Spatial Frailty Survival Model for Infection of Tuberculosis among HIV Infected Individuals

Srinivasan R[1], Ponnuraja C[1], Rajendran S[1], Venkatesan P[2]
[1] *Dept. of Statistics, National Institute for Research in Tuberculosis, ICMR, Chennai, India.*
[2] *Sri Ramachandra University, Chennai, India.*

Abstract

Capturing spatial variation is important while studying disease survival in different geographical regions where regions has its own risk factor associated with disease. To account this risk factor, frailty effect was introduced in this model that captures correlation and variation between neighboring locations using conditionally autoregressive (CAR) prior in Bayesian parametric survival model for studying dual infection of tuberculosis and HIV. National Institute for Research in Tuberculosis data on tuberculosis and HIV were used in this study. Monte Carlo Markov Chain (MCMC) technique was used to estimate the parameter. WinBUGS software was used for Bayesian Survival model estimation. The result of the study revealed that the spatial frailty model accounts higher heterogeneity along with weight at baseline was one of the significant factors associated with death and conclude that there were unmeasured covariates and risk factors influencing death in the Chennai regions.

Keywords: *Bayesian spatial, Survival, CAR, Weibull, random effects*

I. BACKGROUND

Survival analysis used when outcome variable of interest is time until an event occurs for individuals. The time indicates any unit of time for their time to end point. The event of interest in this study is death in an area along with covariates within model, random effects relating to individual heterogeneity and spatial correlation effect between regions[1]. Capturing spatial correlation and variation is important while studying disease survival in different geographical regions where regions has its own risk factor associated with disease. To account this risk factor, frailty or random effect was introduced in this model that captures spatial correlation and variation between geographical regions[2-4]. Normally, in conventional analysis regions are assumed to be independent, but random effects corresponding to region that are spatially arranged and suspect that regions in closer proximity to each other might also be similar in magnitude4[5]. It will affect the robustness of the estimates of the parameter but in Bayesian frailty model incorporates random effects were introduced at neighboring locations that are allowed to exhibit spatial correlation in this model by specifying a conditionally autoregressive (CAR) prior developed by Besag[6] for application in effects in time-to-event data across neighboring units, with the neighbors defined via an adjacency matrix.

Tuberculosis and HIV have been closely linked since the emergence of AIDS. TB is the most common opportunistic infection affecting HIV individuals and it remain the most common cause of death in patients with AIDS. HIV infection has contributed to a significant increase in the worldwide incidence of TB. TB and HIV co-infection each disease speeds up the progress of the other. In addition to HIV infection speeding up the progression from latent to active TB, TB bacteria also accelerate the progress of HIV infection[7]. The risk of developing tuberculosis (TB) is estimated to be between 26 and 31 times greater in people living with HIV (PLHIV) than among those without HIV infection. This work may accounts all this unmeasured risk factors considered in this model along with spatial variation.

In this work, Parametric model of Weibull is explored for survival analysis due to its versatility. It contains shape parameter β, the scale parameter η, and location parameter γ. Also, parametric model with a Weibull formulation for the baseline hazards, placing a univariate Conditional autoregressive (CAR) structure on the frailty intercept terms was used to account the correlation between regions. This CAR model indicates the existence of spatial dependence on the composition of covariance where the CAR parameter distribution stating precision or variance inverse of its random effect distribution. CAR distribution captures correlation across both geographic regions and the random effects for a given region. This model implemented using MCMC computational techniques in a hierarchical Bayesian approach that permits borrowing of strength across regions.

In the literature, Parametric proportional hazards model with a Weibull formulation for the baseline hazards, placing a univariate CAR structure on the frailty intercept terms was studied for infant mortality data and highlighted the importance of spatial model[7]. Bayesian spatial survival models for political

event process were extensively studied in spatial, non-spatial model with in parametric and semiparametric approach and proved that spatial dependence in the random effects also produces changes in the effects of covariates[8]. Spatio-temporal model using CAR prior for tuberculosis disease was extensively studied for other diseases.

