

# Survival Analysis Assignment

Due June 3, 2021

David M. Rocke

May 25, 2021

A clinical trial is intended to test a new treatment for malignant melanoma vs. current standard of care. The main outcome is disease-free survival. Each patient has a time  $t$  in days after start of therapy that either represents the time of recurrence or death (if status = 1) or the end of the study (if status = 0). Each patient has two covariates we can use:  $T_x = 1$  if they are on the new therapy or  $T_x = 0$  if they are on standard of care. There are four subtypes of melanoma, which we will characterize as Type = A, B, C, or D. The effect of the new therapy may differ among the four subtypes.

- 1 We can estimate the survival curve for patients on the new and standard therapy including all four subtypes using the Kaplan-Meier product limit estimator. Suppose at a given time  $t$  (in days after starting therapy), and for a particular subset of the patients, that there are 194 patients whose survival or censoring time is at least  $t$ . Suppose that 3 patients died or relapsed on day  $t$  and 2 patients had a censoring time of  $t$ . By what fraction does the estimated survival curve drop at time  $t$ ? How many patients are in the risk set just before  $t$  and just after  $t$ ?

- 2 Write down the Cox model for predicting survival from Tx, Type, and the Tx by Type interaction including a definition of the coefficients and their relationship to survival. State the important assumptions.
- 3 If there are 20 patients at risk at a given time and 1 of them fails, write down the contribution (factor) to the partial likelihood from that failure time in terms of the model specification. Does it depend on the base hazard?

- 4 Which will generate more accurate coefficient estimates, a study with 1000 patients of whom 900 survive to the end of the study without recurrence, or a study with 500 patients 300 of whom survive? Why?
- 5 Describe the most appropriate hypothesis test for whether the interaction term is required in the model. How the test statistic be calculated? To what specific statistical distribution would the test statistic be compared?

- 6 Suppose that the reference levels of the covariates are  $T_x = 0$  and  $Type = A$ . List all the coefficients in the model (symbolically). In terms of those coefficients, what would be the estimated log hazard ratio of a patient with Type B melanoma on  $T_x = 1$  to a patient with Type C melanoma on  $T_x = 0$ ? What would be the estimated hazard ratio?
- 7 How would you examine the proportionality assumption, graphically and/or with a statistical test?
- 8 If it appears that the different subtypes have non-proportional hazards, how would you change the model so that this could be accommodated?