Test Answers

Chapter 1

True-False Questions:

1. T
2. T
3. T
4. F: Step function.
5. F: Ranges between 0 and 1.
6. T
7. T
8. T
9. T
10. F: Median survival time is longer for group 1 than for group 2.
11. F: Six weeks or greater.
12. F: The risk set at 7 weeks contains 15 persons.
13. F: Hazard ratio.
14. T
15. T
16. $h(t)$ gives the instantaneous potential per unit time for the event to occur given that the individual has survived up to time t;$k(t)$ is greater than or equal to 0; $h(t)$ has no upper bound.
17. Hazard functions
    - give insight about conditional failure rates;
    - help to identify specific model forms (e.g., exponential, Weibull);
    - are used to specify mathematical models for survival analysis.
18. Three goals of survival analysis are:
    - to estimate and interpret survivor and/or hazard functions;
    - to compare survivor and/or hazard functions;
    - to assess the relationship of explanatory variables to survival time.

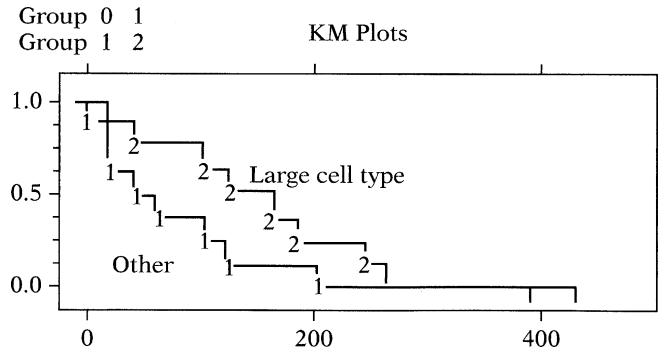
19.

	$t_{(j)}$	m_j	q_j	$R(t_{(j)})$
Group 1:	0	0	0	25 persons survive $\geq$ 0 years
	1.8	1	0	25 persons survive $\geq$ 1.8 years
	2.2	1	0	24 persons survive $\geq$ 2.2 years
	2.5	1	0	23 persons survive $\geq$ 2.5 years
	2.6	1	0	22 persons survive $\geq$ 2.6 years
	3.0	1	0	21 persons suivive $\geq$ 3.0 years
	3.5	1	0	20 persons survive $\geq$ 3.5 years
	3.8	1	0	19 persons survive $\geq$ 3.8 years
	5.3	1	0	18 persons survive $\geq$ 5.3 years
	5.4	1	0	17 persons survive $\geq$ 5.4 years
	5.7	1	0	16 persons survive $\geq$ 5.7 years
	6.6	1	0	15 persons survive $\geq$ 6.6 years
	8.2	1	0	14 persons survive $\geq$ 8.2 years
	8.7	1	0	13 persons survive $\geq$ 8.7 years
	9.2	2	0	12 persons survive $\geq$ 9.2 years
	9.8	1	0	10 persons survive $\geq$ 9.8 years
	10.0	1	0	9 persons survive $\geq$ 10.0 years
	10.2	1	0	8 persons survive $\geq$ 10.2 years
	10.7	1	0	7 persons survive $\geq$ 10.7 years
	11.0	1	0	6 persons survive $\geq$ 11.0 years
	11.1	1	0	5 persons survive $\geq$ 11.1 years
	11.7	1	3	4 persons survive $\geq$ 11.7 years

20. a. Group 1 has a better survival prognosis than group 2 because group 1 has a higher average survival time and a correspondingly lower average hazard rate than group 2.
    b. The average survival time and average hazard rates give overall descriptive statistics. The survivor curves allow one to make comparisons over time.

Chapter 2

1. a. KM plots and the log rank statistic for the cell type 1 variable in the vets.data dataset are shown below.

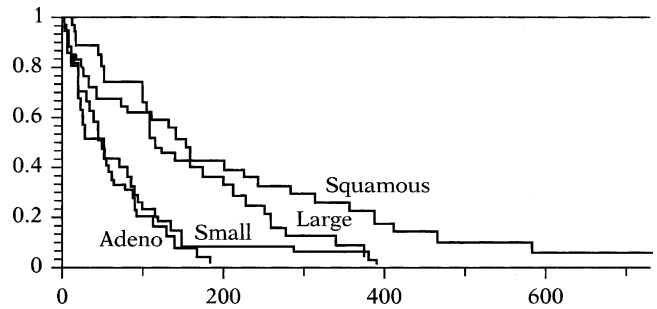


Group	Events observed	Events expected
1	102	93.45
2	26	34.55
Total	128	128.00

Log rank = chi2(1) = 3.02
p-value = Pr > chi2 = 0.0822

The KM curves indicate that persons with large cell type have a consistently better prognosis than persons with other cell types, although the two curves are essentially the same very early on and after 250 days. The log rank test is not significant at the .05 level, which gives somewhat equivocal findings.

b. KM plots and the log rank statistic for the four categories of cell type are shown below.



The KM curves suggest that persons with adeno or small cell types have a poorer survival prognosis than persons with large or squamous cell types. Moreover, there does not appear to be a meaningful difference between adeno or small cell types. Also, persons with squamous cell type seem to have, on the whole, a better prognosis than persons with large cell type.

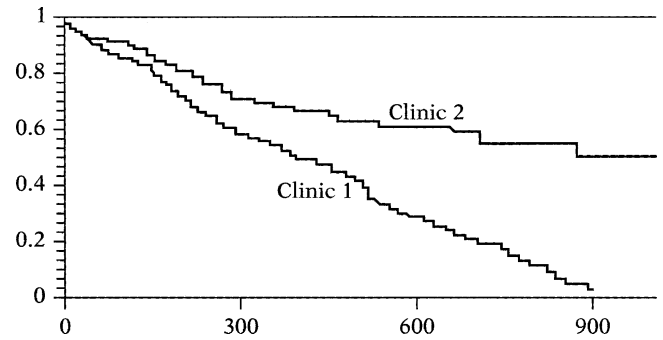
Computer results from Stata giving log rank statistics are now shown.

