Preventive Veterinary Medicine 91 (2009) 2–10

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Statistical approaches to the monitoring and surveillance of infectious diseases for veterinary public health☆

Michael Höhle [a,*], Michaela Paul [b], Leonhard Held [b]

[a] Department of Statistics, University of Munich, Ludwigstr. 33, 80539 Munich, Germany
[b] Biostatistics Unit, Institute of Social and Preventive Medicine, University of Zurich, Switzerland

ARTICLE INFO

Keywords:
Veterinary public health
Outbreak detection
Rabies
GIS

ABSTRACT

This paper covers the aspect of using statistical methodology for the monitoring and surveillance of routinely collected data in veterinary public health. An account of the Farrington algorithm and Poisson cumulative sum schemes for the prospective detection of aberrations is given with special attention devoted to the occurrence of seasonality and spatial aggregation of the time series. Modelling approaches for retrospective analysis of surveillance counts are also described. To illustrate the applicability of the methodology in veterinary public health, data from the monitoring of rabies among fox in Hesse, Germany, are analysed.

1. Introduction

The specific aim of disease monitoring and surveillance, which has a long history in veterinary sciences as described in e.g. Wilkinson (1992), is the early detection of emerging and re-emerging diseases in order to prepare contingency plans to contain those diseases. Following Christensen (2001) we shall differentiate between *animal disease monitoring* as the ongoing efforts directed at assessing the health and disease status of a given population and *disease surveillance*, which describes a more active system and implies that some form of directed action will be taken if the data indicate aberrations in disease level. An example of monitoring could be the process of keeping track of a known disease, e.g. foot an mouth disease, classical swine fever or rabies. This is in contrast to the detection of non-specific signals as part of early detection for specific infections not present in the population. Such surveillance could, e.g. be the monitoring of mortality in poultry triggering sampling and testing for highly pathogenic avian influenza in the affected poultry flocks. However, as described in Doherr and Audigé (2001) it is common to use the term disease monitoring and surveillance system (MOSS) as umbrella term for the two activities and we shall adopt this terminology in what follows.

The focus of this paper is on statistical methodology for performing prospective outbreak detection and retrospective modelling in time series of disease counts resulting from continuous data collection within a disease MOSS. This task has become increasingly important as the amount of data gathered through automatic data collection increases. For a general overview on statistical challenges for survey design and diagnostic testing in veterinary MOS systems, see Salman (2003).

Examples from human epidemiology include the monitoring of notifiable diseases, congenital malformations, surgical outcomes and bioterrorism syndromes (Widdowson et al., 2003; Chen, 1978; Steiner et al., 2000; Bravata et al., 2004). Examples in veterinary epidemiology are the monitoring of salmonella in livestock reports (Kosmider et al., 2006) or abortions in dairy cattle (Carpenter et al., 2007). One issue in adapting statistical outbreak detection methods from humans to animals is the fundamental differences in terms: Animal MOS systems have to deal with diverse species and completely different

☆ This paper is part of a special issue entitled "GisVet 2007", Guest Edited by Annette Kjær Ersbøll.

* Corresponding author.
E-mail address: hoehle@stat.uni-muenchen.de (M. Höhle).

0167-5877/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.prevetmed.2009.05.017

living conditions (e.g. production, wild, or companion animal) resulting in different entities of interest (e.g. individual or herd) and different professionals interacting with the population. As a consequence the possibility and cost of investigation and control strategy depends heavily on character of living and the mobility of the animal population. However, as much as the human and veterinary MOS systems differ, zoonoses like salmonellosis, rabies or emerging zoonoses (e.g. avian flu) underline the need for a comparative and co-operative approach in monitoring and surveillance.

Data quality is a major practical concern in the analysis of MOSS data, which complicates the statistical analysis. Examples are the lack of a clear case definition, imperfect diagnostic tests, under-reporting, reporting delays and reporting of only test positives with no information on the total number of tests conducted (lack of denominator data). In this paper the focus is however on the statistical challenges of analysing the resulting univariate and multivariate time series containing daily, weekly or monthly counts. This task shall – following Lawson and Kleinman (2005) – be denoted as statistical surveillance of count data time series.

