IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Identification Of Different Forms Of Tuberculosis Spatial Clusters In Tamil Nadu Using Discrete Poisson Model And Log-Likelihood Ratio Test

1Dr.K.Chandrasekaran, 2Mrs.R.Mahalaxmi, 3Dr.P.Paulkumaran, 4Mr.P.Palaniyandi, 5Dr.C.Ponnuraja

1Principal Technical Officer, 2Scientist-C, 3Scientist-F, 4Technical Officer

1ICMR-NIRT,

2ICMR-NIRT,

3ICMR-NIRT,

4ICMR-NIRT,

5ICMR-NIRT

Abstract:

Backgound & Objectives: The Government of India has announced its National Strategic Plan (NSP) for controlling TB in India by the year 2025. To achieve NSP, it will be useful for program implementers to identify various forms of cluster locations with high risk of transmission and to initiate appropriate measures to control TB. In this study, our aim is to identify and rank spatial clustering of various forms of tuberculosis among the districts of Tamil Nadu State, India for the years 2016, 2017 and 2018.

Methods: We have considered aggregate (count) data in the form of tuberculosis case notification for the districts of Tamil Nadu state and used discrete Poisson model and Log-Likelihood Ratio test to identify the high risk spatial TB clusters. We have used SaTScan software for this purpose.

Results: We have identified Madurai, Dindigul and Thiruvarur Districts for new sputum positive pulmonary tuberculosis case clustering, Madurai and Thiruvarur Districts for previously treated smear positive pulmonary tuberculosis case clustering (towards the direction of containing the transmission dynamics of tuberculosis), Vellore District for clinically diagnosed tuberculosis case clustering (towards the direction of enhancing diagnostic facilities such as providing more number of X-ray mechines and incresing the number of trained chest physicians) and Vellore and Chennai Districts are for extra-pulmonary TB case clustering (towards the direction of setting up more surgical

facilities to handle such cases) in the study period. Also we have produced thematic maps showing various forms of TB clusters using QGIS software.

Interpretation & Conclusions: This method can be used to identify TB clusters for aggregate data for a given time period for prioritizing enhanced intervention activities.

K**eywords:** TB spatial cluster, Discrete Poisson model, Log-likelihood Ratio Test and Spatial scan statistic.

Introduction

In India, Tuberculosis (TB) is one of the major diseases of concern and India accounts for more than a quarter of global cases. In this background, identifying high risk TB clusters by incorporating spatial component as a variable in a model based approach and understanding the etiology behind the same is essential to bring the disease of concern under control in India. A cluster is being a bounded group of occurances of a disease to occur characterestically, in sufficient size and concentration to be unlikely to have occured by chance and related to each other through some social or biological mechanism / or having a common relationship with some other event or circumstace.

In this section we have discussed the tests available for testing the homogenity of disease risks, its disadvantages and alternative method to overcome the limitation of the said tests. Pearson's Chi-square statistic[1], Potthof-Whittinghill's test[2] and Tango's statistic[3] are available to test for homogeneity of disease risk for a given region. These methods provide general measurement for homogeneity and clustering in the study area, but may fail to detect clustering when the actual clusters are small and scattered. Also they may not detect the location of clusters or answering the question related to their statistical significance. Further, these tests have been showing the relative risks in the study region as non-homogeneous for both higher and lower ones. Openshaw et al[4] have developed a graphical method called the Geographical Analysis Machine (GAM) and it uses multiple overlapping circles of variable size as quadrats. Also Turnbull et al[5] have developed Cluster Evaluation Permutation Procedure (CEPP) for similar purpose and both suffers from multiple hypothesis testing. Consequently these two procedures were not able to assess the overall significance of the results. Besag and Newell[6] developed a method and proposed to "fix the number of cases in a circle and search for those with such a small risk population that make them highly significant". Kulldorff and Nagarwalla[7] and Kulldorff[8] proposed a methodology based on Scan Statistics[8] and it is able to detect clusters even in a small region, identify the exact location of clustters and answer the question related to their statistcal significance. The SaTScan[9] software implements spatial scan statistic which can be used to analyze spatial, temporal and space-time aggregate (count data) data and exact locations of disease occurrence (point data). In this study our objective is to identify various forms of high risk TB clusters such as new smear positive, previously treated, clinically diagnosed, extrapulmonary in Tamilnadu State and considering district as a unit for the years 2016, 2017 and 2018.

