SPATIO-TEMPORAL ANALYSIS OF CHOLERA

A CASE STUDY IN VELLORE, INDIA, 2000-2014

RESEARCH QUESTIONS:

What is the spatial extent of cholera incidence around CMC Vellore?

What were key trends within cholera incidence clusters in 2000-2014?

INTRODUCTION:

Case tracking and source epidemiology provide much needed data to understand the underlying patterns driving emergence, spread and persistence of infectious disease. Detailed, reliable and timely surveillance data can enable rapid source identification and interventions to prevent a local event from spreading. Modern surveillance systems can also be effectively used to characterize and study spatiotemporal patterns and routes of transmission in disease endemic regions. In the state of Tamil Nadu, India, integrated environmental surveillance for *Vibrio cholerae*, the causative agent of cholera, has been established in collaboration with the Christian Medical College of Vellore (CMC) and Tufts University since 2012. Routine regional environmental and health monitoring undertaken by CMC has demonstrated the presence of pathogenic O1 and O139 *V. cholerae* in heavy use water bodies in and around Vellore. Christian Medical College is a destination for patients from all regions of the country seeking medical care. Thus, patients' self-reported addresses do not necessarily reflect the place of exposure. As such, connecting patterns of outbreak in clinical disease to prevailing environmental conditions may demonstrate patterns which do not capture or accurately characterize the drivers of disease outbreak. This study attempts to define the spatiotemporal extent of the Vellore cholera cluster to define a radius of most likely exposure for cholera patients in CMC Vellore, and identify temporal trends within this cluster in 2000-2014.

METHODS:

Data Collection | A total of 1915 cases of cholera were documented during 1992, 1996-1999, and 2000-2014 from electronic databases and logbooks available at CMC during a team research trip to Vellore in January 2015. The following fields were recorded: date received at CMC, patient town and region, age, sex, and pathogenic strain of Vibrio cholerae. All the documented cases exhibited diarrheal disease with cholera-presenting symptoms, and verified V. cholerae isolates (O1, O139, and/or Non-Agglutinating).
Data cleaning | All data were pre-processed for analysis. Acronyms were expanded, correct state information attributed, and ages standardized. Records with missing Date Received and Patient Region fields were discarded. This left 1777 remaining records, which were mapped in this analysis.
ArcGIS statistical analysis | All points were used to generate a Spatial Weights Matrix, based on a space-time window of 30 days and an experimental distance threshold of 300 km. This matrix was used with the following tools: Cluster and Outlier Analysis (Anselin Local Moran's I), Cluster and Outlier Analysis with Rendering, Hot Spot Analysis, Optimized Hot Spot Analysis, and Grouping Analysis.
Census data compilation | In preparation for SaTScan analysis of the data, total population in major agglomerations from 1991, 2001, and 2011, was compiled from the Indian Census. A linear trend was assumed between these years, and annual populations for 2000 – 2014 were interpolated. This data was mapped, and the case incidence points were spatially joined with the nearest major agglomeration's census data. The final population was specific to the patient's listed city, and the year they were admitted into CMC Vellore.
SaTScan statistical analysis | SaTScan™ is a free software that analyzes spatial, temporal and space-time data using spatial, temporal, or space-time scan statistics. It is used primarily by the public health and epidemiology community to analyze spatiotemporal patterns in disease incidence. In SatScan, a Purely Spatial, a Purely Temporal, and a Space-Time Clustering analysis using a discrete Poisson distrubtion for the case data were implemented to analyze clustering patterns in this data.

RESULTS:

Based on both ArcGIS and SaTScan results, it can be concluded that Vellore's "cholera radius" extends to about 175 km from CMC. Thus, the probability of patients traveling from states as far as West Bengal, Assam, Tripura, and Sikkim to CMC for cholera treatment is very low. Patients from these regions are more likely to have been traveling within this radius, where they may have acquired cholera and sought hospitalization at CMC Vellore. It is important to note that this cholera radius includes the city of Chennai, a major industrial, economic, and tourist hub. Located only three hours from Chennai, CMC Vellore may also see a lot of cases from individuals who may not have been exposed to cholera in their home countries or states, but encounter it for the first time while traveling to Chennai.