Group	Events observed	Events expected
1	26	34.55
2	26	15.69
3	45	30.10
4	31	47.65
Total	128	128.00

Log rank = chi2(3) = 25.40
P-value = Pr > chi2 = 0.0000

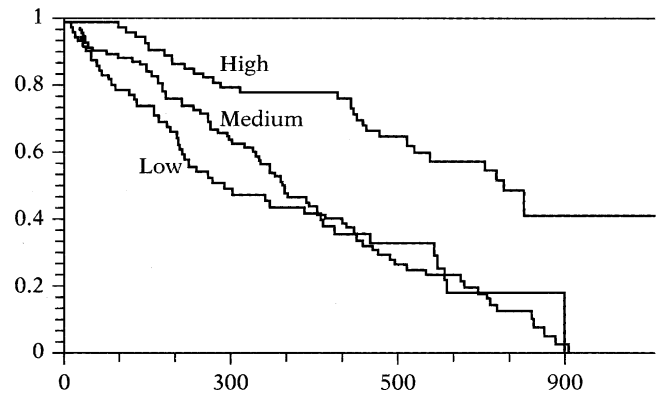
The log-rank test yields highly significant p-values, indicating that there is some overall difference between all four curves; that is, the null hypothesis that the four curves have a common survival curve is rejected,

2. a. KM plots for the two clinics are shown below. These plots indicate that patients in clinic 2 have a consistently better prognosis for remaining under treatment than do patients in clinic 1. Moreover, it appears that the difference between the two clinics is small before one year of follow-up but diverges after one year of follow up.



b. The log rank statistic (27.893) and Wilcoxon statistic (11.63) are both significant well below the .01 level, indicating that the survival curves for the two clinics are significantly different. The log rank statistic is nevertheless much larger than the Wilcoxon statistic, which makes sense because the log rank statistic emphasizes the later survival experience, where the two survival curves are far apart, whereas the Wilcoxon statistic emphasizes earlier survival experience, where the two survival curves are closer together.

c. If methadone dose is categorized into high (70+), medium (55–70) and low (<55), we obtain the KM curves shown below.



The KM curves indicate that persons with high doses have a consistently better survival prognosis (i.e.. maintenance) than persons with medium or low doses. The latter two groups are not very different from each other, although the medium dose group has a somewhat better prognosis up to the first 400 days of follow-up.
The log rank test statistic is shown below for the above categorization scheme.

Group	Events observed	Events expected
0	45	30.93
1	74	54.09
2	31	64.99
Total	150	150.00

Log rank = chi2(2) = 33.02
P-value = Pr>chi2 = 0.0000

The test statistic is highly significant, indicating that these three curves are not equivalent.

Chapter 3

1. a. $h(t,\mathbf{X}) = h_0(t)\exp[\beta_1 T1 + \beta_2 T_2 + \beta_3 PS + \beta_4 DC + \beta_5 BF + \beta_6(T1 \times PS) + \beta_7(T2 \times PS) + \beta_8(T1 \times DC) + \beta_9(T2 \times DC) + \beta_{10}(T1 \times BF) + \beta_{11}(T2 \times BF)]$

   b. **Intervention A: X*** = (1, 0, PS, DC, BF, PS, 0, DC, 0, BE 0)
   Intervention C: X = (− 1, −1, PS, DC, BF, –PS, –PS, –DC, –DC, –BF, –BF)

$$\mathrm{HR} = \frac{h(t,\mathbf{X}^*)}{h(t,\mathbf{X})} = \exp[2\,\beta_1 + \beta_2 + 2\,\beta_6\,\mathrm{PS} + \beta_7\,\mathrm{PS} + 2\,\beta_8\,\mathrm{DC} + \beta_9\,\mathrm{DC} + 2\,\beta_{10}\,\mathrm{BF} + \beta_{11}\,\mathrm{BF}]$$

c. H_0: $\beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = 0$ in the full model.
Likelihood ratio test statistic: $-2 \ln \hat{L}_R - (-2\mathrm{In}\hat{L}_F)$, which is approximately χ^2_6 under H_0, where R denotes the reduced model (containing no product terms) under H_0, and F denotes the full model (given in Part la above)

d. The two models being compared are:
Full model (F): $h(t,\mathbf{X}) = h_0(t)\exp[\beta_1 \mathrm{Tl} + \beta_2 \mathrm{T2} + \beta_3 \mathrm{PS} + \beta_4 \mathrm{DC} + \beta_5 \mathrm{BF}]$

Reduced model (R): $h(t,\mathbf{X}) = h_0(t)\exp[\beta_3 \text{PS} + \beta_4 \text{DC} + \beta_5 \text{BF}]$

H_0: $\beta_1 = \beta_2 = 0$ in the full model

Likelihood ratio test statistic: $-2\ln\hat{L}_R - (-2\ln\hat{L}_F)$, which is approximately χ^2_2 under H_0.

e.

Intervention A:

$$\hat{S}(t,\mathbf{X}) = [\hat{S}_0(t)]^{\exp[\hat{\beta}_1 + (\overline{\text{PS}})\hat{\beta}_3 + (\overline{\text{DC}})\hat{\beta}_4 + (\overline{\text{BF}})\hat{\beta}_5]}$$

Intervention B:

$$\hat{S}(t,\mathbf{X}) = [\hat{S}_0(t)]^{\exp[\hat{\beta}_2 + (\overline{\text{PS}})\hat{\beta}_3 + (\overline{\text{DC}})\hat{\beta}_4 + (\overline{\text{BF}})\hat{\beta}_5]}$$

Intervention C:

$$\hat{S}(t,\mathbf{X}) = [\hat{S}_0(t)]^{\exp[-\hat{\beta}_1 - \hat{\beta}_2 + (\overline{\text{PS}})\hat{\beta}_3 + (\overline{\text{DC}})\hat{\beta}_4 + (\overline{\text{BF}})\hat{\beta}_5]}$$

2. a. $h(t,\mathbf{X}) = h_0(t)\exp[\beta_1 \text{ CHR} + \beta_2 \text{ AGE} + \beta_3(\text{CHR} \times \text{AGE})]$

   b. H_0: $\beta_3 = 0$

   LR statistic $= 264.90 - 264.70 = 0.21$; χ^2 with 1 d.f. under H_0; not *significant*.

   Wald statistic gives a chi-square value of .01, also not significant. Conclusions about interaction: the model should not contain an interaction term.

   c. When AGE is controlled (using the gold standard model 2), the hazard ratio for the effect of CHR is exp(.8051) = 2.24, whereas when AGE is not controlled, the hazard ratio for the effect of CHR (using Model 1) is exp(.8595) = 2.36. Thus, the hazard ratios are not appreciably different, so AGE is not a confounder.

   Regarding precision, the 95% confidence interval for the effect of CHR in the gold standard model (Model 2)is given by exp[.8051 ± 1.96(.3252)] = (1.183, 4.231) whereas the corresponding 95% confidence interval in the model without AGE (Model 1) is given by exp[.8595 ± 1.96(.3116)] = (1.282, 4.350). Both confidence intervals have about the same width, with the latter interval being slightly wider. Thus, controlling for AGE has little effect on the final point and interval estimates of interest.

   d. If the hazard functions cross for the two levels of the CHR variable, this would mean that none of the models provided is appropriate, because each model assumes that the proportional hazards assumption is met for each predictor in the model. If hazard functions cross for CHR, however, the proportional hazards assumption cannot be satisfied for this variable.

e. for $CHR = 1 : \hat{S}(t, X) = \left[\hat{S}_0(t)\right]^{\exp\left[0.8051+0.0856(\overline{\text{AGE}})\right]}$
For $CHR = 0 : \hat{S}(t, X) = \left[\hat{S}_0(t)\right]^{\exp\left[0.0856(\overline{\text{AGE}})\right]}$

f. Using Model 1, which is the best model, there is evidence of a moderate effect of CHR on survival time, because the hazard ratio is about 2.4 with a 95% confidence interval between 1.3 and 4.4, and the Wald text for significance of this variable is significant below the .01 level.

