

Provided for non-commercial research and educational use.
Not for reproduction, distribution or commercial use.

PLISKA

STUDIA MATHEMATICA
BULGARICA

ПЛИСКА

БЪЛГАРСКИ
МАТЕМАТИЧЕСКИ
СТУДИИ

The attached copy is furnished for non-commercial research and education use only.
Authors are permitted to post this version of the article to their personal websites or institutional repositories and to share with other researchers in the form of electronic reprints.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to third party websites are prohibited.

For further information on
Pliska Studia Mathematica Bulgarica
visit the website of the journal <http://www.math.bas.bg/~pliska/>
or contact: Editorial Office

Pliska Studia Mathematica Bulgarica
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Telephone: (+359-2)9792818, FAX:(+359-2)971-36-49
e-mail: pliska@math.bas.bg

ANALYSING AND MONITORING SURVEILLANCE DATA OF MUMPS IN BULGARIA*

M. Mitova-Bobcheva, M. Slavtchova-Bojkova, M. Kojouharova,
A. Kurchatova

This paper aims to present the applicability of the Bayesian methodology for monitoring surveillance data collected for vaccine-preventable diseases, like mumps, measles and possibly others.

We investigate the ability of implementation of the R-Package “surveillance” especially written for monitoring surveillance data to the mumps data kindly provided by the National Center of Infectious and Parasitic Diseases in Bulgaria.

1. Introduction

In an attempt to meet the threats of infectious diseases to society, public health authorities have created comprehensive mechanisms for the collection of disease data. As a consequence, the abundance of data has demanded the development of automated algorithms for the detection of abnormalities and aberrations. Typically, such an algorithm monitors a univariate time series of counts using a combination of heuristic methods and statistical modelling. Prominent examples of surveillance algorithms are the work by Stroup et al. (1989) and Farrington et al. (1996). A comprehensive survey of outbreak detection methods can be found

*The research was partially supported by appropriated state funds for research allocated to Sofia University (contract No 112/2010), Bulgaria.

2000 *Mathematics Subject Classification*: 62P10.

Key words: Bayesian approach, Outbreak, Surveillance of Infectious Diseases.

in (Farrington and Andrews, 2003). The R-package surveillance was written with the aim of providing a test-bench for surveillance algorithms.

The purpose of the analysis in this work is to illustrate the basic functionality of the package with R-code examples. Section 2 contains a short description of how to use the surveillance algorithms and presents the results with description of the data set we used.

2. Bayesian approach for predicting outbreaks

Bayesian approach for predicting outbreaks, implemented in the statistical software R (see [5]) is applied on surveillance data of mumps collected in Bulgaria for the period 2000-2008. A detailed description of the method is beyond the scope of this work and could be seen in Höhle (2005). The official data is kindly provided by the National Center of Infectious and Parasitic diseases at the Ministry of Health, Bulgaria. It has been collected on a weekly base and presents the epidemic picture by regions in the country for 2000-2008 year. The data clearly shows that there was epidemic outbreak in the country in 2007 and 2008 (see Kojouharova M. et al. (2007)). The distribution of infected cases by regions for this period is visualized in Figure 1.

Using R-software we apply the following model on our data. Let us denote by $\{y_t; t = 1, \dots, n\}$ the time series of counts representing the surveillance data. Due to the fact that such data is typically collected on a weekly basis it is also

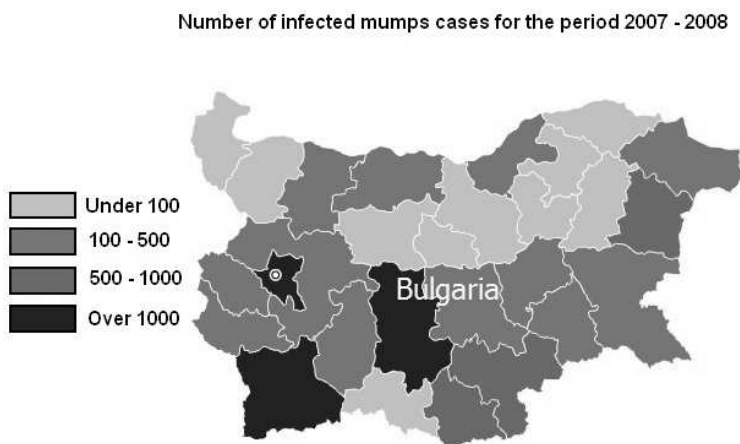


Figure 1: Number of infected mumps cases for the period 2007–2008

convenient to use the following notation $\{y_{i:j}\}$, where $j = \{1, \dots, 52\}$ presents the number of weeks in the year and $i = \{-b, \dots, -1, 0\}$ are the corresponding years. The years have been indexed in such a way that the last year for which we have data has index 0. Let $y_{0:t}$ be the number of cases of the current week (denoted week t in year 0), b the number of years to go back in time and w the number of years around t to include from those previous years. The zero year will be denoted by w_0 . Hence the set of chosen weeks/years for which we want to trace the disease is:

$$R(w, w_0, b) = \left(\bigcup_{i=1}^b \bigcup_{j=-w}^w y_{-i:t+j} \right) \cup \left(\bigcup_{k=-w_0}^{-1} y_{0:t+k} \right).$$