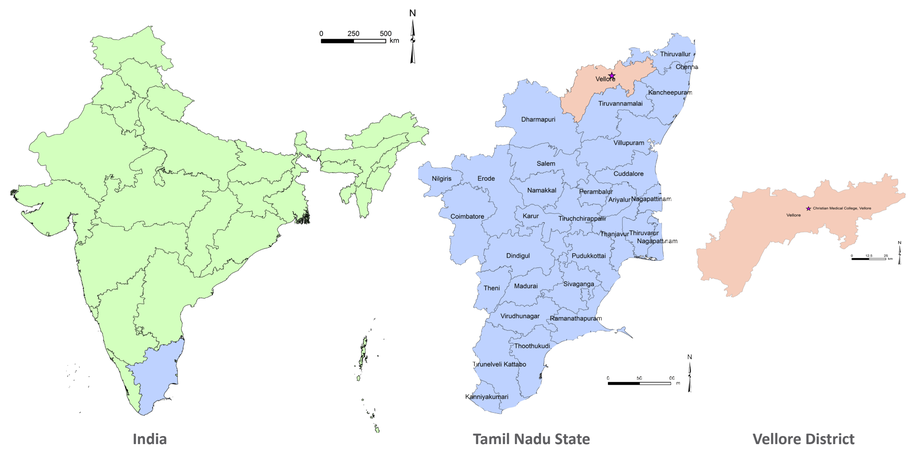
The localized spike observed between May- June 2010 in the purely temporal analysis likely indicates an outbreak, or a particular migratory trend among the local population. The space-time cluster indicates another potential localized outbreak in March- October 2010. This overlap indicates that CMC Vellore sees generally elevated numbers of cholera patients in the summer, which may help plan the hospital's cholera response during these months.

NEXT STEPS:

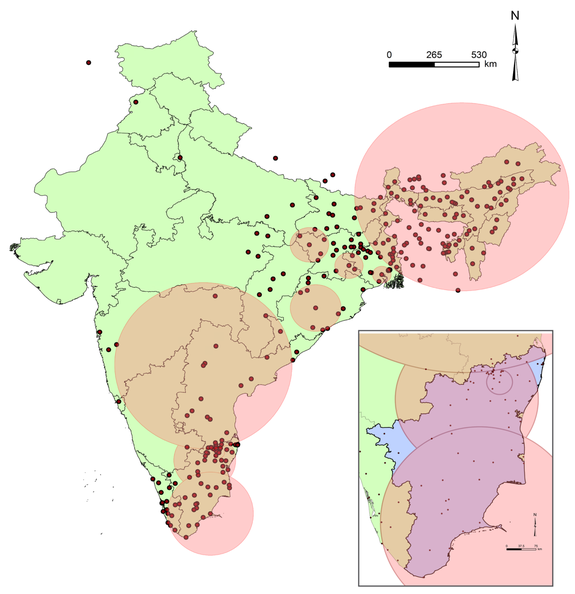
This analysis is the first step towards a collaborative paper regarding the spatial challenges of *Vibrio cholerae* surveillance and source-tracking. Further analytical methods for this data include Bayesian statistics, and predictive spatial and space-time clustering analyses to forecast cholera clusters based on historic trends.

These results demonstrate the need for a standardized national and international reporting mechanism for cholera. Cholera is a mandatory reportable disease in India; however, there is currently no standardized reporting structure, which would be very useful to identify spatiotemporal patterns and outbreaks as they occur. This standardized national survey would be an important step towards increasing reporting rates for cholera, and documenting occurrence for surveillance and source-tracking in case of an outbreak. This survey should include fields for current address, duration of location at current address, and permanent address. The time of onset of initial symptoms, and activities prior to and following symptomatic phases must also be documented for optimal source-tracking.

PROJECT LOCATION



PURELY SPATIAL CLUSTERING



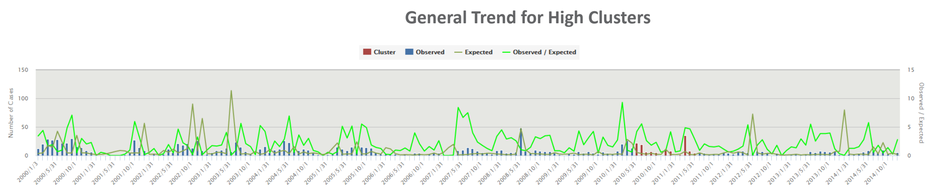
Three major clusters with a p-value of less than 0.05 are found around Vellore.

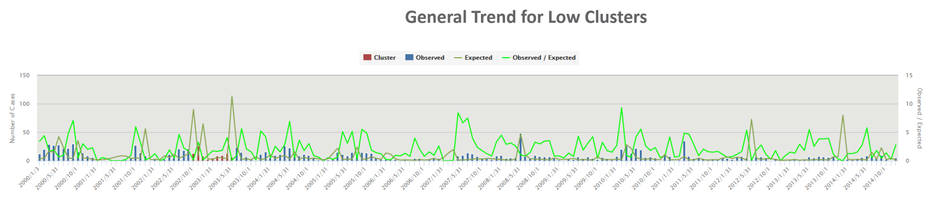
The smallest, most localized cluster around CMC Vellore has a radius of 32 km.

