

1. (a) The proportional hazards model is

$$h(t) = h_0(t)e^{\beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3}$$

where the covariates are coded as

$$z_1 = \begin{cases} 1 & \text{for NHL auto patients} \\ 0 & \text{otherwise} \end{cases}$$

$$z_2 = \begin{cases} 1 & \text{for HOD allo patients} \\ 0 & \text{otherwise} \end{cases}$$

$$z_3 = \begin{cases} 1 & \text{for HOD auto patients} \\ 0 & \text{otherwise} \end{cases}$$

Then,

$$\exp(\beta_1) = \frac{\text{hazard at time } t \text{ for a randomly selected NHL auto patient}}{\text{hazard at time } t \text{ for a randomly selected NHL allo patient}}$$

and the null hypothesis that the survivor functions are the same for the NHL allo and NHL auto patients is  $H_0 : \beta_1 = 0$ . Also,

$$\exp(\beta_2) = \frac{\text{hazard at time } t \text{ for a randomly selected HOD allo patient}}{\text{hazard at time } t \text{ for a randomly selected NHL allo patient}}$$

and the null hypothesis that the survivor functions are the same for the NHL allo and HOD allo patients is  $H_0 : \beta_2 = 0$ . Finally,

$$\exp(\beta_3) = \frac{\text{hazard at time } t \text{ for a randomly selected HOD auto patient}}{\text{hazard at time } t \text{ for a randomly selected NHL allo patient}}$$

and the null hypothesis that the survivor functions are the same for the NHL allo and HOD auto patients is  $H_0 : \beta_3 = 0$ .

- (b) This proportional hazards model is

$$h(t) = h_0(t)e^{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}$$

where the covariates are coded as

$$x_1 = \begin{cases} 1 & \text{for HOD patients} \\ 0 & \text{for NHL patients} \end{cases}$$

$$x_2 = \begin{cases} 1 & \text{for auto transplant patients} \\ 0 & \text{for allo transplant patients} \end{cases}$$

$$x_3 = x_1 x_2$$

Then, the null hypothesis of no interaction between type of lymphoma and type of transplant is  $H_0 : \beta_3 = 0$ . When  $\beta_3 = 0$  the type of lymphoma and type of transplant have additive effects on the logarithm of the hazard function (or multiplicative effects on the hazard function).

(c) Using the proportional hazards model from part (b), we have  $\beta_1 = 2$ ,  $\beta_2 = 1.5$ , and  $\beta_3 = -3$

2. (a) (i) The partial likelihood estimate of  $\beta_1$  is -0.371 with standard error 0.304.

(ii) The partial likelihood estimate of  $\exp(\beta_1)$  is 0.69. This is an estimate of the the hazard ratio of patients treated with thiotepa versus patients treated with placebo. It indicates that at any point in time treatment with thiotepa reduces the conditional probability of tumor recurrence in bladder cancer patients by about 31%.

(iii) An approximate 95% confidence interval is (0.38, 1.25). The confidence interval suggests that the effect of treatment with thiotepa could possibly be as great as a 62% reduction in the hazard of tumor recurrence to as much as a 25% increase. Since the confidence interval contains 1.0, the estimated reduction of 31% could be an artifact of the random assignment of patients to treatments. R code and output are shown below.

```
set1<- read.table("c:/st565/data/bladder1.dat", header=F,
                 col.names=c("patient", "time", "status", "treat", "init", "size"))
set1$treat <- set1$treat-1
```

```
options(contrasts=c("contr.treatment", "contr.poly"))
```

```
bfit1 <- coxph(Surv(time, status) ~ treat, data = set1, x=T, y=T, robust=T,
              method="efron", singular.ok=T)
summary(bfit1)
```

Call:

```
coxph(formula = Surv(time, status) ~ treat, data = set1, method = "efron",
      singular.ok = T, robust = T, x = T, y = T)
```

n= 86

	coef	exp(coef)	se(coef)	robust se	z	p
treat	-0.371	0.69	0.303	0.304	-1.22	0.22

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	0.69	1.45	0.38	1.25

Rsquare= 0.018 (max possible= 0.987 )

Likelihood ratio test= 1.54 on 1 df, p=0.215

Wald test = 1.48 on 1 df, p=0.223

Score (logrank) test = 1.52 on 1 df, p=0.218, Robust = 1.53 p=0.216

(Note: the likelihood ratio and score tests assume independence of observations within a cluster, the Wald and robust score tests do not).

