RESEARCH PAPER | Statistics | Volume : 3 | Issue : 10 | Oct 2013 | ISSN - 2249-555X

Modelling of Breast Cancer Survival Data: A Frailty Model Approach

KEYWORDS: Breast cancer, accelerated failure time models, heterogeneity, shared frailty Model

Pari Dayal L	Leo Alexander T
DRBCCC Hindu College, Chennai	Loyola College, Chennai
Ponnuraja C	**Venkatesan P**
NIRT(ICMR), Chennai	NIRT(ICMR), Chennai

ABSTRACT *The implication of ignoring the unobserved heterogeneity while estimating the parameter of time to event survival model which assumes impliciting the homogeneous population. It is also important to procure and express the analysis for varying treatment effects among different subjects of patients in clinical trials. This paper aims to implement a class of flexible accelerated failure time model with proportional hazard assumptions and random effect. The model is applied to breast cancer data with frailty assumptions and compared with non-frailty models. The results show that the frailty model approach is compared with the other models for survival prediction.*

Introduction

In many applications, particularly in clinical trials, survival analysis implicitly assumes a homogeneous population to be studied Cox (1972). In most of the applications, the study population cannot be assumed to be homogeneous as the effect of drug may be individual specific or group specific or each subjects has its own biological response to the treatment (Aalen,1988). It is a basic observation of medical statistics that individuals are dissimilar (Murphy,1992). The natural course of a disease varies a lot from person to person. This heterogeneity is often termed as biological variation (Aalen,1988) and it is generally recognized as one of the most important sources of variation in medicine and biology. Still, there is a tendency to regard that this variation as a nuisance, and is not as something to be considered seriously on its own right. For instance, a typical clinical trial draws conclusions about the average effect of treatment, and do not say much about how the effect varies between patients. Therefore, there is a mixture of individuals with different hazards. It is not always possible to obtain all relevant covariates related to the study on disease of interest. The heterogeneity may be explainable in terms of observed covariates though there will always be an unexplained residual. Hence, the heterogeneity is considered basically as unobserved and is manifesting only indirectly (Aalen,1988).

Frailty models have been used when groups of subjects have responses that are likely to be dependent in some general way. Liang, et al., (1995) discusses the use of frailty models with multivariate failure time data. If the value of the frailty is assumed to be constant within groups, the models are called as shared frailty models. The shared frailty model has been extensively discussed by many authors (Picklets, et al., 1994; and Yashin, et al., 1995; Vu, et al., 2001; Duchateau, et al.,2002; Rahgozar, et al., 2008). A flexible approach to the parametric analysis of survival data with frailty have been elaborated and discussed by Lambert et al., (2004).

Materials and Methods

In modelling the survival data, handling of censored observations is a key important aspect. In this paper, we consider only right censoring observations. The proportional hazards specification expresses the hazard in terms of a baseline hazard, multiplied by a constant. The hazard function is that of a Weibull model and is reparameterized as (Cox and Oakes,1984)

$$h(t,\beta) = \alpha\ (\alpha t)^{\gamma-1} where\ \alpha = \exp(-x'\beta)$$

The unobserved differences between the observations are modelled through a multiplicative scaling factor ν which is a random variable taking on positive values, with mean normalised to one and finite variance σ^2. It is to desire to have a common value of frailty for a group of observations.

This random effect for the ith cluster, n_i, is incorporated conditionally into the proportional hazard function previously examined:

$$h(t/\nu_i) = \nu_i h_0(t) exp(\beta x_j) \quad [1]$$

which may be re-expressed as

$$h(t/\nu_i) = h_0(t) exp(\beta x_j + \eta_i), \quad [2]$$

showing n_i, actually behaves as an unknown covariate for the ith cluster in the model. η_i is also a random effect term which follows normal distribution and $\exp(\eta_i)$ follows log normal distribution. Using the relationship between the survival and the hazard function, the conditional survival function is

