	short delays		long delays	
		cumulative power	detection delay	cumulative power	detection delay
2	1	0.65	3.97	0.52	4.21
	2	0.96	2.15	0.89	3.20
	3	1	1.11	0.99	2.01
	4	1	0.54	1	1.46
4	1	0.67	4.17	0.63	4.77
	2	0.98	2.51	0.96	3.23
	3	1	1.40	1	2.19
	4	1	0.82	1	1.70

Table 7: Cumulative power at 12 weeks and conditional mean time to detection for $\mu = 0.1$, by outbreak size b, lag m and delay type.

lag m	outbreak size b	short delays		long delays	
		cumulative power	detection delay	cumulative power	detection delay
2	1	0.59	3.84	0.45	4.53
	2	0.90	2.86	0.75	3.87
	3	0.98	1.96	0.90	3.10
	4	1	1.20	0.96	2.50
4	1	0.62	4.49	0.56	4.66
	2	0.89	3.33	0.86	3.75
	3	0.97	2.32	0.96	3.08
	4	1	1.67	0.99	2.52

Supplementary graphs of false positive rate and power, with seasonality
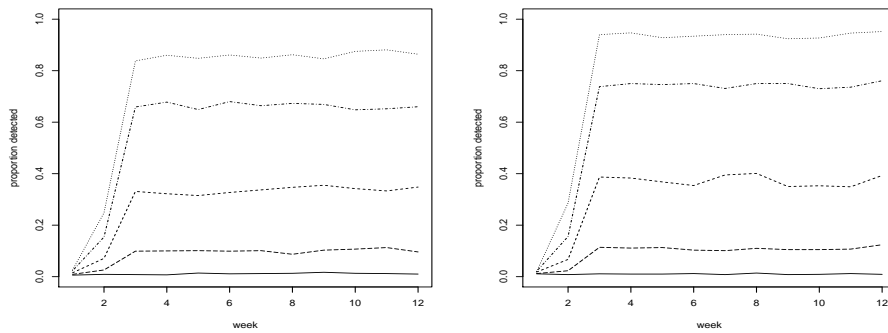
(a) lag $m = 2$
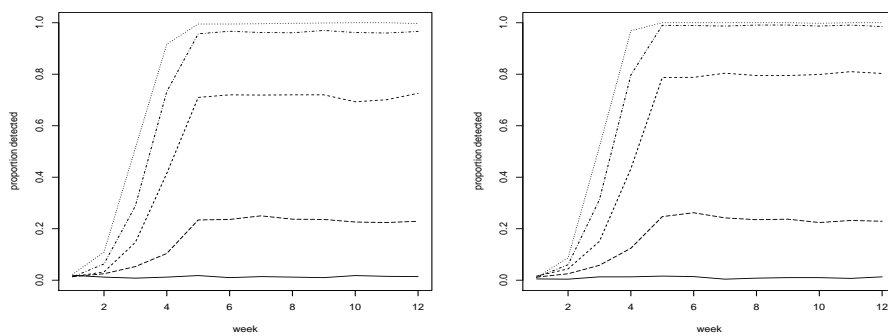
(b) lag $m = 4$

Figure 9: With seasonality: false positive rates (solid lines) and power curves for outbreaks of sizes $b = 1$ (long dashes), 2 (short dashes), 3 (dots and dashes), 4 (dots); outbreaks begin at a trough. μ varying seasonally from 10 to 20 (left) and from 100 to 200 (right), with long delays.
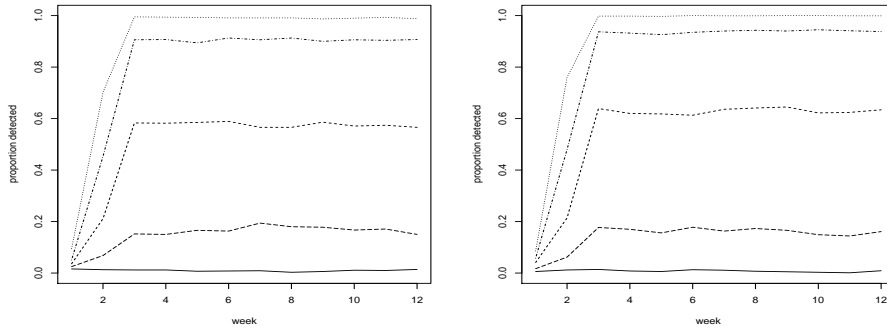
(a) lag $m = 2$
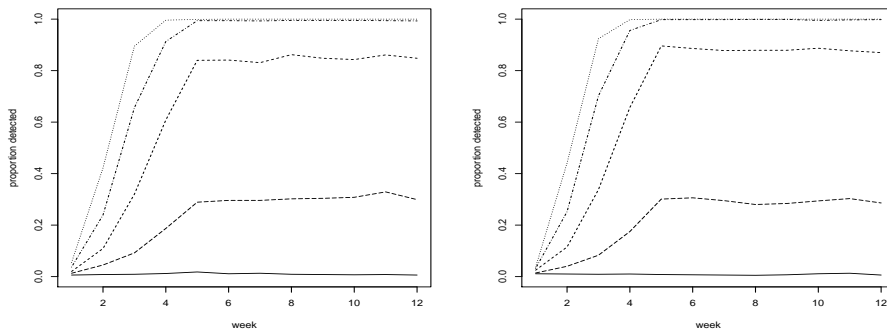
(b) lag $m = 4$