3. a. Full model (F = Model 1): $h(t,\mathbf{X})=h_0(t)\exp[\beta_1 Rx + \beta_2\text{Sex}+\beta_3\log \text{WBC}+\beta_4(Rx \times \text{Sex}) + \beta_5(Rx \times \log \text{WBC})]$
Reduced model (R = model 4):
$h(t,\mathbf{X}) = h_0(t)\exp[\beta_1 Rx + \beta_2\text{Sex}+ \beta_3\text{logWBC}]$
H_0: $\beta_4-\beta_5=0$
LR statistic = 144.218 − 139.030 = 5.19; χ^2 with 2 d.f. under H_0; not significant at 0.05, though signilicant at 0.10. The chunk test indicates some (though mild) evidence of interaction.

b. Using either a Wald test (p-value = .776) or a LR test, the product term $Rx \times$ log WBC is clearly not significant, and thus should be dropped from Model 1. Thus, Model 2 is preferred to Model 1.

c. Using Model 2, the hazard ratio for the effect of Rx is given by HR − $(h(t,\mathbf{X}^*))/(h(t,\mathbf{X})) = \exp[0.405 + 2.013\ \text{Sex}]$

d. Males (Sex = 0) : $\widehat{HR} = \exp[0.405] = 1.499$
Females (Sex = 1) : $\widehat{HR} = \exp[0.405 + 2.013(1)] = 11.223$

e. Model 2 is preferred to Model 3 if one decides that the coefficients for the variables Rx and $Rx \times$ Sex are meaningfully different for the two models. It appears that such corresponding coefficients (0.405 vs. 0.587 and 2.013 vs. 1.906) are different. The estimated hazard ratios tor Model 3 are 1.799 (males) and 12.098 (females), which are different, but not very different from the estimates computed in Part 3d for Model 2. If it is decided that there is a meaningful difference here, then we would conclude that log WBC is a confounder; otherwise log WBC is not a confounder. Note that the log WBC variable is significant in Model 2 ($P = .000$), but this addresses precision and not confounding. When in doubt, as in this case, the safest thing to do (for validity reasons) is to control for log WBC.

f. Model 2 appears to be best, because there is significant interaction of $Rx \times$ Sex ($P = .023$) and because log WBC is a likely confounder (from Part e).

Chapter 4

1. The $P(PH)$ values in the printout provide GOF statistics for each variable adjusted for the other variables in the model These $P(PH)$ values indicate that the clinic variable does not satisfy the PH assumption ($P << .01$), whereas the prison and dose variables satisfy the PH assumption ($P > .10$).
2. The log–log plots shown are parallel. However, the reason why they are parallel is because the clinic variable has been included in the model, and log–log curves for any variable in a PH model must always be parallel. If, instead, the clinic variable had been stratified (i.e., not included in the model), then the log–log plots comparing the two clinics adjusted for the prison and dose variables might not be parallel.
3. The log–log plots obtained when the clinic variable is stratified (i.e., using a stratified Cox PH model) are not parallel. They intersect early on in follow-up and diverge from each other later in follow-up. These plots therefore indicate that the PH assumption is not satisfied for the clinic variable.
4. Both graphs of log–log plots for the prison variable show curves that intersect and then diverge from each other and then intersect again. Thus, the plots on each graph appear to be quite nonparallel, indicating that the PH assumption is not satisfied for the prison variable. Note, however, that on each graph, the plots are quite close to each other, so that one might conclude that, allowing for random variation, the two plots are essentially coincident; with this latter point of view, one would conclude that the PH assumption is satisfied for the prison variable.
5. The conclusion of nonparallel log–log plots in Question 4 gives a different result about the PH assumption for the prison variable than determined from the GOF tests provided in Question 1. That is, the log–log plots suggest that the prison variable does not satisfy the PH assumption, whereas the GOF test suggests that the prison variable satisfies the assumption. Note, however, if the point of view is taken that the two plots are close enough to suggest coincidence, the graphical conclusion would be the same as the GOF conclusion. Although the final decision is somewhat equivocal here, we prefer to conclude that the PH assumption is satisfied for the prison variable because this is strongly indicated from the GOF test and questionably counterindicated by the log–log curves.

6. Because maximum methadone dose is a continuous variable, we must categorize this variable into two or more groups in order to graphically evaluate whether it satisfies the PH assumption. Assume that we have categorized this variable into two groups, say, low versus high. Then, **observed** survival plots can be obtained as KM curves for low and high groups separately To obtain **expected** plots, we can fit a Cox model containing the dose variable and then substitute suitably chosen values for dose into the formula for the estimated survival curve. Typically, the values substituted would be either the mean or median (maximum) dose in each group.
   After obtaining observed and expected plots for low and high dose groups, we would conclude that the PH assumption is satisfied if corresponding observed and expected plots art; not widely discrepant from each other. If a noticeable discrepancy is found for at least one pair of observed versus expected plots, we conclude that the PH assumption is not satisfied.
7. $h(t,\mathbf{X}) = h_0(t)\exp[\beta_1 \text{ clinic} + \beta_2 \text{ prison} + \beta_3 \text{ dose} + \delta_1 (\text{clinic} \times g(t)) + \delta_2 (\text{prison} \times g(t)) + \delta_3 (\text{dose} \times g(t))]$
   where $g(t)$ is some function of time. The null hypothesis is given by H_0: $\delta_1 = \delta_2 = \delta_3 = 0$. The test statistic is a likelihood ratio statistic of the form $LR = -2\ln L_R - (-2\text{In}L_F)$ where R denotes the reduced (PH) model obtained when all δs are 0, and F denotes the full model given above. Under H_0, the LR statistic is approximately chi-square with 3 d.f.
8. Drawbacks of the extended Cox model approach:
   - Not always clear how to specify $g(t)$; different choices may give different conclusions;
   - Different modeling strategies to choose from, for example, might consider $g(t)$ to be a polynomial in t and do a backward elimination to eliminate nonsignificant higher-order terms; alternatively, might consider $g(t)$ to be linear in t without evaluating higher-order terms.
     Different strategies may yield different conclusions.

9. $h(t,\mathbf{X}) = h_0(t)\exp[\beta_1 \text{ clinic} + \beta_2\text{prison} + \beta_3 \text{ dose} + \delta_1(\text{clinic} \times g(t))]$ where $g(t)$ is some function of time. The null hypothesis is given by H_0: $\delta_1 = 0$, and the test statistic is either a Wald statistic or a likelihood ratio statistic. The LR statistic would be of the form $LR = -2 \text{ In } L_R - (-2\text{In } L_F)$, where R denotes the reduced (PH) model obtained when $\delta_1 = 0$, and F denotes the full model given above. Either statistic is approximately chi-square with 1 d.f. under the null hypothesis.
10. $t > 365$ days: $HR = \exp[\beta_1 + \delta_1]$
    $t \leq 365$ days: $HR = \exp[\beta_1]$
    If δ_1 is not equal to zero, then the model does not satisfy the PH assumption for the clinic variable. Thus, a test of H_0: $\delta_1 = 0$ evaluates the PH assumption; a significant result would indicate that the PH assumption is violated. Note that if δ_1 is not equal to zero, then the model assumes that the hazard ratio is not constant over time by giving a different hazard ratio value depending on whether t is greater than 365 days or t is less than or equal to 365 days.

