J Biochem Tech (2018) Special Issue (2): 20-24
ISSN: 0974-2328

Application Space - time Scan Statistics for High-Risk Clusters of Tuberculosis (TB) Disease Incidence in Iran, 2006-2013

Danesh Shahreki Bijan*, Abadi Alireza, Bashiri Yousef and Narges Saber Molashahi

Received: 19 February 2018 / Received in revised form: 24 Jun 2018, Accepted: 28 June 2018, Published online: 05 September 2018

Abstract

Background: Tuberculosis (TB) is currently one of the greatest problems in public health. Mycobacterium tuberculosis infects about one third of the world's population, of whom more than 80% are living in developing countries. The incidence and prevalence of TB are very different in various parts of Iran and also throughout the world. East provinces are one of the areas in the country where the risk of TB rate is the highest because of its ethnically varied population and special location. The present research sought to identify high risk clusters of Tuberculosis with mapping using Space-time Permutation scan statistics. Exact locations of patients, primary residences at the time of diagnoses are routinely collected as part of the TB surveillance program to ability clusters and detect disease outbreaks Tuberculosis early is important in order to decrease morbidity and mortality through timely implementation of disease prevention and control measures. It has been shown for syndromic surveillance data that when exact geographic coordinates of individual patients 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. Materials and Methods: The present research is of descriptive type. The required data were gathered from the registered TB reports of TB Control Office in the Center for Communicable Disease of the Iran Ministry of Health and Medical Education (MOHME). The data were extracted at province level in over the time period of 2006–2013. We apply a variation of Space-time Permutation scan statistic to the TB data in which a patient location. SAT Scan software was used to analyze the data and to identify high risk clusters. ArcGIS10 was utilized to map the distribution of Tuberculosis and to demonstrate high risk clusters. Results: In the purely spatial analyses, the most likely clusters were in Sistan & Balouchestan provinces (in 2006, 2008), the Golestan, Khorasan Razavi, (in 2009), and the Khozestan (in 2010 to 2013). In the space-time analysis, the most likely clusters were Golestan and Khozestan industrial area (in 2010-2013). Accordingly, the most likely clusters and High risk regions included East and West-South provinces, particularly Golestan, Khorasan Razavi, Sistan & Balouchestan and Khozestan. It was statistically significant at the p-value below 0.05. Conclusion: The spatial and space-time Permutation scan statistics are effective ways of describing circular disease clusters. Since, in reality, infectious diseases might form other cluster types, the effectiveness of the method may be limited under actual practice. The sophistication of the analytical methodology, however, is a topic for future study.

Keywords: Space-time Permutation scan statistic; Clusters detection; Mapping; TB disease

Introduction

Tuberculosis (TB) is currently one of the greatest problems in public health. Mycobacterium tuberculosis infects about one third of the world's population, of whom more than 80% are living in developing countries. TB is one of the top four diseases for infection-induced mortality in the world today (Dye and ef al, 2005) There are currently about 54 million people infected with the bacterium Mycobacterium tuberculosis with approximately 8 million new infections occurring each year. TB kills nearly 2.4 million people worldwide annually. The incidence and prevalence of TB are very different in various parts of Iran and also throughout the world. East provinces are one of the areas in the country where the risk of TB rate is the highest because of its ethnically varied population and special location. It has been shown for syndromic data that when exact geographic coordinates of individual patients are used as the spatial base for outbreak detection, higher detection rates and accuracy are achieved compared to when data are aggregated into administrative regions such as zip codes and census tracts. We examine the effect of varying the spatial resolution in the TB data within the Iran population on detection sensitivity, timeliness, and the amount of historical data needed to achieve better performance measures. We aimed to describe the prevalence of TB and related demographic data in all of provinces and spatially in this location. Spatial investigations of disease outbreaks seek to

---

Danesh Shahreki Bijan*
Department of Biostatistics, Zabol University of Medical Sciences, Zabol, Iran

