ISSN: 0975–833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol.3, Issue, 5, pp.096 -100, May, 2011

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

REVIEW ARTICLE

A DIAGNOSTIC CHECK ON COX PH MODEL USING RESIDUALS FOR TB MENINGITIS DATA

[1]Ramadurai, M. and [2]Ponnuraja, C.

[1] Department of Statistics, University of Madras, Chennai 600 005
[2] Tuberculusis Research Centre, ICMR, Chetpet, Chennai 600 031

ARTICLE INFO

Article History:
Received 14th March, 2011
Received in revised form
5th April, 2011
Accepted 29th April, 2011
Published online 14th May 2011

Key Words:
Cox PH model,
Schoenfeld residuals,
TB meningitis

ABSTRACT

In survival analysis, the Cox PH model is a breakthrough in that the effects of the covariates on survival times have been studied, by fitting models using the sample data. Once a model has been fitted, there are a number of aspects of the fit of the model that need to be studied. One of the model checking procedures is based on quantities known as residuals, such as Cox-Snell residuals, Schoenfeld residuals etc. TB Meningitis is one of the diseases which mainly affect children. In this paper, the Cox PH model is fitted for TB Meningitis data collected from the Tuberculosis Research Centre, ICMR, Chennai, which consist of survival times of children along with other covariates. The fitted model is assessed for its proportionality assumption through Schoenfeld residuals.

INTRODUCTION

Survival analysis is the subject which became very popular during the later part of the 20th century. Many models have been proposed, modified and adopted to study the sample data which involves censoring. Cox (1972) proposed a regression model under semi parametric set up called PH (proportional hazards) regression model. This model has found applications in biomedical studies and later, all disciplines like Economics, Sociology, and Psychology, started applying the model in their studies. The model after being fitted has been diagnosed for check using residuals. Schoenfeld residual is one of the popular procedures applied in this context. Section 2 gives the details of the Cox regression model and in Section 3 the concepts on Schoenfeld residuals have been provided. The TB Meningitis is a form of TB prevalent among children. Application of the Cox model and the associated Schoenfeld residuals to TB Meningitis data for diagnostic checking has been highlighted in section 4, along with a discussion on the results.

Cox Proportional Hazard Model

The Cox proportional hazards model (Cox,1972) is the most commonly used approach for analyzing the relationship between survival and one or more covariates (also called predictors). Based on the Cox (1972 and 1975), Collect (2003), Everitt et al (2004), Klein and Moeschberger (1996) and Kalbfleisch and Prentice (1980), we present below a brief description of the Cox proportional hazard model. The general proportional hazards model is given by

$$\lambda_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi})\lambda_0(t) \qquad (1)$$

$$\log\left\{\frac{\lambda_i(t)}{\lambda_0(t)}\right\} = \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi}$$

The dataset, consists of the triplet $(T_i, \delta_i, \underset{\sim}{z}_i); i = 1,2,...,n$, where T_i is the time of the event, δ_i is the indicator and $\underset{\sim}{z}_i$ is the vector of covariates or risk factors for the individual ($\underset{\sim}{z}_i$ may be a function of time) which may affect the survival distribution of T. The relationship between the distribution of event time and the covariates or risk factors can be described in terms of a model, in which the hazard rate at time t for an individual is

$$\lambda(t_i; z) = \lambda_0(t)\exp(\underset{\sim}{\beta}' \underset{\sim}{z}_i) \qquad (2)$$

where

1. i) t_i is the time to the event.
2. ii) $\underset{\sim}{z}_i$ is the $p \times 1$ vector associated with individual i, denoting the covariates.
3. iii) $\lambda_0(t)$ is the baseline hazard rate, an unknown (arbitrary) function.
4. iv) $\underset{\sim}{\beta} = (\beta_1, \beta_2, ..., \beta_p)$ is the vector of coefficients of explanatory variables $z_1, z_2, ..., z_p$.

