International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Semiparametric Multi State Model for Time-To-Event Data

Ramakrishnan .M[1], Viswanathan .N[2]

[1]Assistant professor, Department of Mathematics, RKM Vivekananda College, Chennai, India

[2]Assistant Professor, Department of Statistics, Presidency College, Chennai, India

Abstract: ***Survival Analysis is the study about time-to-event data. It stands apart from classical estimation, in the sense that it has censoring objects with incomplete information to be dealt with. Classical Survival models usually contain two events, of which one is treated as terminal event. Multi State Models (MSM) involve more than two states, of which some may be transient and others absorbing. The multi-state Markov model is a useful way of describing a process in which an individual moves through a series of states in continuous time. Multi-state models can be used to model the movement of patients between different states, such as, hospitalization, recovery, relapse and death. These models may offer a better understanding of the process due to transition specific nature of the events. Also, the estimated transition probabilities from one state to another throws more light on the nature of movements and the possible reasons behind the transitions. In this paper, a multi state model with three states is considered under Semiparametric multi state approach. This method enables us to identify transition specific covariates that throw more light on the entire transition process and the factors influencing the same. A Cox Proportional hazard multi state model is used to derive the necessary estimates and testing procedures are carried out using 'mstate' package of R, a open source software.***

Keywords: Multi State Markov Model, Three State Survival Model, Cox PH model, Transition specific covariates

1. Introduction

Survival analysis is the analysis of data measured from a specific time of origin until an event of interest or a specified endpoint (Collett, 1994). For example, in order to determine the incidence of death due to AIDS among HIV positive patients, every patient will be followed from a baseline date (such as date of diagnosis) until the date of death due to AIDS or study closing date. A patient who dies of HIV during the study period would be considered to have an 'event' at the date of death. A patient who is alive at the end of the study would be considered to be 'censored'. Thus, every patient provides two pieces of information: follow-up time and status (censored status). This model is called as Classical Survival model. In longitudinal studies, patients are observed over time and covariate information is collected at several occasions. In such studies, some state may be partitioned into two or more intermediate (transient) states, each of which corresponds to a particular stage.

2. Multi state Models

The simplest form of MSM is the mortality model for survival analysis with states "alive" and "dead" and only one possible transition. Splitting the "alive" state into two transient states, it is called as simplest progressive three-state model. Both models are special cases of the k-progressive model, illustrated in Figure 2.1

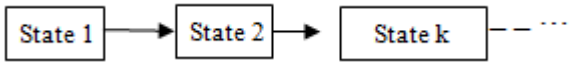

Figure 2.1: k- Progressive Model

Another MSM used for the disease progression is the illness-death model which can be used to study the incidence of the disease and the rate of death. This is illustrated in Figure 2.2.

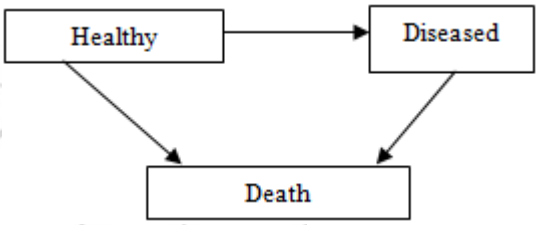

Figure 2.2: Illness-Death Model.

An MSM is a stochastic process (X(t), t ϵT) with a finite state space, where X(t) represents the state occupied by the process at time t $\geq$ 0. In general, the future state transitions of an MSM may depend on past events. However, for the special case of a Markov model the past and future are independent given its present state. Extensive literature on MSMs are available. Main contribution include books by Anderson, Borgan, Gill, and Keiding (1993), Hougaard (1999), Beyersmann, Schumacher and Allignol (2012) and Willekens, F. (2014). Recent reviews on this topic may be found in the papers by Putter, Fiocco, and Geskus (2007), Putter (2014). The 'mstate' package in R for the analysis of multi-state models was developed by Wreede, Fiocco and Putter (2011).

2.1 Preliminaries in Multi State Models

2.2.1Transition intensity If T denotes the time of reaching state j from state i then hazard rate (transition intensity) of the i j transition is

$$\lambda_{ij}(t) = \lim_{\Delta t \to 0} \frac{Prob(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

This definition makes an assumption that the multi-state model is Markovian, which implies that the probability of going to future state depends only on the present state and not on its history.