Material & Methods

Data

The Central TB Division (CTD), Ministry of Health and Family Welfare, Govt. of India is publishing yearly report of Revised National Tuberculosis Control Programme (RNTCP) and renamed as National Tuberculosis Elimination Program (NTEP) in 2020. In this report, for the study period the TB data available are new smear positive (NSP), previously treated (PT) to initiate measures for containing transmission dynamics, clinically diagnosed (CD) for enhancing diagnostic facilities such as x-ray machines and to increase trained chest physicians and extrapulmonary (EPTB) TB cases for enhancing surgical facilities to handle such patients.

RNTCP/NTEP annual reports define NSP, CD, PT and EPTB as follows. The new smear positive (NSP) cases are bacteriologically confirmed TB (from sputum specimen) cases who never had Anti Tuberculosis Treatment (ATT) or has taken the same for <1 month. The clinically diagonosed (CD) TB cases refer to a presumptive TB patients who are not microbiologically confirmed, but has been diagnosed with active TB by a clinician on the basis of X-ray abnormalities, histopathology or clinical signs with a decision to treat the patient with a full course of anti-TB treatment. Previously treated (PT) patients have received one month or more ATT in the past and still smear positive. Extra pulmonary (EPTB) cases are defined as any TB cases with involvement of organs other than the lungs, e.g. pleura, lymph nodes, abdomen, genitourinary tract, skin, joints and bones and meninges.

The RNTCP follows international guidelines for recording and reporting for the TB control program with minor modifications. Epi-Info based EPI-CENTRE software being used for data transmission from the district levels upwards. The data available at district, state and national level were in the aggregated form.

Following are the variables used in this analysis. The observed number of TB notification for NSP, PT, CD and EPTB for the years 2016, 2017 & 2018, population count, expected number of NSP, PT, CD and EPTB and cartesian coordinates of capital city for Tamilnadu districts. Among this the TB Notification were taken from TB India 2017[9], TB Inida 2018[10] and 2019[11] for the years 2016, 2017 and 2018 respectively. The population for all the districts in Tamil Nadu were obtained from the census report from Census Report 2011[12]. The expected number of TB cases were estimated based on the proportion method. The Latitude and Longitude for the district head quarters of Tamilnadu were obtained from Maps of India website[13] and converted the same into cartesian coordinates using online tool[14] (either we can use lattitude and longitudes or cartesian coordinates in SaTScan software).

Methods

Poisson model

Besag and Newell[6] developed a method to identify clusters with size k, that is, "regions are put together to have k number of cases". In this method they have estimated the probability of having more than k cases in the region under the Poisson model using

$$Pr(L_i >= l_i) = p(Numberofcases > k|\theta_i ei, = E_i^*) = 1 - \sum_{s=0}^{k-1} \frac{exp(E_i^*)(E_i^*)^S}{s!}) \quad (2)$$

In this model o_i is the disease incidence, e_i is the expected count and the mean is $\theta_i ei$, where θ_i is defined as the relative risk in the ith region respectively and E_i^* is the expected number of incidences for the l_i regions .

The relative risk (RR) for the ith region is defined as follows.

$$\theta_i = \frac{o_i/e_i}{O-o_i/O-e_i} \quad (1)$$

where O is the total number of observed incidence in the entire study region. A politicaly defined region in the study area is said to be having more or less risk if the relative risk is greater or less than 1. The maximum likelihood estimate for θ_i is given by $\hat{\theta} = oi/ei$. Kulldorff and Naggarwala[7] extended from the above method to handle the following (i) to identify clusters in a multi-dimensional point process (ii) window size is allowed to change, and (iii) can assume inhomogeneous Poisson or Bernoulli process. Also they defined an alternative hypothesis that the window is having more risk and there is an elevated risk within the window as compared to other areas based on the likelihood ratio test.

Log Likelihood Ratio test

Under the Poisson assumption, the likelihood function for a specific window is proportional to

$$(\frac{o_i}{e_i})^{o_i}(\frac{o-o_i}{o-e_i})^{o-o_i} I(z) \quad (2)$$

where $I(z)$ is an indicator function and is equal to 1 when the window has more cases than expected under the null hypothesis and 0 otherwise. The probability of a case falls inside the circle is p and outside is q for every circle. The circle can be viewed as a cluster if p is greater than q. For this, they have considered the following hypothesis.

