Use of Proportional Hazards Model to analyze the data of patients with Hodgkin's lymphoma

Survival Data Analysis
2024-2025

HOMEWORK 2

2nd year Master of Statistics
Hasselt University

Group members:
Christine Katabarwa (2365938)
Mary Grace Lapurga (2365930)
Lulu Asmi Lathifah Sundayana (2365939)

Submission Date: 7 November 2024

Lecturer:
Prof. Tomasz Burzykowski

1 Introduction

This study aimed to evaluate the effect of six factors on the hazard of relapse, second malignancy, or death for 865 Hodgkin's lymphoma patients. These factors are age at diagnosis (in years), sex (0=female, 1=male), clinical stage (1-I, 2-II), use of chemotherapy (0-no, 1-yes), mediastinum involvement (0-none, 1-small, 2-large), and presence of extranodal disease (0-no, 1-yes).

2 Methodology

The following proportional hazards model was initially fitted to the data, with age as a continuous covariate and the other factors as categorical.

$$\lambda(t) = \lambda_0(t)e^{\beta Z}$$

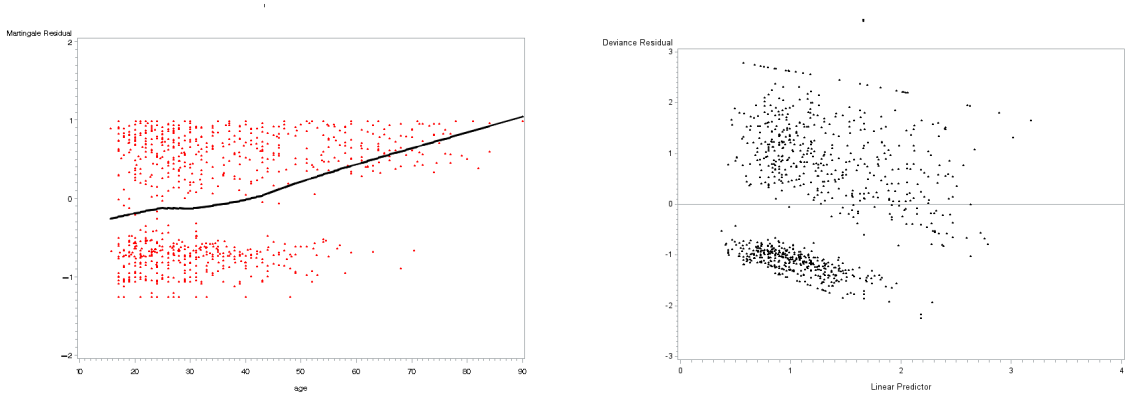
with

$$\begin{aligned}\beta Z =& \beta_1 Age + \beta_2 ClinicalStage + \beta_3 Sex + \beta_4 Chemotherapy + \beta_5 Nodes + \beta_6 Mediastinum_1 \\ &+ \beta_7 Mediastinum_2\end{aligned}$$

The Efron method was used to address tied events in the partial likelihood estimation of model coefficients. Several residual-based methods were applied for model diagnostics. Martingale residuals were used to check the functional form of age as a continuous covariate, ensuring that it is appropriately specified in the model. Deviance residuals were used to assess the overall fit of the model. To test the assumption of proportional hazard, Schoenfeld residuals were used that allowed us to detect any time-dependent effect of covariates. A model with time-dependent covariate was then fitted when the proportionality assumption was not satisfied.

3 Results and Discussion

Figure 1a shows the plot of martingale residuals against the age. The straight line suggests that the age should be included as a linear term in the model. And from Figure 1b, it is observed that the deviance residuals are symmetrically and somewhat randomly spread around zero. However, there are some points lying outside the (-2, +2) interval, which may indicate potential outliers or that the model does not fit the data well.



(a) Functional form of age

(b) Overall fit of the model

Figure 1: Model Diagnostics

Figure 2 shows the scaled Schoenfeld residuals plotted against the observed survival times, along with the p-value of the proportional hazard assumption test for each covariate. It suggests that the coefficient of age may vary with time, while the other covariates satisfy the proportional hazard assumption. Moreover, the p-value of the global test of proportional hazards over all covariates suggests that the assumption does not hold for the model.