2.2.2 Cumulative transition hazard

Cumulative transition hazard for transition i j is defined as

$$\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(s)\, ds$$

2.2.3 Cox Proportional Hazards model

The hazard for the transition i → j for a subject with covariate vector **Z** is

$$\lambda_{ij}(t|\boldsymbol{Z}) = \lambda_{ij,0}(t)\exp(\boldsymbol{\beta}_{ij}^T \boldsymbol{Z})$$

where $\lambda_{ij,0}(t)$ is the baseline hazard of transition i → j, and β_{ij} is the vector of regression coefficients that describe the effect of **Z** on transition i → j.

2.2.4. Test based on the Schoenfeld Residuals

Several formal statistical tests have been proposed for assessment of proportionality of hazards. A simulation study by Ng'andu (1997) described and compared several tests in the Cox PH framework, and concluded that the scaled Schoenfeld residuals test (Grambsch and Therneau,1994), the linear correlation test (Harrell, 1986) and the time-dependent covariate test(Cox, 1972) were the most powerful diagnostic tools for proportionality. The other statistical test of the proportional hazards assumption is based on the Schoenfeld residual. The Schoenfeld residuals are defined for each subject who is observed to fail. If the PH assumption holds for a particular covariate then the Schoenfeld residual for that covariate will be independent of survival time. So this test is accomplished by finding the correlation between the Schoenfeld residuals for a particular covariate and the ranking of individual survival times. The null hypothesis is that the correlation between the Schoenfeld residuals and the ranked survival time is zero. Rejection of null hypothesis concludes that PH assumption is violated.

The scaled residuals scatter in a nonsystematic way around the zero line, and the polygon connecting the values of the smoothed residuals has approximately a zero slope and crosses the zero line several times. Then we conclude that hazard function may be proportional in that covariate. If the polygons connecting the values of smoothed residuals have not zero slope and crosses the zero line only once or it has consistent positive slope, suggesting that the importance of the covariate increases over time and thus has a non proportional hazard.

2.2.5 Approaches in Time Scale

Clock forward: Time t refers to the time since the patient entered the initial state.

Clock reset: Time t in $\lambda_{ij}(t)$ refers to the time since entry in state i. The clock is reset to 0 each time the **patient** enters a new state.

3. Applications to German Breast Cancer Study (GBCS)

3.1 Data and Model Description

GBCS data obtained from the German Breast Cancer Study Group, which they used to illustrate the methods for building prognostic models (Sauerbrei and Royston, 1999) is used for this study. In the main study, a total of 720 patients with primary node positive breast cancer were recruited between July 1984, and December 1989. Data used in this study consists of 686 subjects with complete data contain three events primary node positive, recurrence and death respectively. Using Purposeful selection of covariates method, the following prognostic factors were identified and their distribution of the values is shown in table 3.1

387(299) 128(171)

Figure 3.1 Transition numbers of events and Censored observations
(Figures inside the bracket denotes the events)

Table 3.1: Distribution of Prognostic factors

Progonastic factor	*Categories*	*Number*	*in %*
Hormone Therapy	No	440	64
	Yes	246	36
Tumour Grade	1	81	12
	2	444	65
	3	161	23
Number of Nodes	< 4	376	55
	≥ 4	310	45
Age	(20,46]	181	26
	(46,53]	182	27
	(53,61]	155	23
	(61,81]	168	24
Tumour Size in mm	(2,25]	353	51
	(25,120]	333	49
Number of Progesterone receptors	< 33	343	50
	≥ 33	343	50

The above table indicates that the covariate values are evenly spread among different categories.

The model development and testing are carried out using 'mstate' package in R and the corresponding outputs are provided below:

Table 3.2: Parameter estimates using 'Clock forward' approach

covariates	coef	exp(coef)	SE(coef)	z	p
hormone.1	-0.4241	0.6543	0.1283	-3.3100	0.0010
hormone.2	-0.1633	0.8493	0.1882	-0.8700	0.3856
grade.1	0.1686	1.1836	0.1106	1.5200	0.1274
grade.2	0.1536	1.1660	0.1691	0.9100	0.3639
nodescut.1	0.9228	2.5163	0.1219	7.5700	0.0000
nodescut.2	0.2311	1.2600	0.1901	1.2200	0.2241
agecut1.1	-0.1555	0.8560	0.1649	-0.9400	0.3458
agecut1.2	0.5653	1.7600	0.2263	2.5000	0.0125
agecut2.1	0.1718	1.1874	0.1632	1.0500	0.2926
agecut2.2	0.3503	1.4195	0.2396	1.4600	0.1437
agecut3.1	0.0530	1.0544	0.1628	0.3300	0.7449
agecut3.2	0.4018	1.4945	0.2379	1.6900	0.0913
sizecut.1	0.0895	1.0936	0.1210	0.7400	0.4593
sizecut.2	0.2905	1.3372	0.1730	1.6800	0.0931
progcut.1	-0.6552	0.5194	0.1276	-5.1400	0.0000
progcut.2	-0.8421	0.4308	0.2122	-3.9700	0.0001