Abadi Alireza
Assistant Professor, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Bashiri Yousef
Social Determinants of Health Research Center, Yasuj University of Medical Sciences, Yasuj, Iran.
Department of Biostatistics, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Narges Saber Molashahi
Management of medical Sciences, Zabol University of Medical Sciences, Zabol, Iran

***E-mail:** bijan.shahreki@zbmu.ac

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. When exact patients, locations have been used in cancer surveillance, the detection sensitivity was not appreciably higher than that obtained with larger and more conventional regional partitions such as census block groups (Dye & et al, 1999), though the benefit of localized rate variation (i.e. geographic excess or shortfalls in cancer incidence for small areas) was shown in an earlier study (Murray & Lopez, 1998). In other works, there were few performance differences observed for larger aggregation comparisons such as block group, census tract, zip code, and town (Lopez & et al, 2001; Japan Anti-Tuberculosis Association, 2005), More recently, Olson et al, using the method of Space scan statistic (Japan Anti-Tuberculosis Association, 2005) applied to syndromic data, has shown that when patients exact locations are used, higher detection sensitivity is achieved as compared to center points of larger geographical regions such as zip codes and census tracts (Driver & et al, 2006). The authors demonstrated that the advantage in using higher resolution cluster detection results primarily in a reduced distortion effect that is induced by the use of large detection windows (i.e. spatial scanning windows), as compared to smaller detection windows. This problem occurs when, for example, two cases are geographically close to one another, while they reside in two separate zip codes (or census tracts). In such situations, if the geographic partitioning is by zip code (or census tract), the detection window has to be rather large to encompass both administrative regions because the cases are represented by the cancroids of these regions, while a smaller detection window can capture such localized cases when the exact individual addresses are used. In Space Time Permutation surveillance of disease outbreaks, however, the interdependency between both time and space are manifested by disease clusters that are localized in time and space. Such disease localizations can be investigated through dynamic partitioning of the underlying geographic regions, where different degrees of spatial resolution can be coupled with varying levels of temporal scale to examine both the detection sensitivity and timeliness (i.e. the speed by which an outbreak is detected after it has occurred). While the benefits of using higher spatial resolutions, such as patients’ individual addresses, have been examined in the context of spatial epidemiology, the spatiotemporal effects of disease localizations have not been studied under different degrees of spatial aggregation. As such, the effect of varying degrees of spatial aggregation on detection timeliness has not been investigated. At the same time, any detection method must rely on a pool of historical data to both establish a baseline of normal disease variability and estimate the model parameters. However, the amount of available data varies across surveillance programs, and historical data are often in short supply. Therefore, in addition to detection sensitivity and timeliness, the dependency on the amount of historical data must also be examined when varying spatial resolution. In this work, we use a modification of a Space-time Permutation scan statistic (Fukuoka Prefectural Government, 2005; Geng & et al, 2002; Barnes & et al, 1996; Barnes & et al, 1997; Mosquera, Rodrigo & Gonzálvez, 1994) to examine the effect of varying degrees of spatial resolution (census tracts of patients’ residences versus exact locations) on the detection sensitivity, detection timeliness, and the amount of data needed for training the model, applied to the TB data on Iran population for 2006–2013, using both simulated and confirmed outbreaks. This study is population-based and prospective, where historical data and documented outbreaks are used, but the detection algorithm is applied to the data as would be run in a real-time environment. Exact locations of patients, primary residences at the time of diagnoses are routinely collected as part of the TB surveillance program to ability clusters and detect disease outbreaks Tuberculosis early is important in order to decrease morbidity and mortality through timely implementation of disease prevention and control measures. It has been shown for syndromic surveillance data that when exact geographic coordinates of individual patients 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. Many national, state, and local health departments are launching disease surveillance systems with daily analyses of hospital emergency department visits, ambulance dispatch calls, or pharmacy sales for which population-at-risk information is inaccessible or inappropriate.

