BAYESIAN MULTIVARIATE CONDITIONAL AUTOREGRESSIVE MODEL FOR TUBERCULOSIS AND HIV IN INDIA

Article *in* Indian Journal of Applied Research · January 2019

CITATIONS
2

READS
394

3 authors:

Srinivasan Ramalingam
Indian Council of Medical Research
27 PUBLICATIONS **216** CITATIONS
SEE PROFILE

Chinnaiyan Ponnuraja
Indian Council of Medical Research
97 PUBLICATIONS **753** CITATIONS
SEE PROFILE

Perumal Venkatesan
Sri Ramachandra University
222 PUBLICATIONS **3,479** CITATIONS
SEE PROFILE

Original Research Paper

Volume-8 | Issue-1 | January-2018 | ISSN - 2249-555X | IF : 4.894 | IC Value : 86.18

Statistics

BAYESIAN MULTIVARIATE CONDITIONAL AUTOREGRESSIVE MODEL FOR TUBERCULOSIS AND HIV IN INDIA

Srinivasan, R. National Institute for Research in Tuberculosis, ICMR, Chennai – 600031.

Ponnuraja. C National Institute for Research in Tuberculosis, ICMR, Chennai – 600031.

Venkatesan, P National Institute for Research in Tuberculosis, ICMR, Chennai – 600031.

ABSTRACT

Background: Multivariate disease mapping is a collection of two or more disease, each corresponding to the same geographic region, in orders to know information from the joint distribution of disease. This joint model gives understanding of diseases dynamics and relationships between diseases of tuberculosis and HIV incidence jointly, rather than mapping of each disease separately.

Objective: The objective of the study is to construct the Bayesian multivariate CAR(MCAR) model for studying tuberculosis and HIV in India.

Material and methods: National Family Health Survey data on tuberculosis were used in this study. Monte Carlo Markov Chain(MCMC) simulation techniques was used to estimate the parameter. WinBUGS software was used for disease mapping of MCAR model.

Results and Conclusion: The results of the study revealed that spatial autocorrelation between TB and HIV exist and conclude that Bayesian method is proved to be a useful tool for disease modeling of multiple diseases.

KEYWORDS : MCAR, MCMC, Bayesian, Autocorrelation

Background

Bayesian disease mapping studies have been used at univariate level that is considering relative risk estimation for one disease[1-2]. The multivariate disease mapping is a collection of two or more diseases, each corresponding to the same geographic region, in orders to know information from the joint distribution of disease[3-8]. These spatial models have been proven to be an effective tool for analyzing spatially related multidimensional data arising from a common underlying spatial process. This joint model gives understanding of diseases dynamics and of the relationships between diseases incidence jointly, rather than mapping of each disease separately[9-10]. The merit of joint modeling can be high if the considered diseases share risk factors or if the presence of a disease encourages the occurrence of other diseases.

TB is an opportunistic disease it will affect HIV patients easily and a growing consciousness is that millions of people will become infected in the coming years, needs swift and appropriate action immediately[000]. For comparison these diseases together, Bayesian approach using the Markov Chain Monte Carlo (MCMC) method were used to estimate the disease incidence of both diseases in all the States of India. This models considering fixed and random effects with covariate effects, interregional variability and the spatial variability are all considered. Bayesian analysis is the method where prior selection plays an important role in the inference. The default prior is non-informative but results in a proper posterior on the related parameter spaces. This method not only provides robust inference, but also provides improved estimation. Bayesian computation is providing more stable, efficient and produced stable estimates for each region in the spatially arranged regions. It also allowed for unexplained heterogeneity to be investigated in the disease maps[11-12].

Multivariate Condictional Autoregressive(MCAR) model proposed by Besag et al.[4] incorporates both spatially structured and unstructured random effects in a single model. Several authors were explored multivariate spatial models for lattice data, adopting the Bayesian framework as the natural inferential approach[13-15]. Venkatesan and Srinivasan proposed several spatial models for multivariate data based on general univariate conditional autoregressive (CAR) model for Tuberculosis and HIV disease mappings[16-17].

