Statistical Models for Biosurveillance of Multiple Organisms

Doyo G. Enki*[1], Angela Noufaily[1], C. P. Farrington[1], Paul H. Garthwaite[1], Nick Andrews[2], André Charlett[2] and Chris Lane[2]

[1]Mathematics and Statistics, The Open University, Milton Keynes, United Kingdom; [2]Health Protection Agency, London, United Kingdom

Objective

To look at the diversity of the patterns displayed by a range of organisms, and to seek a simple family of models that adequately describes all organisms, rather than a well-fitting model for any particular organism.

Introduction

There has been much research on statistical methods of prospective outbreak detection that are aimed at identifying unusual clusters of one syndrome or disease, and some work on multivariate surveillance methods (1). In England and Wales, automated laboratory surveillance of infectious diseases has been undertaken since the early 1990's. The statistical methodology of this automated system is described in (2). However, there has been little research on outbreak detection methods that are suited to large, multiple surveillance systems involving thousands of different organisms.

Methods

We obtained twenty years' data on weekly counts of all infectious disease organisms reported to the UK's Health Protection Agency. We summarized the mean frequencies, trends and seasonality of each organism using log-linear models. To identify a simple family of models which adequately represents all organisms, the Poisson model, the quasi-Poisson model and the negative binomial model were investigated (3,4). Formal goodness-of-fit tests were not used as they can be unreliable with sparse data. Adequacy of the models was empirically studied using the relationships between the mean, variance and skewness. For this purpose, each data series was first subdivided into 41 half-years and de-seasonalized.

Results

Trends and seasonality were summarized by plotting the distribution of estimated linear trend parameters for 2250 organisms, and modal seasonal period for 2254 organisms, including those organisms for which the seasonal effect is statistically significant.

Relationships between mean and variance were summarized as given in Figure 1.

Similar plots were used to summarize the relationships between mean and skewness.

Conclusions

Statistical outbreak detection models must be able to cope with seasonality and trends. The data analyses suggest that the great majority of organisms can adequately – though far from perfectly – be represented by a statistical model in which the variance is proportional to the mean, such as the quasi-Poisson or negative binomial models.

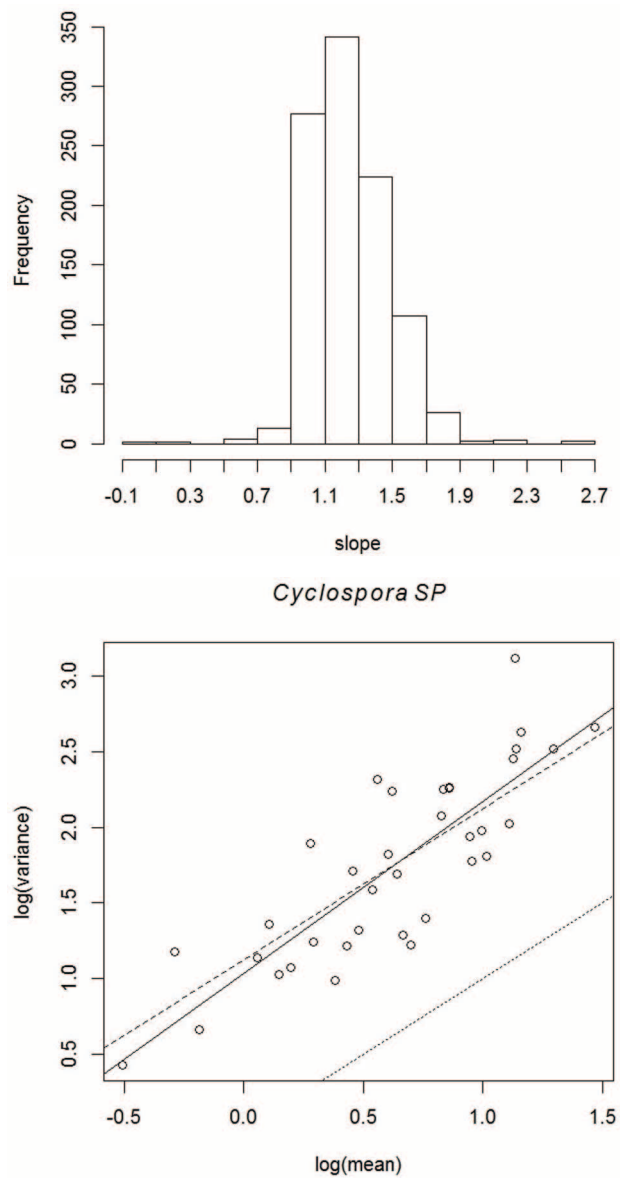

Figure 1. Relationships between mean and variance. (top) Histogram of the slopes of the best fit lines for 1001 organisms; the value 1 corresponds to the quasi-Poisson model; (bottom) log of variance plotted against log of mean for one organism. The full line is the best fit to the points; the dashed line corresponds to the quasi-Poisson model; the dotted line corresponds to the Poisson model.

ISDS Annual Conference Proceedings 2012. This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License (http://creativecommons.org/licenses/by-nc/3.0/), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Keywords

Biosurveillance; Public Health Surveillance; Data Analysis; Infectious Disease Outbreaks; Statistical Model

Acknowledgments

This research was supported by a project grant from the UK Medical Research Council, and by a Royal Society Wolfson Research Merit Award.

References

1. Unkel S, Farrington CP, Garthwaite PH, Robertson C, Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. J. R. Statist. Soc. A 2012; 175:49-82.
2. Farrington CP, Andrews NJ, Beale AD, Catchpole MA. A statistical algorithm for the early detection of outbreaks of infectious disease. J. R. Statist. Soc. A 1996; 159: 547-563.
3. McCullagh P, Nelder JA. Generalized Linear Models. 2nd ed. London: Chapman & Hall; 1989.
4. Hastie TJ, Tibshirani RJ. Generalized Additive Models. London: Chapman & Hall; 1990.

*Doyo G. Enki
E-mail: d.gragn@open.ac.uk

ISDS Annual Conference Proceedings 2012. This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License (http://creativecommons.org/licenses/by-nc/3.0/), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.