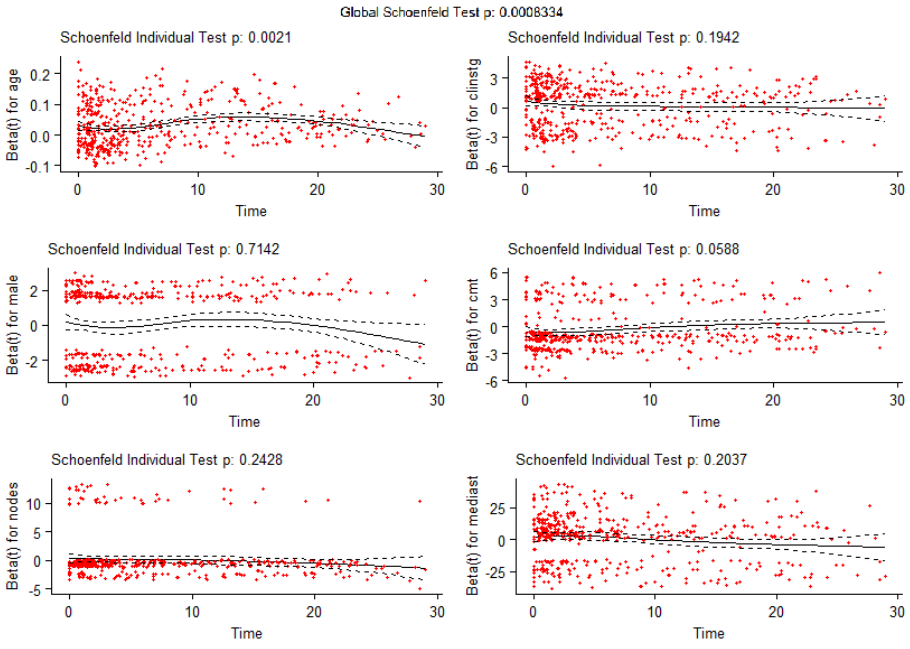


Figure 2: Test of Proportional Hazard using Schoenfeld Residuals

In view of the violation of proportionality assumption, time-dependent age was introduced

in the model. Table 1 presents the parameter estimates from fitting the models with and without time-dependent age variable, for comparison. The two models show that age, clinical stage, and use of chemotherapy significantly affect the hazard of relapse, second malignancy, or death. A slight difference was observed from both models in terms of parameters estimates, except for age. Model 1 and Model 2 indicate that hazard is higher by 34.2% and 35.7%, respectively, for patients with stage 2 lymphoma than those with stage 1, for fixed values of other covariates. Meanwhile, the hazard is lower by 28.2% and 29.1% for patients using chemotherapy. In terms of the effect of age, Model 1 suggests that for every year increase in patient's age, the hazard increases by 3.3% and the age effect is constant over time. On the other hand, Model 2 suggests that for every year increase in age, the effect on hazard is 2.5% at baseline (month=0) and it is not constant over time. Specifically, the effect at month 12 is estimated to be 4.2%; 6% at month 24; and 7.8% at month 36.[1]

Since the proportionality assumption does not hold for Model 1, interpreting the effect of age as constant over time can be incorrect. Also, the log-likelihood ratio test between Model 1 and Model 2 shows a significant result (p-value = 0.002) indicating that the time-dependent age should be in the model. Hence, Model 2 is deemed more appropriate.

Table 1: Parameter estimates from Cox's models

Parameter	Model 1: without time-dependent age			Model 2: with time-dependent age		
	Coefficient	Exp(Coeff)	p-value	Coefficient	Exp(Coeff)	p-value
Age	0.0321*	1.0326	<0.0001	0.0247*	1.0251	<0.0001
Clinical stage 2	0.2941*	1.3419	0.0077	0.3056*	1.3574	0.0056
Sex (1=male)	0.0278	1.0282	0.7670	0.0286	1.0291	0.7605
Chemotherapy (1=yes)	-0.3308*	0.7184	0.0060	-0.3438*	0.7091	0.0044
Nodes (1=present)	0.0689	1.0713	0.6817	0.0794	1.0826	0.6367
Mediastinum (small)	-0.1060	0.8994	0.3721	-0.1118	0.8943	0.3466
Mediastinum (large)	0.0627	1.0647	0.7204	0.0563	1.0579	0.7480
t(Age)[a]				0.0014*	1.0014	0.0022

*significant at 1% level; [a] - t(Age) is a time-transform term that allows age to vary dynamically in time.

4 Conclusion

In this study of Hodgkin's lymphoma patients, we find that age, clinical stage, and use of chemotherapy have a significant effect on the hazard of relapse, second malignancy, or

[1] At baseline, the effect of every unit increase in age is exp(0.0247)=1.025;
At month=12: exp(0.0247+0.0014*12)=1.042;
At month=24: exp(0.0247+0.0014*24)=1.060;
At month=36: exp(0.0247+0.0014*36)=1.078;

death. Age shows a time-dependent effect on hazard with increasing effect over time. On the other hand, there is no sufficient evidence to conclude that sex, presence of extranodal disease, and mediastinum involvement have an effect on the hazard.

5 Reference

Therneau, T.M., & Grambsch, P.M. (2000) Modeling Survival Data: Extending the Cox Model. *Springer, New York*.

Therneau, T. M., & Lumley, T. (2022). coxph function: Cox Proportional Hazards Model. In survival package (Version 3.6-4) [R package documentation]. RDocumentation. Retrieved on 11 November 2024, from https://www.rdocumentation.org/packages/survival/versions/3.6-4/topics/coxph

Appendix

R and SAS Codes

```
#Fitting the Cox's models
model1 <- coxph(Surv(firsttime, event) ~ age + clinstg + male + cmt + nodes + mediast,
data=data, ties="efron")
model2 <- coxph(Surv(firsttime, event) ~ age + clinstg + male + cmt + nodes + mediast
+ tt(age), data=data, tt=function(x,t,...)x*t, ties="efron")

#Schoenfeld test and plot
PHfit <- cox.zph(model1, transform="identity"); ggcoxzph(PHfit)

# Martingale and Deviance residuals
proc gplot data=smoothed;
   plot DepVar*age Pred*age / overlay;
   symbol1 value=dot interpol=none height=0.5;
   symbol2 value=none interpol=join line=1;
   title "";  run; quit;

#Deviance plot
plot(residuals_df$fitted, residuals_df$resdev,
     xlab = "Fitted Values",
     ylab = "Deviance Residuals",
     pch = 19, col = "black",cex=0.5)
abline(h = 0, col = "red", lty = 2)
```