Accurate detection of arbitrarily-shaped spatial disease clusters

Shannon Wieland, John Brownstein, Ph.D., Bonnie Berger, Ph.D., Kenneth Mandl, M.D., M.P.H.

Children's Hospital Informatics Program at the Harvard-MIT Division of Health Sciences and Technology

OBJECTIVE

A method for detecting spatial clusters of diseases of any shape based on the Euclidean minimum spanning tree (EMST) is described and compared to the circular scan statistic.

BACKGROUND

Methods for locating spatial clusters of diseases are typically variations of the circular scan statistic method [1]. They restrict the number of potential clusters by considering all circular, rectangular [2], or elliptical [3] regions, and then apply a likelihood ratio test to evaluate the statistical significance of each potential cluster. Because disease outbreaks may have variable shapes, there has been recent interest in developing methods to detect irregularly-shaped clusters. Starting with a neighborhood graph of the administrative regions in the study area, certain subgraphs are evaluated. These include all connected subgraphs within a circular window [4] and subgraphs of the minimum spanning tree of a weighted neighborhood graph formed by deleting one edge [5]. These methods restrict the maximum cluster size or identify large clusters having greater likelihood ratios than true clusters in the data, suggesting a limitation of using the likelihood ratio to detect arbitrarily-shaped clusters [5].

METHODS

We consider data sets consisting of point locations of disease cases and underlying population counts for each administrative region in a study area. We first transform the study region into a density-equalizing cartogram [6] to reduce the problem to the uniform population density case. Using a mathematical definition of a potential cluster based only on intercase distances, we prove that there is a one-to-one correspondence between potential clusters and a special class of subsets of the EMST formed by iteratively deleting the longest edge of the tree. We develop a test statistic based on the weight of the potential cluster subgraph which accounts for the multiple hypotheses tested.

RESULTS

We applied the EMST method and SatScan [1] to data sets containing synthetic rectangular clusters superimposed on a background population drawn from census data in Boston, Massachusetts. Each data set contained 500 cases, and the rectangular clusters had various height-to-width ratios (1, 2, 4, 8 or 16) and densities relative to the background (1 or negative control, 2, 3, 4, or 5). The EMST method had greater average power than SatScan for 17 of 20 cluster types (see Figure 1). Given that the cost associated with a false negative case is at least the cost of a false positive, the EMST method always had a lower mean cost than SatScan if the relative cluster density was at least three and the height-to-width ratio was at least four. The EMST cost was usually lower for other density and ratio parameters.

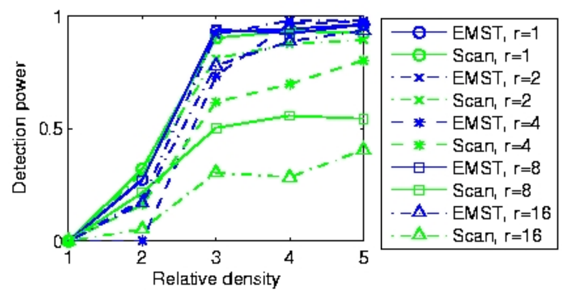

Figure 1 – Power comparison of the EMST and SatScan (Scan) methods. The power was defined to be the mean fraction of patients correctly identified in the cluster. The EMST method had higher power when the relative cluster density was above two for all height-to-width ratios (r). .

CONCLUSIONS

The EMST method may be a powerful alternative to existing methods when the precise case locations are known. Although the EMST method is not optimized to detect clusters of a particular shape, it accurately located rectangular clusters of various height-to-width ratios. We anticipate that it will also detect clusters of other shapes; further study is needed.

REFERENCES

[1] Kulldorff M, A spatial scan statistic, Commun Statist: Theory and Methods. 1997; 26(6):1481-1496.

[2] Neill D, Moore A and Sabhnani M, Detecting elongated disease clusters, MMWR. 2005; 54(Suppl): 197.

[3] Kulldorff M, Huang L, Pickle L and Duczmal L, An elliptical spatial scan statistic, Stat Med. 2006 (in press).

[4] Tango T and Takahashi K, A flexibly shaped spatial scan statistic for detecting clusters, Int J Health Geogr. 2005; 4:11.

[5] Assunção R, Costa M, Tavares A, and Ferreira S, Fast detection of arbitrarily shaped disease clusters, Stat Med. 2006; 25:723-742.

[6] Merrill D, Use of a density equalizing map projection in analyzing childhood cancer in four California counties, Stat Med. 2001; 20:1499-1513.