$$H_0: p = q, \qquad H_1: p > q$$

depending upon the total cases notified, the maximum likelihood ratio is being calculated under the Poisson model that is equivalent to consider the following statistic

$$K.N = max_z \epsilon z \frac{L(z)}{L_0} \text{ where} \quad (3)$$

$$L_0 = \frac{O^O(P-O)^P - O}{N^N} \quad (4)$$

$$L(z) = \left(\frac{O_z^{O_z}(P_z - O_z)^{P_z - O_z}}{P_z^{P_z}}\right) \frac{(O-O_z)^{O-O_z}(P-P_z-(O-O_z))^{P-P_z-(O-O_z)}}{(P-P_z)^{P-P_z}} if \frac{O_z}{P_z} > \frac{O-O_z}{P-P_z} \quad (5)$$

$$= \frac{O^O(P-O)^{P-O}}{N^N} if \frac{O_z}{P_z} <= \frac{O-O_z}{P-P_z} \quad (6)$$

where O_z is defined as the total observed cases for regions whose centroids falls inside the circle z and P_z is the population of the the regions whose centroids falls within circle z. The likelihood function is maximized over all window locations and sizes, and the circle with the maximum likelihood constitutes the most likely cluster and it is least likely to occur by chance. This analytic exercise is being repeated on a large number of random replications of the data set under the null hypothesis to obtain the maximum likelihood ratio test statistic distribution. Monte Carlo hypothesis testing[17] is used to obtain p-value by considering the ranks of the maximum likelihood from the real and random data sets. The p-value[15] is obtained using

$$p = \frac{R}{1 + number\ of\ simulations}$$

where R is the rank of the maximum likelihood's and the number of simulations is restricted to 999. Further the null hypothesis can be accepted based on p-value.

In this method, the study region is devided into circles with pre defined radious to have maximum fifty percent of total study population size and are called as cells. For our study we have considered the radius as ten percent of study population. Large number of circles are generated and each circle is defined as a zone. Zones can have irregular boundaries depend on the size and shape of the neighbouring areas. The Individuals belong to the centroids in each circle are included irrespective of the regions boundary to estimate maximum likelihoods. In this study we have used the software SaTScan[8] for identification of various TB clusters in Tamilnadu state.

The above statistic needs the data on the following variables such as observed & expected number of disease cases, population count for the study area and cartesian coordinates for the centroid of the politically defind area. To estimate the expected count we can use either simple raito or age-sex adjusted ratio, if available and in this study we have used the simple ratio method.

Results

The output in SaTScan software are produced in such a way that the clusters are reported based on the ranking of the LLR value (from highest to lowest). The cluster reported as having the highest LLR is considered as primary and other clusters are considered as secondary. In this, we have considered the clusters which are statistically significant (<1% level) and having the Relative Risk 1.25 and above among the districts, so as to include only clusters which are having atleast twenty five percent more risk.

We have given various forms of TB clusters in Table 1 and used the terms LLR for Log-Likelihood Ratio and RR for Relative Risk. The primary clusters are mentioned in bold letters.

For new sputum positive cases, the districts Madurai, Dindigul and Thiruvarur are clusters (either primary or secondary) in all the three years. Here, the districts Madurai and Dindigul are neighbours and they are situated in southern part of the Tamil Nadu State and Thiruvarur district situated in central part of the state. For people of southern districts, Madurai city is preferable destination since Madurai Rajaji Hospital as a public health facility is very old one and having more facilities for diagnosing and treatments of various diseases. Apart from this, many more private hospitals are available in Madurai city. Further stuides may reveal whether the NSP clustering phenomena is due to enhanced medical facilities or the clustering is realy exists in Madurai district.

The districts Madurai and Thiruvarur were the clusters in all three years for previously treated TB cases. Further studies may throw more light on whether the clustering phenomena is due to low level of treatment adherence and more drug resistant cases are available as compared to other districts.

For clinically diagnosed TB cases, Vellore is the cluster in all the three years and situated in northern part of the state. Here a internationaly reputed medical college is functioning for a long time.

For extra-pulmonary TB cases, Vellore and Chennai were the clusters in all the three years. Vellore district is nearer to Chennai and both sitated in northern part of the State. Morover the Chennai is the capital city of Tamil Nadu state where many public and private medical colleges are functioning.

Further stuides may reveal whether the clustering phenomena is due to some unknown etiology or enhanced medical facilities in the above mentioned districts.

We have produced choropleth map of TB clusters with high risk, for Tamil Nadu during the years 2016-18 for NSP in Figure 1, for PT in Figure 2, for CD in Figure 3 and for EPTB in Figure 4. we have used QGIS software version 3.6.3. for producing choropleth map of the TB clusters.