Corresponding author: cponnuraja@gmail.com

It may be noted that $\exp(\beta' z)$ may be replaced by a non-negative function $g(z,\beta)$. Special assumptions on $\lambda_0(t)$ leads to parametric models, such as exponential regression models, Weibull regression models etc. But the advantage of Cox's model is that such assumptions can be avoided, and hence is known as semi parametric model. The Cox model is called as a proportional hazards model since the ratio of the hazard rates of two individuals with covariate values z and z^* is $\frac{\lambda(t;z)}{\lambda(t;z^*)} = \exp\beta'(z - z^*)$ an expression that does not depends on t. The likelihood of the estimator is given by

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp(\beta' x_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta' x_l)} \qquad (3)$$

Schoenfeld residuals

Schoenfeld residual has an advantage over the Cox-Snell residual, so we use this residual to asses the adequacy of the model fitted. Let there be *p* covariates with *n* independent observations of time, covariates and censoring indicator which are denoted by the triplet (t_i, x_i, c_i), where $i = 1,2,...,n$ and $c_i = 1$ for uncensored observation is zero otherwise. Schoenfeld (1982) proposed the residuals which are based on the individual hazards model. This derivative for the k^{th} *covariate* in the log of the partial likelihood (Cox, 1972) is

$$\frac{\partial L_p(\beta)}{\partial \beta_k} = \sum_{i=1}^{n} c_i \left\{ x_{ik} - \frac{\sum_{j \in R(t_i)} x_{jk} e^{x_j'\beta}}{\sum_{j \in R(t_i)} e^{x_j'\beta}} \right\} = \sum_{i=1}^{n} c_i \left\{ x_{ik} - \bar{x}_{w_i k} \right\} \qquad (4)$$

where

$$\bar{x}_{w_i k} = \frac{\sum_{j \in R(t_i)} x_{jk} e^{x_j'\beta}}{\sum_{j \in R(t_i)} e^{x_j'\beta}} \qquad (5)$$

The estimator of the Schoenfeld residual for the i^{th} subject on the k^{th} covariate is obtained form (5) by substituting the partial likelihood estimator of the coefficient, $\hat{\beta}$, and is given by

$$\hat{r}_{ik} = c_i (x_{ik} - \bar{x}_{w_i k}) \qquad (6)$$

where

$$\hat{\bar{x}}_{w_i k} = \frac{\sum_{j \in R(t_i)} x_{jk} e^{x_j'\hat{\beta}}}{\sum_{j \in R(t_i)} e^{x_j'\hat{\beta}}}$$

Since the partial likelihood estimator of the coefficient, $\hat{\beta}$, is equal to zero, the sum of the Schoenfeld residuals is zero. The scaling the Schoenfeld residuals by an estimator of its variance yields a residual with greater diagnostic power than the unscaled residuals was suggested by Therneau and Grambsch (2000). Let the vector of *p* Schoenfeld residuals for the i^{th} subject be denoted as

$$\hat{r}_i' = (\hat{r}_{i1}, \hat{r}_{i2}, ..., \hat{r}_{ip}),$$

where $\hat{r}_{ik}$ is the estimator given in (6). The vector of scaled Schoenfeld residuals is the product of the inverse of the covariance matrix times the vector of residuals, as

$$\hat{r}_i^* = \left[Var\hat{r}(\hat{r}_i) \right]^{-1} \hat{r}_i \qquad (7)$$

For the i^{th} subject, the diagonal elements in this matrix are

$$Var\hat{r}(\hat{r}_i)_{kk} = \sum_{j \in R(t_i)} \hat{w}_{ij} \left(x_{jk} - \hat{\bar{x}}_{w_i k} \right)^2$$

and the off diagonal elements are

$$Var\hat{r}(\hat{r}_i)_{ki} = \sum_{j \in R(t_i)} \hat{w}_{ij} \left(x_{jk} - \hat{\bar{x}}_{w_i k} \right) \left(x_{ji} - \hat{\bar{x}}_{w_i i} \right)$$

where

$$\hat{w}_{ij} = \frac{e^{x_j'\hat{\beta}}}{\sum_{j \in R(t_i)} e^{x_j'\hat{\beta}}}$$

$Var\hat{r}(\hat{r}_i)$, tends to be fairly constant over time. If this matrix is constant, its inverse may be approximated coefficients by multiplying the estimator of the covariance matrix of the estimated coefficients by the number of events.

$$\left[Var\hat{r}(\hat{r}_i) \right]^{-1} = m Var\hat{r}(\hat{\beta})$$

where *m* is the observed number of uncensored survival times. Hence the Schoenfeld residuals are in the form as shown below

$$\hat{r}_i^* = m Var\hat{r}(\hat{\beta}) \hat{r}_i \qquad (8)$$

Application to the TB Meningitis data: Results and Discussion.