Chapter 5

1. By fitting a stratified Cox (SC) model that stratifies on clinic, we can compare adjusted survival curves for each clinic, adjusted for the prison and dose variables. This will allow us to visually describe the extent of clinic differences on survival over time. However, a drawback to stratifying on clinic is that it will not be possible to obtain an estimate of the hazard ratio for the effect of clinic, because clinic will not be included in the model.
2. The adjusted survival surves indicate that clinic 2 has a better survival prognosis than clinic 1 consistently over time. Moreover, it seems that the difference between the effects of clinic 2 and clinic 1 increases over lime.
3. $h_g(t, \mathbf{X}) = h_{0_g}(t)\exp[\beta_1 \text{ prison} + \beta_2 \text{ dose}], \; g = 1, 2$
   This is a no-interaction model because the regression coefficients for prison and dose are the same for each stratum.
4. Effect of prison, adjusted for clinic and dose: $\widehat{HR} = 1.475$, 95% CI: (1.059, 2.054). It appears that having a prison record gives a 1.475 increased hazard for failure than not having a prison record. The p-value is 0.021, which is significant at the 0.05 level.
5. Version 1: $h_g(t,\mathbf{X}) = h_{0_g}(t)\exp[\beta_{1g} \text{ prison} + \beta_{2g} \text{ dose}], \; g = 1, 2$
   Version 2: $h_g(t,\mathbf{X}) = h_{0_g}(t)\exp[\beta_1 \text{ prison} + \beta_2 \text{ dose} + \beta_3 (\text{clinic} \times \text{prison}) + \beta_4 (\text{clinic} \times \text{dose})], \; g = 1, 2$

6. $g = 1$ (clinic 1):
   $h_1(t,\mathbf{X}) = h_{01}(t)\exp[(0.502)\text{prison} + (-0.036)\text{ dose}]$
   $g = 2$ (clinic 2):
   $h_2(t,\mathbf{X}) = h_{02}(t)\exp[(-0.083)\text{prison} + (-0.037)\text{dose}]$
7. The adjusted survival curves stratified by clinic are virtually identical for the no-interaction and interaction models. Consequently, both graphs (no-interaction versus interaction) indicate the same conclusion that clinic 2 has consistently larger survival (i.e., retention) probabilities than clinic 1 as time increases.
8. H_0: $\beta_3 = \beta_4 = 0$ in the version 2 model (i.e., the no-interaction assumption is satisfied). $LR = -2\text{In } L_R - (-2 \text{ In } L_F)$ where R denotes the reduced (no-interaction) model and F denotes the full (interaction) model. Under the null hypothesis, LR is approximately a chi square with 2 degrees of freedom.
   Computed $LR = 1195.428 - 1193.558 = 1.87$; p-value = 0.395; thus, the null hypothesis is not rejected and we conclude that the no interaction model is preferable to the interaction model.

Chapter 6

1. For the chemo data, the –log-log KM curves intersect at around 600 days; thus the curves are not parallel, and this suggests that the treatment variable does not satisfy the PH assumption.
2. The P (PH) value for the Γx variable is 0, indicating that the PH assumption is not satisfied for the treatment variable based on this goodness-of-fit test.
3. $h(t,\mathbf{X}) = h_0(t)\exp[\beta_1(T\,x)g_1(t) + \beta_2(T\,x)g_2(t) + \beta_3(T\,x)g_3(t)]$
   where
   $$g_1(t) = \begin{cases} 1 & \text{if } 0 \le t < 250 \text{ days} \\ 0 & \text{if otherwise} \end{cases}$$
   $$g_2(t) = \begin{cases} 1 & \text{if } 25\,0 \le t < 500 \text{ days} \\ 0 & \text{if otherwise} \end{cases}$$
   $$g_3(t) = \begin{cases} 1 & \text{if } t \ge 500 \text{ days} \\ 0 & \text{if otherwise} \end{cases}$$
4. Based on the printout the hazard ratio estimates and corresponding p-values and 95% confidence intervals are given as follows for each time interval:

	Haz. Ratio	p> |z|	[95% Conf. Interval]	
$0 \leq t < 250$ days:	0.221	0.001	0.089	0.545
$250 \leq t < 500$ days:	1.629	0.278	0.675	3.934
$t \geq 500$ days:	1.441	0.411	0.604	3.440

The results show a significant effect of treatment below 250 days and a nonsignificant effect of treatment in each of the two intervals after 250 days. Because the coding for treatment was 1 = chemotherapy plus radiation versus 2 = chemotherapy alone, the results indicate that the hazard for chemotherapy plus radiation is 1/0.221 = 4.52 times the hazard for chemotherapy alone. The hazard ratio inverts to a value less than 1 (in favor of chemotherapy plus radiation after 250 days), but this result is nonsignificant. Note that for the significant effect of 1/0.221 = 4.52 below 250 days, the 95% confidence interval ranges between 1/0.545= 1.83 and 1/0.089= 11.24 when inverted, which is a very wide interval.

5. Model with two Heaviside functions:
$h(t,\mathbf{X}) = h_0(t)\exp[\beta_1(Tx)g_1(t) + \beta_2(Tx)g_2(t)]$
where

$$g_1(t) = \begin{cases} 1 & \text{if } 0 \leq t < 250 \text{ days} \\ 0 & \text{if otherwise} \end{cases}$$

$$g_2(t) = \begin{cases} 1 & \text{if } t \geq 250 \text{ days} \\ 0 & \text{if otherwise} \end{cases}$$

Model with one Heaviside function:
$h(t,\mathbf{X}) = h_0(t)\exp[\beta_1(Tx) + \beta_2(Tx)g_1(t)]$
where $g_1(t)$ is defined above.

6. The results for two time inteivals give hazard ratios that are on the opposite side of the null value (i.e., 1). Below 250 days, the use of chemotherapy plus radiation is, as in the previous analysis, 4.52 times the hazard when chemother apy is used alone. This result is significant and the same confidence interval is obtained as before. Above 250 days, the use of chemotherapy alone has 1.532 times the hazard of chemotherapy plus radiation, but this result is nonsignificant.