Interpretation and Conclusions

The Spatial scan statistic softwate (SaTScan) using poisson model and log-likelihood ratio may be used to identify spatial clusters for aggregate data of the morbidity or mortality cases for the study region. This method can be applied to identify spatial clusters for time duration such as monthly, quarterly and annual data sets. The limitation of this statistic is that it does not able to identify the sinuous cluster since it relies on circles as the basic shape for this analyzes. Further studies on etilogy behind these clusters may reveal more light and thereby useful to plan the future strategy to acheive the National Strategic Plan to control TB.

Table 1. Various forms of TB clusters in Tamil Nadu for the years 2016, 2017 and 2018

SI.No	District Name	NSP			PT			CD			EPTB		
		2016	2017	2018	2016	2017	2018	2016	2017	2018	2016	2017	2018
		LLR (RR)	LLR (RR)	LLR (RR)	LLR (RR)	LLR (RR)	LLR (RR)	LLR (RR)	LLR (RR)	LLR (RR)	LLR (RR)	LLR (RR)	LLR (RR)
1	Chennai	-	-	-	**82 (1.4)**	-	-	**220 (1.5)**	-	-	210 (2.0)	30 (1.3)	94 (1.4)
2	Coimbatore	-	-	-	-	-	-	-	-	-	-	-	44 (1.3)
3	Cuddalore	-	-	-	-	-	-	-	-	-	-	-	-
4	Dharmapuri	-	61 (1.3)	-	-	-	-	-	25 (1.5)		-	23 (1.4)	-
5	Dindigul	**229 (1.4)**	45 (1.4)	81 (1.4)	-	-	-	-	56 (1.6)	109 (1.7)	-	-	44 (1.3)
6	Erode	-	72 (1.5)	-	-	**100 (2.2)**	80 (1.8)	-	16 (1.3)	-	-	-	-
7	Kancheepuram	-	-	-	-	-	-	-	-	-	-	-	28 (1.3)
8	Kanniyakumari	-	-	-	-	-	-	-	-	-	-	-	-
9	Karur	-	-	-	-	-	-	-	-	-	-	-	-
10	Krishnagiri	-	-	-	-	-	-	-	-	-	-	23 (1.4)	-
11	Madurai	**229 (1.4)**	48 (1.3)	**815 (1.9)**	47 (1.4)	38 (1.6)	**225 (2.2)**	-	-	**150 (1.7)**	-	-	81 (1.6)
12	Nagapattinam	-	209 (1.8)	81 (1.3)	-	51 (1.7)	-	-	-	-	-	-	-
13	Namakkal	-	61 (1.3)	122 (1.3)	-	-	-	-	-	-	-	-	-
14	Perambalur	-	-	-	-	-	-	-	-	-	-	-	-
15	Pudukkottai	-	-	-	-	-	-	-	-	-	-	-	-
16	Ramanathapuram	-	48 (1.3)	132 (1.3)	-	-	-	-	-	-	-	-	-
17	Salem	-	61 (1.3)	122 (1.3)	-	-	-	-	-	-	-	-	-
18	Sivaganga	-	-	-	-	-	-	-	-	-	-	-	-
19	Thanjavur	-	-	-	15 (1.3)	-	-	-	-	-	-	-	-
20	The Nilgiris	-	-	-	-	-	-	-	-	-	-	-	-
21	Theni	**229 (1.4)**	-	-	-	-	19 (1.3)	-	-	-	-	-	-
22	Thiruvallur	-	**262 (2.0)**	68 (1.3)	-	53 (1.8)	11 (1.3)	-	51 (1.6)	-	-	76 (1.9)	18 (1.3)
23	Thiruvarur	90 (1.6)	209 (1.8)	81 (1.3)	28 (1.5)	51 (1.7)	24 (1.5)	-	-	-	-	-	-
24	Thoothukudi	-	68 (1.5)	132 (1.3)	-	-	-	-	-	-	-	-	-
25	Tiruchirappalli	-	-	-	-	-	-	-	-	-	-	20 (1.4)	27 (1.3)
26	Tirunelveli	-	-	132 (1.3)	-	-	-	-	-	-	-	-	-
27	Tiruppur	-	-	-	-	-	-	-	-	-	-	-	-
28	Tiruvannamalai	-	-	-	-	-	-	-	-	-	-	-	-
29	Vellore	-	41 (1.3)	-	-	-	-	207 (1.5)	**129 (1.7)**	100 (1.5)	**244 (1.9)**	**272 (2.3)**	**276 (2.0)**
30	Viluppuram	-	-	-	-	-	-	197 (1.6)	38 (1.4)	-	-	-	-