Table 3.3: Test of the proportional hazards assumption based on the Schoenfeld residual using 'Clock forward' approach

covariates	rho	Chi-square	p
hormone.1	-0.0116	0.0607	0.8054
hormone.2	0.0442	0.9429	0.3315
grade.1	-0.0864	2.9429	0.0863
grade.2	-0.0235	0.2602	0.6100
nodescut.1	-0.0838	3.0185	0.0823
nodescut.2	0.0142	0.1079	0.7426
agecut1.1	0.0440	0.8661	0.3520
agecut1.2	0.0214	0.2144	0.6433
agecut2.1	0.1270	7.1586	0.0075
agecut2.2	0.0170	0.1243	0.7244
agecut3.1	0.1009	4.5632	0.0327
agecut3.2	-0.0588	1.5368	0.2151
sizecut.1	0.0086	0.0347	0.8523
sizecut.2	-0.0024	0.0027	0.9588
progcut.1	0.1420	8.8152	0.0030
progcut.2	0.0873	3.6615	0.0557
GLOBAL	NA	38.3936	0.0013

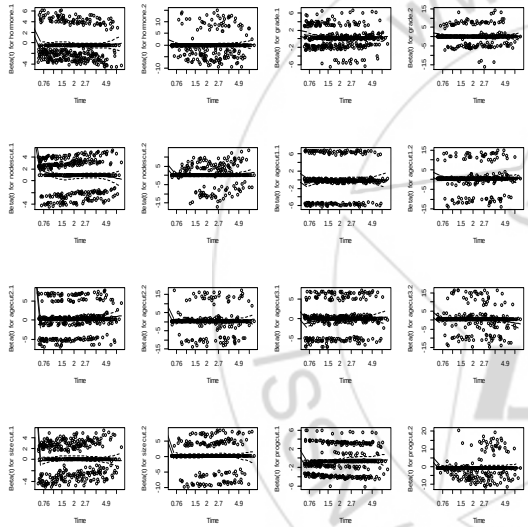

Figure 3.2: Scatter plot of scaled Schoenfeld residuals using 'Clock forward' approach

Table 3.4: Parameter estimates using 'Clock reset' approach

Parameters	coef	exp(coef)	se(coef)	z	p
hormone.1	-0.4241	0.6543	0.1283	-3.3100	0.0009500*
hormone.2	-0.1536	0.8576	0.1713	-0.9000	0.3698
grade.1	0.1686	1.1836	0.1106	1.5200	0.1274
grade.2	0.1925	1.2122	0.1573	1.2200	0.2212
nodescut.1	0.9228	2.5163	0.1219	7.5700	0.0000000*
nodescut.2	0.2833	1.3275	0.1721	1.6500	0.0998
agecut1.1	-0.1555	0.8560	0.1649	-0.9400	0.3458
agecut1.2	0.6289	1.8756	0.2167	2.9000	0.0037100*
agecut2.1	0.1718	1.1874	0.1632	1.0500	0.2926
agecut2.2	0.4134	1.5120	0.2290	1.8100	0.0710
agecut3.1	0.0530	1.0544	0.1628	0.3300	0.7449
agecut3.2	0.4912	1.6342	0.2228	2.2000	0.0275100*
sizecut.1	0.0895	1.0936	0.1210	0.7400	0.4593
sizecut.2	0.2148	1.2396	0.1616	1.3300	0.1838
progcut.1	-0.6552	0.5194	0.1276	-5.1400	0.0000003*
progcut.2	-0.8387	0.4323	0.1945	-4.3100	0.0000160*

*Denotes significance at 5% level.

Table 3.5: Test of the proportional hazards assumption based on the Schoenfeld residual using 'Clock reset' approach

Parameters	rho	chi-square	p
hormone.1	-0.0108	0.0549	0.8147
hormone.2	-0.0048	0.0117	0.9137
grade.1	-0.0748	2.3100	0.1287
grade.2	-0.0101	0.0507	0.8219
nodescut.1	-0.0757	2.5800	0.1082
nodescut.2	0.0302	0.4950	0.4818
agecut1.1	0.0386	0.7000	0.4028
agecut1.2	-0.0366	0.6280	0.4281
agecut2.1	0.1110	5.7300	0.01670*
agecut2.2	-0.0153	0.1070	0.7435
agecut3.1	0.0882	3.6500	0.0562
agecut3.2	-0.0975	4.3400	0.0373
sizecut.1	0.0092	0.0412	0.8392
sizecut.2	-0.0001	0.0000	0.9990
progcut.1	0.1200	6.5900	0.01020*
progcut.2	0.0169	0.1500	0.6984
GLOBAL	NA	29.6000	0.0204

*Denotes significance at 5% level.