$$S(t/\nu_i) = Exp[\nu_i \Lambda_0(t) exp(\beta x_j)\ ] \quad [3]$$

and the conditional likelihood is

$$L(\gamma,\beta \mid \nu_i) = \prod_{i=1}^{I} \prod_{j=1}^{n_i} \left(h(t_j \mid \nu_i)^{\delta^j}\ S(t_j \mid \nu_i)\right) \quad [4]$$

where there are I clusters, ith one being of size n_i and g and b represents baseline hazard and regression parameters, respectively. On substitution it gives

$$L(\gamma,\beta \mid \nu_i) = \prod_{i=1}^{I} \prod_{j=1}^{n_i} \left([h_0(t)\nu_i exp(\beta X_j)]^{\delta_j}\ exp[-\nu_i \Lambda_0(t) exp(\beta X_j)]\right)$$
$$= \prod_{i=1}^{I} \prod_{j=1}^{n_i} \left(\frac{\phi}{\Phi}\right)^{\delta_j} \Pi\, exp(-\Phi)(\Phi \quad [5]$$

where $\phi = \nu_i\, exp(\beta' X_i)\, exp(\alpha' W_j)\, \pi_{ij}^{r-1}$

$$\Phi = \nu_i\, exp(\beta' X_i) \sum_{j=1}^{n_j} exp(\alpha' W_j)\, \pi_j^r = \nu_i\, exp(\beta' X_i)\, e_i$$

The marginal likelihood $L\ \gamma,\beta)$ is obtained through integration of the random effect distribution which is also independent of ν_i.

In a shared frailty model, the ith group and fixed observed covariate vector x_{ij}, is assumed as

$$h_{ij}(t \mid y_i, x_j) = y_i h_0(t) exp(x_j'\beta) \quad i = 1,\ldots,n \quad , j = 1,\ldots,n_i \quad [6]$$

Volume : 3 | Issue : 10 | Oct 2013 | ISSN - 2249-555X

where $h_0(t)$ is an unknown baseline hazard function which is common to every subject and β is the vector of fixed effect parameters. The shared frailty variable Y_i is assumed to be independent and identically distributed for groups of patients. To model a frality,the most commonly used is Gamma distribution,which is given by

$$f_{Y_i}(y_i) = \frac{1}{\Gamma\theta}\theta^{\theta} y_i^{\theta-1} exp(-\theta y_i), \quad i = 1,...,n \qquad [7]$$

The higher values of $(\theta - 1)$ signifies larger variances for y_i, consequently greater heterogeneity among different groups of patients. The role of shared frailty model is most useful when we consider multivariate survival times.

Results

We consider the database consisting of 368 breast cancer women patients diagnosed at Cancer Institute (WIA), Chennai, India and follow-up period up to 180 months. The event of interest was time to death. Overall 187(51%) cases have experienced the event and 63% of 130 are of stage 3B cases.

The demographic and disease characteristics of the patients are given in table 1

Table 1: Classification of death according to Stages and Age group

	Stages			Age groups	
	Stage2B N (%)	**Stage3A** N (%)	**Stage3B** N (%)	**Age <50 years** N (%)	**Age ³ 50 years** N (%)
Death Status					
Alive	61 (55)	72 (56)	48 (37)	115 (53)	66 (44)
Dead	**49 (45)**	**56 (44)**	**82 (63)**	**103 (47)**	**84 (56)**
Total	110	128	130	218	150

From the table1, we see that death increases with the severity of stages and ages. The event experienced cases among age group in more than 50 years is higher than the less than 50 years (Pari Dayal et al., 2013). The linear predictor is set equal to the intercept in the reference group (stage = 3); this defines the baseline hazard. The corresponding distribution of survival time is Gamma(Cox and Oakes, 1984).

The following *Table2* illustrates the accelerated failure time model and estimate the cumulative distribution function of time to death among breast cancer patients.