31	Virudhunagar	**229 (1.4)**	-	**815 (1.9)**	-	-	19 (1.3)	-	-	-	-	-	-

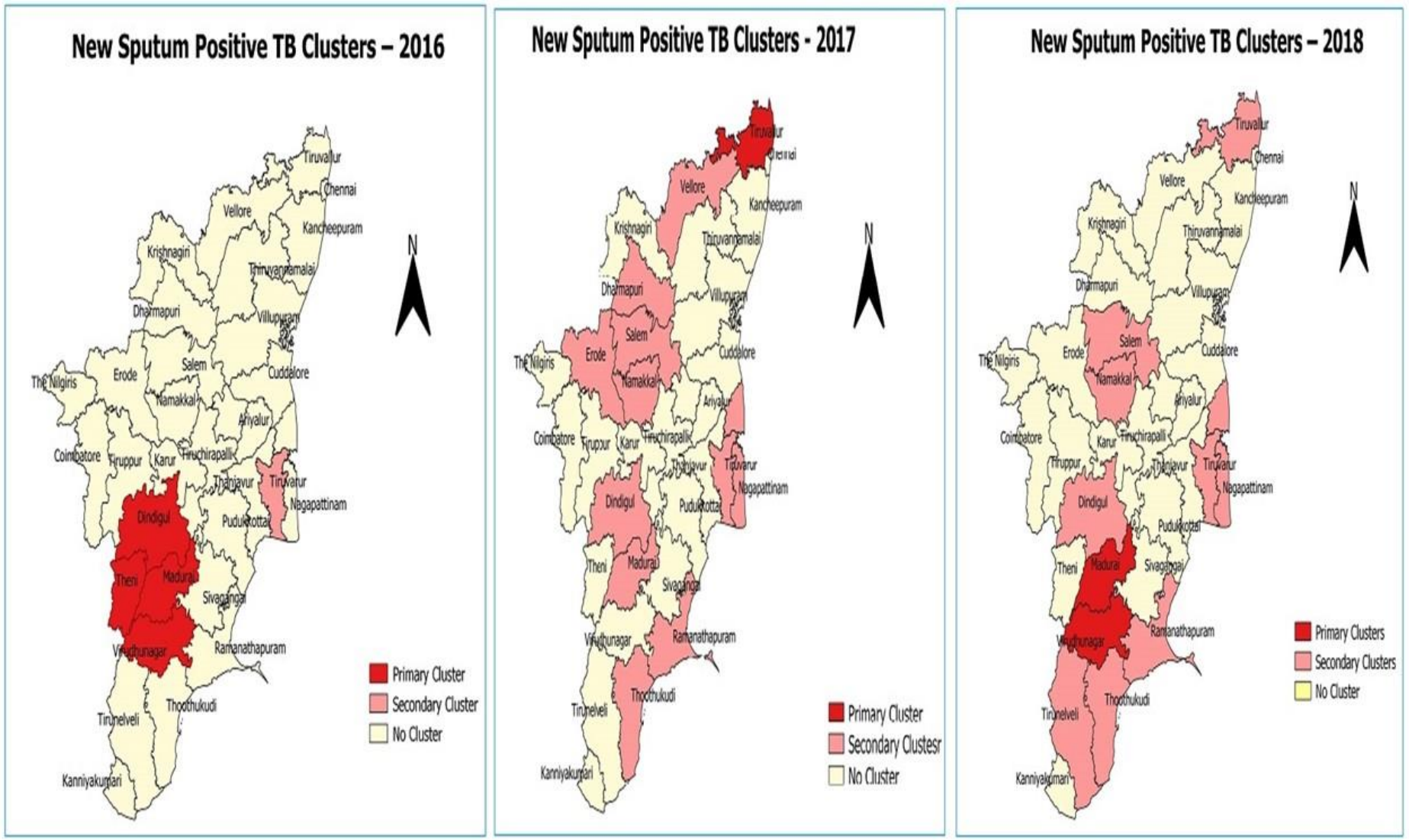

Figure 1: NSP TB clusters in Tamilnadu for the years 2016, 2017 and 2018.