Tuberculosis (TB) meningitis occurs when tuberculosis bacteria **(**Myobacterium tuberculosis**)** invade the membranes and fluid surrounding the brain and spinal cord. The infection usually begins elsewhere in the body, usually in the lungs, and then travels through the bloodstream to the *meninges* where small abscesses (called *microtubercles*) are formed. When these abscesses burst, TB meningitis is the result. Of all form of TB, TB Meningitis carries a high mortality and morbidity. In areas where TB prevalence is high, TB meningitis is most common in children aged 0 - 6 years, and in areas where TB prevalence is low, most cases of TB meningitis are in adults. This data have been collected from the Tuberculosis Research Centre, ICMR, Chennai (Ramachandran *et al.,* 1997). The data presented here is obtained from randomized clinical trial and it consists of 225 cases of TB meningitis among children. The event of interest is occurrence of death over a period of five years. The following are the covariates were also considered including time and status (event-1, censored-0):The Treatments (treatment1, treatment2), Mantoux (mm), History of contact TB (yes-1, no-0), Clinical stages (conscious-1, semi-conscious-2 and un-conscious-3), Weight at baseline (in kg), Chest X-ray (normal-0, abnormal-1), Sputum Smear (positive-1,negative-0), Sputum Culture (positive-1, negative-0), Protein level (normal-0, abnormal-1), Sex (male-1, female-0) and Age (in months). The Cox PH model is fitted for the

TB Meningitis data and the following are the deviance and parameter estimates for the regression model are given in the following tables (Table 1a and Table 1b).

Table 1a. Crude method of Cox PH model

Variable	-2LL	d.f.	Chi-Square	Significance
None	1114.092	-	-	-
Age	1110.645	1	3.447	0.063
Smear	1110.664	1	3.428	0.064
Cl. Stage	1110.901	2	3.191	0.074
Weight	1112.912	1	1.179	0.277
Protein	1113.953	1	0.139	0.710
Treat	1113.956	1	0.136	0.713
Culture	1114.05	1	0.042	0.838
X-ray	1114.062	1	0.030	0.862
Sex	1114.072	1	0.020	0.888
Family History	1114.075	1	0.017	0.896
Mantoux	1114.078	1	0.014	0.905

Table 1b. Adjusted method of Cox PH model with all covariates

Variable	Parameter Estimate	SE	HR	95% CI	
				Lower Limit	Upper Limit
Treatment	-0.064	0.175	1.066	0.757	1.503
Mantoux	-0.001	0.011	0.999	0.977	1.021
Family History	0.025	0.192	1.025	0.704	1.492
Clinical Stage	-0.341	0.188	0.711	0.492	1.028
Weight	0.028	0.026	1.029	0.978	1.082
X-ray	-0.030	0.176	0.970	0.688	1.368
Smear	-1.105	0.718	0.331	0.081	1.353
Culture	-0.044	0.215	0.957	0.628	1.458
Protein	0.072	0.193	1.074	0.736	1.568
Sex	0.025	0.176	1.025	0.726	1.447
Age	0.006	0.003	1.006	1.000	1.012