Materials and Methods

Data sources

The datasets were used in the study (Table 1) consists of comprehensive information on all of the forty one Health and Medical university in provinces of Iran over the time period of 2006–2013 provided by the Iran Ministry of Health and Medical Education (MOHME), Centre Communicable Diseases Management TB and Lung Diseases, routinely collects comprehensive information on TB cases and their contacts including demographics (e.g., age, gender, race), personal risk factors (e.g., intravenous drug use, HIV status, alcohol intake), laboratory results (e.g., skin test, chest x-ray), time of diagnosis, and primary residence addresses. Specifically, tests for spatial randomness were applied to six different geographic data sets (Table 1). Data on TB cases were obtained from the TB surveillance system. This information can aid in outbreak investigation to identify patterns and hubs of transmission often hidden in a network of complex interactions between primary infected cases and their contacts.

Methods and Findings

We apply a variation of Space-time Permutation scan statistic to the TB data in which a patient location is either represented by its exact longitude and latitude or by the cancroids of its census tract and propose a prospective Space-time Permutation scan statistic for the early detection of Tuberculosis and anymore disease outbreaks that uses only case numbers, with no need for population-at-risk data. The data were collected according to the International Classification of Diseases (ICD-10-CM) Diagnosis

Code A15.0 in Iran. In this study, the cases of occurrence in the whole country were between 2006 and 2013 in 41 Health and Medical University to 31 provinces.

For rare diseases such as the different types of cancer or Tuberculosis, it is assumed that the number of patients in each region follows a Poisson distribution considering the fact that the data are numerical and discrete. Spatial scan statistics are used to identify regions where the number of patients is significantly higher than the expected. Therefore, in spatial scan statistics, the assumption of the sameness of disease risk inside and outside the cluster is tested against the assumption of a higher risk of disease inside the cluster than outside. Then, based on the maximum value of logarithm likelihood ratio would be identified the high risk cluster. Space-time Permutation scan statistics is a simple generalization of spatial scan statistics. Space-time Permutation scan statistics is defined as one having a cylinder like window and a circular base and a height corresponding to time. This window moves in time and space for each geographical region in the whole country. It includes all possible time intervals for the occurrence of a disease from the time the study begins till the present time. The size of the scan window is decided upon by the importance of the target disease (Neill and Moore, 2004). In the present study, its size is set to be 50% of the whole population at risk. The required data were firstly entered into Excel software. Then, they were prepared for statistical analysis by means of SAT Scan software (Kulldorff, 2014). Once the high-risk clusters were spotted, the test of cluster significance was run using Monte Carlo method. SAT Scan software was used to analyze the data and to identify high risk clusters. The primary residences of these individuals were used to identify their geographical coordinates (latitudes and longitudes) using Arc GIS v10.0. ArcGIS10.0 was utilized to map the distribution of Tuberculosis and to demonstrate high risk clusters. It makes minimal assumptions about the time, geographical location, or size of the outbreak and it adjusts for natural purely spatial and purely temporal variation. We show that the detection sensitivity and timeliness of the method generally improve when exact locations are used to identify both simulated and real TB outbreaks; however, better performance measures were attained under simulated cases as compared to actual outbreaks. The new method was evaluated using daily analyses of hospital emergency department visits in New York City. Four of the five strongest signals were likely local precursors to citywide outbreaks due to rotavirus, nor virus, and influenza. The number of false signals was at most modest.

Results

Before reporting the results, we refer to the meaning of background risk for TB. Even in a society without known risk factors for TB, some TB cases (i.e., the expected number) are observed. The TB clusters are calculated based on the expected numbers. Since the absence of background risk does not necessarily indicate the absence of disease, we must again be careful in interpreting the results. The results of the purely spatial analysis of TB data from 2006 to 2013 are shown in Tables 2 and 3, and in Figure 2. The most likely significant clusters for a high occurrence of TB are listed in Table 2, and depicted on the map in Figure 2. For the purely spatial analyses, the most likely clusters were in Sistan & Balouchestan provinces (in 2006, 2008), the Golestan, Khorasan Razavi, (in 2009), and the Khozestan (in 2010 to 2013). In the space-time analysis, the most likely cluster were Golestan and Khozestan industrial area (in 2010-2013). In the purely spatial analyses, the most likely clusters and High risk regions included East and West-South provinces, particularly Golestan, Khorasan Razavi, Sistan & Balouchestan and Khozestan. It was statistically significant at the p-value below 0.05. We also detected several secondary clusters, which were not significant.

