Early Detection of Tuberculosis Outbreaks among the San Francisco Homeless

Mojdeh Mohtashemi[1, 2], Ph.D., Brandon Higgs[1], Ph.D., Jennifer Grinsdale[3], MPH, Masae Kawamura[3]

[1]MITRE; [2]MIT CS and AI Lab; [3]TB Control Section, San Francisco Dept. of Public Health

OBJECTIVE

The objective of this study is to investigate the effect of varying the spatial resolution in a variant of space-time permutation scan statistic applied to the tuberculosis (TB) data on the San Francisco homeless population on detection sensitivity, timeliness, and the amount of historical data needed for training the model.

BACKGROUND

San Francisco has the highest rate of TB in the US. Although in recent years the incidence of TB has been declining in the San Francisco general population, it has remained relatively constant in the homeless population. Spatial investigations of disease outbreaks seek to identify and determine the significance of spatially localized disease clusters by partitioning the underlying geographic region. The level of such regional partitioning can vary depending on the available geospatial data on cases including towns, counties, zip codes, census tracts, and exact longitude-latitude coordinates. It has been shown for syndromic surveillance data that when exact patients' geographic coordinates are used, higher detection rates and accuracy are achieved compared to when data are aggregated into administrative regions such as zip codes and census tracts [1]. While the benefits of using a finer spatial resolution, such as patients' individual addresses, have been examined in the context of spatial epidemiology, the effect of varying spatial resolution on detection timeliness and the amount of historical data needed have not been investigated.

METHODS

The data are provided by the San Francisco Department of Public Health (SFDPH), Tuberculosis Program and include confirmed outbreaks of TB of a specific strain during 1991-2002. The geospatial information in the data consists of precise locations of 392 homeless individuals infected with TB. We mapped each individual location to the corresponding census tract using ArcGIS v9.2 (ESRI). There were a total of 76 census tracts in the catchment of SFDPH. Variations to two space-time scan statistic methods [2, 3] were implemented and applied to the TB data aimed at identifying both simulated and confirmed TB outbreaks and comparing the results using two spatial resolutions: census tracts and exact patient's coordinates.

RESULTS

When exact patients' locations are used and the method is applied to real TB outbreaks, both the detection sensitivity and timeliness are improved and smaller amount of historical data is required compared to when census tracts are used as the spatial base for geographic partitioning. When outbreaks are simulated, while the detection timeliness is improved and smaller amount of historical data is required for training the model, the detection sensitivity results are not always consistent. When the size of the scanning window is sufficiently small (<0.2 km), irrespective of the number of census tracts to which the simulated cases are assigned, the detection method performs better using exact locations. However, if the size of the scanning window is large (>0.2 km) and the simulated cases are placed in one or two census tracts, the detection method performs better using census tracts.

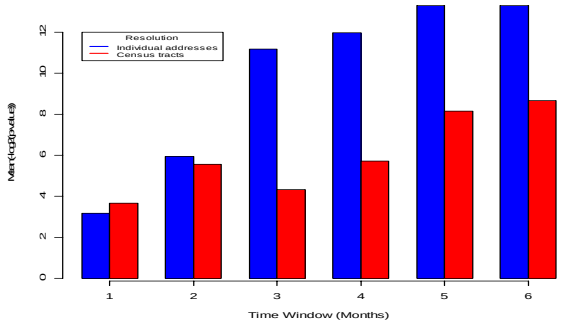

Figure 1 – Detection timeliness and sensitivity. The average -log2 transformed p-value distribution for individual addresses versus census tracts with an increasing window size.

CONCLUSIONS

We investigated the relationship between using exact locations of TB patients, the timeliness of identifying real TB outbreaks, and the amount of historical data required for early detection. We demonstrated that using higher spatial resolution generally results in higher detection sensitivity, but more importantly in timely detection of TB outbreaks even when the amount of available data is relatively small. Trading higher spatial resolution for better performance, however, is ultimately a tradeoff between maintaining patient confidentiality and improving public health when sharing data. Understanding such tradeoffs is critical to managing the complex interplay between public policy and public health. This study is a step forward in this direction.

REFERENCES

[1] Olson KL, Grannis SJ, Mandl KD. Privacy protection versus cluster detection in spatial epidemiology. *American Journal of Public Health* 2006; 96: 2002-2008.

[2] Kulldorff M, Heffernan R, Hartmann J, *et al*. A space-time permutation scan statistic for disease outbreak detection. *PLoS Med* 2005; 2(3).

[3] Neill DB, Moore AM. A fast multi-resolution method for detection of significant spatial disease clusters. *Advances in Neural Information Processing Systems* 2003; 16.