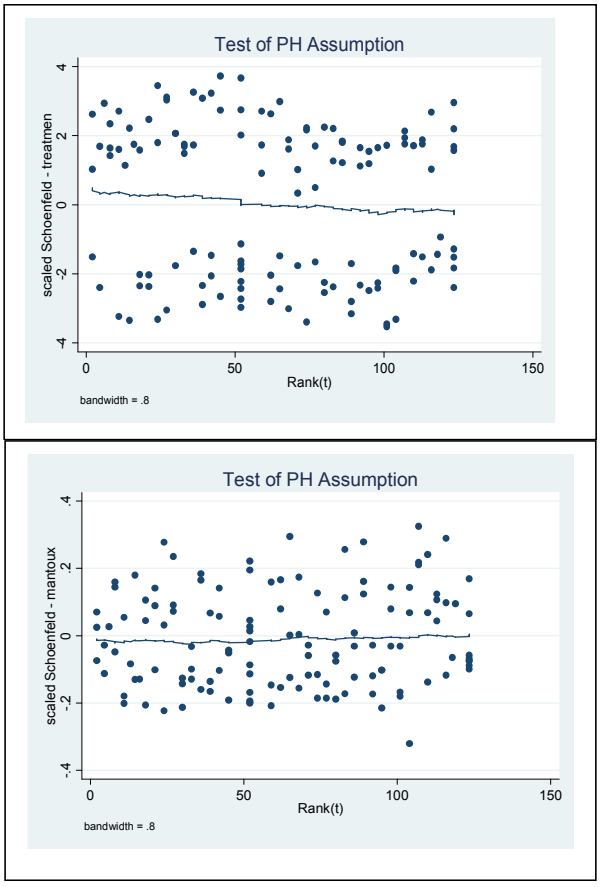
It is clear from Table.1a that the covariates age, smear, clinical stages influence the time to death of patients. Whereas Table.1b are clearly indicates that none of the variables are not significant while fitting the adjusted method of Cox PH model. This clearly indicates that the number of covariates considered here simultaneously adjusts with each other. How a statistical test of the proportional hazards assumption is performed for a given covariate using the TB meningitis dataset is demonstrated in the following. This is accomplished by finding the correlation between the Schoenfeld residuals for a particular covariate and the ranking of individual failure (death) times as established by Harrel and Lee (1986). If the proportional hazard assumption is met then the correlation should be near zero. The p-value for testing this correlation can be obtained from the Schoenfeld residuals. The ranking is calculated for those who experience of events. The null hypothesis framed is that the PH assumption is not violated. The ordering of the names corresponds to the order of the independent variables in the model statement. The actual variable names are arbitrary. The names of variables in Table 2 are given after ranking through Schoenfeld residuals. The p-values for treatment, mantoux, family history of contact, clinical stages, weight, x-ray, sputum smear, sputum culture, protein level, sex and age are given in Table 2, suggesting that the PH assumption is not violated for any of the above variables and all are seems reasonable. The sample correlations *rho* with their corresponding p -values shown in Table 3.

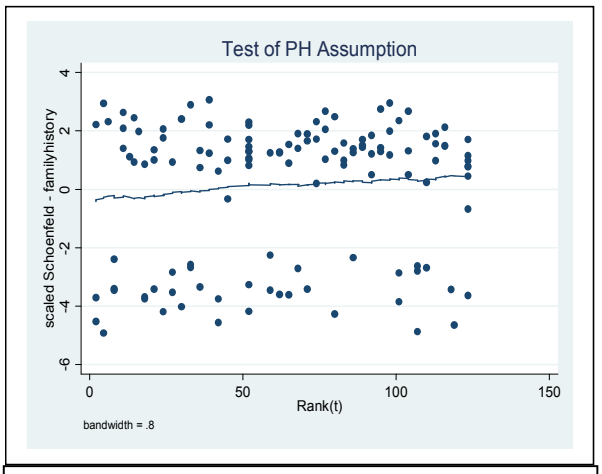
Table 2. The correlations between the ranked failure time variables and the variables containing the Schoenfeld residuals

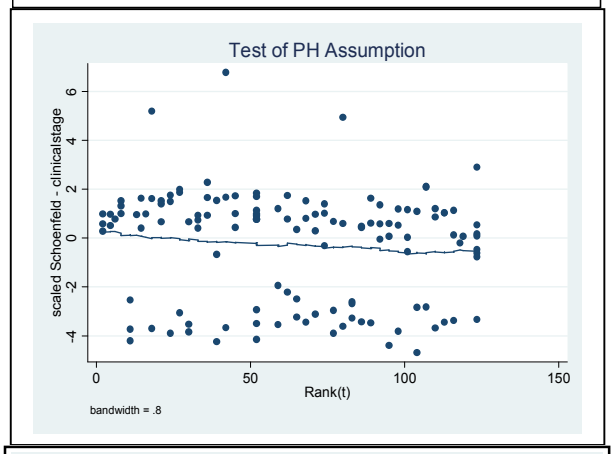
Variables	Time Rank	Rank for Variable time
rtreatment	-0.12531	0.1522
rmantoux	0.14511	0.0969
rfamily	0.15574	0.0746
rclinstages	-0.15027	0.0855
rweight	-0.00526	0.9523
rxray	-0.00179	0.9838
rsmear	-0.01574	0.8579
rculture	-0.05099	0.5615
rprotein	-0.13569	0.1208
rsex	0.02751	0.7541
rage	0.02271	0.7960