Figure 10: With seasonality: false positive rates (solid lines) and power curves for outbreaks of sizes $b = 1$ (long dashes), 2 (short dashes), 3 (dots and dashes), 4 (dots); outbreaks begin at a peak. μ varying seasonally from 10 to 20 (left) and from 100 to 200 (right), with short delays.
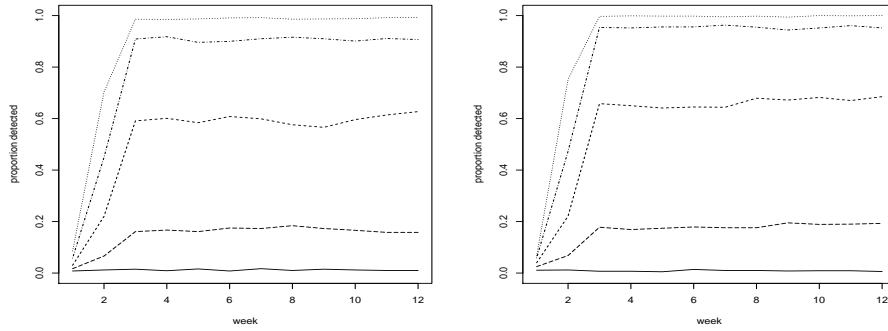
(a) lag $m = 2$
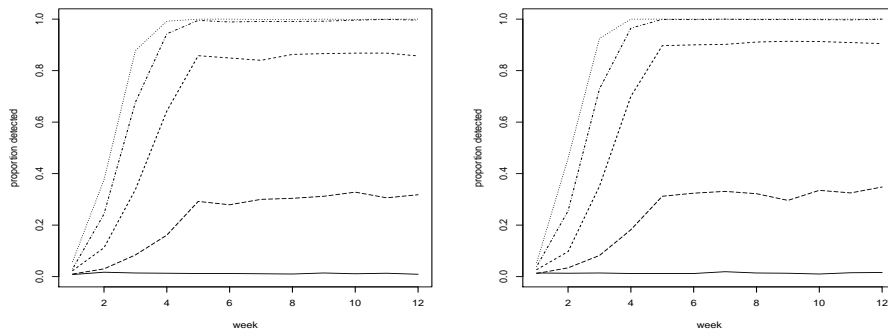
(b) lag $m = 4$

Figure 11: With seasonality: false positive rates (solid lines) and power curves for outbreaks of sizes $b = 1$ (long dashes), 2 (short dashes), 3 (dots and dashes), 4 (dots); outbreaks begin at a trough. μ varying seasonally from 10 to 20 (left) and from 100 to 200 (right), with short delays.