Table 2: Iteration History for Fixed-Effects Model (without random effect)

Iteration History					
Iteration	evaluations	NegLog-Like	Diff	MaxGrad	Slope
1	2	4525.01	9162.29	18100000000.0	-31380000.0
5	31	3362.22	205.23	29351809.0	-11.4
10	41	2318.78	210.86	93508.6	-12.0
15	55	1773.42	29.19	4716.1	-64.7
20	67	1536.56	10.46	1210.4	-14.4
25	77	1402.25	34.70	743.2	-130.3
30	86	1180.29	32.98	213.5	-41.5
35	96	1011.98	10.36	64.5	-36.7
40	106	976.94	0.02	3.4	-0.1
41	108	976.94	0.00	0.1	0.0
42	110	976.94	0.00	0.0	0.0
Parameter Estimates					
Parameter	Estimate	SE	p Value	lower	upper
gamma	3.0934	0.1892	<.0001	2.7214	3.4654
b0	4.863	0.0241	<.0001	4.8155	4.9104
b1(stage)	-0.01183	0.0314	0.7065	-0.07358	0.04991
-2 log likelihood is 1953.8					

The fixed effects failure time model is used the Dual Quasi-Newton method for optimization technique. There is no initial values assigned for *gamma*, *beta0* and *beta1*. The "Iteration History" in Table2 shows that the procedure converges after 40 iterations and 110 evaluations of the objective function along with the parameter estimates. The parameter estimates and their standard errors are shown in *Table2* with confidence limits. The deviance (-2 log likelihood) of this model is 1953.8. Since the slope estimate is negative with *p*-value of 0.7065, there is no significant difference between stages, but the estimated probabilities give information about patient-to-patient variation within and between stages. Further, the deaths are more in the higher stages than compared to the lower stages.

Table 3: Iteration History for Random-Effects Failure Time Model and parameter estimates

Iteration History					
Iteration	evaluations	Neg-gLog-Like	Diff	MaxGrad	Slope
1	2	1586.79	150.99	90.61	-2610.21
5	13	1129.33	140.89	138.22	-48.75
10	23	982.36	6.36	31.80	-15.54
15	32	976.61	0.32	9.88	-0.30
20	42	976.34	0.00	0.80	-0.01
25	51	976.34	0.00	0.29	0.00
26	52	976.34	0.00	0.09	0.00
27	54	976.34	0.00	0.00	0.00
Parameter Estimates(Random Effect)					
Parameter	Estimate	SE	p Value	lower	upper
gamma	3.0834	0.1882	<.0001	2.6214	3.2254
b0	4.2571	0.0124	<.0001	4. 1050	4.4153
b1(stage)	-0.0101	0.0241	0.5235	-0.0558	0.0346
logsig	-0.3443	0.0330	0.7180	-0.0800	0.4000
-2 log likelihood is 1952.7					

The *Table3* shows that the objective function is computed by adaptive Gaussian quadrature because of the presence of random effects and also reports that nine quadrature points are being used to integrate over the random effects. It models the hazard for patient with random effect. The random effect is V in the linear predictor. This term enables to accommodate and estimate patient-to-patient variation in their status by introducing random effects into a subject's hazard function. The empirical Bayes estimates of the random effect and the estimated cumulative distribution function are also saved to subsequently graph the patient-specific distribution functions. The procedure converges in less than 15 iterations and 54 evaluations (*Table3*). The achieved -2 log likelihood is 1952.7 marginally less than the fixed effect model. The AIC, BIC and AICC are similar between fixed effect and random effect models.

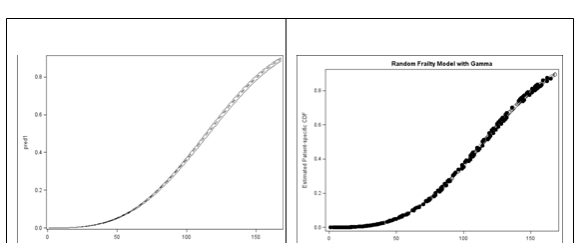

Figure1 : Patient-Specific Cumulative Distribution Functions and Predicted Values for all stages. Stage 2B: Solid Lines with filled Circles(above); Stage3A: Dotted Middle Lines with Open Circles, Stage 3B: Solid Lines with filled Circles(below). Pre plotted(Left) and plotted(Right)

The predicted values and patient-specific survival distributions is plotted. The separation of the distribution functions by stages is marked in *figure1*. Most of the distributions of patients in the stage2B are to the left of the distributions in the stage3A and stage3B. The separation is not absolute. However, several patients who are in the higher stages experience the event more in amount than patients in the beginning stage.