II. BAYESIAN SPATIAL MODEL SPECIFICATION

Weibull model was estimated using WinBUGS software to model the hazard rate of death or relapse and significance of factors associated with the same.

Cox proportional hazard model is of the form

$$h(t_i;\boldsymbol{x}_{ij}) = h_0(t_i)\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i) \quad 1a)$$

$$h(t_{ij};\boldsymbol{x}_{ij}) = h_0(t_{ij})\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i) \quad 1b)$$

where t_{ij} is the time to death or censoring for individual i in the ward j, $i = 1,\cdots,n_j j = 1,\cdots,K$,$\boldsymbol{x}_{ij}$ is a vector of individual-specific covariates and $\boldsymbol{\beta}$ is a vector of parameters,h_0 is the baseline hazard. Extending this model to include the spatial dependency, we propose the model

$$h(t_i;\boldsymbol{x}_i) = h_0(t_i)\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i + W_i + V_i) \quad 2a)$$

$$h(t_{ij};\boldsymbol{x}_{ij}) = h_0(t_{ij})\exp(\boldsymbol{\beta}^T\boldsymbol{x}_i W_i + V_i) \quad 2b)$$

The likelihood for the Cox model with spatial and non-spatial frailties is;

$$L(\beta,W;t,x,\gamma) \propto \prod_{j=1}^{k}\prod_{i=1}^{n_j}\{h_0(t_{ij};x_{ij})\}^{\gamma_{ij}}\exp\{-H_0(t_{ij})\exp(\beta' x_{ij} + W_i + V_i)\} \quad 3a)$$

Where W_i is the spatial frailties and V_i is non-spatial frailties for the same region. The event of interest is death during treatment as well as follow up period. The event is coded as 1 and 0 is coded for censoring. There are other important covariates are included in the analysis such as age, sex, treatment group, weight at baseline. *W* represents CAR model distribution supporting the possibility of correlated random effects at the wards level $W_i \sim N(0,1/\sigma)$ and *V* represents non-frailty random effect and this exchangeable prior $V_i \sim N(0,1/\tau)$.

III. WEIBULL MODELS WITH SPATIAL FRAILTIES

The joint posterior distribution for the Bayesian parametric Weibull model is

$$p(\beta,W,\rho,\lambda \mid t,x,\gamma) \propto L(\beta,W,\rho;t,x,\gamma)\,p(W \mid \lambda)\,p(\beta)\,p(\rho)\,p(\lambda) \quad (1)$$

Where ρ is the shape parameter for the baseline hazard in the Weibull model. The likelihood for the Bayesian Weibull model with spatial individual frailties is;

$$L(\beta,W,\rho;t,x,\gamma) \propto \prod_{i=1}^{I}\rho t_i^{\rho-1}\exp(\beta^T x_i + W_i)^{\gamma_i}\exp\{-t_i^{\rho}\exp(\beta^T x_i + W_i)\} \quad (2)$$

The individual Weibull frailties are completed by assigning suitable prior for the parameter. The likelihood for the Bayesian Weibull model with spatial individual frailties is;

$$L(\beta,W,\rho;t,x,\gamma) \propto \prod_{i=1}^{I}\rho t_i^{\rho-1}\exp(\beta^T x_i + W_i)^{\gamma_i}\exp\{-t_i^{\rho}\exp(\beta^T x_i + W_i + V_i)\} \quad (3)$$

Where V_i represents the non-spatial frailties with $V_i \sim N(0,1/\tau)$

IV. MATERIAL AND METHODS

The data have been collected from National Institute for Research in Tuberculosis, Chennai. The data consists of 105 HIV infected Tuberculosis patients admitted in clinical trial who were treated with three types of treatments of 6 months to 8 months duration. The covariates considered for analysis are age (in years), sex (Male-1 and Female-0), treatment group, weight at baseline (in kg). The event of interest is death during treatment and follow up period. The event of interest is coded as1 and censoring is coded as 0.