The larger cluster around CMC Vellore has a radius of 178 km.

The largest cluster, which includes the lower portion of Tamil Nadu state, has a radius of 250 km. The cases in all three clusters exceeded exptected values, and had a relative risk of 6 or greater. Therefore, individuals in these three spatial clusters are most likely to come to CMC Vellore for cholera treatment, compared to the other clusters.

PURELY TEMPORAL CLUSTERING

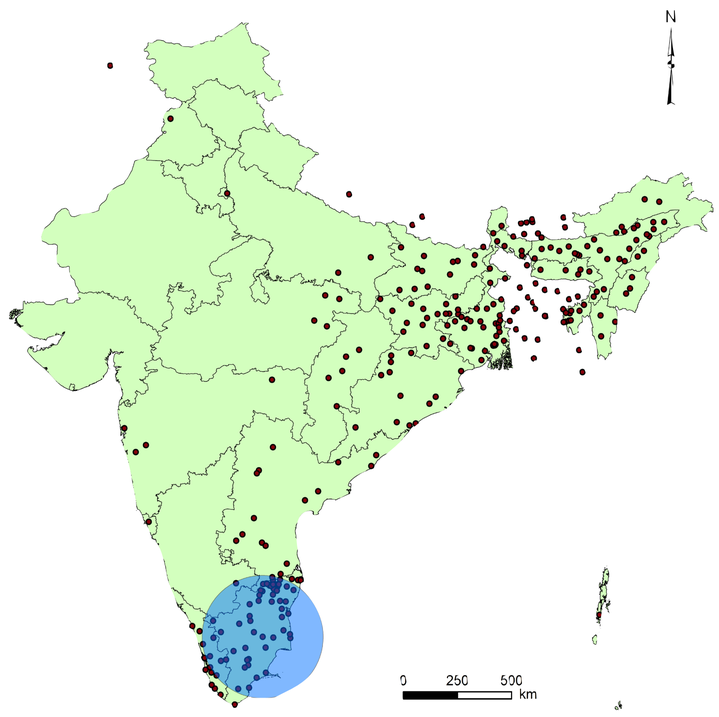




Iterative purely temporal analysis with an average annual decrease of 4% indicates the following clustering trends:

- 2002/10/1 to 2003/6/10 had lower than expected cases of cholera. 294.04 cases were expected, and only 90 were documented. This trend was particularly significant during 2003/5/31 to 2003/6/30, when 113.49 cases were expected, and only 7 were documented at CMC Vellore.
- Between 2010/5/31 and 2011/6/30, a spike in cholera cases was observed. 46.20 cases were expected; however, 128 cases were documented.

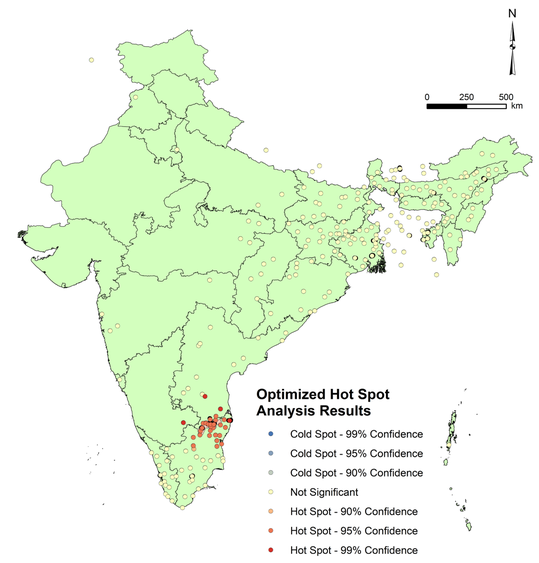
SPACE-TIME CLUSTERING



Only one significant space-time cluster is detected from this record. The temporal extent of this cluster is from 2000/3/1 to 2000/10/30. During this time, only 17 cases were expected from the region; however, 131 cases of cholera were documented at CMC Vellore. The relative risk during this time is 8.33, which is significantly higher than the average of 6.20 that was generally observed during the six months before and after this period. This cluster has a p-value of almost 0, with a radius of 274.29 km.

