STATISTICAL APPROACHES TO ENVIRONMENTAL SPATIAL SURVEILLANCE

Dolores Catelan[1,2], Corrado Lagazio[3], Annibale Biggeri[1,2]

1 Department of Statistics "G. Parenti" – University of Florence
2 Biostatistics Unit – CSPO Institute for Cancer Prevention
3 Department of Statistical Science – University of Udine

Disease mapping is a cornerstone of Epidemiologic Surveillance. Small scale of analysis is used to scan for abnormal rates associated to environmental exposures or to search for health effect of putative sources. Maximum Likelihood risk estimates show large variability and p-values are inconsistent with estimates. Multiple testing lead to poor control of false positives. In the contest of disease Mapping shrinkage estimators (Empirical Bayes or Full Bayes) are popular. They are conservative. Trade-off between control for false positive and false negative is suggested and takes advantage of posterior probabilities.

We aimed to develop a simple procedure to control False Positive Rate in the context of Disease Mapping. Indeed multiple testing in Epidemiological Studies is rarely addressed and control of Family Wise Error Rate (FWER) is generally used in surveillance framework. Low sensitivity of Bayesian analysis has been criticized and the control of False Discovery Rate (FDR) has been advocated in the literature. Two different philosophies underlined FWER vs FDR. In particular, control of FWER is important when our decision based on the whole set of individual inferences - is likely to be erroneous when at least one of them is. This may be the case when several areas are analysed, and a decision on each single significant areas is to be taken. However, this is not the more common case. The overall conclusion that the region has areas at higher/lower risk need not be erroneous even if some of the null hypotheses are falsely rejected.

For our purpose we analyzed lung cancer death certificates for males resident in the 287 municipalities of the Tuscany Region (Italy) for the period 1995-1999 have. Data were made available by the Regional Mortality Register. A set of reference rates (Tuscany, 1971-1999) have been used to compute expected number of cases for each municipality, following indirect standardization and classifying the population by 18 age classes (0-5,…, 85 or more). The goal is to identify municipalities with a risk significantly different from the general mean.

We considered conjugate Poisson-Gamma model, BYM convolution model and a three level hierarchical model to screen high risk areas. To this purpose we used posterior probabilities $Prob(\theta_i>1|Y)$ coming from the two first mentioned Bayesian models, and the posterior classification probabilities obtained from the three level hierarchical Bayesian model.

For matter of comparison we calculated the Bayesian probabilities (q-values) obtained for each area from exact p-values. We discuss the results in the framework of Optimal Discovery Procedure.

We point out that posterior probabilities from Hierarchical Bayesian models for Disease Mapping are useful as exploratory tool but FDR control is recommended. A three-level Hierarchical Bayesian model allow to obtain appropriate false discovery rate and model-based estimates of q-values.