An Empirical Comparison of Spatial Scan Statistics for Outbreak Detection

Daniel B. Neill, Ph.D.
Heinz School of Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213

Objective

We present a systematic empirical comparison of five recently proposed expectation-based scan statistics, in order to determine which methods are most successful for which spatial disease surveillance tasks.

Background

Expectation-based scan statistics [1] extend the traditional spatial scan statistic approach [2] by using historical data to infer the expected counts for each spatial location, then detecting regions with higher than expected counts. Here we consider five recently proposed expectation-based statistics: the expectation-based Poisson (EBP), expectation-based Gaussian (EBG), population-based Poisson (PBP), population-based Gaussian (PBG), and robust Bernoulli-Poisson (RBP) methods. We also consider five different time series analysis methods used to predict the expected counts (including the Holt-Winters method and moving averages optionally adjusted for day of week and seasonality), giving a total of 25 methods to compare. All of these methods are detailed in the full paper [3].

Methods

We first compared the detection power of the 25 methods on four different datasets from Allegheny County: respiratory ED visits and OTC cough/cold, antifever, and thermometer sales. For each dataset, we injected three simulated outbreak types of varying size and severity. 1000 outbreaks were injected for each combination of dataset and outbreak type, giving a total of 12,000 simulated outbreaks for evaluation. Second, to further investigate the relationship between outbreak size and the relative performance of the different methods, we computed the average number of injected cases needed for each method to detect each outbreak type on a given day, as a function of the number of affected zip codes. Finally, we compared two methods of calibration: obtaining p-values by randomization testing, and using the empirical distribution of maximum region scores.

Results

The primary results of our comparison are as follows:

1) Relative performance of different detection methods varied significantly depending on the dataset and outbreak size (Fig. 1). PBP and PBG had high detection power for small outbreaks and low power for large outbreaks, while the reverse was true for RBP.

2) EBP achieved high detection performance across a wide range of datasets and outbreak sizes, and is thus

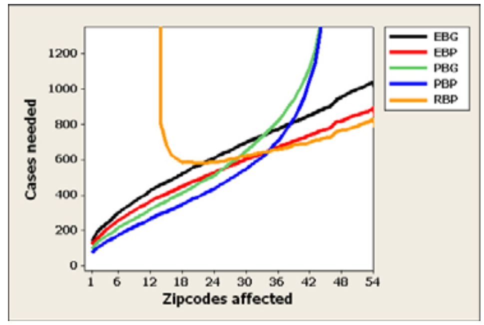

Figure 1: Number of injected cases needed to detect versus outbreak size, OTC cough/cold dataset.

recommended for use in typical detection scenarios where the outbreak characteristics are unknown. Other methods outperform EBP for specific combinations of dataset and outbreak size, and thus may be preferable in these specific outbreak scenarios.

3) Moving average baseline methods outperformed Holt-Winters. Seasonal adjustments were necessary for the cough/cold data, and day-of-week adjustments improved detection on cough/cold and antifever data.

4) The p-values produced by randomization were not properly calibrated for the three OTC datasets (i.e. the proportion of false positives at $\alpha = .05$ was much higher than 5%). Reporting the highest scoring regions gave higher detection power than reporting the regions with lowest p-values, suggesting that randomization testing is not necessary for these datasets.

Conclusions

Our results demonstrate the importance of evaluating methods over a wide range of datasets and outbreak sizes, using a framework such as the one given here, instead of drawing conclusions based on performance for a single dataset or outbreak type. This evaluation framework can be applied to compare other methods as well, and EBP (the only detection method that achieved consistently high performance across all trials) can be used as a baseline for comparison.

References

[1] Neill DB, Moore AW, Methods for detecting spatial and spatio-temporal clusters. Handbook of Biosurveillance, 2006, 243-254.

[2] Kulldorff M, A spatial scan statistic. Communications in Statistics: Theory and Methods, 1997, 26(6): 1481-1496.

[3] Neill DB, An empirical comparison of spatial scan statistics for outbreak detection. Submitted for publication.

Further Information:
Daniel B. Neill, neill@cs.cmu.edu
www.cs.cmu.edu/~neill