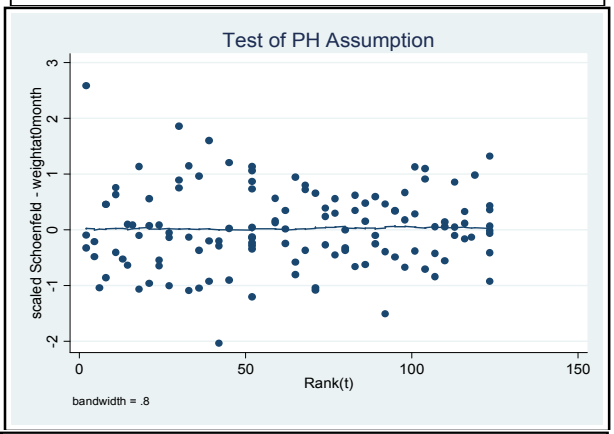
Table 3. Test of proportional hazards assumption using global test and Schoenfeld scaled residuals for separate tests with each covariate

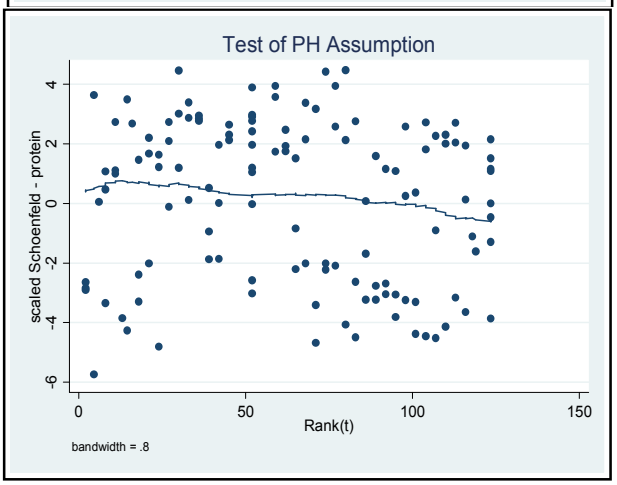
	rho	χ^2	Prob> χ^2
Treatment	-0.12467	2.20	0.1379
Mantoux	0.07927	0.81	0.3691
Family History	0.11221	1.70	0.1920
Clinical Stage	-0.13128	2.47	0.1160
Weight	0.03612	0.21	0.6440
X-ray	0.03888	0.20	0.6523
Smear	-0.07833	0.85	0.3576
Culture	-0.00825	0.01	0.9207
Protein	-0.12011	2.07	0.1498
Sex	0.00807	0.01	0.9238
Age	0.01092	0.02	0.8859
Global Test		12.79	0.3073