Material and methods

The number of cases for tuberculosis Y_{1i} and HIV Y_{2i} occurring in area S_i is recorded, where the set of areas $\{S_i\}$, i = 1,..,n represents a partition of the region under study. For each area S_i, the expected number of cases E_{1i} and E_{2i} is computed using reference rates for the disease prevalence. The multivariate collection of TB and HIV for disease mapping, for the same geographic region will give the joint distribution of disease pattern for the same area in order to understanding of diseases relationships jointly.

For comparison these disease together, spatial HIV and TB data obtained from National AIDS Control Society (NACO), National Family Health Survey (NFHS) data used for multivariate modeling. It consists of information on persons in the household with Tuberculosis and HIV, Literacy, Incomes and TB awareness. The population of all States are taken from Census 2001.

Let Y_{1i}, Y_{2i} from the two diseases in area i. We assume that these counts arise from two Poisson distribution, $Y_{ki} \sim \text{Poisson}(E_{ki}\theta_{ki})$, k=1,2, i=1,..,n where E_{1i}, E_{2i} are the expected count for the Tuberculosis and HIV/AIDS disease. The convolution model of MCAR is;

$$\log(\theta_{ik}) = \log(E_{ik}) + \alpha_k + \alpha_{ik} Inc + \alpha_{2k} Edu + S_{ik} + U_{ik} \quad (1)$$

here S_{ik} and U_{ik} are the structured and unstructured random effect in this model. E_{ik} is expected count for diseases in area i, $k\alpha$ is an intercept term representing the baseline RR of both diseases across the study region, S_{ik} is the area and disease specific log relative risk of cases. log RR of both diseases are spatially correlated across areas, and within area i, relative risks for TB and HIV/AIDS are also correlated due to dependence on shared area-level unmeasured risk factors. Correlation assumptions using an intrinsic bivariate CAR prior used for S_{ik} and U_{ki} values. The MCAR random effects require a flat prior for intercepts. For Income and Education, the normal independent $N(0,\sigma^2)$ prior were assigned. The unstructured random effects are assumed to have the distribution of $MVNor(0, \sigma^2_u)$, whereas the structured random effects S_{ik} are assumed to follow the MCAR model with conditional variance σ^2_c Independent InvGamma (0.01, 0.0001) priors are used for the hyper parameters.

The software used for multivariate conditional autoregressive model analysis is WinBUGS that fits spatial models[18]. It implements MCAR models for data that are collected within discrete regions and joint distribution of this smoothing is done based on Markov random field models for the neighborhood structure of the regions relative to each other. To fit the model in WinBUGS, observed count, expected count and covariate of education and income with adjancy matrix for India were included for analysis. The prior for this model is hyper prior i.e., Gamma prior which is distributed with a small precision, thus taking a larger neighborhood structure into account. Results are based on an MCMC simulation of 30,000 draws and an inverse distance-matrix.

RESULTS

The posterior summaries are presented in the following Table 1 and 2 for TB and HIV for 32 States