Chapter 7

1. F: They are multiplicative models, although additive on the log scale.
2. T
3. T
4. F: If the AFT assumption holds in a log-logistic model, the proportional odds assumption holds.
5. F: An acceleration factor greater than one suggests the exposure is beneficial to survival.
6. T
7. T
8. T
9. F: ln(T) follows an extreme value minimum distribution.
10. F: The subject is right-censored.
11. $$\gamma = \frac{\exp[\alpha_0 + \alpha_1(2) + \alpha_2 PRISON + \alpha_3 DOSE + \alpha_4 PRISDOSE]}{\exp[\alpha_0 + \alpha_1(1) + \alpha_2 PRISON + \alpha_3 DOSE + \alpha_4 PRISDOSE]}$$
$$= \exp(\alpha_1)$$
$$\hat{\gamma} = \exp(0.698) = 2.01$$
$$95\%\ \text{CI} = \exp[0.698 \pm 1.96(0.158)] = (1.47, 2.74)$$

    The point estimate for the acceleration factor (2.01) suggests that the survival time (time off heroin) is double for those enrolled in CLINIC = 2 compared to CLINIC = 1. The 95% confidence interval does not include the null value of 1.0 indicating a statistically significant preventive effect for CLINIC = 2 compared to CLINIC = 1.
12. $$\text{HR} = \frac{\exp[\beta_0 + \beta_1(2) + \beta_2 PRISON + \beta_3 DOSE + \beta_4 PRISDOSE]}{\exp[\beta_0 + \beta_1(1) + \beta_2 PRISON + \beta_3 DOSE + \beta_4 PRISDOSE]}$$
$$= \exp(\beta_1)$$
$$\hat{HR} = \exp(-0.957) = 0.38$$
$$95\%\ \text{CI} = \exp[-0.957 \pm 1.96(0.213)] = (0.25, 0.58)$$

    The point estimate of 0.38 suggests the hazard of going back on heroin is reduced by a factor of 0.38 for those enrolled in CLINIC = 2 compared to CLINIC = 1. Or from the other perspective: the estimated hazard is elevated for those in CLINIC = 1 by a factor of exp (+0 957) = 2.60.
13. $\beta_1 = -\alpha_1 p$ for CLINIC, so $\hat{\beta}_1 = -(0.698\text{X}1.370467) = -0.957$, which matches the output for the PH form of the model.

14. The product term PRISDOSE is included in the model as a potential confounder of the effect of CLINIC on survival. It is not an effect modifier because under this model the hazard ratio or acceleration factor for CLINIC does not depend on the value of PRISDOSE. The PRISDOSE term would cancel in the estimation of the hazard ratio or acceleration factor (see Questions 11 and 12). On the other hand, a product term involving CLINIC would be a potential effect modifier.
15. Using the AFT form of the model:

$$\frac{1}{\lambda^{1/p}} = \exp[\ \alpha_0 + \ \alpha_1 \text{CLINIC} + \ \alpha_2 \text{PRISON} + \ \alpha_3 \text{DOSE} + \alpha_4 \text{PRISDOSE}]$$

Median survival time for CLINIC = 2, PRISON = 1, DOSE = 50, PRISDOSE = 100:

$$\text{t} = [-\ln\ \text{S(t)}]^{1/p} \times \frac{1}{\lambda^{1/p}} = [-\ln(0.5)]^{1/p} \times \exp[\beta_0 + 2\beta_1 + \beta_2 + 50\beta_3 + 100\beta_4]$$

$\hat{t}$ (median) = 403.66 days (obtained by substituting parameter estimates from output).

16. Using the same approach as the previous question: Median survival time for CLINIC = 1, PRISON = 1, DOSE = 50, PRISDOSE = 100:
$\text{t} = [-\ln(0.5)]^{1/p} \times \exp[\beta_0 + \text{l}\beta_1 + \beta_2 + 50\beta_3 + 100\delta_4]$
$\hat{t}$ (median) = 200.85 days.
17. The ratio of the median survival times is 403.66/200.85 = 2.01. This is the estimated acceleration factor for CLINIC = 2 vs. CLINIC = 1 calculated in Question 11. Note that if we used any survival probability (i.e., any quantile of survival time), not just S(ι) -= 0.5 (the median), we would have obtained the same ratio.
18. The addition of the frailty component did not change any of the other parameter estimates nor did it change the log likelihood of −260.74854.
19. If the variance of the frailty is zero (theta = 0), then the frailty has no effect on the model. A variance of zero means the frailty (α) is constant at 1. Frailty is defined as a multiplicative random effect on the hazard $h(t|\alpha) = \alpha h(t)$. If $\alpha = 1$ then $h(t|\alpha) = h(t)$, and there is no frailty.

Chapter 8

1. a. **Survival time (say, in weeks) to the first event (stratum 1)**:

$t_{(f)}$	n_f	m_f	q_f	$R(t_{(f)})$
0	2	0	0	{B,L}
12	2	1	0	{B,L}
20	1	1	0	{L}

b. For each approach, the observation for the first event is identical.

c. **Survival rime (say, in weeks) from the first to the second event (stratum 2) using the Stratified CP approach**:

$t_{(f)}$	n_f	m_f	q_f	$R(t_{(f)})$
0	0	0	0	–
16	1	1	0	{B}
23	1	1	0	{L}

d. **Survival time (say, in weeks) from the first to the second event (stratum 2) using the Gap Time approach**:

$t_{(f)}$	n_f	m_f	q_f	$R_{(f)}$
0	2	0	0	{B,L}
3	2	1	0	{B,L}
4	1	1	0	{B}

e. **Survival time (say, in weeks) from the first to the second event using the Marginal approach**:

$t_{(f)}$	n_f	m_f	q_f	$R(t_{(f)})$
0	2	0	0	{B,L}
16	2	1	0	{B.L}
23	1	1	0	{L}

f. Correct choice is iii.
Bonnie is at risk for a second event between times 12 to 16.
Lonnie is at risk for a second event between times 20 to 23.
Neither is in the risk set for the other's second event.

g. Correct choice is ii.
Bonnie is at risk for a second event between times 0 to 4.
Lonnie is at risk for a second event between times 0 to 3.
Bonnie is in the risk set when Lonnie gets her second event.

h. Correct choice is i.
Bonnie is at risk for a second event between times 0 to 16.
Lonnie is at risk for a second event between times 0 to 23.
Lonnie is in the risk set when Bonnie gets her second event.

2. a. Cox PH Model for **CP** approach to Defibrillator Study:

$$h(t,\mathbf{X}) = h_0(t)\mathbf{exp}[\beta\ \mathbf{tx} + \gamma\ \mathbf{smoking}]$$

where **tx** = 1 if treatment A, 0 if treatment B. **smoking status** = 1 if ever smoked, 0 if never smoked.

b. Using the **CP** approach, there is no significant effect of treatment status adjusted for smoking. The estimated hazard ratio for the effect of treatment is 1.09, the corresponding P-value is 0.42 and a 95% CI for the hazard ratio is (0.88, 1.33).

c. No-interaction SC model for **Marginal** approach:

$$h_g\{t,\mathbf{X}) = h_{0g}(t)\mathbf{exp}[\beta\ \mathbf{tx} + \gamma\ \mathbf{smoking}],\ g = 1, 2,3$$

Interaction SC model for **Marginal** approach:

$$h_g\{t,\mathbf{X}) = h_{0g}(t)\mathbf{exp}[\beta_g\ \mathbf{tx} + \gamma_g\ \mathbf{smoking}],\ g = 1, 2, 3$$

d. **LR**$= -2\ln L_R - (-21n\ L_F)$ is approximately χ^2 with 4 df under
$\mathbf{H_0}$:no-interaction SC model is appropriate, where R denotes the reduced (no interaction SC) model and **F** denotes the full (interaction SC) model

e. The use of a no-interaction model does not allow you to obtain stratum-specific HR estimates, even though you are assuming that strata are important.