Table4: Patient-Specific Cumulative Distribution Functions and Predicted Values for all stages: pcdf1 for Stage 2B(110 cases); pcdf2 for Stage3A(128 cases); pcdf3 for

Volume : 3 | Issue : 10 | Oct 2013 | ISSN - 2249-555X

Stage 3B(130 cases)

pt	pcdf1	pcdf2	pcdf3	pt	pcdf1	pcdf2	pcdf3	pt	pcdf1	pcdf2	pcdf3
1	0.62419	0.61075	0.39650	46	0.79220	0.75133	0.00150	91	0.64197	0.64602	0.64122
2	0.76390	0.82579	0.38755	47	0.78531	0.16889	0.34359	92	0.07088	0.60177	0.00002
3	0.00066	0.00616	0.01674	48	0.74140	0.35381	0.83265	93	0.01240	0.01736	0.04961
4	0.87482	0.00310	0.00007	49	0.41920	0.00012	0.21653	94	0.56941	0.01588	0.00255
5...	0.41920	0.60177	0.00401	50...	0.06710	0.52841	0.00298	95...	0.00001	0.02417	0.03368
20...	0.20176	0.00416	0.80178	65	0.66808	0.72853	0.07737	110...	0.27807	0.61075	0.62385
35	0.82470	0.68842	0.00216	80	0.01367	0.77311	0.00298	125	.	0.00310	0.47864
36	0.20176	0.25382	0.00018	81	0.00722	0.65466	0.02154	126	.	0.54695	0.00298
37	0.31153	0.76597	0.00079	82	0.56941	0.00544	0.00181	127	.	0.57454	0.00181
38	0.00161	0.50979	0.00181	83	0.65946	0.00127	0.46023	128	.	0.01588	0.00047
39	0.04710	0.61966	0.78856	84	0.65946	0.50047	0.00255	129	.	.	0.26910
40...	0.74902	0.00874	0.05583	85...	0.11596	0.72853	0.00460	130	.	.	0.00079

In Table 4 the predicted values of patient-specific cumulative distribution functions are calculated for all stages of breast cancer. The stage 2B is having 110 patients, stage 3A is 128 patients and stage 3B is 130 patients. The 'pcdf1', 'pcdf2', 'pcdf3' are in table4 is for stage2B, stage3A and stage3B respectively. The preferred prediction values for all three types are illustrated under the heading of 'pcdf1', 'pcdf2' and 'pcdf3' in Table 4.

Discussion and conclusion

The parameter estimates of both fixed effects and random effects failure time models are identifiying the same risk factor for time to death breast cancer patients. Venkatesan et al. (2009) modeled the predictions based on survival as well as regression approaches for breast cancer patients and he also **highlighted** the essential to adjust for patient-related factors that could potentially affect the survival time of breast cancer patients**(Venkatesan et al.,2011)**. Swaminathan et al. (2010) projected the utility of a mixture model to estimate the cure fraction when PH assumptions is violated for patients with breast cancer. Kong et al.(2010) expressed the parametric frailty models provide viable ways to study the relationship between exposure variables and clustered survival outcome that is subject to random censoring. But we documented through this paper that the role of frailty model showed the advantages than the all other models fits the breast cancer data and compared the same in the absence of frailty factor too. The frailty factor varies from individual to individual and is proved to be observable. Here in, it focuses the term frailty and its enables to accommodate and estimate patient-to-patient variation in their status by introducing random effects into a subject's hazard function. The same was concentrated and proved at last. Since the hazard function is non-negative, frailty factor must be restricted to non-negative values. The frailty model for per subject basis assumes the random effect and it varies between individuals. The shared frailty model for group factor assumes an unexplained heterogeneity and is shared by related individuals with frailty in common with several individuals. The random effect model predicted values for stages using adaptive Gaussian quadrature in the presence of frailty factor. The deviance between these two appraoches are highlighting the same. Though the deviance of random effect model is marginaly lesser than the fixed effect model, but it is ultimately appreciable. In most of the cases, a frailty model can only imply a positive correlation within group. This is substandard in some situations, because it is not reflecting the reality.