Table 1 Posterior Summaries of Relative Risk for Tuberculosis and HIV/AIDS

	Tuberculosis				HIV			
	Mean	SD	Credible Interval		Mean	SD	Credible Interval	
RR1[1]	0.793	0.119	0.576,	1.040	1.004	0.147	0.720,	1.270
RR1[2]	5.760	3.120	3.060,	8.250	1.460	0.994	0.228,	3.781
RR1[3]	0.412	0.102	0.240,	0.642	3.850	4.047	2.095,	6.970
RR1[4]	0.879	0.960	0.052,	3.506	0.842	0.883	0.091,	3.132
RR1[5]	0.699	0.200	0.385,	1.163	3.022	2.510	0.398,	10.080
RR1[6]	0.604	0.160	0.365,	1.007	1.426	1.777	0.217,	7.111
RR1[7]	1.826	0.443	1.185,	2.844	1.055	0.836	0.240,	3.427
RR1[8]	0.176	0.047	0.110,	0.298	0.914	0.739	0.154,	2.841
RR1[9]	1.347	1.493	0.041,	2.458	0.278	0.260	0.048,	1.010
RR1[10]	0.182	0.040	0.116,	0.275	3.060	3.640	2.537,	5.280
RR1[11]	3.742	2.266	2.490,	5.670	3.800	3.530	2.095,	6.570
RR1[12]	4.760	3.885	.299,	23.710	0.708	0.513	0.144,	2.016
RR1[13]	1.209	1.300	0.064,	4.718	0.726	0.623	0.131,	2.147
RR1[14]	0.316	0.067	0.208,	0.465	4.390	5.420	3.180,	5.600
RR1[15]	0.380	0.085	0.230,	0.579	2.612	2.318	1.976,	3.370
RR1[16]	0.168	0.038	0.111,	0.256	0.785	0.616	0.137,	2.461
RR1[17]	0.793	0.119	0.576,	1.040	1.004	0.147	0.720,	1.270
RR1[18]	0.049	0.017	0.024,	0.087	0.233	0.137	0.059,	0.574
RR1[19]	0.252	0.065	0.175,	0.423	4.773	3.774	0.984,	7.860
RR1[20]	3.462	10.400	0.008,	26.030	1.252	3.394	0.013,	7.960
RR1[21]	1.306	0.228	0.968,	1.904	1.756	1.098	0.527,	4.538
RR1[22]	1.181	0.208	0.877,	1.710	0.496	0.341	0.159,	1.461
RR1[23]	0.490	0.132	0.310,	0.835	0.288	0.210	0.078,	0.819
RR1[24]	2.266	0.255	1.839,	2.798	0.393	0.139	0.176,	0.737
RR1[25]	4.554	6.210	3.023,	6.370	0.348	0.195	0.117,	0.739
RR1[26]	2.112	0.396	1.656,	3.104	3.666	10.170	1.944,	6.020
RR1[27]	0.145	0.037	0.092,	0.235	1.958	1.274	0.625,	5.797
RR1[28]	3.262	0.947	2.092,	6.279	0.449	0.362	0.063,	1.347
RR1[29]	1.681	0.349	1.141,	2.474	1.172	0.500	0.516,	2.420
RR1[30]	0.391	0.084	0.256,	0.612	0.337	0.291	0.062,	1.184
RR1[31]	1.459	0.273	1.052,	2.010	0.163	0.127	0.036,	0.494
RR1[32]	0.203	0.050	0.118,	0.309	0.275	0.194	0.035,	0.709

Table 1 shows the posterior expected relative risk for both diseases for all the state based on MCAR model. It includes mean, standard deviation, and credible interval. From the joint distribution of both diseases, the relative risk of TB varying from 0.05 to 5.76 and for HIV the relative risk is varying from 0.16 to 4.39. There is 15 states comes under relative risk of greater than one for HIV and TB. The largest RR for HIV found in State 14 and for TB in State 2.

