ANALYSIS OF SPATIAL PATTERN & DISTRIBUTION OF TUBERCULOSIS IN THE DISTRICT OF TAMILNADU USING GEOGRAPHICAL INFORMATION SYSTEM (GIS)

Avinaba Banerjee, PG Student, Bijesh Yadav, Faculty
Department of Biostatistics, Christian Medical College, Vellore, Tamil Nadu

ABSTRACT

Abstract
Various regression models including both linear & non linear has been used to analyse NFHS-4 TB data. However, most of these studies didn't include space into account while analysing them.
This study tried to assess the relationship between TB rate & various high risk factors of TB having a direct or indirect effect on it, in Tamil Nadu by incorporating spatial parameters and neighbouring locations using **GIS(geographical information system)**. We calculated **Global Moran's I** to detect any clustering pattern of TB at the global level. After accounting for Global autocorrelation we assessed **local clusters** of TB in the district level & used spatial modelling which included various **weights matrice**s to find out the best model among all the spatial and non spatial models (Spatial lag , error & the Durbin) . The Moran's I value of 0.215, was statistically significant suggesting that TB rate was dependent on its neigh boring TB rate. The Moran's I test for residuals was also significant. We used the LM tests, Elhorst approach of LR tests & also the R^2, RMSE & (AIC) to choose the best model among all the spatial & non spatial models. All these values suggested that Spatial Lag Model (R^2 = 0.45; Log Likelihood = 33.61, AIC = -15.34) identified the spatial relationship best. The study resulted in having 3 influencing factors, mainly people using mud & incomplete roof, mud & incomplete floor & people using solid cooking fuel suggesting that people coming from lower socio economic status were more prone to having the disease.
Adaptive Randomization

OBJECTIVE

- Examine the relationship between the predictor variables and TB rate using non-spatial(OLS)& spatial autoregressive models (SAR)
- Check whether the relation changes in the presence of various spatial weights matrices

METHODS AND MATERIALS

Source : NFHS-4 survey
Participants: 101108 people across 32 districts of Tamil Nadu
Tobler's law : "Everything interacts with everything, but two nearby objects are more likely to do so than two distant objects"

- We used , Rook's weight, w_{ij} = 1 if i, j th region share a boundary &
  Queen's weight, w_{ij} =1 if i, j th region share a boundary or vertex to define our neighbours.
- Spatial similarity over the entire study area of neighbours was assessed by Global Moran's Index
- Local measure of similarity between nearby regions was assessed by Local Moran's Index

MODELS

Linear regression :
Assess relationship between two variables by fitting a linear equation to the observed data. Defined as,
$Y=\beta_0+\beta_1X+\varepsilon$ or $Y= X\beta+ \varepsilon$,where, β_0= intercept, β_1=slope, ε=error in the model.

SPATIAL AUTOREGRESSIVE MODEL(SAR):
Applied when autocorrelation is observed in the dependent variable
$Y= \rho W Y +X \beta +\varepsilon$
ρ is the spatial lag coefficient or spatial autoregressive parameter,
WY is the spatially lagged dependent variable

SPATIAL ERROR MODEL(SEM)
Used when spatial autocorrelation is present in the error terms. Spatial weights matrix is applied to the error terms, not the dependent variable, SEM is defined by,

$$Y = X \beta + u$$
$$\text{Where , } u = \lambda Wu + \varepsilon$$

SPATIAL DURBIN MODEL
Special form of Spatial autoregressive model (SAR), which assumes that the auto regression occurs both in dependent & the independent variables General form,

$$Y = \rho W Y+ X \beta +W X \theta+ \varepsilon$$
OR
$$(I_n- \rho W) Y = Z \beta^* + \varepsilon$$