f. The **CP** approach makes sense for these data because recurrent defibrillator (shock) events on the same subject are the same kind of event no matter when it occurred.

g. You might use the **Marginal** approach if you determined that different recurrent events on the same subject were different because they were of different order.

h. The number in the risk set (n_f) remains unchanged through day 68 because every subject who failed by this time was still at risk for a later event.

i. Subjects 3,6, 10,26, and 31 all fail for the third time at day 98 and are not followed afterwards.

j. Subjects 9, 15, and 28 fail for the second time at 79 days, whereas subject #16 is censored at 79 days.

k. Subjects 4, 14, 15, 24, and 29 were censored between days 111 and 112.

l. Subject #5 gets his first event at 45 days and his second event at 68 days, after which he drops out of the study. This subject is the first of the 36 subjects to drop out of the study, so the number in the risk set changes from 36 to 35 after 68 days.

m. None of the above.

n. The product limit formula is not applicable to the **CP** data; in particular, $P(T > t|T \geq t)$ does not equal "# failing in time interval /# in the risk set at start of interval."

o. Use the information provided in Table T.2 to complete the data layouts for plotting the following survival curves.

i. $S_1(t) = Pr(T_1 > t)$ where T_1 = time to first event from study entry

$t_{(f)}$	n_f	m_f	q_f	$S(t_p) = S(t_{f-1}) \times Pr(T_1 > t \mid T_1 \geq t)$
0	36	0	0	1.00
33	36	2	0	0.94
34	34	3	0	0.86
36	31	3	0	0.78
37	28	2	0	0.72
38	26	4	0	0.61
39	22	5	0	0.47
40	17	1	0	0.44
41	16	1	0	0.42
43	15	1	0	0.39
44	14	1	0	0.36
45	13	2	0	0.31
46	11	2	0	0.25
48	9	1	0	0.22
49	8	1	0	0.19
51	**7**	**2**	**0**	**0.19 × 5/7 = 0.14**
57	**5**	**2**	**0**	**0.14 × 3/5 = 0.08**
58	**3**	**2**	**0**	**0.08 × 1/3 = 0.03**
61	**1**	**1**	**0**	**0.03 × 0/1 = 0.00**

ii. **Gap Time** $S_{2c}(t) = Pr(T_{2c} > t)$ where T_{2c} = time to second event from first event.

$t_{(f)}$	n_f	m_f	q_f	$S_2(t_{(f)}) = S_2(t_{(f-1)}) \times Pr(T_2 > t \mid T_2 \geq t)$
0	36	0	0	1.00
5	36	1	0	0.97
9	35	1	0	0.94
18	34	2	0	0.89
20	32	1	0	0.86
21	31	2	1	0.81
23	28	1	0	0.78
24	27	1	0	0.75

(*Continued on next page*)

(*Continued*)

$t_{(f)}$	n_f	m_f	q_f	$S_2(t_{(f)}) = S_2(t_{(f-1)}) \times \Pr(T_2 > t \mid T_2 \geq t)$
25	26	1	0	0.72
26	25	2	0	0.66
27	23	2	0	0.60
28	21	1	0	0.58
29	20	1	0	0.55
30	19	1	0	0.52
31	18	3	0	0.43
32	15	1	0	0.40
33	14	5	0	0 26
35	9	1	0	0.23
39	8	2	0	0.17
40	**6**	**2**	**0**	**0.17 × 4/6 = 0.12**
41	**4**	**1**	**0**	**0.12 × 3/4 = 0.09**
42	**3**	**1**	**0**	**0.09 × 2/3 = 0.06**
46	**2**	**1**	**0**	**0.06 × 1/2 = 0.03**
47	**1**	**1**	**0**	**0.03 × 0/1 = 0.00**

iii. **Marginal** $S_{2m}(t) = \Pr(T_{2m} > t)$ where T_{2m} = time to second event from study entry.

$t_{(f)}$	n_f	m_f	q_f	$S(t_{(f)}) = S_2(t_{(f-1)}) \times \Pr(T_2 > t \mid T_2 \geq t)$
0	36	0	0	1.00
63	36	2	0	0.94
64	34	3	0	0.86
65	31	2	0	0.81
66	29	3	0	0.72
67	26	4	0	0.61
68	22	2	0	0.56
69	20	1	0	0.53
70	19	1	0	0.50
71	18	1	0	0.47
72	17	2	0	0.42
73	15	1	0	0.39
74	14	1	0	0.36
76	13	1	0	0.33
77	12	1	0	0.31
78	11	2	0	0.25
79	**9**	**3**	**1**	**0.25 × 6/9 = 0.17**
80	**5**	**2**	**0**	**0.17 × 3/5 = 0.10**
81	**3**	**2**	**0**	**0.10 × 1/3 = 0.03**
97	**1**	**1**	**0**	**0.03 × 0/1 = 0.00**

p. The survival curves corresponding to the above data layouts will differ because they are describing different survival functions. In particular, the composition of the risk set differs in all three data layouts and the ordered survival times being plotted are different as well.

Chapter 9

1. Cause-specific no interaction model for local recurrence of bladder cancer (event = 1):

$$h_1(t,\mathbf{X}) = h_{01}(t)\exp[\beta_{11}tx + \beta_{21}num + \beta_{31}size]$$

2. Censored subjects have bladder metastasis (**event** = 2) or other metastasis (**event** = 3).
3. Cause-specific no-interaction model for bladder metastasis (**event** =2):

$$h_2(t,\mathbf{X}) = h_{02}(t)\exp[\beta_{12}tx + \beta_{22} + \beta_{32}size]$$

where censored subjects have local recurrence of bladder cancer (**event** = 1) or other metastasis (**event** = 3).

4. A sensitivity analysis would consider worst-case violations of the independence assumption. For example, subjects censored from failing from events = 2 or 3 might be treated in the analysis as either all being event-free (i.e., change event status to 0 and time to 53) or all experiencing the event of interest (i.e., change event status to 1 and leave time as is).
5. a. Verify the CIC_1 calculation provided at failure time t_f=8 for persons in the treatment group (**tx** = 1):

$$\begin{aligned}\hat{\mathbf{h}}_1(8) &= 1/23 = 0.0435\\ \hat{\mathbf{S}}(4) &= \hat{\mathbf{S}}(3)\mathbf{Pr}(T > 4|T \geq 4) = 0.9630(1 - 2/26)\\ &= 0.9630(0.9231) = 0.8889\\ \hat{\mathbf{i}}_1(8) &= \hat{\mathbf{h}}_1(8)\hat{\mathbf{S}}(4) = 0.0435(.8889) = 0.0387\\ \mathbf{CIC}_1(8) &= \mathbf{CIC}_1(4) + 0.0387 = 0 + 0.0387 = 0.0387\end{aligned}$$