(b) (i) From output produced by the PHREG procedure in SAS, the partial likelihood estimates are as follows:

Variable	DF	Parameter	Standard	Chi-Square	Pr > ChiSq	Hazard Ratio
		Estimate	Error			
x	1	-0.52593	0.31583	2.7730	0.0959	0.591
init	1	0.23824	0.07588	9.8579	0.0017	1.269
size	1	0.06963	0.10156	0.4700	0.4930	1.072

(ii) There are three tests of  $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$  available on the output:

```

Likelihood ratio test= 9.92 on 3 df, p=0.0193
Wald test              = 10.8 on 3 df, p=0.0126
Score (logrank) test = 11.1 on 3 df, p=0.0111

```

Each of the tests rejects the null hypothesis at the .05 level of significance. This suggests that at least one of the coefficients is not zero.

The “likelihood ratio test” is computed as  $LR = -2\log\left(\frac{L(\hat{\beta}_{H_0})}{L(\hat{\beta}_{H_A})}\right)$  where  $L(\hat{\beta}_{H_0})$  is the maximum value of the partial likelihood under the null hypothesis (i.e. evaluated at  $\beta_1 = \beta_2 = \beta_3 = 0$ ) and  $L(\hat{\beta}_{H_A})$  is the maximum value of the partial likelihood under the alternative hypothesis (i.e. evaluated at  $\hat{\beta}_1 = -0.52593$ ,  $\hat{\beta}_2 = 0.23624$ ,  $\hat{\beta}_3 = 0.06963$ ).

The score test is computed as  $S^2 = [U(\hat{\beta}_{H_0})]^T I^{-1}(\hat{\beta}_{H_0}) [U(\hat{\beta}_{H_0})]$  where  $[U(\hat{\beta}_{H_0})]$  is the score function for the full (or alternative) model evaluated at the mle under the null hypothesis (i.e. evaluated at  $\beta_1 = \beta_2 = \beta_3 = 0$ ) and  $I(\hat{\beta}_{H_0})$  is the information matrix for the full model evaluated at at the mle under the null hypothesis (i.e. evaluated at  $\beta_1 = \beta_2 = \beta_3 = 0$ ). Note the covariance matrix for the large sample normal approximation to the distribution of  $U(\hat{\beta}_{H_A})$  is estimated as  $I(\hat{\beta}_{H_A})$ .

The Wald test is computed as  $X^2 = (\hat{\beta}_{H_A} - 0)^T I(\hat{\beta}_{H_A}) (\hat{\beta}_{H_A} - 0)$  where  $\hat{\beta}_{H_A}$  is the mle for the parameters under the full (or alternative) model and  $I(\hat{\beta}_{H_A})$  is the information matrix for the full model evaluated at at the mle under the full model (i.e. evaluated at  $\hat{\beta}_1 = -0.52593$ ,  $\hat{\beta}_2 = 0.23624$ ,  $\hat{\beta}_3 = 0.06963$ ). Note the covariance matrix for the large sample normal approximation to the distribution of  $\hat{\beta}_{H_A}$  is estimated as  $I^{-1}(\hat{\beta}_{H_A})$ .

For large sample sizes, the distribution of each of these test statistics is well approximated by a central chi-square distribution when the null hypothesis is true. The degrees of freedom is the difference in the dimension of the parameter spaces for the alternative and null hypotheses, or the number of restrictions placed on the parameters in the full model to obtain the null hypothesis. When the null hypothesis is false, these test statistics will have thicker right tails than the corresponding central chi-square distribution and large

values of the test statistic provides evidence against the null hypothesis. The tests can have different relative power for different types of departure from the null hypothesis. No one of these three tests is uniformly more powerful than the other. The score test often provides the largest test statistic, but the chi-square approximation to the score statistic tends to yield inflated type I error levels for small samples when the null hypothesis is true. Chi-square approximations to the finite sample distributions for the likelihood ratio and Wald tests are generally better at keeping the true type I error level close to the desired level. Consequently, the likelihood ratio and Wald tests are generally recommended over the score test. None of the three tests will yield reliable p-values if the sample size is too small.