Supplementary graphs for 9 organisms
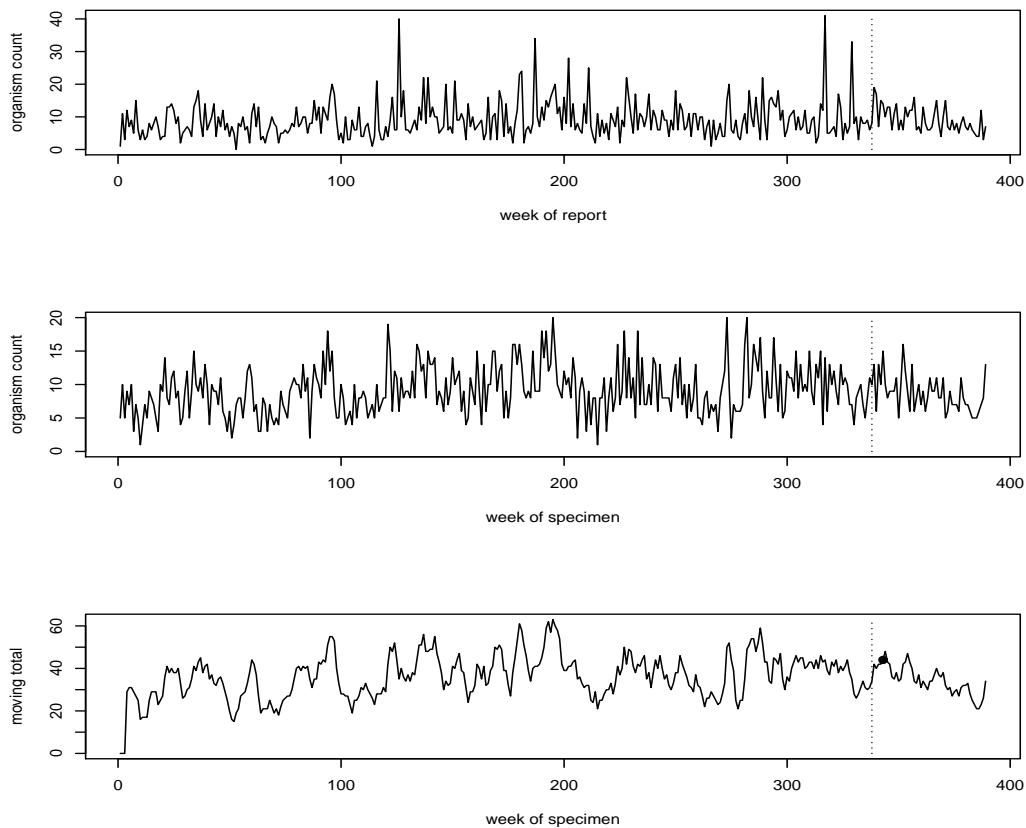
Figure 12: *Acinetobacter baumanii*: Time series of counts (full lines); no weeks were flagged. Top: standard algorithm applied prospectively by week of report; middle: standard algorithm applied retrospectively by week of specimen (gold standard); bottom: new algorithm applied prospectively, line represents 4-point moving totals (see main text). The algorithms were applied to the 52 weeks to the right of the vertical dotted line.
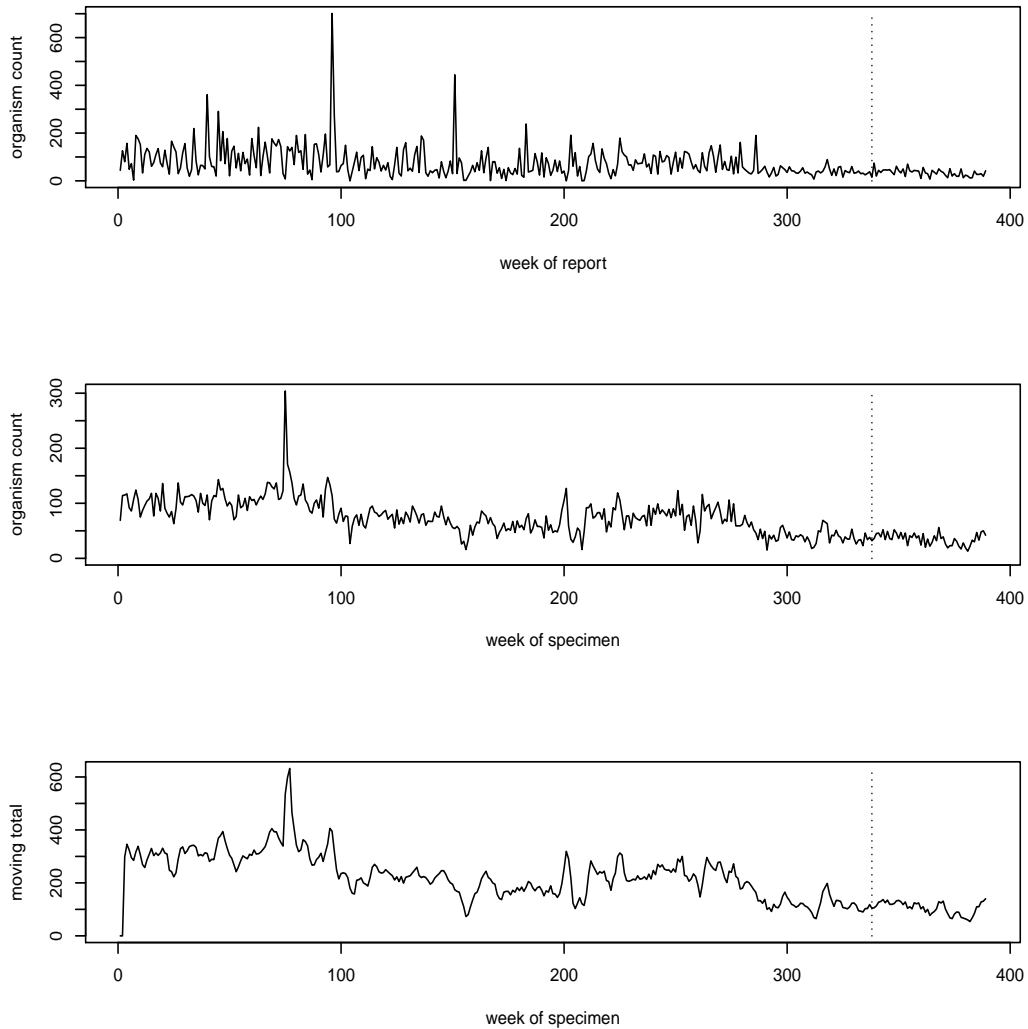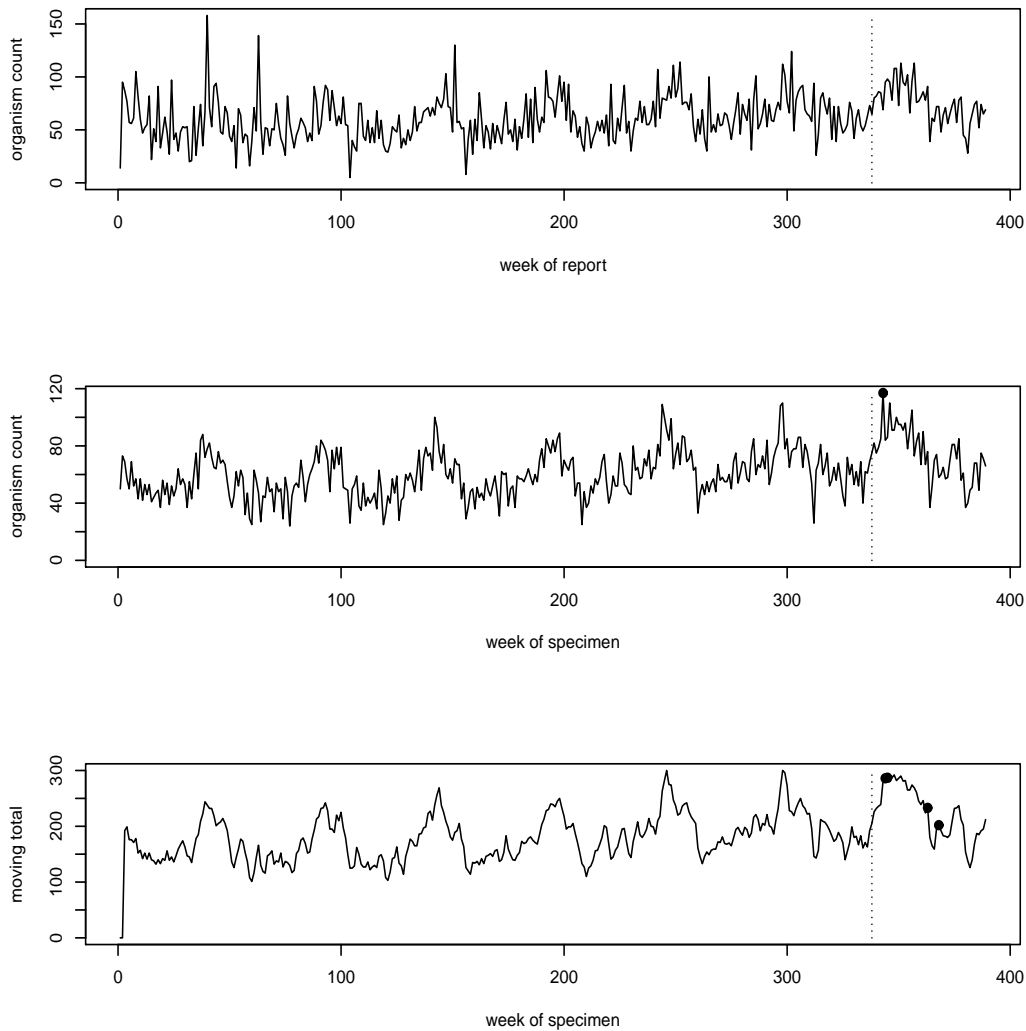
Figure 13: *Chlamydia* sp: Time series of counts (full lines); no weeks were flagged. Top: standard algorithm applied prospectively by week of report; middle: standard algorithm applied retrospectively by week of specimen (gold standard); bottom: new algorithm applied prospectively, line represents 3-point moving totals (see main text). The algorithms were applied to the 52 weeks to the right of the vertical dotted line.
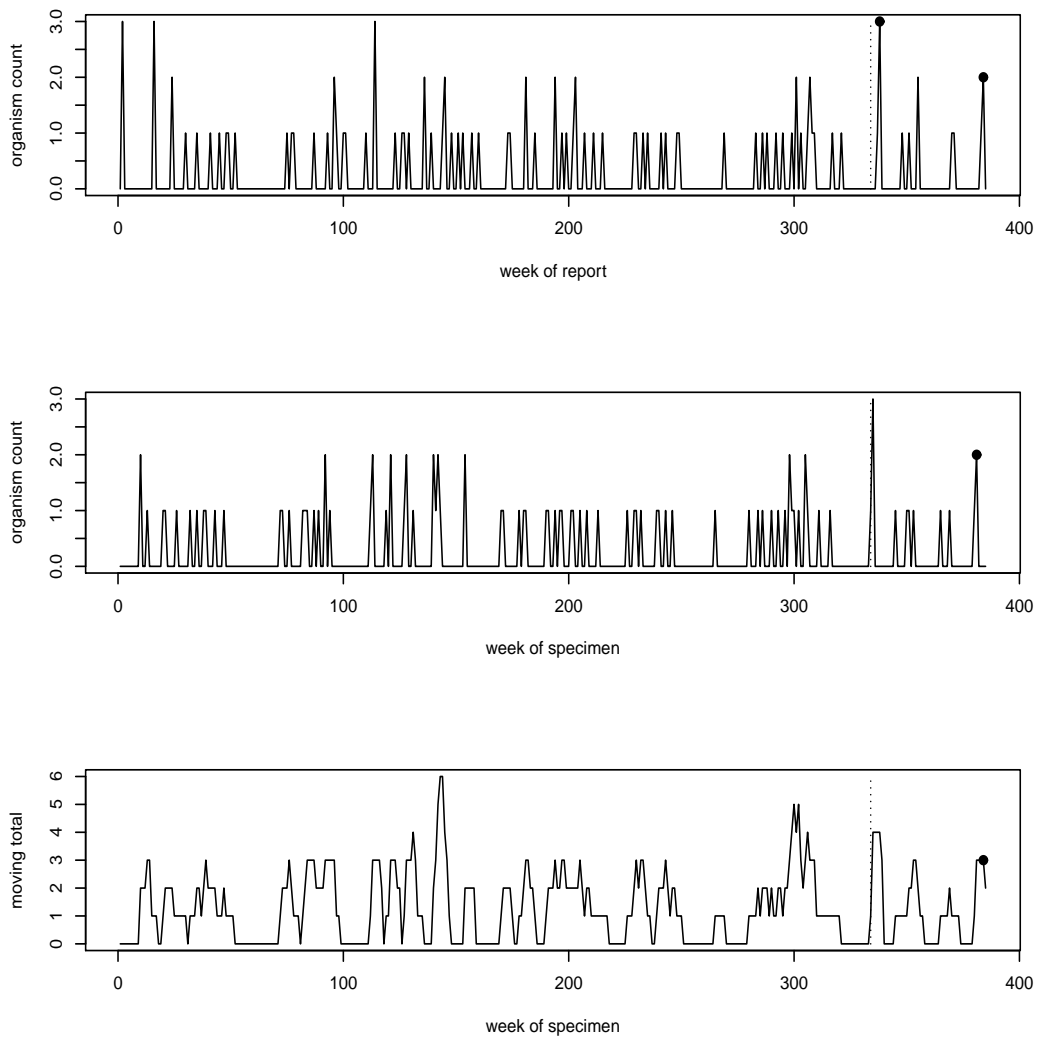
Figure 14: *Giardia lamblia*: Time series of counts (full lines) and weeks flagged (dots). Top: standard algorithm applied prospectively by week of report; middle: standard algorithm applied retrospectively by week of specimen (gold standard); bottom: new algorithm applied prospectively, line represents 3-point moving totals (see main text). The algorithms were applied to the 52 weeks to the right of the vertical dotted line.