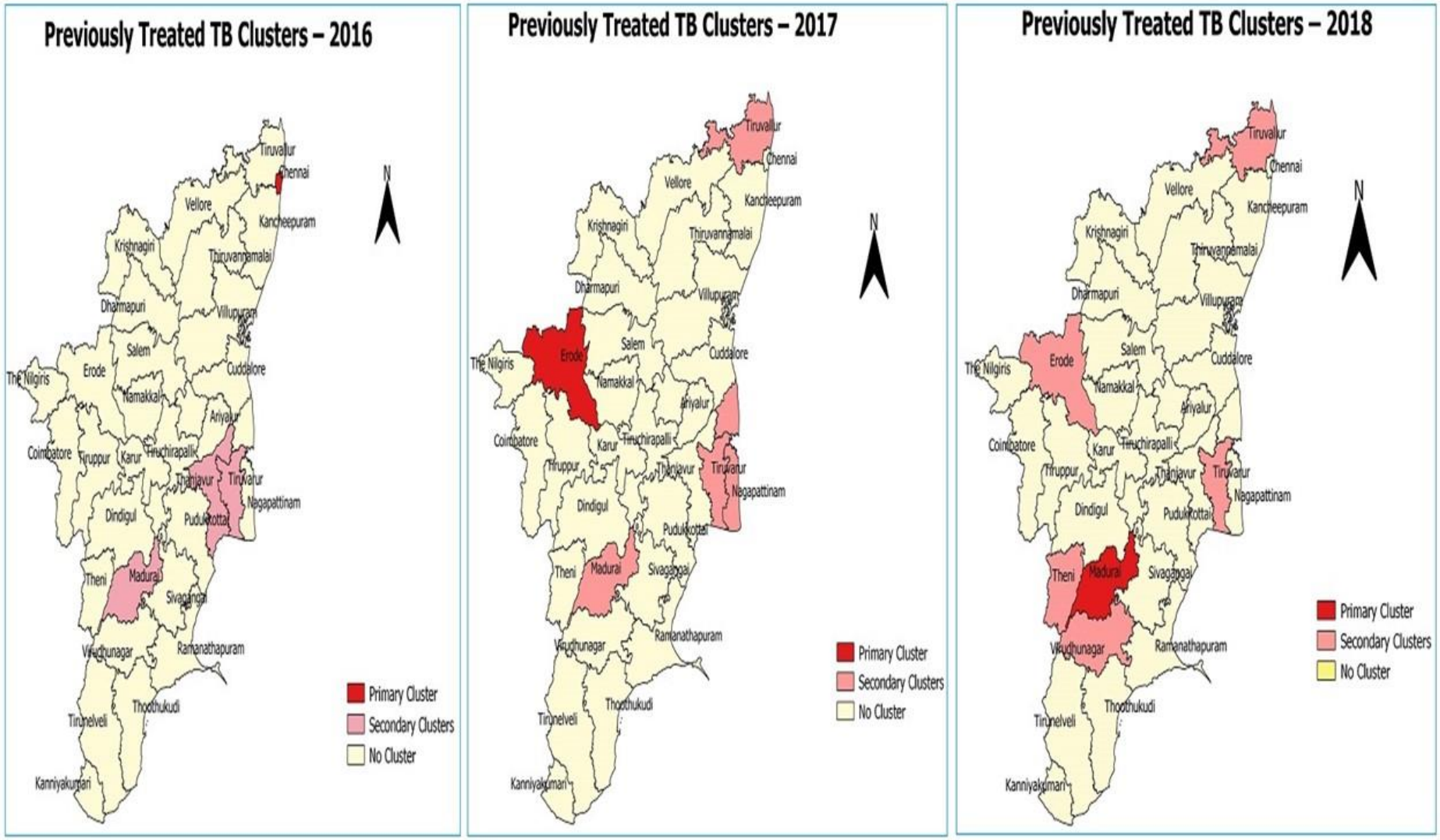

Figure 2: PT TB clusters in Tamilnadu for the years 2016, 2017 and 2018.