(iii) After adjusting for the number of initial tumors and the diameter of the largest initial tumor, the effect of treatment with thiotepa is not significant at the .05 level. The p-value is 0.0959, so the effect is significant at the .10 level.

(iv)  $\exp(\hat{\beta}_1) = 0.591$  is the estimated hazard ratio of patients treated with thiotepa vs. patients treated with placebo when initial number of tumors and diameter of largest initial tumor are the same for both patients. This suggests that for any given set of values for initial number of tumors and diameter of largest initial tumor, treatment with thiotepa provides about a 40% reduction in the conditional probability of tumor recurrence at any point time point. Since this effect was not significant at the .05 level, it could be an artifact of the random assignment of bladder cancer patients to treatments. The confidence interval in the R output shown below, indicates that the effect is likely to be as great as a 62% reduction to a 10% increase in the conditional probability of tumor recurrence.

$\exp(\hat{\beta}_2) = 1.269$  is the estimated hazard ratio when the initial number of tumors is increased by one while treatment and diameter of the largest initial tumor are kept the same. Each additional initial tumor is associated with an increase in the conditional probability of tumor recurrence of about 27%. This effect is significant at the .05 level (p-value=.0014). The approximate 95% confidence interval in the R output shown below, indicates that the effect is likely to be as small as a 9.6% reduction to as much as a 47% increase in the conditional probability of tumor recurrence.

$\exp(\hat{\beta}_3) = 1.072$  is the estimated hazard ratio when the diameter of the largest initial tumor is increased by 1 cm while the treatment and the initial number of tumors stay the same. This suggests that for either treatment and any given value for the initial number of tumors, each additional increase of 1 cm in the diameter of the largest initial tumor is associated with about a 7% increase in the conditional probability of tumor recurrence at any point time point. Since this effect was not significant at the .05 level, it could be an artifact of the random assignment of bladder cancer patients to treatments. The confidence interval in the R output shown below, indicates that the effect is likely to be as much as a 10% reduction to a 28% increase in the conditional probability of tumor recurrence.

R code:

```
library(survival)
bfit2 <- coxph(Surv(time, status) ~ treat+init+size,
              data = set1, x=T, y=T, robust=T,
              method="efron", singular.ok=T)
summary(bfit2)
```

Call:

```
coxph(formula = Surv(time, status) ~ treat + init + size, data = set1, method
      = "efron", singular.ok = T, robust = T, x = T, y = T)
```

n= 86

	coef	exp(coef)	se(coef)	robust se	z	p
treat	-0.5259	0.591	0.3158	0.3152	-1.668	0.0950
init	0.2382	1.269	0.0759	0.0746	3.194	0.0014
size	0.0696	1.072	0.1016	0.0886	0.786	0.4300

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	0.591	1.692	0.319	1.10
init	1.269	0.788	1.096	1.47
size	1.072	0.933	0.901	1.28

Rsquare= 0.109 (max possible= 0.987 )

Likelihood ratio test= 9.92 on 3 df, p=0.0193

Wald test = 10.8 on 3 df, p=0.0126

Score (logrank) test = 11.1 on 3 df, p=0.0111, Robust = 9.96 p=0.0189

(Note: the likelihood ratio and score tests assume independence of observations within a cluster, the Wald and robust score tests do not).

(v) An approximate 95 % confidence interval for  $\beta_1$  is  $-0.526 \pm 1.95 * 0.3158 \Rightarrow (-1.145, 0.093)$ . Applying the exponential function to the endpoints of that intervals yields  $(e^{-1.145}, e^{0.093}) \Rightarrow (0.319, 1.10)$  as an approximate 95 % confidence interval for  $exp(\beta_1)$ .

- (c) Display your plot.
- (d) Estimated median times to first tumor recurrence are 16 and 32 months for the placebo and thiotepa patients, respectively, who have one initial tumor of size 2cm.
- (e) Including the interaction terms in the Cox model, the maximum partial likelihood estimates of the coefficients are:

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
----------	----	--------------------	----------------	------------	----------	--------------

x	1	0.23233	0.73149	0.1009	0.7508	1.262
init	1	0.28440	0.14799	3.6932	0.0546	1.329
size	1	0.16454	0.11735	1.9660	0.1609	1.179
xinit	1	-0.03115	0.17276	0.0325	0.8569	0.969
xsize	1	-0.35656	0.25086	2.0203	0.1552	0.700

The two interaction terms are not significant at  $\alpha = 0.05$ . Therefore, this model provides little evidence of a significant interaction.