Table 2 Posterior Summaries for Parameters under MCAR Model

Parameter	Mean	SD	Median	Credible Interval	
$\alpha[TB]$	-0.2864	0.1645	-0.2847	-0.6095,	0.0232
$\alpha[HIV]$	-0.0235	0.1556	0.0188	-0.3041,	0.2370
α_1	0.0060	0.0035	0.0056	0.0000,	0.0119
Corr.structured	0.3448	0.2301	0.2481	-0.2220,	0.6863
Corr.unstructured	0.1922	0.6699	0.3994	-0.9274,	0.9663
Corr.sum	0.2786	0.2039	0.2377	-0.1962,	0.5987
$\omega.structured[TB]$	0.0826	0.0263	0.0791	0.0431,	0.1449
$\omega.structured[HIV]$	0.1966	0.1619	0.1496	0.0733,	0.7352
$\sigma.unstructured[TB]$	0.1979	0.0979	0.1819	0.0669,	0.4271
$\sigma.unstructured[HIV]$	0.8739	0.2581	0.8308	0.5002,	1.4870
$\sigma 2.structured[TB]$	15.1000	4.5520	14.3000	8.6390,	26.2400
$\sigma 2.structured[HIV]$	7.6970	3.2140	7.4190	1.8960,	14.8800
$\sigma 2.unstructured[TB]$	0.0487	0.0493	0.0331	0.0045,	0.1824
$\sigma 2.unstructured[HIV]$	0.8303	0.5224	0.6902	0.2502,	2.2110
$\tau.[TB]$	113.5000	92.8100	16.9300	85.7500,	362.5000
$\tau.[HIV]$	6.1090	6.4170	0.9190	3.9360,	23.9400

From Table 2, the posterior correlation between the spatially structured risk components for TB and HIV is 0.34; (95% CI: -0.22, 0.68), although correlation between the unstructured risk components is 0.19 ; (95% CI: -0.92, 0.96). The correlation between the total random effect for TB and HIV is 0.27; (95% CI: -0.19, 0.59), suggesting shared geographical pattern of risk between the two diseases. The posterior mean for α for TB is -0.28, compared to posterior mean for for HIV is -0.02. The posterior mean for α_1 is 0.006.

Spatially unstructured heterogeneity and structured heterogeneity models are considered, in which the structured model normally distributed with multivariate normal (MVCAR) prior for the spatial random effects. The precision tau for multivariate normal prior is high for both diseases comparing to MVCAR prior for the spatial random effects model. The credible Interval for MVCAR model is narrow and posterior median also close with posterior mean and adding random effects gives smoothed relative risk based on neighborhood structure of the States. This model pulls relative risk estimates towards the local mean, but in the Bayesian spatial unstructured heterogeneity model assumes MV normal prior that the variance is fixed in all the States and it pulls relative risk towards the global mean. It reflected in the map also. The relative risk in unstructured heterogeneity model smoothed towards in global mean.

The posterior distribution for precision of $\omega.structured$ [*TB*] is 0.08 while mean of $\tau.[TB]$ is 113 with standard deviation of 0.02 and 92.81 respectively. The posterior distribution for precision of $\omega.structured$ [*HIV*] is 0.19 while for precision of $\tau.[HIV]$ is 6.11 with standard deviation of 0.16 and 6.4 respectively. There was much shrinkage in the estimates from the structured heterogeneity on both the diseases. It was also observed in the maps. The variability of the relative risk is attributed more to the unstructured model than to the spatially structured model.

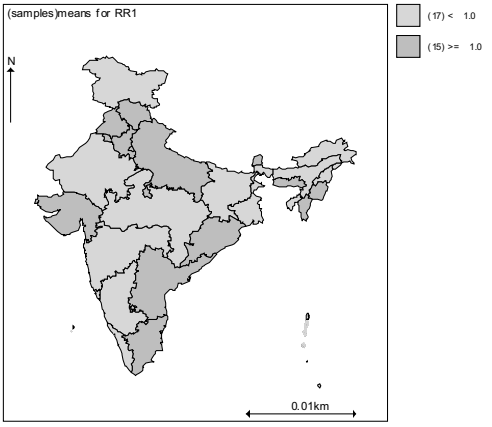

Figure 1 Posterior Expected RR for Tuberculosis under MCAR model

A map of the RR for TB (Fig.1), again classified into 2 groups in which area estimates larger than 1 relative risk appear in dark shade area. Lighter shades reflect the area of lower risk, and there is 17 States wards come under low risk area and 15 wards come under higher risk group which is scattered throughout India.