38

Figure 15: *Salmonella abony*: Time series of counts (full lines) and weeks flagged (dots). Top: standard algorithm applied prospectively by week of report; middle: standard algorithm applied retrospectively by week of specimen (gold standard); bottom: new algorithm applied prospectively, line represents 5-point moving totals (see main text). The algorithms were applied to the 52 weeks to the right of the vertical dotted line.
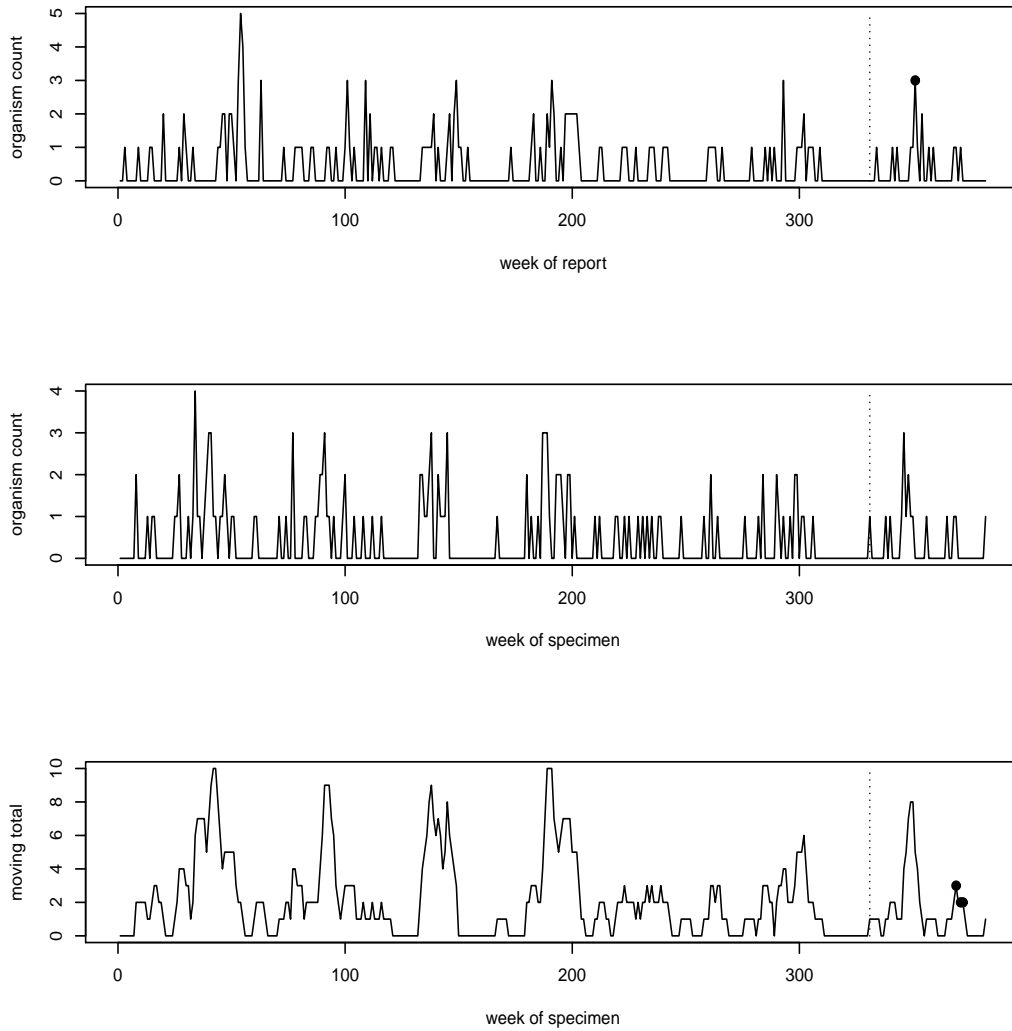
Figure 16: *Salmonella brandenburg*: Time series of counts (full lines) and weeks flagged (dots). Top: standard algorithm applied prospectively by week of report; middle: standard algorithm applied retrospectively by week of specimen (gold standard); bottom: new algorithm applied prospectively, line represents 5-point moving totals (see main text). The algorithms were applied to the 52 weeks to the right of the vertical dotted line.
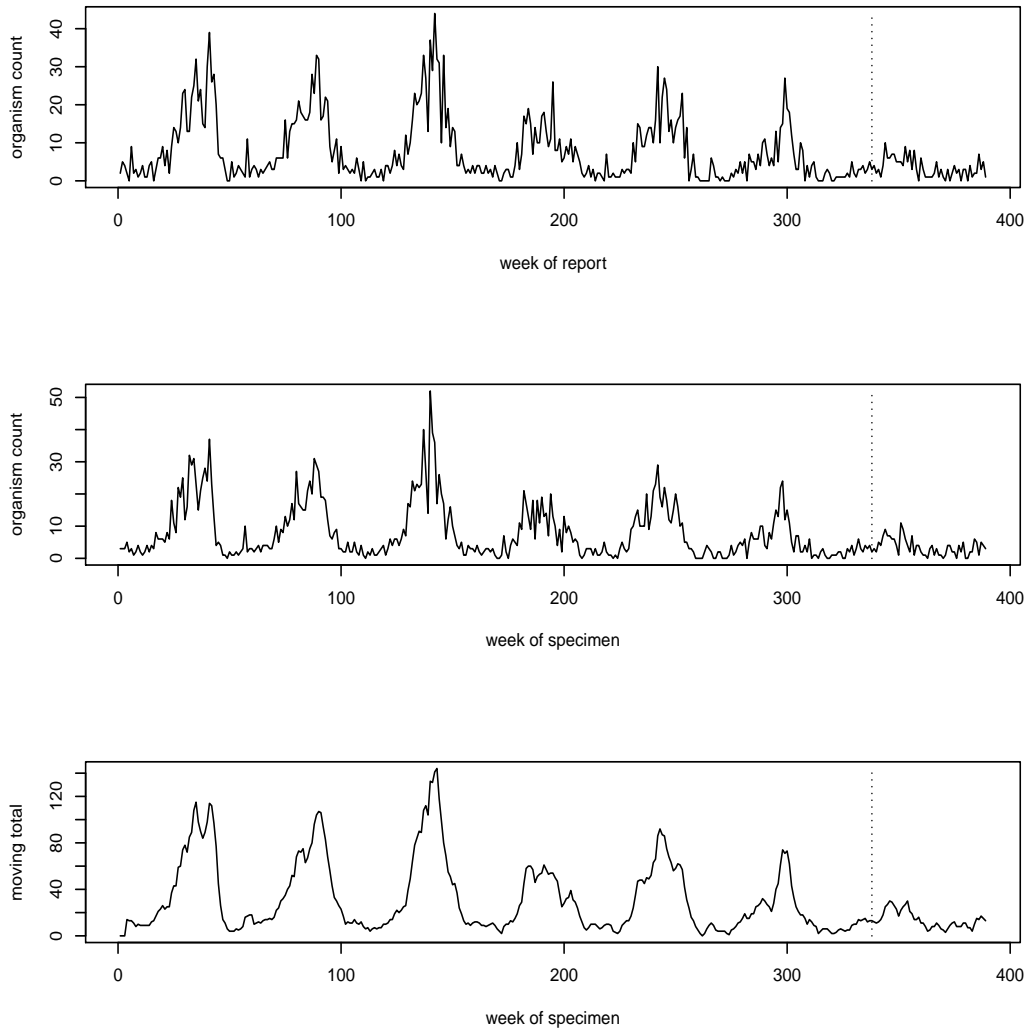
40

