Evaluating the Performance of a Spatial Scan Statistic Using Simulated Outbreak Characteristics

Shaun Grannis MD MS,[1] James Egg BS,[1] Christopher A. Cassa, MEng,[2] Karen Olson PhD,[2] Ken Mandl MD PhD[2], J. Marc Overhage MD PhD[1]

[1]The Regenstrief Institute and Indiana University School of Medicine; [2]Children's Hospital Informatics Program at the Harvard-MIT Div of Health Sciences and Technology

OBJECTIVE

To characterize the performance of a spatial scan statistic, we used SaTScan™ to measure the sensitivity and positive predictive value for detecting simulated outbreaks having varying size, case density, and syndrome type.

BACKGROUND

Research evaluating the use of spatial data for surveillance purposes is ongoing and evolving. As spatial methods evolve, it is important to characterize their effectiveness in real-world settings. Assessing the performance of surveillance systems has been difficult because there has been a paucity of data from real bioterrorism events.[1] Recent efforts to assess surveillance system performance have focused on injecting synthetic outbreak data (signal) into actual background visit data.[1,2,3] These studies focused on either temporal data, a single syndrome category, or a single bioterrorism agent. We are unaware of prior studies evaluating the performance of spatial outbreak detection for multiple syndrome categories in an operational surveillance system.

METHODS

We extracted respiratory, gastrointestinal, and constitutional syndromes from two year's of actual emergency department surveillance data from 16 Indianapolis hospitals in the Indiana Network for Patient Care. Patient's home addresses were converted to latitude and longitude using ESRI ArcGIS 8.3. For each week of visit data we used a cluster creation tool[4] to insert one of 360 unique simulated outbreaks that varied by the number of cases (10, 25, 40), cluster radius (0.25, 0.5, 1, 3 km), and distance from the center of Indianapolis (8, 16, and 24 km). We used SaTScan version 6.0 to detect the single simulated outbreak in each of 35,280 1-week datasets for each syndrome.

We calculated sensitivity and positive predictive value (PPV) for various combinations of syndrome category, cluster size, and cluster density. We performed sensitivity analyses using p-values of 0.005, 0.01, 0.1, 0.2, and 0.5. We defined a true positive SaTScan cluster as having at least 50% simulated cases with a scan statistic p-value below the pre-specified cut-off. False positive clusters were defined as having less than 50% synthetic points and a p-value below the pre-specified cut-off.

RESULTS

Average sensitivity and PPV for all clusters was 0.97 ± 0.08 and 0.92 ± 0.10, respectively. Table 1 shows averages for SaTScan sensitivity and PPV for all synthetic clusters stratified by syndrome.

Table 1: SaTScan accuracy stratified by syndrome

Syndrome	Sensitivity	PPV
CONST	0.99 ± 0.04	0.95 ± 0.08
GI	0.96 ± 0.09	0.90 ± 0.11
RESP	0.95 ± 0.10	0.90 ± 0.11

While these results reflect acceptable accuracy, substantially lower performance is noted for subsets of clusters with unfavorable parameters. Figure 1 illustrates sensitivity and PPV for outbreaks having 10 simulated cases within a 3000 m radius. At a p-value of 0.01, SaTScan sensitivity is 0.7 for respiratory syndromes.

Accuracy varies substantially across syndromes. This variation is in part explained by the each syndrome's prevalence, or background visit rate. As prevalence increases, the ability to accurately detect a signal decreases. Constitutional prevalence is lowest in our dataset, while respiratory prevalence is highest.

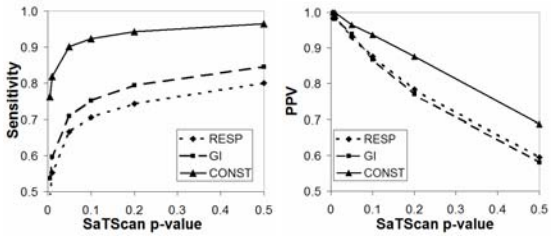

Figure 1: Sensitivity and PPV for simulated clusters having 10 simulated cases and 3000 m radius.

CONCLUSIONS

Our results show how detection accuracy varies with outbreak size, density, and background prevalence. Such data sheds light on the characteristics of outbreaks that can be reasonably detected by spatial scan statistics. Although overall detection accuracy was acceptable, the simulated outbreak parameters may not reflect real-world outbreaks. Further work is needed to refine outbreak parameter estimates (e.g. cluster size) for potential outbreaks by exploring threat models such as geographical areas served by public schools. Such refinements will inform future surveillance system performance expectations.

REFERENCES

1. Nordin J, et al. Simulated Anthrax Attacks and Syndromic Surveillance. Emerg Inf Dis 2005;11:1394-1398.
2. Mandl KD, et al. Measuring outbreak detection performance using controlled feature set simulations. MMWR. 2004;53 Suppl:130-136.
3. Kulldorff M, et al. Benchmark power calculations for evaluating disease outbreak detection methods. MMWR 2004;53 Suppl:144–51.
4. Cassa CA, et al. A software tool for creating simulated outbreaks to benchmark surveillance systems. BMC Med Inform Decis Mak. Jul 14 2005;5(1):22-28.