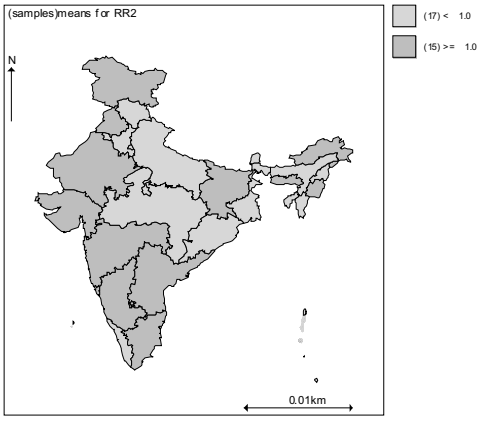

Figure 2 Posterior Expected RR for HIV under MCAR Model

A map of the RR for HIV (Fig.2), classified into 2 groups in which area estimates larger than 1 relative risk appear in dark shade area. Lighter shades reflect the area of lower risk, and there is 17 States wards come under low risk area and 15 wards come under higher risk group which is scattered throughout India.

Table 3 Goodness of fit (DIC) for MCAR

No	Models	$\overline{D}$	$\hat{D}$	D_p	DIC
1	$\alpha+\beta_{Edu}*Edu+\beta_{Inc}*Inc+S_i+U_i$, $S\sim MVCAR, U\sim MVNOR$	257.95	220.35	37.59	295.54
2	$\alpha+\beta_{Edu}*Edu+\beta_{Inc}*Inc+S_i$, $S\sim MVCAR$,	259.74	222.68	37.06	296.79
3	$\alpha+\beta_{Edu}*Edu+S_i$, $S\sim MVCAR$,	256.66	220.98	35.68	292.34
4	$\alpha+\beta_{Inc}*Inc+S_i$, $S\sim MVCAR$,	257.92	221.63	36.29	294.21
5	$\alpha+\beta_{Edu}*Edu+\beta_{Inc}*Inc+U_i$, $U\sim MVNOR$	262.63	222.76	39.87	302.50
6	$\alpha+\beta_{Edu}*Edu+\beta_{Inc}*Inc$	1088.58	1086.55	2.03	1090.61

The models comparison using DIC in Table 3, spatial structured heterogeneity model has the lower DIC than without unstructured heterogeneity model. The spatial structured heterogeneity effect with education model gives smaller DIC value for this data which implies that covariate education and spatial MVCAR plays important role. The fixed effect model is very high DIC comparing with spatial random effect model which shows the advantage and importance of Bayesian random effect in disease modeling.

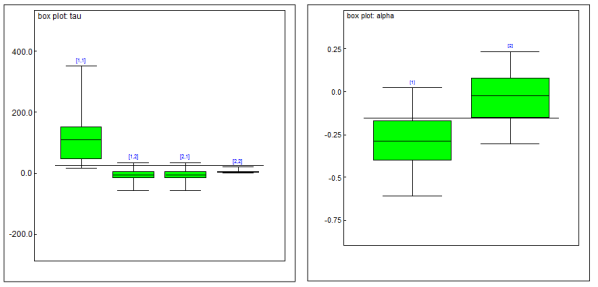

Figure 3 Box Plot for Tau and Alpha

The box plot for tau and alpha shows the variation of two diseases and the precision tau represents the effect of TB, HIV and shared effect.

Conclusion

In this work, modeling of multivariate disease mapping is explored for HIV and TB. This joint disease model combines HIV and TB disease maps, each corresponding to the same geographic region and the joint distribution of disease has the advantage of not only being able to rely on covariate but also borrows strength between observation vectors as well. Structured random effects gives smoothed relative risk based on neighborhood structure of the States and this model pulls relative risk estimates towards the local mean, but in the spatial unstructured model pulls relative risk towards the global mean.