RESULTS

VARIABLES	OBS	MEAN	SD	MIN	Q25	MED	Q75	MAX
Tb cases(study population)	32	10.66	4.9	2.00	6.00	11	13.50	20
Tb rate	32	0.0033	0.00158	0.00054	0.001952	0.003316	0.004542	0.006632
Proportion of male cases	32	0.69	16.91	0.33	0.5	0.7	0.81	1.0
Proportion of no potability of water	32	0.55	27.4	0.00	0.36	0.62	0.71	1.0
Proportion of smoker	32	0.5	20.97	0.12	0.35	0.93	0.6	1.0
Proportion of no separate kitchen	32	0.22	1.8	0.00	0.13	0.65	0.33	0.55
Proportion of overcrowding	32	0.55	19.25	0.1	0.44	0.54	0.56	1.0
Proportion of solid cooking fuel	32	0.38	23.4	0.00	0.2	0.38	0.63	0.8
Proportion of main wall material mud & incomplete	32	0.14	14.2	0.00	0.0	0.14	0.23	0.5
Proportion of main floor material mud & incomplete	32	0.11	10.5	0.00	0.0	0.11	0.2	0.37
Proportion of main roof material mud & incomplete	32	0.16	13.5	0.00	0.04	0.17	0.25	0.44

Figure 1. Descriptive Table

Model	OLS	SEM	SAR	SDM
Lambda		0.34		
Row			0.5	0.22
No of observations	32	32	32	32
Adjusted R^2	0.29	0.43	0.45	0.43
AIC	-17.158	-23.69	-15.34	-25.23
RMSE	0.13	0.11	0.08	0.1
Log likelihood	19.57	21.874	33.61	21.46

Figure 2. Comparison of 4 models

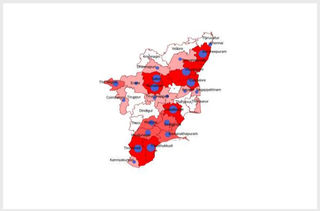

Figure 3. TB Cases in Tamil Nadu

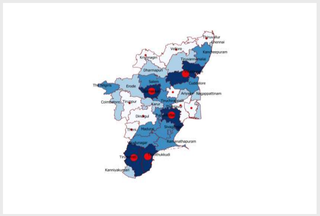

Figure 4: TB Rates in Tamil Nadu.

Figure 5: Local Clusters of TB in Tamil Nadu

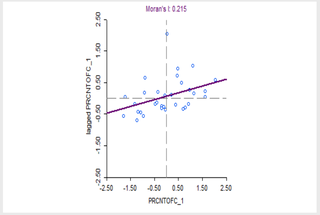

Figure 6: Spatial Auto Correlation of TB in Tamil Nadu

CONCLUSION

- The Moran's I value was close to 0.215, which was statistically significant suggesting that there was almost 22 percent dependence of TB rates on its neigh boring TB rate. The Moran's I test for testing residuals was also significant.
- The LM tests, Elhorst approach & also the R^2, RMSE & AIC all suggested Spatial Lag as the best model among all the models.(R^2 = 0.45; Log Likelihood = 33.61, AIC =-15.34)
- 4 apparently close districts, namely Sivaganga, Ramanathapuram,, Kanniyakumari and Thoothukkudi formed a cluster of high tb rate & it was significant whereas far from them Vellore showed a pattern of very low cases of local clustering of TB rate.
- The study resulted in having 3 influencing factors ,mainly people using mud & incomplete roof , mud & incomplete floor & people using solid cooking fuel i.e. people coming from lower socio economic status were more prone to having the tb.

DISCUSSION

- Recent evidence from the previous studies supported the claim that those living in the poor-quality housing are at higher risk of Tuberculosis.
- Spatial Durbin error model (not center attention of our study) turned out to be the best because & as we all know tuberculosis is a local disease which is more dangerous to people living in the nearby regions i.e. it has a local spillover effect rather than a global spillover effect.
- The study left out missing cases
- this study left out several environmental factors related to tb disease,.

Therefore, a study focusing on a spatiotemporal analysis along with the correlation with environmental factors would be an improvement over this study

REFERENCES

- Lou, M.; Zhang, H.; Lei, X.; Li, C.; Zang, H. Spatial Autoregressive Models for Stand Top and Stand Mean Height Relationship in Mixed *Quercus Mongolia* Broadleaved Natural Stands of Northeast China. *Forests* 2016, *7*, 43
- Elhorst JP. Applied spatial econometrics: raising the bar. Spatial economic analysis. 2010 Mar 1;5(1):9-28.
- Anselin L, Griffith DA. Do spatial effects really matter in regression analysis?. Papers in Regional Science. 1988 Jan;65(1):11-34