Aditya Chauhan
Indian Institute of Technology Mandi
aditya_chauhan@students.iitmandi.ac.in

RESEARCH STATEMENT

INTRODUCTION

My research interest revolved around data analysis and management. Primarily to analyse and mange the data of tuberculosis outbreak in India over four quarter of a year and than provide a relevant result compared to past research in this field. There are approximately 2 million cases annually of Tuberculosis in world and out of it 20% are from India. This is equal to approx. 40% of Indian population affected by TB. It is also estimated by *World Health Organisation (WHO)* that approximately 300,000 people die of TB in India every year.

In India TB care is provided by *Revised National TB Control Programme(RNTCP)* and also some private sector health facilities. Government's *RNTCP* started in 1997 and expanded in entire nation by March 2006. This programme follows WHO recommended DOTS guidelines and reaches over a billion people in around *632 districts/reporting units.* Data for the research too is collected from all these centres and analyses is done using Scan statistics algorithm.

The spatial scan statistics offers several advantages in getting results to good accuracy in short time analysis. Temporal,spatial and space-time scan statistics are now commonly used for disease detection and evaluation for many diseases. The detection of these clusters may be highly useful in surveillance of the disease, finding the factors behind the spread of the disease, and making suitable policies to control these factors.

METHODOLOGY

The algorithm mentioned above is used to find statistically significant clusters of TB and to identify their approximate location. Spatial analysis takes only area into account no time frame analysis is done, this is majorly used for generating TB cluster in study region. Spatial statistics imposes a circular window on the map and lets the centre of the circle move over the surrounding region. If the are contains the centroid of the data than whole are is included in window. For each circle centroid,the radius of circular window is varied from 0 to max radius such that window never include more than 50% of total population at risk. The spatial scan statistics is based on likelihood ratio test. As likelihood reaches max value it identifies most likely cluster. It's P-value is obtained using *Monte Carlo Hypothesis* testing technique*.* To find the distribution of test statistics, 999 random Monte Carlo replicates of the data set under the null hypothesis of no significant clusters are generated, calculating the test statistics of each replica.

Recognition of clusters was done using *Poisson Probability* model assumption. The maximum spatial cluster size was set to include up to 50% of population for both excess and deficits. For geographic and cartographic outputs GIS softwares were used.

RESULTS

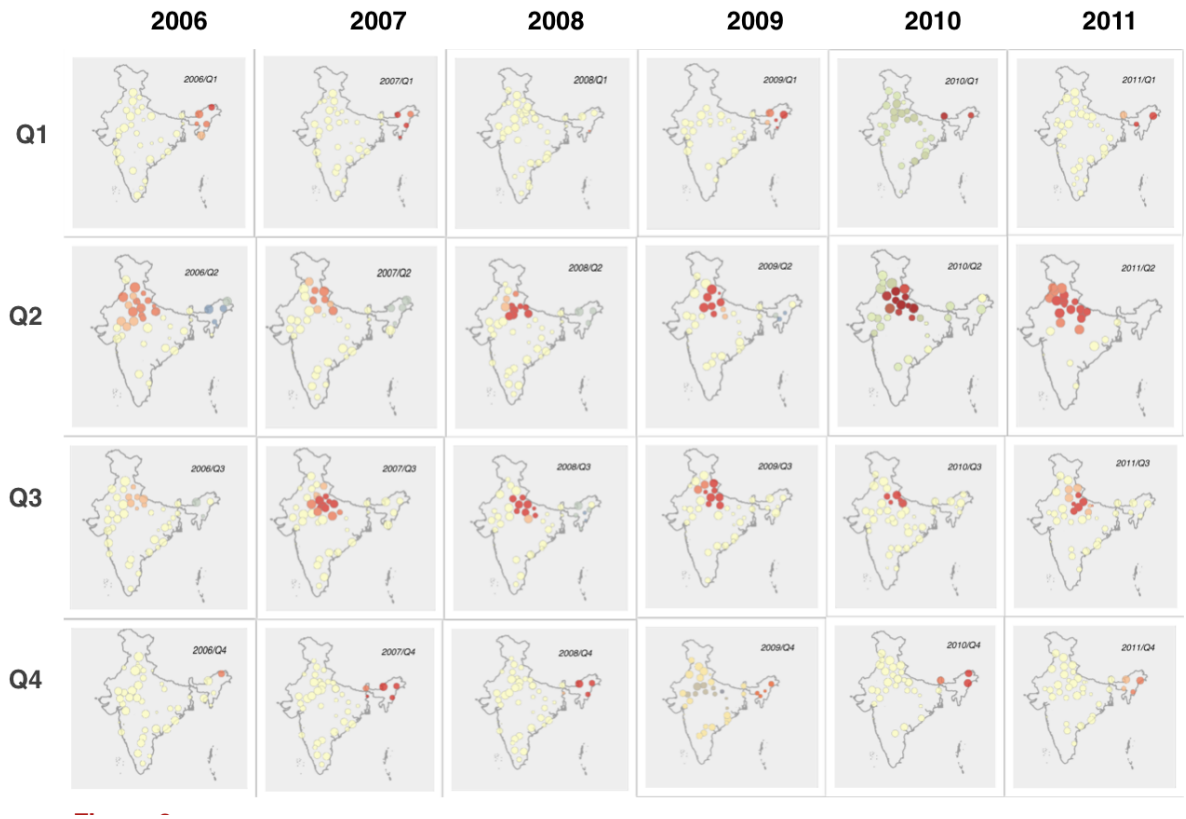
With spatial cluster size to be ≤ 50% of the total population. The spatial cluster analysis identified the most likely cluster for high occurrence of TB in northeastern states for winters(quarter 1 and quarter 4) and norther Indian regions for summers(quarter 2 and quarter 3). The region for all the years from 2006 to 2011 are common though clusters move a bit or scatter a bit.

Relative risk is a major factor used to make above analysis as it helps in statistical analysis of binary outcomes where outcome of interest has relatively low probability. It is most suited method

for clinical trials combined with scan statistics it gives a decent result to consider for analysis. Scan statistics is implemented using mean spatial Semivariogram ($\gamma_s(h)$)

$$\gamma_s(h) = \frac{1}{2NtNh} \sum_{j=1}^{Nt} \sum_{i=1}^{Nh} \left[Z(x_i, t_j) - Z(x_{i+h}, t_j) \right]^2$$

Here Nt is number of time periods and Nh is number of pair of sub districts at distance h.



Above figure shows the movement of red clusters(major clusters) for four quarters through time period 2006 to 2011. To analyse spatial scan statistics is used but for plotting the clusters with their variable likelihood is done by calculation likelihood using relative risk factor.

REFERENCES

[1] Mukherjee, A "Outcomes of different subgroups of smear-positive retreatment patients under RNTCP in rural West Bengal, India", Rural and Remote Health www.ncbi.nlm.nih.gov/pubmed/19260766 - See more at: http://www.tbfacts.org/tb-india/#sthash.CAWS6siF.dpuf

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

[3] Wallenstein S: A test for detection of clustering over time. *Am J Epidemiol* 1980, 111:367–372

[4] Weinstock MA: A generalized scan statistic test for the detection of clusters. *Int J Epidemiol* 1981, 10:289–293.

[5] http://www-users.cs.umn.edu/~gandhi/courses/CS8701/g4_e2_semivariogram.pdf