Figure 17: *Salmonella enteritidis* PT21: Time series of counts (full lines) and weeks flagged (dots). Top: standard algorithm applied prospectively by week of report; middle: standard algorithm applied retrospectively by week of specimen (gold standard); bottom: new algorithm applied prospectively, line represents 4-point moving totals (see main text). The algorithms were applied to the 52 weeks to the right of the vertical dotted line.
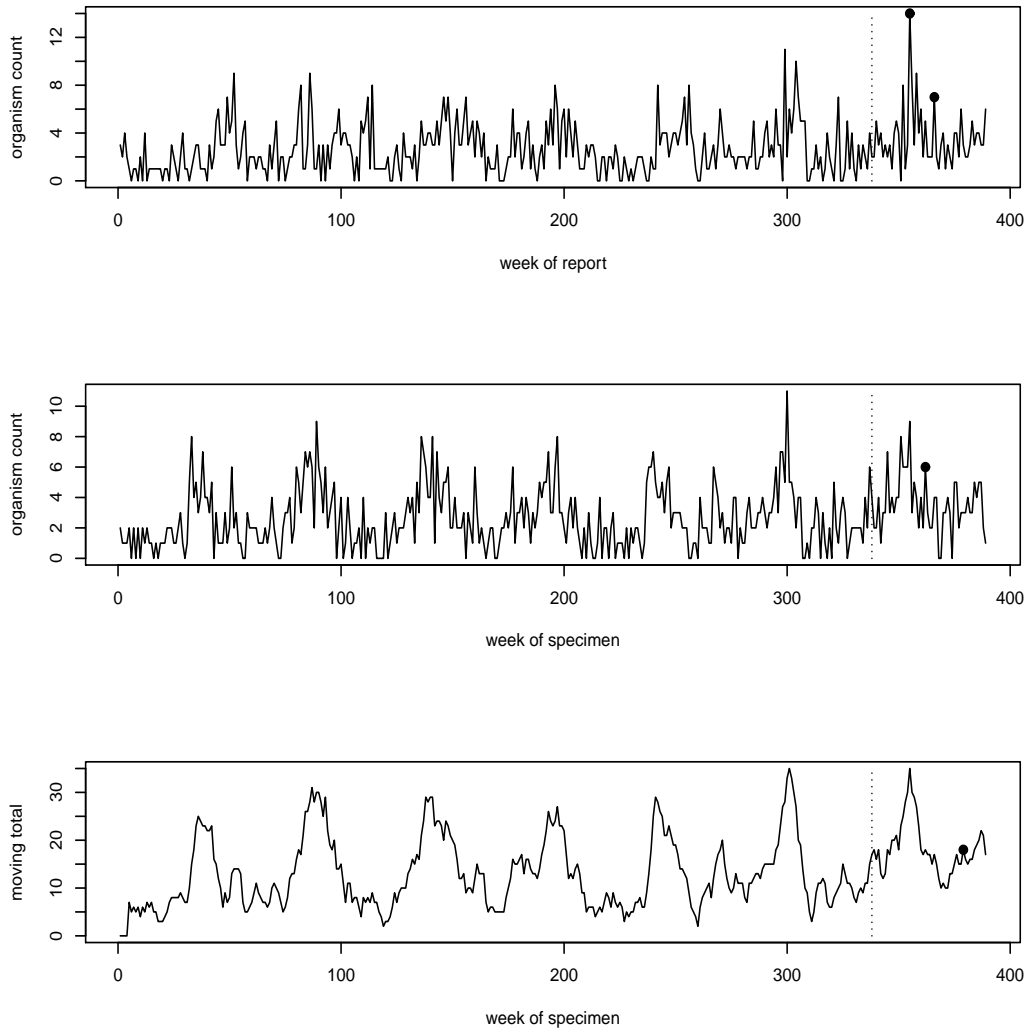
41