Conclusion

The spatial and Space-time Permutation scan statistics are effective ways of describing circular disease clusters. Since, in reality, infectious diseases might form other cluster types, the effectiveness of the method may be limited under actual practice. The sophistication of the analytical methodology, however, is a topic for future study.In summary, using data from 2006 to 2013, we used the Space-time Permutation scan statistic analysis technique, to examine of TB in Iran. The most likely significant clusters and high-risk regions occurrence of TB was identified which includes 4 provinces: Golestan, Khorasan Razavi, Sistan and Balouchestan in East and Khozestan in west - South. Since the efficacy of TB control measures in specific areas could be evaluated by reviewing the control measures and a longitudinal change in TB prevalence, the space-time scan statistic can contribute to health program evaluation. In the context of a disease such as TB with chronic characteristics, we denote an outbreak is detected early if it is identified within months (less than a year) from the start of the outbreak. If such results hold up over longer study times and in other locations, the Space-time Permutation scan statistic will be an important tool for local and national health departments that are setting up early disease detection surveillance systems. It was statistically significant at the p-value below 0.05. Although the method used here could help prioritize the assignment and investigation of diseases, the applicability of our methodology might be limited. We need to develop more sophisticated analytical methodology for future studies.

Table 1- Summary of the data sets

Data type	Ages	Geographic area	Spatial resolution	Locations (n)	Year	Total cases(n)
Incidence	All	Iran	provinces	41	2006-2007	10693
Incidence	All	Iran	provinces	41	2007-2008	11086
Incidence	All	Iran	provinces	41	2008-2009	10987
Incidence	All	Iran	provinces	41	2009-2010	10555

Incidence	All	Iran	provinces	41	2010-2011	10044
Incidence	All	Iran	provinces	41	2011-2012	10132
Incidence	All	Iran	provinces	41	2012-2013	10412

Source. Iran Ministry of Health and Medical Education (MOHME): published of TB 2014.

Table 2- High-risk Region Cluster for the Occurrence of Tuberculosis in Iran using Space-Time Permutation Scan Statistics

Provinces	Time span	Observe cases	Expected cases	Relative risk	P-value
Golestan, Khorasan Razavi, Sistan & Balouchestan, Khozestan	2006-2013	53365	45704	1.61	.000

Table 3- The Number of Case and Incidence rate of Tuberculosis in Iran between 2006 –2013

Year	2006 -2007	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013
Number of Case	10693	11086	10987	10555	10044	10132	10412
Incidence rate diseases	14.34	14.63	14.43	13.73	12.90	13.99	14.59

* Incidence rate diseases per hundred thousand (100,000)

