

Reading Assignment: Collett, Chapters 3, 4 and 8

Written Assignment: Due in class on Tuesday, October 11, 2005

- Starting with the relationship $h(t) = f(t)/S(t)$ that we derived in class from the definition of the hazard function $h(t)$, show that if the hazard function satisfies the proportionality criterion $h(t) = h_o(t)g(x, \beta)$, for some function $g(x, \beta)$, then $S(t) = [S_o(t)]^{g(x, \beta)}$.
- The data given in the following table are from a sample of 101 patients with advanced acute myelogenous leukemia reported to the International Bone Marrow Transplant Registry. Fifty-one of those patients had received an autologous (auto) bone marrow transplant in which, after high doses of chemotherapy, their own bone marrow was reinfused to replace their destroyed immune system. Fifty patients had an allogeneic (allo) bone marrow transplant where marrow from an HLA (Histocompatibility Leukocyte Antigen) matched sibling was used to replace their destroyed immune system. An important question in bone marrow transplantation is the comparison of the effects of these two methods of transplant on the distribution of leukemia-free survival times. An event time is recorded if either the patient dies or the patient exhibits signs of leukemia, whichever occurs first. Leukemia-free survival times (in months) are shown in the following table. An asterisk denotes a right-censored time.

Lukemia free survival times for allo transplant patients			Lukemia free survival times for auto transplant patients		
0.030	6.941	26.776*	0.658	8.717	17.303*
0.493	6.941*	28.717*	0.822	9.441*	17.664*
0.855	7.993*	28.717*	1.414	10.329	18.092
1.184	8.882	32.928*	2.500	11.480	18.092*
1.283	8.882	33.783*	3.322	12.007	18.750*
1.480	9.145*	34.221*	3.816	12.007*	20.625*
1.776	11.480	34.770*	4.737	12.237	23.158
2.138	11.513	39.539*	4.836*	12.401*	27.730*
2.500	12.105*	41.118*	4.934	13.059*	31.184*
2.763	12.796	45.033*	5.033	14.474*	32.434*
2.993	12.993*	46.053*	5.757	15.000*	35.921*
3.224	13.849*	46.941*	5.855	15.461	42.237*
3.421	16.612*	48.289*	5.987	15.757	44.638*
4.178	17.138*	57.401*	6.151	16.480	46.480*
4.441*	20.066	58.322*	6.217	16.711	47.467*
5.691	20.329*	60.625*	6.447*	17.204*	48.322*
5.855*	22.368*		8.651	17.237	50.086

- A. Plot the Kaplan-Meier estimator of the survivor function for allo transplant patients.
 - (i) Estimate the probability of leukemia free survival for at least 18 months with the allo transplant and report a standard error for your estimate.
 - (ii) Construct a 95% confidence interval for the median leukemia free survival time for allo transplant patients.
- B. Repeat part A for the auto transplant patients.
- C. Construct a 95% confidence interval for the difference in the median leukemia free survival times for allo and auto transplant patients and state your conclusion.
- D. Test the null hypothesis that the survivor curves are the same for allo and auto transplant patients.
 - (i) State your results and conclusion from the log-rank test
 - (ii) State your results and conclusion from the Wilcoxon test
- E. It is well known that patients given the allogeneic transplant tend to have more
 - (iii) complications early in the recovery process. The most critical of those complications is acute graft-versus-host disease, which tends to occur within the first 100 days after transplant and is often lethal. Because patients given an autologous transplant are not at risk of developing acute graft-versus-host disease, they tend to have a higher survival rate during the first 100 days. Of primary interest to most investigators is the comparison of leukemia free survival rates among long-term survivors. Taking account of this concern, should either of the tests in part D be preferred to the test using Fleming and Harrington weights with $p=0$ and $q=1$. Those weights, $W(t_i) = 1 - \hat{S}(t_{i-1})$, produce a chi-square test statistic with value 4.20 and $p\text{-value}=0.0404$. Justify your answer.

- 3. A. Fit a Weibull model to the distribution of leukemia free survival times for the allo transplant patients in problem 1. If you can, plot the estimated survivor curve on the same plot as the Kaplan-Meier estimator, otherwise turn in a separate plot.
 - (i) Estimate the probability of leukemia free survival for at least 18 months and report a standard error for your estimate.
 - (ii) Construct a 95% confidence interval for the median leukemia free survival time.
- B. Repeat part A for the auto transplant patients in problem 1.
- C. Construct a 95% confidence interval for the difference in the medium leukemia free survival times for the allo and auto transplant patients. State your conclusion.