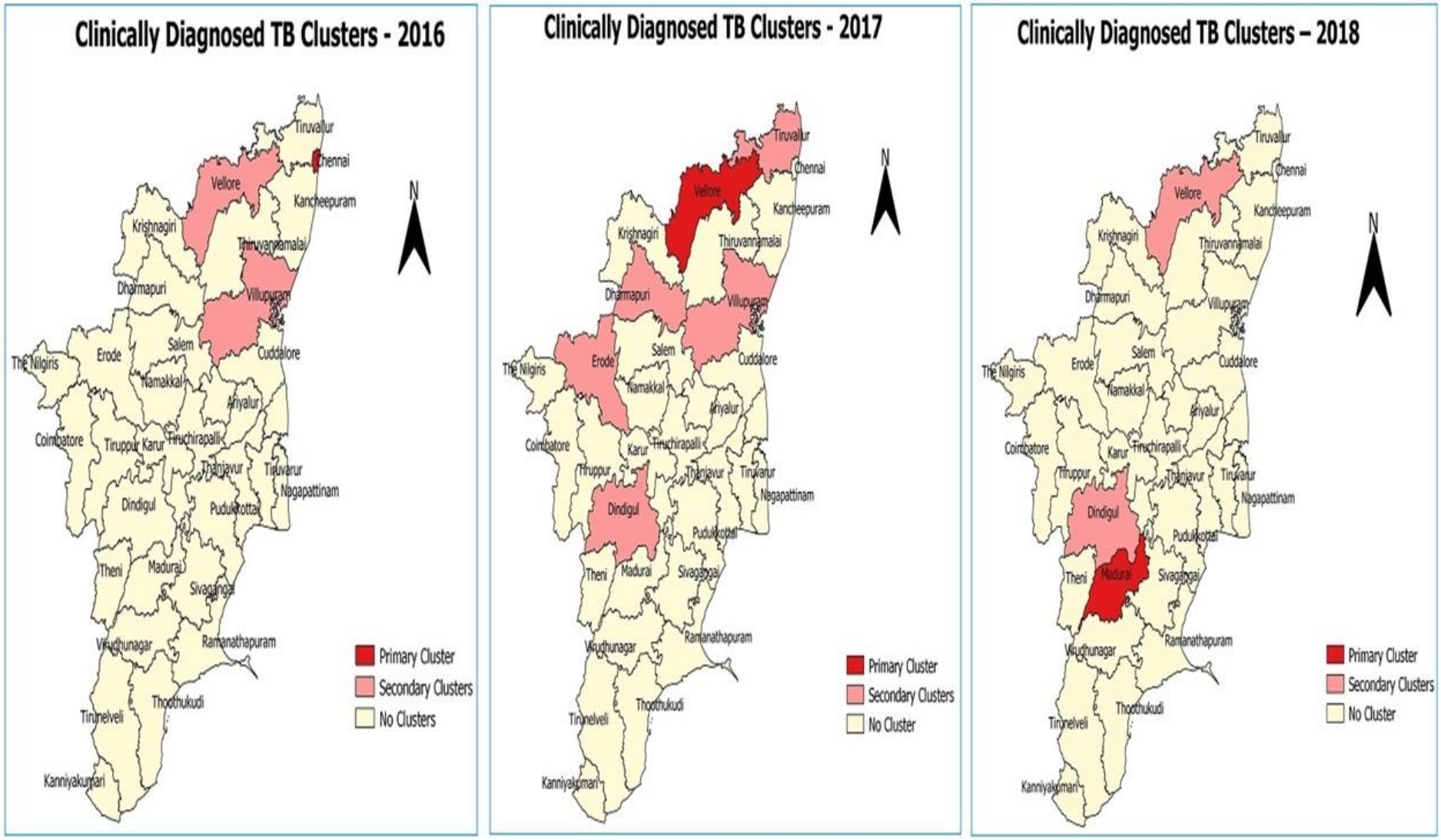

Figure 3: CD TB clusters in Tamilnadu for the years in 2016, 2017, 2018.