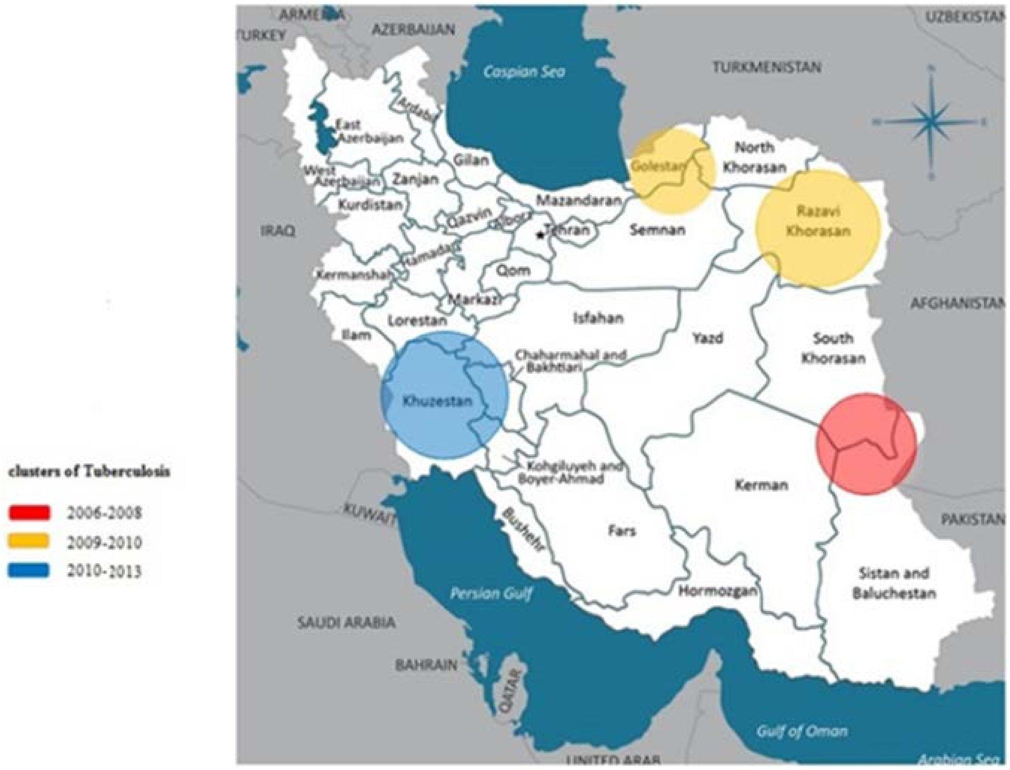

Figure 1: Locations of the detected clusters of TB cases, using data Source from 2006 to 2013 based on the Space-time Permutation analysis

References

Barnes, P. F., El-Hajj, H., Preston-Martin, S., Cave, M. D., Jones, B. E., Otaya, M., & Eisenach, K. D. (1996). Transmission of tuberculosis among the urban homeless. *Jama*, *275*(4), 305-307.

Barnes, P. F., Yang, Z., Preston-Martin, S., Pogoda, J. M., Jones, B. E., Otaya, M., & Cave, M. D. (1997). Patterns of tuberculosis transmission in Central Los Angeles. *Jama*, *278*(14), 1159-1163.

Driver, C. R., Macaraig, M., McElroy, P. D., Clark, C., Munsiff, S. S., Kreiswirth, B., & Zhao, B. (2006). Which patients' factors predict the rate of growth of Mycobacterium tuberculosis clusters in an urban community?. *American journal of epidemiology*, *164*(1), 21-31.

Dye C, Scheele S, Dolin P, Pathania ystematic analysis of population health data. *The Lancet*, *367*(9524), 1747-1757.

Dye, C., Watt, C. J., Bleed, D. M., Hosseini, S. M., & Raviglione, M. C. (2005). Evolution of tuberculosis control and prospects for reducing tuberculosis incidence, prevalence, and deaths globally. *Jama*, *293*(22), 2767-2775.

Fukuoka Prefectural Government . Statistics of TB 2005 in Fukuoka Prefecture. Fukuoka Prefectural Government; 2005.

Geng, E., Kreiswirth, B., Driver, C., Li, J., Burzynski, J., DellaLatta, P. ... & Schluger, N. W. (2002). Changes in the transmission of tuberculosis in New York City from 1990 to 1999. *New England Journal of Medicine*, *346*(19), 1453-1458.

Japan Anti-Tuberculosis Association: Statistics of TB 2005 in Japan Japan Anti-Tuberculosis Association; 2005. http://www.jata.or.jp/english/

Mosquera, J. A., Rodrigo, L., & Gonzálvez, F. (1994). The evolution of pulmonary tuberculosis in coal miners in Asturias, Northern Spain. *European journal of epidemiology*, *10*(3), 291-297.