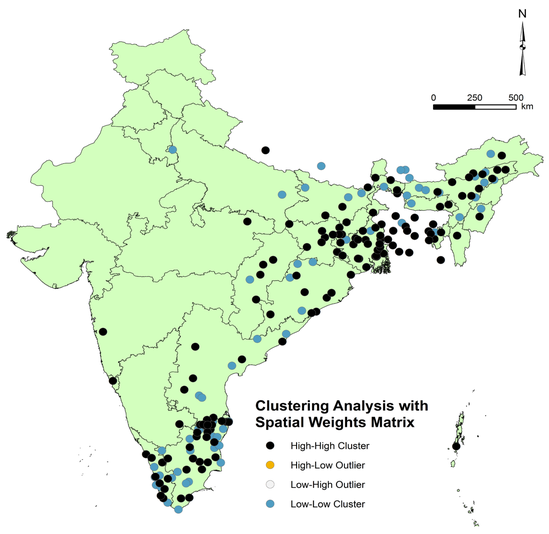
ARCMAP CLUSTERING ANALYSIS

Optimized Hot Spot Analysis



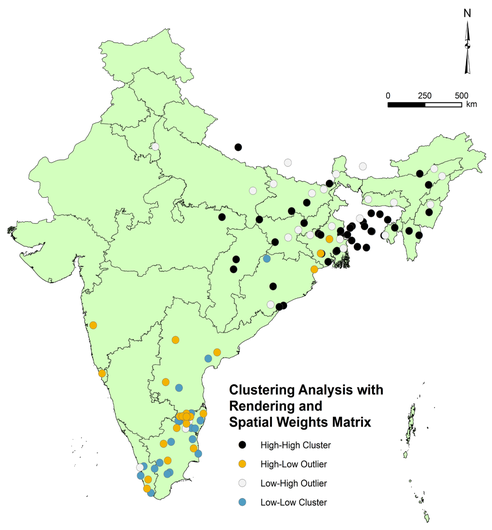
One major cluster around CMC Vellore is observed, and most other regional clusters are categorized as insignificant due to low spatial weights.

Cluster & Outlier Analysis (Anselin Local Morans I)



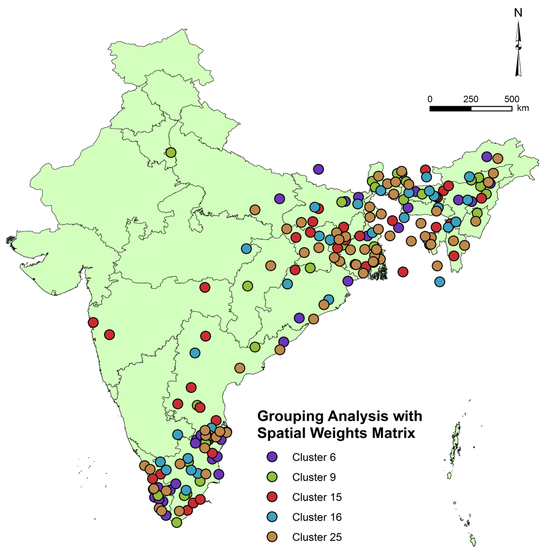
One cluster is observed around CMC Vellore, and additional regional clusters are also observed around Kerala, West Bengal, and other north-eastern states.

Cluster & Outlier Analysis with Rendering



One major High-High cluster is observed in West Bengal, and a few High-Low clusters are observed around Vellore. However, this does not reflect the highly localized cluster around CMC Vellore, which would be predicted to be a High-High cluster.

Simple Hot Spot Analysis



The ideal number of clusters was identified to be 42, since the data is very unevenly distributed across the country. Clusters 6, 9, 15, 16, and 25 had the highest number of points in them; however, a spatially distinct cluster around Vellore is not observed in this analysis.

Data Sources | Christian Medical College Vellore hospital records | Kulldorff M. and Information Management Services, Inc. SaTScanTM v8.0: Software for the spatial and space-time scan statistics. http://www.satscan.org/, 2009 | Kulldorff M. A spatial scan statistic. Communications in Statistics: Theory and Methods, 26:1481-1496, 1997 | Osei, F. (2010). Chapter 3: Spatial and space-time clustering of cholera. In Spatial statistics of epidemic data: The case for cholera epidemiology in Ghana.
Acknowledgements | *Innovative Public Health Engineering Strategies to Reduce Waterborne Disease Burden in Developing Countries* grant, sponsored by the Tufts Institute for the Environment. PIs: E.N. Naumova, K.D. Pennell, D.M. Gute; Faculty: N. Cápiro, A. Camilli, H. Ward | I would also like to thank Tania Alarcon, Hanna Ehrlich, and Meghan Hartwick for their involvement in and contribution to this project.

Tufts University | Institute for Innovation
Cartographer | Aishwarya Venkat
Date Revised | May 6, 2015