Weibull distribution assumed for the time distribution and a number of covariate is considered for the individual level and at a higher spatial level the wards within which the individual was distributed is considered. Weibull time to endpoint model is defined as

$$f(t_i) = \rho\mu_i t_i^{\rho-1} \exp(-\mu_i t_i^{\rho}) + W_i + V_i \quad (4)$$

Where μ_i is modeled as

$$\log(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + W_i + V_i \quad (5)$$

where X_1 – Age, X_2 – Sex, X_3 – Treatment group, X_4 – Weight at baseline, W – CAR model distribution supporting the possibility of correlated random effects at the wards level and V – represents non-frailty random effect and this exchangeable prior $V_i \sim N(0,1/\tau)$. The WinBUGS software used for Weibull model analysis for which observed time, censoring time with other covariate mentioned above and adjancy matrix for Chennai wards 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. The software used for Cox model analysis is WinBUGS 14.0. 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. The non-informative priors were considered where $\beta \sim N(0,0.000001)$. Using a burn in of 10000 samples and additional of 10000, from 10000 to 30000 Gibbs samples were drawn, posterior estimates of β's given in the table. The comparisons of the posterior estimates indicate that the convergence has achieved in 30000 iterations.

V. RESULTS

The Table 1 shows the non-frailty and frailty Weibull model for the HIV infected TB data considering the covariates age, treatment (cat), sex and wt were given in the table. The analysis was carried over by STATA software for these models.

Table 1 Summary Statistics for Weibull model

	Weibull model without frailty				Weibull model with frailty			
Cov.	Beta	SE	95 % CI		Beta	SE	95 % CI	
Age	0.003	0.015	-0.03,	0.033	-0.028	0.040	-0.107,	0.051
Sex	0.41	0.333	-0.24,	1.061	0.693	0.708	-0.695,	2.081
Cat	-0.16	0.194	-0.54,	0.219	0.300	0.603	-0.882,	1.482
wt0m	-0.049*	0.047	-0.08,	-0.02	-0.112 *	0.013	-0.195,	-0.029
Cons	-1.02	0.9	-2.78,	0.748	0.765	0.934	-3.026,	4.555

*P<0.05

The above table shows the Weibull models for frailty and non-frailty in which weight is one factors associated with death for both model, but SE for frailty model is substantially reduced for all the covariates. The heterogeneity accounts between areas are 2.37 in this frailty model.