Figure 18: *Salmonella infantis*: Time series of counts (full lines) and weeks flagged (dots). Top: standard algorithm applied prospectively by week of report; middle: standard algorithm applied retrospectively by week of specimen (gold standard); bottom: new algorithm applied prospectively, line represents 5-point moving totals (see main text). The algorithms were applied to the 52 weeks to the right of the vertical dotted line.
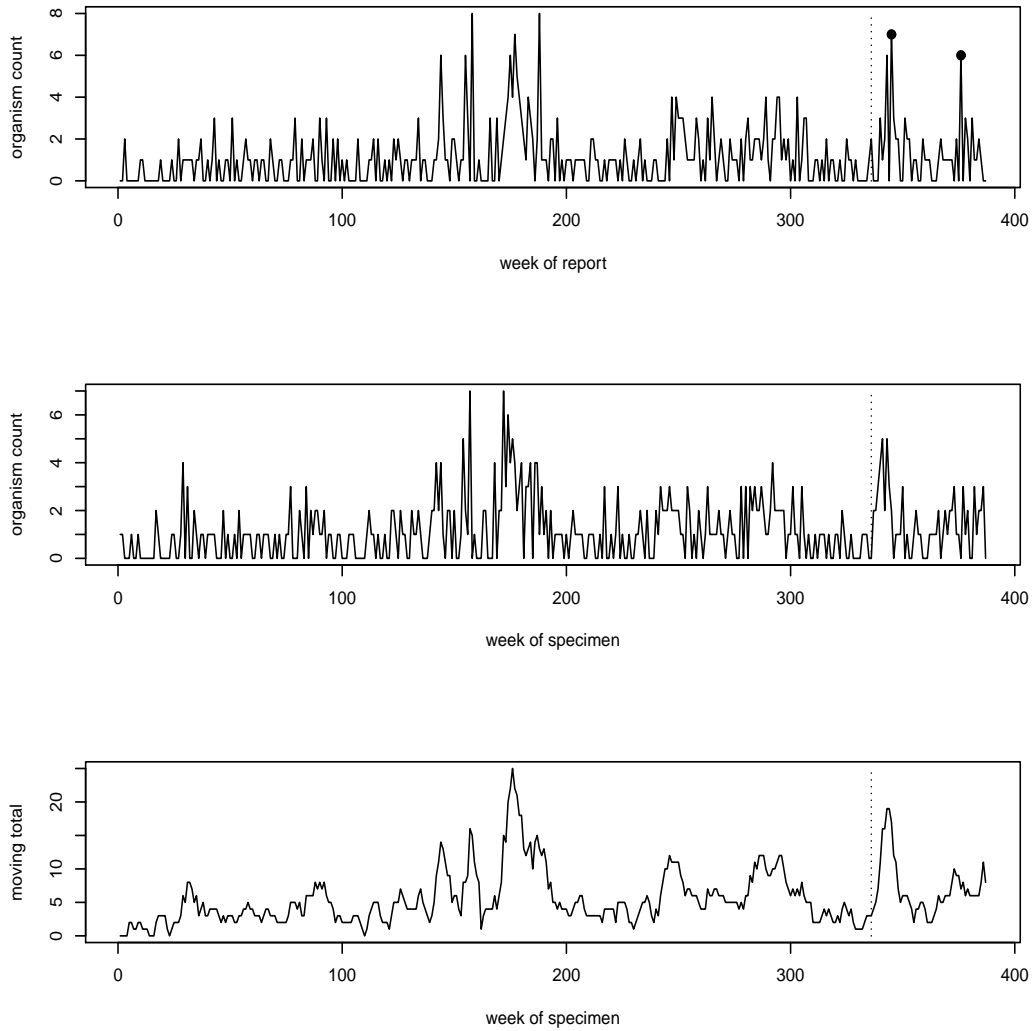
Figure 19: *Salmonella senftenberg*: Time series of counts (full lines) and weeks flagged (dots). Top: standard algorithm applied prospectively by week of report; middle: standard algorithm applied retrospectively by week of specimen (gold standard); bottom: new algorithm applied prospectively, line represents 5-point moving totals (see main text). The algorithms were applied to the 52 weeks to the right of the vertical dotted line.
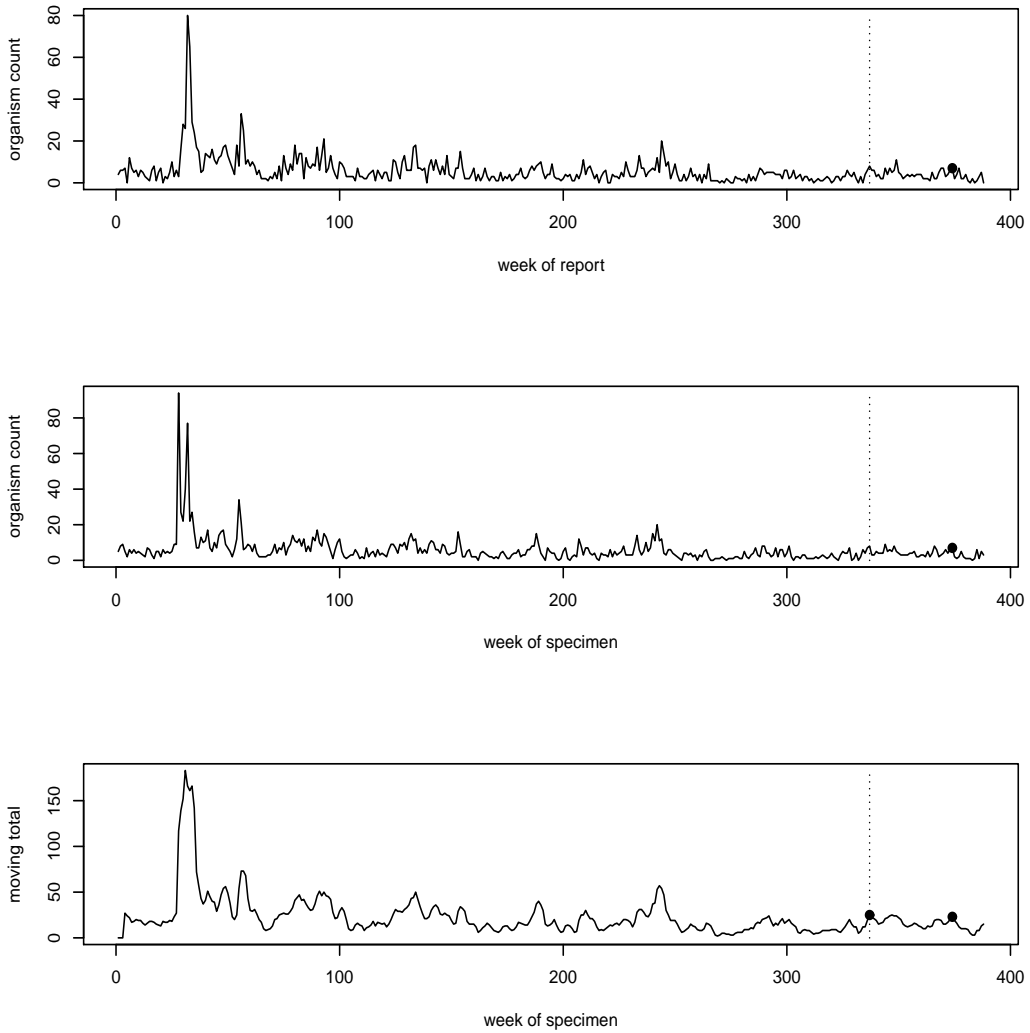
43