- D. Using a proportional hazards Weibull model, test the null hypothesis that the survivor curves are the same for allo and auto transplant patients. State your conclusion.
- E. What are the potential advantages of using Weibull models? What are potential advantages of using a nonparametric approach like Kaplan-Meier estimation?
4. Now we will assess the suitability of the Weibull models and the proportional hazards condition considered in problem 3. Expressing the survivor function for a Weibull model as

$$S(t) = \exp(-\theta t^\alpha)$$

the log-cumulative hazard is

$$\log(-\log(S(t))) = \log \theta + \alpha \log(t)$$

Using the Kaplan-Meier estimators of the survivor functions from problem 1, plot $\log(-\log(\hat{S}(t)))$ against $\log(t)$ for both the allo and auto transplant patients. Overlay these plots on the same graph.

- (i) A pair of straight lines indicates that Weibull distributions are appropriate for survival times for the two groups.
- (ii) Parallel curves suggest that a proportional hazards model is appropriate.

State your conclusions from the plot. Is there any indication that a pair of exponential models would be appropriate? (Note: You could construct a formal goodness-of-fit test from squared correlation coefficients computed from points on a single Weibull probability plot, but you may have to simulate the distribution of the squared correlation coefficient under the null hypothesis that the Weibull model is appropriate. You are not asked to do that here.)

5. In some time to event situations, the event may never occur for part of the population. In a study of the effectiveness of treatments for getting smokers to quit, for example, one could examine time from quitting until the subject resumes smoking (time to recidivism). Some subjects may permanently refrain from smoking (the “cured” subjects). It would be desirable to have a model that incorporates both time to recidivism for “non-cured” subjects and the percentage of cured subjects. Right censoring of recidivism times is a potential problem because the data would not reveal if right censored times belong to subjects who are “cured” or if the subjects would eventually return to smoking. Suppose n subjects were enrolled in a study of the effectiveness of a smoking cessation program and each subject quit smoking at the beginning of the program. The responses of interest are time until a subject returns to smoking, for those subjects who return to smoking (recidivists), and the proportion of subjects who were successful in permanently giving up smoking.

- A. Let π denote the proportion of the population of smokers who would be successful at permanently quitting smoking with this program. Propose a model that includes a Weibull distribution for recidivism times and allows for right censoring of recidivism times and the proportion of cured subjects.
- B. Write down the likelihood function for your model in part A.
- C. Let \mathbf{X}_i denote a set of covariates for the i -th individual in the study that describes conditions under which that individual responds, such as previous smoking history, current family support, and other drug use. Indicate how effects of covariates could be incorporated into your model from part A.
- D. If it is possible, explain how π , the proportion of the population of smokers who would be successful at permanently quitting smoking with this program, could be estimated with a nonparametric Kaplan–Meier approach.
6. A study is being planned to examine the possible effects of the anti-viral drug, interferon, to delay infections in patients with chronic granulomatous disease (CGD). CGD is a group of inherited disorders of the immune system that render patients vulnerable to certain infections. CGD patients will be randomized to either a placebo or the interferon treatment. The end-point of interest in this study is time to first serious infection. Past experience with untreated CGD patients suggests that survival probabilities for the placebo group will approximately be as follows:

<u>Time</u> <u>(In months)</u>	<u>Survival</u> <u>Probability</u>
0	1.00
1	0.90
2	0.80
3	0.70
3	0.60
6	0.45
8	0.25
12	0.15

- (a) In planning the study, the researchers believe that treatment with interferon will result in a survival distribution with the following survival times.

<u>Time</u> <u>(In days)</u>	<u>Survival</u> <u>Probability</u>
0	1.00
1	0.91
2	0.82
3	0.75
4	0.68
6	0.59
8	0.35
12	0.25

Patients will be recruited for during a 3 month accrual period. This will be followed by an 8 month follow-up period. What is the number of patients needed to have power .90 of observing a significance difference between the estimated survivor functions using a log-rank test at the $\alpha = .05$ level of significance? Assume that patients are recruited uniformly across time during the accrual period and 50% of the patients are randomized to each treatment group. CGD patients who do not exhibit an infection by the end of the follow-up period have right censored times. In addition, assume an exponential distribution for random loss to follow-up with rate 0.05 patients per day for each treatment group.

- (b) Repeat part (a) for the Wilcoxon test.
- (c) Repeat part (a) for the log-rank and Wilcoxon tests with the follow-up time extended to 12 months. Add a line at $t = 16$ months with $S(t) = .05$ for the placebo group and $S(t) = .15$ for the CGD treated patients.
- (d) Repeat part (a) for the log-rank and Wilcoxon tests with the original follow-up time of 8 months, but change the accrual time to 2 months. Change the last line of each set of project survival times from 12 to 10 months with survival probabilities of 0.20 and 0.30 for the placebo and CGD groups respectively.
- (e) Which is the better thing to do, shortening the accrual time or lengthening the follow-up time? Explain.