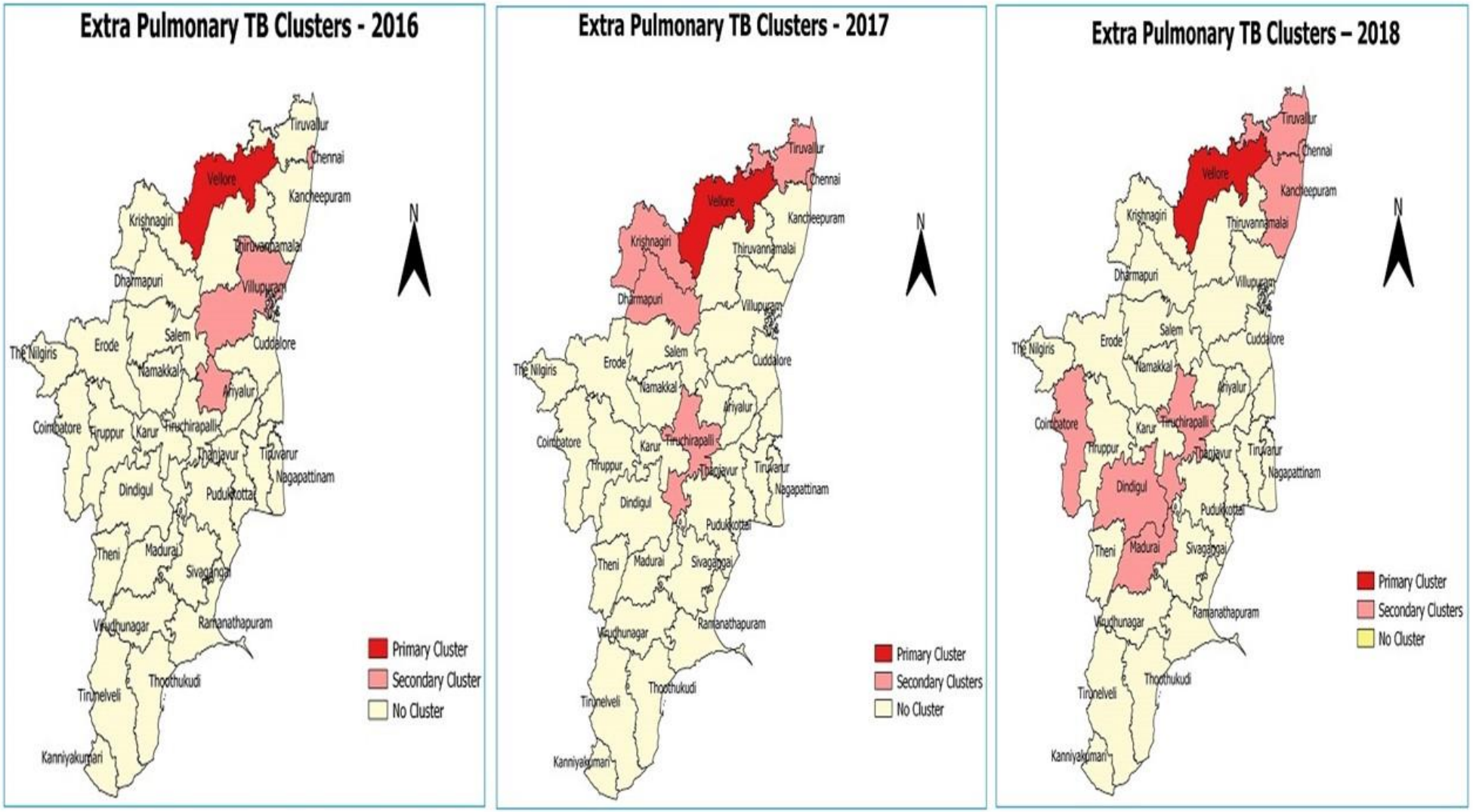

Figure 4: EPTB clusters in Tamilnadu for the years 2016, 2017 and 2018.

References

1. Pearson, Karl (1900). "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling". Philosophical Magazine. Series 5. 50 1990; 302: 157–75.

2. Potthoff R.F, Whittingil M. Testing for homogeneity. II. The Poisson Distribution Biometrica 1966; 53: 183–90.

3. Tango A. A class of tests for detecting 'general' and 'focused' clustering of rare diseases . Stat. Med. 1995; 14: 2323–34.

4. Openshaw S., Charton M., Wymer C., Craft A. Geographical Analysis Machine for the automated analysis of point data sets. International Journal of Geographical Information Sciences 1987; **1**(13): 335–58.

5. Turnbull B. W., Iwano E.J., Burnett W.S., Howe H.L. and Clark L.C. (1990). Monitoring of clustering of disease application. American Journal of Epidemiology. 1990; 132: S136–S43.

6. Besag J, Newll J. (1991). The detection of clusters in rare diseases. Statist. Soc.Ser. A, 1991; 154:143–55.

7. Kulldorff M., Nagarwalla N. Spatial disease clusters: detection and inference. Stat. Med., 1995; 14: 799–810.

8. Kulldorff M. Bernoulli, Discrete Poisson and Continuous Poisson Models: A spatial scan statistic. Communications in Statistics: Theory and Methods. 1997; 26:1481-96.

9. Central TB Division.TB India 2017. Available from https://tbcindia.gov.in/index1.php?lang=1&level=2&sublinkid=4728&lid=3275, accessed on 10.08.2021.

10. Central TB Division.TB India 2018. Available from https://tbcindia.gov.in/showfile.php?lid=3314, accessed on 25.08.2021

11. Central TB Division.TB India 2019. Available from https://tbcindia.gov.in/WriteReadData/India%20TB%20Report%202019.pdf

12. Census Organisation of India. Census Report: 2011.

https://www.census2011.co.in/census/state/districtlist/tamil+nadu.html

13. Maps of India.com. Avialble from https://www.mapsofindia.com/lat_long/Tamil Nadu/

14. University of Minnasota. Polar Geospatial Center. Available from www.pgc.umn.edu/tools/conversion, accessed on 15.02.2022.

15. Dwass M. Modified randomization tests for non-parametric hypothesis. Annals of Mathematical Statistics. 1957; 28: 181–87.