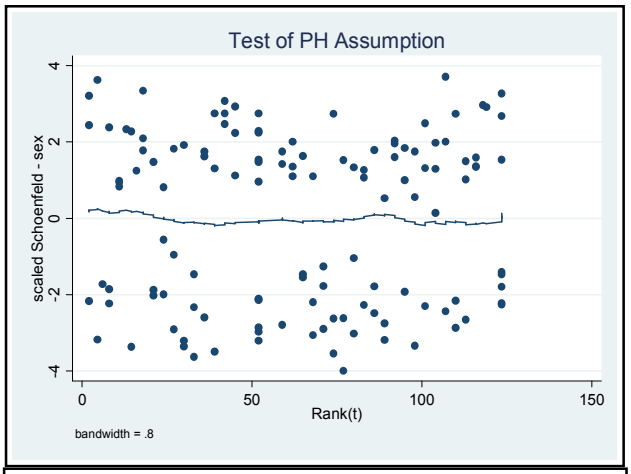








The PH global test is performed for all the covariates simultaneously and the test also can be carried out for each covariate separately. The idea behind the PH test is that if the PH assumption is satisfied, then the residuals should not be correlated with survival time (or ranked survival time). On the other hand, if the residuals tend to be positive for subjects who become events at a relatively early time and negative for subjects who become events relatively at a later time, then there is evidence that the hazard ratio is not constant over time; implying that PH assumption is violated. But the global test is not significant and it reflects that the PH assumption is not at all violated. Moreover the Schoenfeld scaled residuals for separate tests with each covariate suggest that the PH assumption is not violated for all covariates such as treatment, mantoux, family history of contact, clinical stages, weight, xray, sputum smear, sputum culture, protein level, sex and age. Rank plots of the scaled Schoenfeld residuals for all covariates against their ranking of survival time are given in the following figure. On performing the rank plot of the scaled Schoenfeld residuals for all variables against survival time ranking it is seen that the PH assumption is not violated for every plot. It is also seen that majority of the fitted curves give the impression of being horizontal and very few slightly upward but that too satisfy the PH assumption. Even though the reference line of the variable protein is slightly deviated from the originality, this variable is also satisfying the PH assumption numerically. The other supportive evidences for the PH assumption and related Schoenfeld residuals are given in Table 4.



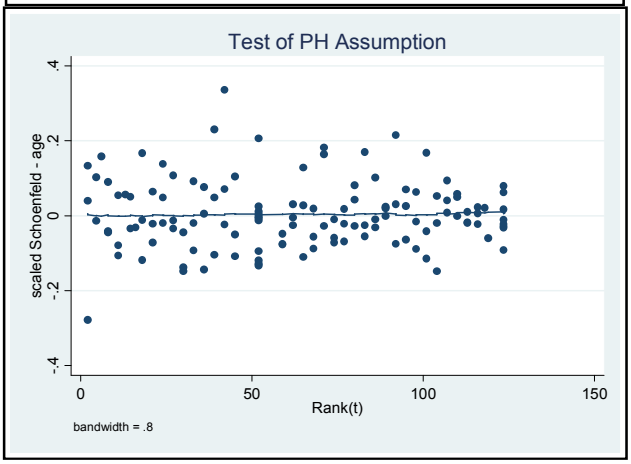


Table.4. The first 25 observations of all the variables containing the Schoenfeld residuals (only for those experienced the event)