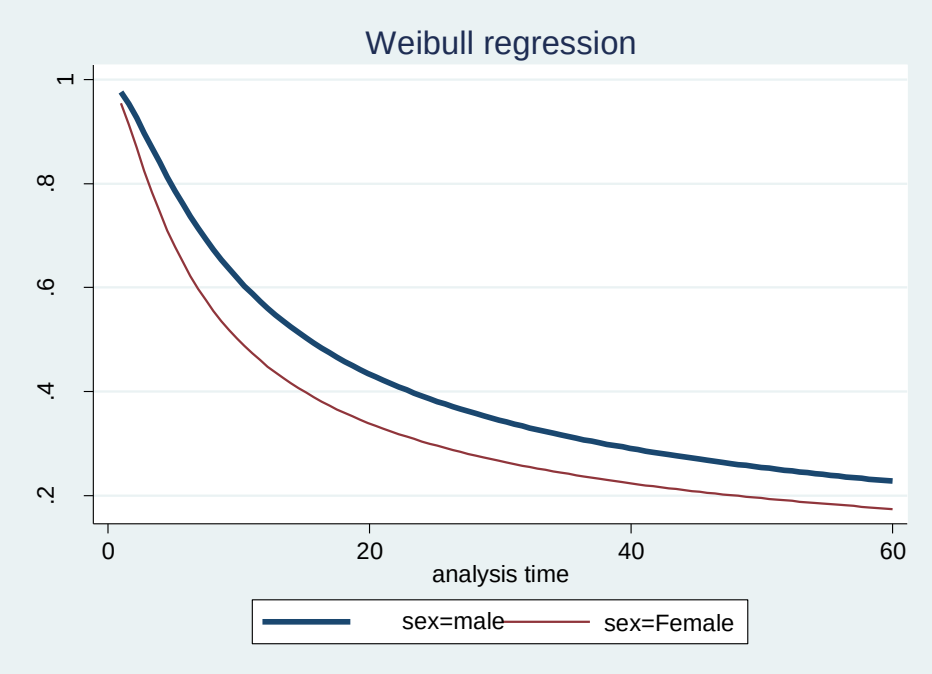

Figure 1 Survival curve for Weibull model for gender

Figure 1 shows the survival curve for male and female in which male are having better survival pattern than female.

Table 2 Posterior Summaries for Spatial Weibull frailty model

Parameter	Spatial Weibull frailty Model			
	Mean	MC error	Credible Interval	
β [Tr.Cat]	-0.129	0.0047	-0.469,	0.198
β [Age]	-0.011	0.0006	-0.039,	0.017
β [Sex]	0.381	0.0088	-0.206,	1.021
β [Wt]	-0.044	0.0017	-0.074,	-0.015
β [Const]	-1.039	0.2455	-2.615,	0.492
γ	0.004	0.0032	0.001,	0.164
σ	0.739	0.0198	0.428,	1.218
τ	2.276	0.0613	0.674,	5.472

The table 2 shows the Bayesian spatial Weibull frailty model for the HIV-TB data. The Bayesian spatial Weibull frailty model's MC error is less for all the covariates. The credible Interval for spatial Weibull frailty model is also. The precision for τ is 2.276 to the Weibull model.

Table 3 Goodness of Fit for Weibull Models

No.	Model	Bayesian Weibull frailty			
		Spatial		Non-Spatial	
		pD	DIC	pD	DIC
1	$\beta_0 + \beta_{Tr.Cat} * Tr.Cat + \beta_{Age} * Age + \beta_{Sex} * Sex + \beta_{Wt} * Wt + W_i + V_i$	89.67	701.525	101.92	719.291

2	$\beta_0 + \beta_{Tr.Cat} * Tr.Cat + \beta_{Age} * Age + \beta_{Sex} * Sex + W_i + V_i$	13.755	864.435	123.61	896.164
3	$\beta_0 + \beta_{Tr.Cat} * Tr.Cat + \beta_{Age} * Age + \beta_{Wt} * Wt + W_i + V_I$	15.096	855.437	119.34	881.235
4	$\beta_0 + \beta_{Tr.Cat} * Tr.Cat + \beta_{Sex} * Sex + \beta_{Wt} * Wt + W_i + V_i$	14.977	854.624	91.341	876.473
5	$\beta_0 + \beta_{Age} * Age + \beta_{Sex} * Sex + \beta_{Wt} * Wt + W_i + V_i$	14.893	853.856	87.128	875.932
6	$\beta_0 + \beta_{Tr.Cat} * Tr.Cat + \beta_{Age} * Age + \beta_{Sex} * Sex + \beta_{Wt} * Wt$	25.992	973.724	67.243	1051.18

The Deviance Information Criterion(DIC) was used to assess the model fit for this data in which first model contains full model with correlated effect were captured through (W) and uncorrelated effects were captured through (V). The model with lowest DIC value was considered to be better model. From the above six models, spatial frailty models fits better than the other non-spatial frailty model and the minimum value of DIC 701.5 in model(1) of Weibull model better than other spatial model. The model(5) consists of age, sex and weight had the next lowest DIC(853.8) after ignoring Treatment regimen, indicates that the effect of Treatment regimen was very less impact on death. The model(4) and model(3) are almost having same DIC value viz, 854.6 and 855.4 respectively. The model(2) contains treatment regimen, age and sex were having high value DIC(864.4) and ignoring weight in this model indicates that weight at baseline is important factor for death in our analysis. The last model(6) is a fixed effect model, with highest DIC value of 973.7, indicates that random effects with in region and spatial autocorrelation between ward were important factors for estimating any parameter in spatial analysis.