REFERENCE

1. Aalen O. O. (1988). Heterogeneity in Survival Analysis. Statistics in Medicine, 7, 1121-37. | 2. Cox, D.R. (1972). Regression model and life tables (with discussion). Journal of the Royal Statistical Society(B), 34, 187–220. | 3. Cox, D.R. (1975). Partial Likelihood. Biometrika, 62, 269-76. | 4. Cox, D.R and Oakes, D. (1984), Analysis of Survival Data. London Chapman and Hall. | 5. Duchateau L, Janssen P, Lindsey P, Legrand C, Nguti R and Sylvester R (2002) The shared frailty model and the power for heterogeneity tests in multicenter trails. Comp. Stat. Data Analysis. 40, 603–620. | 6. Kong X, Archer K.J, Moultan L.H, Gray R.H, Wang M.C (2010). Parametric Frailty Models for Clustered Data with Arbitrary Censoring: Application to Effect of Male Circumcision on HPV Clearance. BMC Medical Research Methodology 2010, 10-40. | 7. Lambert P, Collett D, Kimber A, Johnson R: Parametric accelerated failure time models with random effect and an application to kidney transplant survival. Statistics in Medicine 2004, 23:3177-3192. | 8. Liang, K.-Y., Self, S. G., Bandeen-Roche, K. J., and Zeger, S. L. (1995). Some recent developments for regression analysis of multivariate failure time data. Lifetime Data Analysis, 1, 403-415. | 9. Murphy SA (1992) Consistency in a proportional hazards model incorporating a random effect. Annals Stat. 22, 712–731. | 10. Pari Dayal L, Leo Alexander T, Ponnuraja C, Ranganathan Rama, Venkatesan P. Modelling of time to event breast cancer data using Accelerated failure time (AFT) in south India women. Global Research Analysis, 2013, 2(5), 192-194. | 11. Picklets, A., Crouchley, R., Simonoff, E., Eaves, L., Meyer, J., Rutter, M., Hewitt, J., Silberg, J. (1994). Survival Models for Developmental Genetic Data: Age of Onset of Puberty and Antisocial Behavior in Twins. Genetic Epidemiology, 11, 155 – 70. | 12. Ponnuraja, C., Venkatesan, P.(2010). Correlated frailty model: an advantageous approach for covariate analysis of tuberculosis data. Indian Journal of Science and Technology, 3(2), 151-155 | 13. Rahgozar M, Faghihzadeh S, Rouchi GB and Peng Y (2008). The power of testing a semi-parametric shared gamma frailty parameter in failure time data. Stat. Med. 27, 4328-4339. | 14. Rama R, Swaminathan R, Venkatesan P.(2010). Cure models for estimating hospital-based breast cancer survival. Asian Pac J Cancer Prev. 2010;11(2):387-91 | 15. SAS/STAT 9.2 User's Guide. The NLMIXED Procedure: SAS Institute Inc; Chapter16:4421-4431. | 16. Venkatesan P, Raman TT, Ponnuraja C(2011). Survival Analysis of Women with Breast Cancer under Adjuvant Therapy in South India. Asian Pacific Journal of Cancer Prevention.12, 1-4 | 17. Venkatesan P, Suresh M.L.(2009). Breast Cancer Survival Prediction using Artificial Neural Network. International Journal of Computer Science and Network Security. 9(5), 169-174. | 18. Vu H, Segal MR, Knuiman MW and James IR (2001), Asymptotic and small sample statistical properties of random frailty variance estimates for shared gamma frailty models. Comm. Stat—Simulation Comp. 30, 581–595. | 19. Yashin, A. I., Vaupel, J. W. and Iachine, IA (1995). Correlated individual frailty: an advantageous approach to survival analysis of bivariate data, Mathematical Population Studeis, 5, 145-159. |