Figure 20: *Salmonella typhimurium* DT104: Time series of counts (full lines) and weeks flagged (dots). Top: standard algorithm applied prospectively by week of report; middle: standard algorithm applied retrospectively by week of specimen (gold standard); bottom: new algorithm applied prospectively, line represents 4-point moving totals (see main text). The algorithms were applied to the 52 weeks to the right of the vertical dotted line.

44

References

Becker, N. G., Watson, L. F., and Carlin, J. B. (1991). A method of non-parametric back-projection and its application to AIDS data. *Statistics in Medicine*, 10:1527–1542.

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

Brookmeyer, R. and Gail, M. H. (1988). A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *Journal of the American Statistical Association*, 83:301–308.

Brookmeyer, R. and Liao, J. (1990). The analysis of delays in disease reporting: methods and results for the acquired immunodeficiency syndrome. *American Journal of Epidemiology*, 132:355–365.

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

Cox, D. R. and Medley, G. F. (1989). A process of events with notification delay and the forecasting of AIDS. *Philosophical Transactions of the Royal Society of London*, 325:135–145.

Donker, T., van Boven, M., van Ballegooijen, W. M., van't Klooster, T. M., Wielders, C. C., and Wallinga, J. (2011). Nowcasting pandemic influenza A/H1N1 2009 hospitalizations in the Netherlands. *European Journal of Epidemiology*, 26:195–201.

Enki, D. G., Noufaily, A., Garthwaite, P. H., Andrews, N. J., Charlett, A., Lane, C., and Farrington, C. P. (2013). Automated biosurveillance data from England and Wales, 1991-2011. *Emerging Infectious Diseases*, 19:35–42.

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

Green, T. A. (1998). Using surveillance data to monitor trends in the AIDS epidemic. *Statistics in Medicine*, 17:143–154.

Guillou, A., Kratz, M., and Strat, Y. L. (2014). An extreme value theory approach for the early detection of time clusters. A simulation-based assessment and an illustration to the surveillance of *Salmonella*. *Statistics in Medicine*, 33:5015–5027.

Hohle, M. and an der Heiden, M. (2014). Bayesian nowcasting during the STEC O104:H4 outbreak in Germany, 2011. *Biometrics*, 70:9931002.

Hulth, A., Andrews, N., Ethelberg, S., Dreesman, J., Faensen, D., van Pelt, W., and Schnitzler, J. (2010). Practical usage of computer-supported outbreak detection in five European countries. *Eurosurveillance*, 15:article 6.

Jones, G., Hello, S. L., da Silva, N. J., Vaillant, V., de Valk, H., Weill, F. X., and Strat, Y. L. (2014). The french human salmonella surveillance system: evaluation of timeliness of laboratory reporting and factors associated with delays, 2007 to 2011. *Eurosurveillance, Surveillance and Outbreak Reports*, 19(1):206–264.

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

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*, 100:61–69.

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

Pagano, M., Tu, X. M., Gruttola, V. D., and MaWhinney, S. (1994). Regression analysis of censored and truncated data: estimating reporting-delay distributions and AIDS incidence from surveillance data. *Biometrics*, 50:1203–1214.

Rosenberg, P. S. and Gail, M. H. (1991). Backcalculation of flexible linear models of the human immunodeficiency virus infection curve. *Applied Statistics*, 40:269–282.

Salmon, M., Schumacher, D., Stark, K., , and Hohle, M. (2015). Bayesian outbreak detection in the presence of reporting delays. *Biometrical Journal*, 00:117.

Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*. Princeton: Van NostrandReinhold.

Shmueli, G. and Burkom, H. (2010). Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics*, 52:3951.

Sonesson, C. and Bock, D. (2003). A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society, Series A*, 166:5–21.

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*, 175:49–82.