VI. CONCLUSION

In the standard approach on frailty modeling, the random effects are treated as independent. But in the Bayesian spatial survival modeling approach, spatial dependence between neighboring effects are modeled with a spatial prior. The conditionally autoregressive prior allows to incorporate this spatial dependence in their survival models. The random effect model considers the other unmeasured factors influencing death in the region. Also, it was found that weight at baseline, was one of the factor associated with death in our study and hazard rate of male having two fold high as comparing to female. The spatial correlation was diverse in Chennai ward and north-eastern wards in Chennai having spatial high spatial dependence. Spatial random effect model accounts higher heterogeneity (2.3) in our model which indicates that the regional variation and other environmental factors influencing survival pattern of disease in this study. Weibull spatial frailty model fits better other than the non-spatial Weibull frailty model. Spatial frailty accounts higher heterogeneity between regions. It is clear that a non-spatial and non-frailty model understates the unexplained heterogeneity in the data. Hence, spatial frailty model captures the unexplained spatial heterogeneity and it draws accurate inferences about other spatial covariates of interest.

REFERENCES

[1] Banerjee S, Carlin B and Gelfand A E (2004): Hierarchical Modeling and Analysis for Spatial Data. Boca Raton: Chapman & Hall.

[2] Banerjee S, Wall M M and Carlin B P (2003): Frailty modeling for spatially-correlated survival data with application to infant mortality in Minnesota. Biostatistics, 4: 123-142.

[3] Besag J (1974): Spatial interaction and the statistical analysis of lattice systems. Journal of Royal Statistical Society, Series-B, 36: 192-236.

[4] Besag J and Kooperberg C (1995): On conditional and intrinsic autoregressions. Biometrika, 82: 733-746.

[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] Carlin B P and Banerjee S (2003): Hierarchical multivariate CAR models for spatio-temporally correlated survival data, Bayesian Statistics 7. Oxford: Oxford University Press.

[7] Carlin B P and Hodges J S (1999): Hierarchical proportional hazards regression models for highly stratified data. Biometrics, 55: 1162-1170.

[8] Cox D and Oakes D (1984): Analysis of Survival Data. London: Chapman and Hall.

[9] Gelfand A Diggle P and Guttorp P (2010): Handbook of spatial statistics, Chapman & Hall / CRC.

[10] GeoBUGS User Manual (2004): GeoBUGS User Manual. Version 1.2.
[11] Lawson A, Browne W and Vidal Rodeiro C (2003): Disease mapping with WinBUGS and MLwiN. John Wiley & Sons Ltd.
[12] Li Y and Ryan L (2002): Modeling spatial survival data using semiparametric frailty models. Biometrics, 58, 287-297.
[13] Spiegelhalter D, Thomas A, Best N and Lunn D (2002b): WinBugs User Manual Version 1.4. Cambridge, England: MRC Biostatistics Unit.
[14] Venkatesan P and Srinivasan R (2008): Bayesian model of HIV/AIDS in India: A spatial Analysis. Applied Bayesian Statistical Analysis : 51-56.
[15] Venkatesan P and Srinivasan R (2011): Bayesian intrinsic conditional autoregressive random effect model of tuberculosis : An application, Recent trends in Statistics and computer application, Journal from Manonmanian Sundaranar University : 195-200.
[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] Xia H and Carlin B (2005): Multivariate parametric spatiotemporal models for country level breast cancer survival data, Life time Data analysis, 11: 107–120.