   b. Verify the $\mathbf{CIC}_1$ calculation provided at failure time t_f= 25 for persons in the placebo group (**tx** = 0):

$$\begin{aligned}\hat{\mathbf{h}}_1(25) &= 1/6 = 0.1667\\ \hat{\mathbf{S}}(23) &= \hat{\mathbf{S}}(21)\mathbf{Pr}(T > 23|T \geq 23) = 0.4150(1 - 1/8)\\ &= 0.4150(0.875) = 0.3631\\ \hat{\mathbf{I}}_1(25) &= \hat{\mathbf{h}}_1(25)\hat{\mathbf{S}}(23) = 0.1667(.3631) = 0.0605\\ \mathbf{CIC}_1(25) &= \mathbf{CIC}_1(23) + 0.0605 = 0.2949 + 0.0605\\ &= 0.3554\end{aligned}$$

   c. interpret the $\mathbf{CIC}_1$ values obtained for both the treat ment and placebo groups at $t_f = 30$.
   For tx = 1, $\mathbf{CIC}_1(t_f = 30) = 0.3087$ and for tx = 0, $\mathbf{CIC}_1(t_f = 30) = 0.3554$.

Thus, for treated subjects (tx = 1), the cumulative risk (i.e., marginal probability) for local bladder cancer recurrence is about 30.1 % at 30 months when allowing for the presence of competing risks for bladder metastasis or other metastasis.

For placebo subjects (tx = 1), the cumulative risk (i.e., marginal probability) for local bladder cancer recurrence is about 35.5% at 30 months when allowing for the presence of competing risks for bladder metastasis or other metastasis.

The placebo group therefore has a 5% increased risk of failure than the treatment group by 30 months of follow-up.

d. Calculating the $\mathbf{CPC_1}$ values for both treatment and placebo groups at $t_f = 30$:

The formula relating CPC to CIC is given by $\mathbf{CPC_c} = CIC_c/(1 - CTC_{c'})$ where CIC_c = CIC for cause-specific risk event = 1 and $CIC_{c'}$ = CIC from risks for events = 2 or 3 combined

For tx = l, $CIC_1(t_f = 30) = 0.3087$ and for tx = 0, $CIC_1(t_f = 30) = 0.3554$.

The calculation of $CIC_{c'}$ involves recoding the event variable to 1 for subjects with bladder metastasis or other metastasis and 0 otherwise and then computing $CIC_{c'}$. Calculation of $\mathbf{CIC_{c'}}$ involves the following calculations.

tx = 1 (Treatment A)

t_f	n_f	d_{1f}	$\hat{h}_1(t_f)$	$\hat{S}(t_{f-1})$	$\hat{I}_1(t_f)$	$CIC_{1'}(t_f)$
0	27	0	0	—	—	—
2	27	1	**.0370**	**l**	**.0370**	**.0370**
3	26	2	**.0769**	**.9630**	**.0741**	**.1111**
4	24	0	0	.8889	0	**.1111**
8	**23**	**1**	**.0435**	**.8889**	**.0387**	**.1498**
9	21	1	**.0476**	**.8116**	**.0386**	**.1884**
10	20	1	**.0500**	**.7729**	**.0386**	**.2270**
15	17	1	**.0588**	**.7343**	**.0432**	**.2702**
16	15	1	**.0667**	**.6479**	**.0432**	**.3134**
18	14	0	0	.6047	0	.3134
22	12	0	0	.6047	0	.3134
23	11	0	0	.5543	0	.3134
24	8	0	0	.5039	0	.3134
26	7	0	0	.4409	0	.3134
28	4	1	**.2500**	**.3779**	**.0945**	**.4079**
29	2	0	0	.2835	0	.4079
30	1	0	0	.2835	0	.4079

tx = 0 (Placebo)

t_f	n_f	d_{1f}	$\hat{h}_1(t_f)$	$\hat{S}(t_{f-1})$	$\hat{I}_1(t_f)$	$CIC_{\mathbf{1}'}(t_f)$
0	26	0	0	—	—	—
1	26	0	0	1	0	0
2	24	0	0	.9615	0	0
3	23	0	0	.9215	0	0
5	21	1	**.0476**	**.8413**	**.0400**	**.0400**
6	20	2	**.1000**	**.8013**	**.0801**	**.1201**
7	18	1	**.0556**	**.7212**	**.0401**	**.1602**
10	16	1	**.0625**	**.6811**	**.0426**	**.2028**
12	15	1	**.0667**	**.6385**	**.0426**	**.2454**
14	13	0	0	6835	0	.2454
16	12	1	**.0833**	**.5534**	**.0461**	**.2915**
17	10	0	0	.4612	0	.2915
18	9	0	0	.4150	0	.2915
21	8	1	**.1250**	**.4150**	**.0519**	**.3434**
23	7	0	0	.3632	0	.3434
25	6	1	**.1667**	**.3632**	**.0605**	**.4039**
29	4	0	0	.2421	0	.4039
30	2	0	0	.2421	0	.4039

From these tables, for tx = 1, $CIC_{\mathbf{1}'}((t_f) = 30) = 0.4079$, and for tx = 0, $CIC_{\mathbf{1}'}((t_f) = 30) = 0.4039$.
Thus, for tx =1, $\mathbf{CPC_1}((t_f)=30) = 0.3087/(1 - 0.4079) = \mathbf{0.5213}$, and for tx = 0, $\mathbf{CPC_1}((t_f) = 30) = 0.3554/(1 - 0.4039) = \mathbf{0.5962}$.

6. a. $HR_1(\text{tx} = 1 \text{ vs. tx} = 0) = 0.535 (= 1/1.87)$, p-value = 0.250, N.S.
   b. $HR_2(\text{tx} = 1 \text{ vs. tx} = 0) = 0.987$, p-value = .985, N.S.
   c. $HR_3(\text{tx} = 1 \text{ vs. tx} = 0) = 0.684\ (= 1/1.46)$, p-value = .575, N.S.
7. a. Hazard model formula for the LM model:

$$h^*_g(t, \mathbf{X}) = h^*_{0g}(t) \exp[\beta_1 \text{tx} + \beta_2 \text{num} + \beta_3 \text{size} + \delta_1(\text{txd2}) + \delta_2(\text{numd2}) + \delta_3(\text{sized2}) + \delta_4(\text{txd3}) + \delta_5(\text{numd3}) + \delta_6(\text{sized3})]$$

$g = 1, 2, 3$

where
d2 = 1 if bladder metastasis and 0 otherwise, and
d3 = 1 if or other metastasis and 0 otherwise

b. Hazard ratios for the effect of each of the 3 cause-specific events:

$$\begin{aligned}
HR_1(tx = 1 \text{ vs. } tx = 0) &= \exp(-0.6258) \\
&= 0.535(= 1/1.87) \\
HR_2(tx = 1 \text{ vs. } tx = 0) &= \exp(-0.6258 + .6132) \\
&= 0.987(= 1/1.01) \\
HR_3(tx = 1 \text{ vs. } tx = 0) &= \exp(-0.6258 + .2463) \\
&= 0.684(= 1/1.46)
\end{aligned}$$

c. Corresponding HRs are identical.