- (f) Using the METHOD=SCORE option in the PHREG procedure in SAS to examine all possible subsets of the variables in this model, however, does reveal evidence of interaction. The best three variable model contains an interaction between treatment and the number of initial tumors.

2	10.7257	x init
2	10.1919	x xinit
3	13.1219	init size xsize
3	11.9113	size xsize xinit
4	13.2552	x init size xsize
4	13.1224	init size xsize xinit
5	13.3420	x init size xsize xinit

Use PHREG in SAS or the coxph function in R to investigate some of these models.

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
x	1	-0.51214	0.31299	2.6774	0.1018	0.599
init	1	0.23084	0.07541	9.3695	0.0022	1.260

In this model the coefficient for the treatment variable, x, represents a conditional effect, given the initial number of tumors, but averaging over the diameter of the largest tumor and any other factors not included in the model. This treatment effect is not significant at the .05 level. A larger number of initial tumors is associated with a larger hazard, or conditional probability, of tumor recurrence (p-value=.0022). For this model, AIC=367.237 and SBC=370.938.

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
init	1	0.26122	0.07839	11.1033	0.0009	1.299
size	1	0.14444	0.09998	2.0871	0.1485	1.155
xsize	1	-0.29818	0.14357	4.3135	0.0378	0.742

This model has a significant interaction between treatment and size of initial tumor (p-value=.0378). The effect of treatment with thiotepa on reducing the conditional probability of tumor recurrence increases as the size of the initial tumor increases. When the largest initial tumor is 1 cm in diameter, treatment with thiotepa reduces the risk of tumor recurrence by about 25%. When the largest initial tumor is 2 cm in diameter, treatment with thiotepa reduces the risk of tumor recurrence by about 45%. (Note that the hazard ratio is  $\exp(-.59636)=0.5508$ ). The coefficient on SIZE represents the effect of the size of the largest initial tumor on the hazard for patients given the placebo (when  $x=0$  and  $xsize=0$ ). This association is not significant (p-value=.1485). The exponential function applied to the sum of the coefficients for SIZE and XSIZE gives the estimate of the hazard ratio for a 1 cm increase in the diameter of the largest initial tumor for patients treated with thiotepa (As an additional exercise, you can compute this and obtain a confidence interval.) The lack of a treatment factor,  $x$ , in this model indicates that there is essentially no difference in treatment with thiotepa and the placebo when initial tumors are small. For this model, AIC=364.286 and SBC=369.837, which are small improvements over the previous model. These two models describe the data about equally well, although the descriptions are different. The second model indicates that there may be a treatment effect, and it is more pronounced for patients with larger initial tumors. This is certainly something for the researchers to consider.

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
x	1	0.14523	0.54820	0.0702	0.7911	1.156
init	1	0.26162	0.07824	11.1825	0.0008	1.299
size	1	0.15696	0.10973	2.0464	0.1526	1.170
xsize	1	-0.35175	0.25007	1.9785	0.1595	0.703

For this model, AIC=366.217 and SBC=374.617, which are not improvements over the previous model. In this model the coefficient on  $x$  represents a treatment effect when the size of the largest initial tumor is zero cm, that is, when the patient did not have bladder tumors and would not be in the study. Such uninformative extrapolation could be avoided by replacing the SIZE variable with SIZE-2), for example. Then the coefficient on  $x$  would represent a treatment effect for patients for whom the largest initial tumor was 2 cm in diameter. The confounding between the  $x$ , size, and xsize variables results in a model in which none of the partial regression coefficients are significant at the .05 level.

There is no **correct** model for this study. Any model is a simplification of the truth and will not be entirely accurate. Some models may be useful, however, if they provide useful approximations to the truth or reveal relationships of important practical value. Do not let a stepwise procedure choose your model. Model selection procedures should be used to alert you to interesting possibilities. There may be several models that provide equally good descriptions of the data. The data will provide no information for choosing between those possibilities (you will have exhausted the information in the data), but considering several models may lead to some interesting ideas.