S.No.	rtreatmen	rmantoux	rfamily	rclinstage	rweight	rxray	rsmear	rculture	rprotein	rsex	rage
1	0.481695	-5.13682	0.322884	0.261321	-3.33922	0.511825	-0.00766	-0.21113	0.278872	-0.56948	-21.3077
2	0.463298	10.31835	0.299807	0.297152	-2.16951	0.49907	-0.007	-0.19423	0.277245	-0.56127	-23.8453
3	0.470135	7.074123	-0.65749	0.220226	5.314181	-0.50083	-0.00905	-0.23151	-0.72519	0.453222	-20.1223
4	0.470135	-4.925880	0.342506	0.220226	-1.68582	0.499168	-0.00905	-0.23151	-0.72519	-0.54678	-8.12227
5	0.475075	-5.376860	-0.68866	0.280047	-1.89907	0.500548	-0.00683	-0.20501	0.283224	0.416911	2.441632
6	-0.528600	-5.356040	0.297913	-0.69756	-0.00573	0.504693	-0.0073	-0.20221	0.284245	0.406603	13.54849
7	0.475075	6.623143	-0.68866	0.280047	0.100927	0.500548	-0.00683	-0.20501	-0.71678	-0.58309	-15.5584
8	-0.518300	-5.136820	-0.67712	0.261321	2.160781	-0.48818	-0.00766	-0.21113	-0.72113	0.430523	26.69235
9	-0.536700	-5.681650	0.299807	0.297152	-3.66951	0.49907	-0.007	-0.19423	-0.72276	0.438734	-23.8453
10	0.463985	-5.590190	0.304225	0.291268	-2.20084	0.500281	-0.00799	-0.1911	-0.72037	0.432771	-12.1301
11	0.471397	-5.356040	0.297913	-0.69756	-2.00573	0.504693	-0.0073	0.797792	0.284245	0.406603	-34.4515
12	0.471397	6.643956	0.297913	-0.69756	1.994274	-0.49531	-0.0073	-0.20221	0.284245	0.406603	1.548494
13	0.481567	-5.299420	0.311304	0.269593	0.136786	-0.4891	-0.0079	0.786858	0.277502	-0.57264	26.66663
14	-0.529860	-4.925880	-0.65749	0.220226	8.314181	-0.50083	-0.00905	0.768492	0.274805	0.453222	75.87773
15	0.475075	6.623143	-0.68866	0.280047	0.100927	0.500548	-0.00683	0.794994	0.283224	-0.58309	-15.5584
16	-0.526130	7.862059	0.34645	-0.7676	-3.41731	-0.45333	-0.01608	-0.23109	0.266874	0.401053	-13.8027
17	0.427795	16.113520	-0.62321	0.19114	-2.34033	0.536107	-0.01749	-0.25133	0.264276	-0.56382	-13.7649
18	-0.539790	12.057310	0.329577	0.174313	3.533541	-0.44616	-0.01881	-0.23906	0.261415	-0.55235	33.2862
19	-0.553060	7.321925	0.320896	0.149423	-0.34765	-0.49274	-0.02078	-0.26402	0.261388	0.417307	-1.06714
20	0.464640	15.091410	0.356567	0.117754	-0.12752	0.48339	-0.02309	0.706632	0.242165	0.415415	-11.8185
21	-0.540440	-5.730200	-0.70205	-0.69225	3.708031	-0.489	-0.00749	-0.2018	0.276364	0.440426	11.22214
22	-0.546360	6.215259	0.288513	0.3124	-5.273	0.513138	-0.00765	-0.19926	0.276526	0.450024	-30.7314
23	0.448823	11.260420	0.287334	0.318473	-4.31545	-0.48948	-0.0078	0.796867	0.276098	-0.55382	-31.0903
24	0.450366	4.217665	0.285358	0.316336	-1.31207	-0.49217	-0.00802	-0.20185	0.273806	0.441674	-12.8474
25	0.453123	-5.634640	0.27679	-0.68394	-1.37921	-0.49587	-0.0082	-0.19882	0.279997	-0.56336	-25.3162

In Table 4, the Schoenfeld residuals for the first 25 observations which experiences event of interest (death) in TB Meningitis are given for brevity. It is seen from the above analysis that for TB Meningitis data, the proportionality assumption has been satisfied. The Schoenfeld residual check plays a dominant role in validating the diagnostic check on the Cox PH model.

REFERENCES

Collett, D. 2003. Modeling of Survival Data in Medical Research. *Chapman Hall*, London, U. K.

Cox, D.R. 1972. Regression model and life tables (with discussion). *Journal of the Royal Statistical Society(B),* 34, 187–220.

Cox, D.R. 1975. Partial Likelihood. *Biometrika,* 62, 269-276.

Cox, D.R. and Snell,E.J. 1968. A general definition of residuals. *Journal of Royal Statistical society*, B. 248-275.

Everitt, B.S. and Pickles, A. 2004. Statistical Aspects of the Design and Analysis of Clinical Trials. *Imperial College Press*. London.

Harrel, F. and Lee, K. 1986. Proceedings of the eleventh annual SAS user's group International. 823-828.

Kalbfleisch, J.D and Prentice, R.L 1980. The statistical analysis of failure time data, *John Wiley and Sons,* London.

Klein and Moeschherger,1996. Survival Analysis: A Self-Learning Text, *Springer*, New York.

Oakes, D. 1983. Comparison of models for survival data, *Statistics in Medicine,*2, 305-311.

Ramachandran, P., Duraipandian, M. and Reetha, M.1997. A 5 year follow-up study of children treated for Tuberculosis Meningits with short course Chemotherapy, *Indian Journal Tuberculosis*, 44,125-127.

Schoenfeld, D. 1982. Partial residuals for the proportional hazards model. Biometrika, 69:51-55.

Therneau,T.M. and Grambsch,P.M. 2000. Modeling Survival Data: Extending the Cox Model. *Springer Inc*. USA.

****