8. a. Hazard model formula for the LM_{alt} model:

$$\begin{aligned}
h'_g(t, \mathbf{X}) \quad &= h'_{0g}(t) \exp[\, \delta'_{11}\, \text{txd1} + \delta'_{12}\, \text{numd1} + \delta'_{13}\, \text{sized1} \\
g = 1, 2, 3 \qquad &+ \delta'_{21} \text{txd2} + \delta'_{22} \text{numd2} \\
&+ \delta'_{23} \text{sized2} + \delta'_{31} \text{txd3} \\
&+ \delta'_{32} \text{numd3} + \delta'_{33} \text{sized3}]
\end{aligned}$$

where

d1 = 1 if local bladder cancer recurrence and 0 otherwise

d2 = 1 if bladder metastasis and 0 otherwise, and

d3 = 1 if or other metastasis and 0 otherwise

b. Hazard ratios for the effect of each of the three cause-specific events:
output.

$$\begin{aligned}
HR_1(tx = 1 \text{ vs. } tx = 0) &= \exp(-0.6258) \\
&= 0.535(= 1/1.87) \\
HR_2(tx = 1 \text{ vs. } tx = 0) &= \exp(-0.0127) \\
&= 0.987(= 1/1.01) \\
HR_3(tx = 1 \text{ vs. } tx = 0) &= \exp(-0.3796) \\
&= 0.684(= 1/1.46)
\end{aligned}$$

c. Corresponding hazard ratios arc identical.

9. No interaction SC LM model:

$$\begin{aligned}
h^*_g(t, \mathbf{X}) \quad &= h^*_{0g}(t) \exp[\, \beta_1\, tx + \beta_2\, num + \beta_8\, size] \\
g = 1, 2, 3
\end{aligned}$$

Assumes $HR_1(X) = HR_2(X) = HR_3(X)$ for any X variable e.g., Rx = 0 vs. Rx = 1:
$HR_1(tx) = HR_2(tx) = HR_3(tx) = \exp[\beta_1]$

10. Carry out the following likelihood ratio test:

$$H_0: \ \delta_{gj} = 0 \quad g = 2, 3; \quad j = 1, 2, 3$$

where δ_{gj} is coefficient of $D_g X_j$ in the interaction SC LM model
$LR = 2\log L_R - (-2\text{Log}L_F)$ approx χ^2_6 under H_0
R = no-interaction SC (reduced) model
F = interaction SC (full) model

Chapter 10

1. Example: A=2, F=2, so Mt=A/2 + F =3, R=2
   α=0.05, β=0.10
   $\lambda_0 = 0.10$, $\lambda_1 = 0.05$, $\Delta = \lambda_0/\lambda_1 = 2$

   $N_{EV} = \{(1.96 + 1.282)[2(2) + 1]/[\sqrt{2}(2 - 1)]\}^2$
   $= 131.382 \approx 132$

   Using Formula 1:

   $$N = \frac{131.382}{\frac{2}{2+1}\{1 - e^{-2(0.05)(3)}\} + \frac{1}{2+1}\{1 - e^{-(0.05)3}\}}$$

   N_{EV}

   $$= \frac{131.832}{0.2192} = 601.4 \approx 602$$

   P_{EV}

   $N_1 = [2/3]601.4 = 400.93 \approx 401$ and
   $N_0 = 400.9/2 = 200.45 \approx 200$

2. Nev =131.383 = 132 from question 1.

   $$P_{EV1} = 1 - \frac{1}{(0.05)(2)}\left[e^{-(0.05)(2)} - e^{-0.05)(2+2)}\right]$$
   $$= 1 - 0.8611 = 0.1389$$

   $$P_{EV0} = 1 - \frac{1}{(0.10)(2)}\left[e^{-(0.10)(2)} - e^{-(0.10)(2+2)}\right]$$
   $$= 1 - 0.7421 = 0.2579$$

   $$N = \frac{131.382}{\frac{2}{2+1}(0.1389) + \frac{1}{2+1}(0.2579)}$$
   $$= \frac{131.832}{0.1786} = 738.14 \approx 739$$

   $N_1 = [2/3]738.14 = 492.09 \approx 492$ and
   $N_0 = 492.09/2 = 246.04 \approx 246$.

3. The results using Formulae 1 and 2 are somewhat different since Formula 1 yields N=602 whereas Formula 2 yields N=739. Formula 1 uses the median follow-up time M_F in the computation of p_{EVi} whereas Formula 2 computes p_{EVi} by assuming that the time X at which any subject enters the study has the uniform distribution over the accrual period.
4. $N_{LOFadj} = 739/(1 - 0.25) = 985.33 \approx 986$
5. $N_1 = [2/3]985.33 = 656.89 \approx 657$ and $N_0 = 656.89/2 = 328.44 \approx 328$
6. $N_{ITTadj} = 986/(1 - 0\ 05\ -0.10)^2 = 1364.71 \approx 1365$

7. $N_1 = [2/(2+l)]1364.71 = 909.81 \approx 910$ and $N_0 = 909.81/2 = 454.91 \approx 455$
8. From question 6, the required accrual rate is $r = N/A = 1365/2 = 682.5 \approx 683$ subjects per year. If this accrual rate is not feasible, but r* was considered feasible, then you can adjust your sample size by reducing the accrual period to $A^* = N/r^*$. For example, if the maximum for r is $r_{max} = 600$/yr, then the required accrual period is modified from A=2 to $A^* = 2.275$ years.

   Now suppose, we keep N_{EV} (=131.382), F=2, R=2, α=0.05, β=0.10, $\lambda_0 = 0.10$, $\lambda_1 = 0.05$, and $\Delta = \lambda_0/\lambda_1 = 2$ all constant, but increase the accrual time to A*=2.275 years. Then we would need to re-compute p_{EV1}, p_{EV0} and N to obtain p_{EV1}= 0.2677, p_{EV0}= 0.1447, and N = 579.541 (prior to adjusting for LOF and Crossovers), which is modified to N* = 1069.51 after adjusting for 25% LOF rate, 5% d_c rate and 10% d_t, rate. For this modified sample size, the modified required accrual rate is $r^* = N^*/A^* = 1069.71/2.275 = 470.11$, which is less than $r_{max} = 600$, so that the study is feasible.

   Note, however, it is also possible to obtain a feasible study if the accrual period remains at A=2, but the follow-up period increases to, say F=4, again keeping N_{EV}(=131.382), R = 2, $\alpha = 0.05$, $\beta = 0.10$, $\lambda_0$0.10, $\lambda_1 = 0.05$, and $\Delta = \lambda_0/\lambda_1 = 2$ all constant. This will require re-computing p_{EV1}, p_{EV0} and N again, followed by adjustments for LOF and Crossovers. In particular, if F is increased (to say F=4), then p_{EV1} and p_{EV0} should correspondingly increase from previously calculated values because the probability for an event occurring should increase if follow-up time is increased.