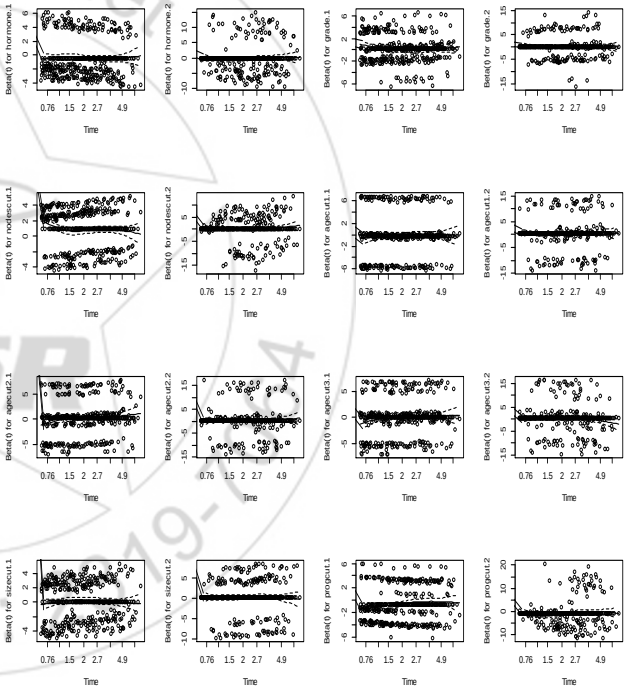

Figure 3.3: Scatter plot of scaled Schoenfeld residuals using 'Clock reset' approach

4. Summary and Conclusion

Multi state model identifies significance of each state and their respective transitions in survival analysis. In this study, subjects who have undergone hormone therapy have recurrence at a rate 35% lower than subjects who do not undergo hormone therapy. In transition 1, subjects who have more than three tumor nodes have higher hazard than those with lesser number of nodes. In both transitions, subjects having higher (greater than 33) number of Progesterone receptors have lower hazard compared to those with relatively lower number of Progesterone receptors. Difference between 'clock forward' and 'clock reset' approaches is quite small with regard to the estimated

regression coefficients. No evidence of non-proportionality of the base line transition intensities are seen in the two transitions (P value > 0.05), except for a few subclasses in the covariates. Multistate model, when compared to classical survival models, brings out extra features that aid in analyzing survival structures.

References

[1] Aalen, O.O. (1978). Non parametric inference for afamily of counting processes. *Annals of Statistics*, 6, 701-726.
[2] Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistical Association, 53*, 457-481.
[3] Meira-Machado L, Cadarso Suarez C and Una-Alvarez J. (2008). Inference in the progressive three-state model. *International journal of Mathematical Models and Methods in Applied Scicnces*,**2(3)**, 447-454.
[4] Meira-Machado L, Cadarso Suarez C, Una-Alvarez J and Andersen, K.(2009). Multi-state models for the analysis of time-to-event data. *Statistical Methods Medical Research*,**18(2)**,195-222.
[5] Nelson.W. (1972). Theory and Applications of Hazard Plotting for Censored Failure Data. *Technometrics*, 14, 945-966.
[6] Putter, H, Fiocco, M and Geskus, R.B. (2006). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*,**26**, 2389-2430.
[7] Sadagopan, JM, Ben-Porat, L, Berwick, M, Robson, M, Kutler, D and Auerbach, AD. (2004). A note on competing risks in survival data analysis. *British Journal of Cancer*, **91**, 1229-1235.
[8] Willekens, F. (2014). Multi state Analysis of Life Histories with R. New York: Springer series.
[9] Wreede,L.C, Fiocco, M and Putter, H. (2011). mstate:An R Package for the Analysis of Competing Risks and Multi-State Models. *Journal of Statistical Software*, **38(7)**, 1-30.

Author Profile

Ramakrishnan M has received his Doctorate degree in Mathematics-Statistics under University of Madras in 2015 with specialization in Survival Analysis and served as Assistant Professor in Mathematics in RKM Vivekananda College, Chennai from 2008 to 2016.

Viswanathan. N has received his Doctorate degree in Statistics under University of Madras in 2016 with specialization in Survival Analysis and served as Assistant Professor in Statistics in Presidency College, Chennai from 2001 to 2016.