Note that the number of cases of the current week is not a part of $R(w, w_0, b)$. The aim of the surveillance algorithm described above is to create a prediction $\hat{y}_{t:0}$ for the current week of the process. This prediction is then compared to the actual observed value $y_{0:t}$. If the observed value is much higher than the predicted one we get an alert, which warns us to investigate further the reasons for this. The red triangles on the graphs represent the alarms.

We apply the $R(6, 6, 0)$ model for the data we have for the regions of Plovdiv and Sofia as the epidemic outbreak in these two regions was significant. The histograms at Figure 3 give us a clear idea of the spread of infected mumps cases in these regions for the period 2000-2008. As we see, there were very few infected cases during 2001-2006. In 2007 and 2008, the number of infected cases significantly increased and grew into epidemic outbreak in the end of 2007 and in the beginning of 2008.

In order to apply the Bayesian model on our data we first need to format the data in such a way so that it can be used by the R DisProg class. Then we draw infection spread charts for the period 2000-2008. We choose week 7 of 2008 and week 50 of 2007 for outbreak peak for Sofia and Plovdiv correspondingly. We apply the Bayesian model that uses data for the last six weeks, i.e. $R(w, w_0, b) = R(6, 6, 0)$ on our data. The chart in Figure 2 is produced by applying $R(6, 6, 0)$ on aggregated data of the country. Figures 2. and 2. represent the curve of the expected infected cases for the regions of Sofia and Plovdiv, respectively.

The results we get for Sofia and Plovdiv are quite close. This can be explained with the fact that in both regions there are enough outbreaks that occurred in a short time period. On the other hand, as these two outbreaks are the two biggest in the country, they describe the epidemic picture in Bulgaria very well. It's essential to clarify here that as the epidemic situation in the country quite differs from one region to another, it's always better to make predictions for the disease

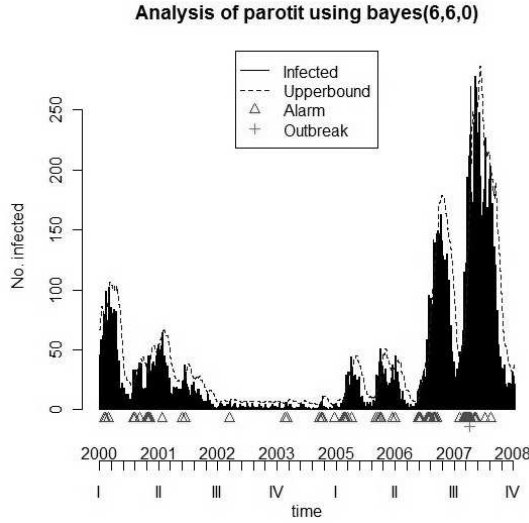


Figure 2: Bayesian approach for predicting outbreaks

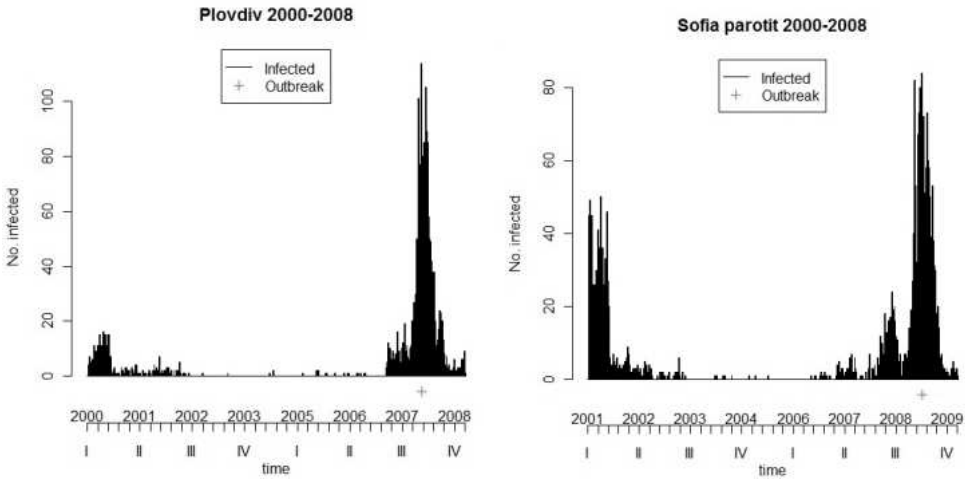


Figure 3: Analysis of Sofia and Plovdiv

spread, based on the data collected for the separate regions. In this way we will get a more clear and precise idea as to which regions are threatened with epidemic. It's also important to take into account the increasing mobility nowadays, that could lead to importing infected cases from one region into another.