The first part of the article deals with prospective statistical surveillance. Several known methods from the literature are treated, but also new methodological developments such as weighting timeliness of data is presented. One important question in the MOSS is deciding on the best level of aggregation. To this end a new scheme for hierarchical disease detection is introduced. Our presentation thus emphasizes the time series nature of the data as an alternative to spatial and spatio-temporal cluster detection methods, e.g. scan statistics (Kulldorff, 2001; Rogerson, 2001) or point-process oriented approaches (Diggle et al., 2005). The second part of the text describes a stochastic model for the analysis of multivariate surveillance data. This model can be used to detect temporal and spatio-temporal dependencies in multivariate time series of MOSS counts. Applicability of the presented methods is illustrated throughout the text with data from the monitoring database of the WHO rabies surveillance program (WHO Collaboration Centre for Rabies Surveillance and Research, 2007).

2. Prospective surveillance

In this section, statistical methods for univariate and hierarchical disease surveillance are discussed with a focus on outbreak detection for count data with seasonality. Broader surveys of outbreak detection methods can be found in Farrington and Andrews (2003); Sonesson and Bock (2003); Lawson and Kleinman (2005); Buckeridge et al. (2005).

2.1. Univariate surveillance

For many surveillance problems univariate time series of counts are readily available. If not additional preprocessing is performed, e.g. by aggregating geo-referenced outbreak data to an appropriate level or by time-wise aggregation of event time data. We denote the resulting univariate time series by $\{y_t; t = 1, 2, \ldots\}$. Prospective outbreak detection can be seen as a classification task: based on the observed values $y_1, \ldots, y_n$ it is to be decided if there is an aberration at time n or not. In what follows two classes of methods that address the problem are described.

2.1.1. Farrington method

The core of the method by Farrington et al. (1996) is to predict the observed value y_n using a set of reference values taken from the observed values $y_1, \ldots, y_{n-1}$. To handle long-term trends and seasonality, only values from a window of size $2w + 1$ around time n upto b years back in time are taken. Thus, the set of reference values consists of recent values with similar conditions as at time n and can formally be defined as

$$R(w, b) = \left(\bigcup_{i=1}^{b} \bigcup_{j=-w}^{w} y_{n-i\cdot r+j} \right),$$

where r is the period of the observations, e.g. for monthly data r is 12. Thus no observations from the current year are used. Poisson regression with overdispersion is then used to model the $(2w + 1)b$ reference values, i.e. for $y_t \in R(w, b)$

$$\mathrm{E}(y_t) = \mu_t, \quad \text{with} \quad \log \mu_t = \alpha + \beta t \quad \text{and} \quad \mathrm{Var}(y_t) = \phi \mu_t.$$

Based on the estimated model a one-sided $(1 - \kappa) \times 100\%$ prediction interval for y_n can be formed. The classical way to compute such a prediction interval is based on the normal distribution, however, as the skewness of the Poisson distribution with mean μ is $1/\sqrt{\mu}$, for low valued μ this is a bad approximation. Therefore a (2/3)-power transformation is applied to normalize the distribution before computing the interval. The resulting back-transformed upper limit of the prediction interval for y_n is then

$$U_n = \hat{\mu}_n \left\{ 1 + \frac{2}{3} z_{1-\kappa} \cdot \sqrt{\frac{\hat{\phi}\hat{\mu}_n + \mathrm{Var}(\hat{\mu}_n)}{\hat{\mu}_n^2}} \right\}^{3/2},$$

where $\hat{\mu}_n = \exp(\hat{\alpha} + n\hat{\beta})$ and $z_{1-\kappa}$ is the $100(1 - \kappa)\%$ quantile of the standard normal distribution. Subsequently, if $y_n > U_n$ an alarm is sounded. To ease exposition some details of the algorithm have been left out in the above description; e.g. the linear trend is only included if it is significant at the 5% level and a second round of estimation is performed with observations weighted by their inverse residuals. The latter corrects for possible past outbreaks in the reference values. Furthermore, protection against preposterous alarms is made by post-processing alarms and only reporting those where enough cases have been seen.

One virtue of the Farrington method is its simple yet flexible modelling depending on only one user specified parameter κ. Hence, virtually no time series specific tuning is required, which becomes advantageous when applying the method to multiple surveillance time series. One shortcoming is that only a moving window of historical values is taken for estimation with no values taken from the current year. A simple extension would be to include