

Bayesian Methods for Syndromic Surveillance

David Banks, Gauri Datta, James Lynch, and Francisco Vera
National Institute of Statistical Sciences

OBJECTIVE

This paper describes a Bayesian approach to syndromic surveillance. The method provides more interpretable inference than traditional frequentist approaches. Bayesian methods avoid many of the problems associated with alpha levels and multiple comparisons, and make better use of prior information. The technique is illustrated on simulated data.

BACKGROUND

Syndromic surveillance needs to be (1) transparent, (2) actionable, and (3) flexible. Traditional frequentist approaches to syndromic surveillance, such as cusum charts [1] and scan statistics [2], tend to fail on all three criteria. First, the validity of the assumptions is generally difficult to check and the methods are hard to modify; second, the false positive rate makes it impossible to be both sensitive to true signal and resistant to spurious signal; and third, the implementation usually requires significant hand-tinkering to adjust background rates for known seasonal affects and other identifiable influences.

In contrast, Bayesian methods use explicit models that can be scrutinized by all. Also, the Bayesian analysis gives a clean probability statement concerning whether a disease outbreak has occurred. And the Bayesian machinery is a mathematically simple way to incorporate new information. Berger [3] provides an introduction to Bayesian statistics.

METHODS

To be concrete, we present the simplest realistic version of Bayesian syndromic surveillance. Thus we assume that only one kind of syndrome is reported (so that the data are univariate counts), and we do not detail all of the issues in model specification.

Suppose hospital i reports the number of cases of fever on day t , denoted by $Y_i(t)$. The natural model for the counts is a Poisson distribution (but more complicated models can also be considered, such as the overdispersed Poisson---the Bayesian method applies there as well, in an analogous way).

The mean of the Poisson is the sum of the background rate for time period t and the product of an indicator variable and an epidemic rate. If the indicator variable is 0, then there is no outbreak; thus the product is 0 and one expects counts that are in line with the background rate for high fever. But if the indicator variable is 1, then the count is the sum of the background rate and the epidemic rate.

The indicator variable is the key element. It is a latent variable, because it cannot be observed directly, but the information in the counts permit the analyst to calculate the probability the value of indicator variable is 1. If this probability is large, then there is a very good chance that an epidemic condition exists.

There are many details. First, one needs to use the generalized linear model to estimate the background rate for fever at hospital i on day t . This entails the use of historical data and covariate information about weekend/weekday effects, flu season, and so forth. (Or one could do a fully Bayesian analysis with the covariates, but the computational burden grows quickly.) Second, one needs to take account of the neighborhood structure. If all hospitals in a city show a moderate elevation in their fever counts, this probably means that there is outbreak; the strategy for combining the information uses an two-state Markov process model [4]. The state (disease or no disease) at one hospital affects the probability that neighboring hospitals are in a specific state. There are other issues; one that is less important than one might guess is the model for epidemic rate---in outbreak detection, almost any sensible model works.

RESULTS AND CONCLUSIONS

Simulations of reasonable Bayesian models indicate that the method can perform well. It is impossible to make a direct comparison to frequentist methods unless one has a decision rule on when to declare that an outbreak has occurred, or unless one has real data.

REFERENCES

- [1] Miller, B., Kassenborg, H., Dunsmuir, W., Griffith, J., Hadidi, M., Nordin, J.D., Danila, R. (2004). "Syndromic Surveillance for influenza-like illness in an ambulatory care network," *Emerging Infectious Diseases*, **10**. 1806-1811.
- [2] Kulldorff, M. (1997). "A Spatial Scan Statistic," *Communications in Statistics: Theory and Methods*, **26**, 1481-1496.
- [3] Berger J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York, N.Y.
- [4] Kindermann, R., and Snell, J. L. (1980). *Markov Random Fields and Their Applications*, American Mathematical Society, Providence, R.I.

Further Information:

David Banks, banks@stat.duke.edu

www.stat.duke.edu/~banks