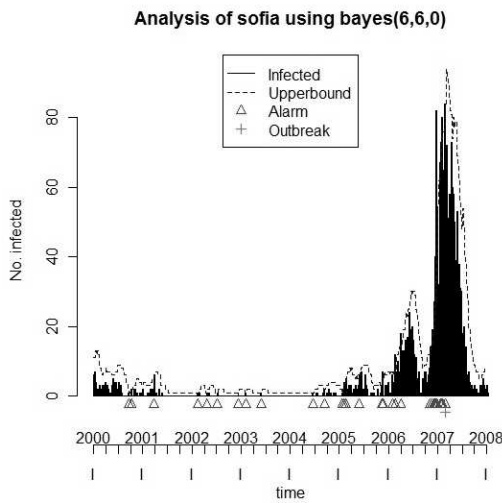


Figure 4: Analysis of Sofia using Bayes(6,6,0)

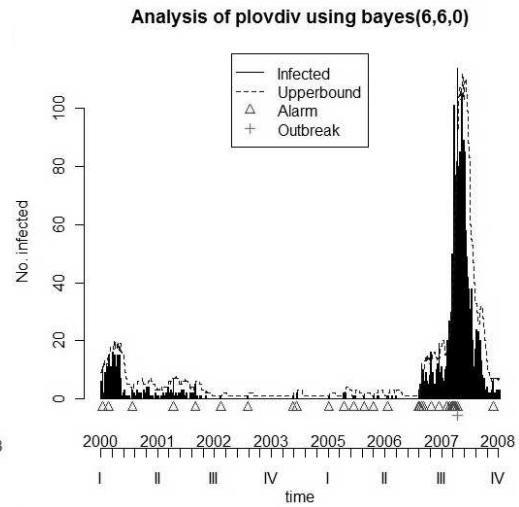


Figure 5: Analysis of Plovdiv using Bayes(6,6,0)

3. Conclusion

The statistical methods we applied here provide a simple framework for analyzing surveillance data of infectious diseases. The quality of the collected data is of a particular importance for the adequate application of the statistical methods. It would lead to an added value to the results of the applied measures from the surveillance of infectious diseases, in general. The extra vaccinations as applied in the case of mumps in Bulgaria were inevitable and could be optimized as to gain a desired effect.

Acknowledgments. The authors are grateful to the referee for the comments and suggestions on an earlier version of the paper.

REFERENCES

- [1] C. FARRINGTON, N. ANDREWS. Monitoring the Health of Populations. Chapter Outbreak Detection: Application to Infectious Disease Surveillance, Oxford University Press, 2003.
- [2] C. FARRINGTON, N. ANDREWS A. BEALE, M. CATCHPOLE. A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society. Series A.* **159** (1996), 547—563.

- [3] M. KOJOUHAROVA, A. KURCHATOVA, L. MARINOVA, T. GEORGIEVA. Mumps outbreak in Bulgaria, 2007: a preliminary report. *Eurosurveillance* **12**, (2007).
- [4] R. HÖHLE. The R-package “surveillance”. *Sonderforschungsbereich* **386**, (2005).
- [5] R Software. <http://www.r-project.org/>
- [6] D. STROUP, G. WILLIAMSON, J. HERNDON, J. KARON. Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Statistics in Medicine* **8** (1989), 323–329.

M. Mitova-Bobcheva
Online Market Research Specialist
FlyResearch BG Ltd (M.S.)
e-mail: miroslava.mitova@gmail.com

A. Kurchatova
National Center for
Infectious and Parasitic Diseases
Ministry of Health, Bulgaria
e-mail: akurchatova@ncipd.org

Maroussia Slavtchova-Bojkova
Faculty of Mathematics and Informatics
Sofia University “St. Kl. Ohridski”
5, J. Bourchier Blvd
1164 Sofia, Bulgaria
e-mail: bojkova@fmi.uni-sofia.bg

M. Kojouharova
National Center
for Infectious and Parasitic Diseases
Ministry of Health, Bulgaria
e-mail: mkojouharova@ncipd.org

and
Department of Probability and Statistics
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Acad. G. Bontchev Str., Bl. 8
1113 Sofia, Bulgaria
e-mail: bojkova@math.bas.bg