The spatial autocorrelation between TB and HIV is 0.38 suggesting that shared geographical pattern of risk between the two diseases exist. The precision of structured model for TB is 0.08 while unstructured for TB is 113 with standard deviation of 0.02 and 92.81 respectively. The precision of structured for HIV is 0.19 while unstructured for HIV is 6.11 with standard deviation of 0.16 and 6.4 respectively. However, the precision value of unstructured model is high for both diseases comparing to structured model. The credible Interval for structured random effects model is narrow for all the values and posterior median also close with posterior mean. There was much shrinkage in the estimates from the structured model on both the diseases. It shows that the estimates of structured random effects models provide shrinkage estimates and the variability of the relative risk is attributed more to the uncorrelated heterogeneity than to the spatially structured effects.

The models comparison using DIC, spatial structured random effect model has the lower DIC than without unstructured heterogeneity model. The spatial structured heterogeneity effect with education model gives smaller DIC value for this data which implies that covariate education plays important role. The fixed effect model is very high DIC comparing with structured random effect model which shows the advantage and importance of Bayesian MCAR with random effect in disease modeling.

REFERENCES

1. Banerjee S, Carlin B and Gelfand A E (2004): Hierarchical Modeling and Analysis for Spatial Data. Boca Raton: Chapman & Hall.
2. Besag J and Kooperberg C (1995): On conditional and intrinsic autoregressions. Biometrika, 82: 733-746.
3. Gelfand A and Vounatsou P (2003): Proper multivariate conditional autoregressive models for spatial data analysis. Biostatistics, 4 : 11-25.
4. Gelman A, Carlin J, Stern H and Rubin D (2004): Bayesian Data Analysis, 2nd ed. Boca Raton, FL: Chapman and Hall/CRC Press.
5. Besag J, York J C and Mollie A (1991): Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics, 43: 1-59.
6. Xia H and Carlin B (2005): Multivariate parametric spatiotemporal models for country level breast cancer survival data, Life time Data analysis, 11: 107–120.
7. Jin X, Carlin B, Banerjee S (2005): Generalized hierarchical multivariate CAR models for areal data. Biometrics, 61: 950-61.
8. Jin X, Banerjee S and Carlin B P (2007): Order-free co-regionalized areal data models with application to multiple-disease mapping. Journal of the Royal Statistical Society 65: 817-838.
9. Gelfand A Diggle P and Guttorp P (2010): Handbook of spatial statistics, Chapman & Hall / CRC.
10. Besag J (1975): Statistical Analysis of Non-Lattice Data. Statistician, 24: 179–195.
11. Lawson A (2009): Bayesian disease mapping: Hierarchical modeling in spatial epidemiology, Chapman & Hall/ CRC.
12. Lawson A, Browne W and Vidal Rodeiro C (2003): Disease mapping with WinBUGS and MLwiN. John Wiley & Sons Ltd.
13. Ghosh M, Natarajan K, Waller L and Kim D (1999): Hierarchical Bayes for the Analysis of Spatial Data: An Application to Disease Mapping. Journal of Statistical Planning and Inference, 75: 305–318.
14. Sain S and Cressie N (2002): Multivariate lattice models for spatial environmental data. In Proc. A.S.A. Section on Statistics and the Environment. Alexandria, VA: American Statistical Association, 2820-2825.
15. Sain S and Cressie N (2007): A spatial model for multivariate lattice data, Journal of Econometrics,140: 226-259.
16. Venkatesan P, Srinivasan R and Dharuman C (2012): Bayesian conditional autoregressive model for mapping tuberculosis prevalence in India, International Journal of Pharmaceutical Studies and Research, Vol.III, online journal.
17. Venkatesan P and Srinivasan R (2008): Bayesian model of HIV/AIDS in India: A spatial Analysis. Applied Bayesian Statistical Analysis : 51-56.
18. GeoBUGS User Manual (2004): GeoBUGS User Manual. Version 1.2.