Homework 08

Spencer Pease

3/13/2020

(Q1) Bladder Cancer Survival

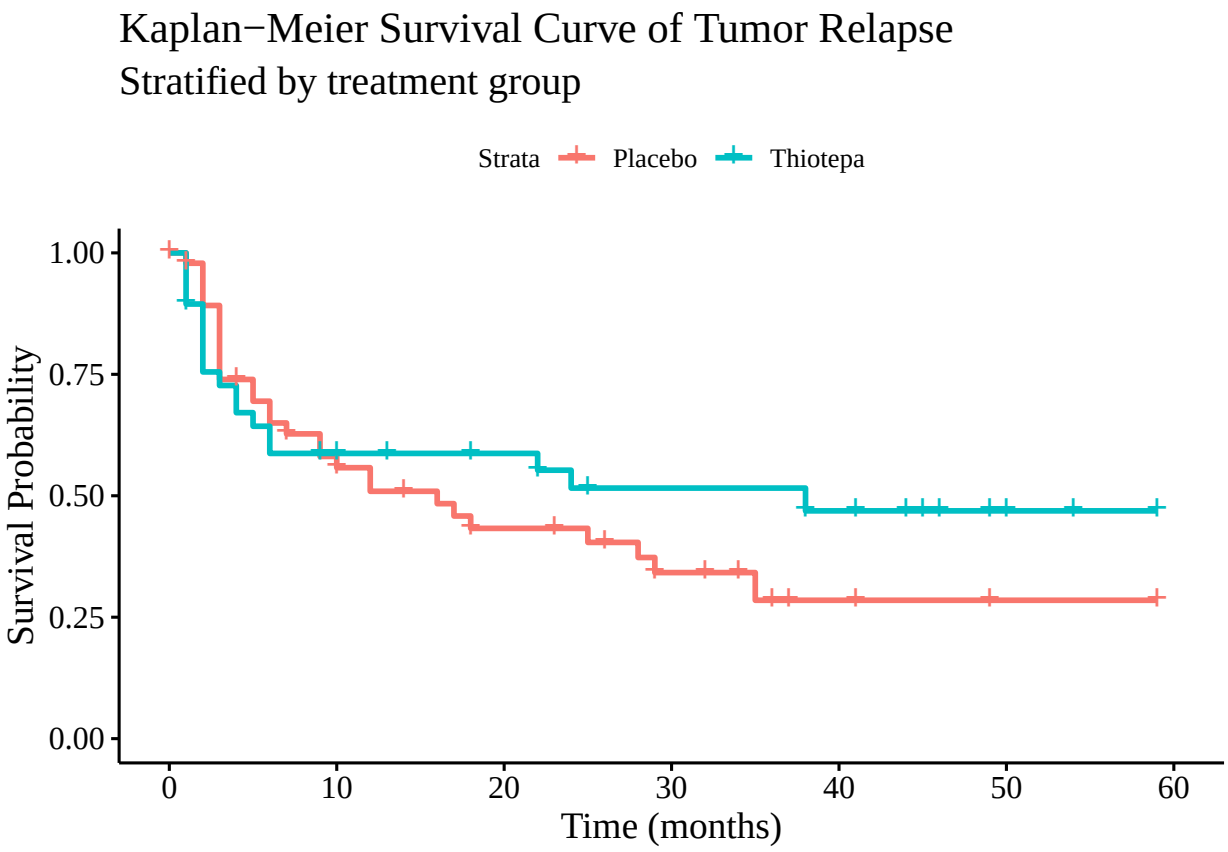

Table 1: Logrank test of Kaplan-Meier Curve by treatment group

Chi-Sq	df	P-value
0.809	1	0.368

A logrank test greater than 0 shows there is some difference in the observed vs expected values of remission by treatment group. A p-value over .05 suggests this difference is not statistically significant.

(Q2)

(Q2.a)

Table 2: Summary of Cox-PH model with predictor 'group'

Predictor	coef	exp(coef)	SE	z-stat	Pr(>|z|)	exp(95%L)	exp(95%H)
group	-0.266	0.767	0.303	-0.877	0.38	0.423	1.388

The subset of the population in the thiotepa treatment group is associated with being 0.767 (95% CI: 0.423, 1.388) times as likely to come out of remission, on average, as the subset of the population in the placebo treatment group. A p-value of 0.38 suggests this difference is not statistically significant.

(Q2.b)

- Cox-PH p-value: 0.368
- Kaplan-Meier Curve logrank test p-value: 0.38

Both of these p-values are on the same order of magnitude and indicate that their results are not statistically significant. The difference comes from what the two methods are testing. The logrank test is only testing for a difference in expected vs observed value between groups, while the Cox-PH regression tests for an association between the response and predictor. In this case, since both the response and predictor are binary variables, the two test are getting at a similar relationship in different ways.

(Q3)

The difference between using *group* and GX as predictors in the model, since they both look at the treatment status of the patients, is the implicit ordering of binary variable. GX encodes "placebo" as the higher treatment group, while *group* encodes "thiotepa" as the higher treatment group. Since all that's changing is the direction, the parameter estimate β_1 will change sign between the two groups, and the respective hazard ratios will be inverses of each other.

Table 3: Summary of Cox-PH model with predictor 'GX'

Predictor	coef	exp(coef)	SE	z-stat	Pr(>|z|)	exp(95%L)	exp(95%H)
GX	0.266	1.304	0.303	0.877	0.38	0.72	2.362

(Q4) Primary Biliary Cirrhosis Survival

Table 4: Summary of Cox-PH model with predictor 'albumin'

Predictor	coef	exp(coef)	SE	z-stat	Pr(>|z|)	exp(95%L)	exp(95%H)
albumin	-1.796	0.166	0.209	-8.576	9.85e-18	0.11	0.25

Two subsets of the population differing by $1\frac{g}{dl}$ albumin are associated with hazards of death 0.166 (95% CI: 0.11, 0.25) times as high, on average, in the higher group. A p-value of 9.85e-18 suggests this difference is statistically significant.

(Q5)

(Q5.a)

Table 5: Estimated log(hazard) difference and HR for differences in 'albumin'

reference	value	log(hazard) diff	hazard ratio
3	2.5	0.898	2.454
3	3.5	-0.898	0.407
3	4.0	-1.796	0.166

(Q5.b,c)

The hazard ratio for $2.5\frac{g}{dl}$ to $3\frac{g}{dl}$ is the inverse of the hazard ratio for $3.5\frac{g}{dl}$ to $3\frac{g}{dl}$.

Since the difference between 3.5 and 2.5 is 1, the estimated hazard ratio would be the same as the exponentiated parameter estimate (e^{β_1}).

(Q6) CVD Smoking and Death

(Q6.a)

We should adjust for age in this analysis, since we are interested in the association between smoking and death. It's a given that older people are more likely to die for many reasons, but we are interested in the direct effect smoking has on mortality. Adjusting for age allows us to get at this relationship, controlling for the additional mortality associated with age.

(Q6.b)

Since there are significantly more people in the sample who don't smoke, and they are on average older than those who do smoke, we would expect the age adjusted hazard ratio for smoking to be higher than the unadjusted hazard ratio. Older people are generally more likely to die, so the age-related deaths in the higher ages made the smoking-related deaths in the younger ages more similar, hiding the real hazard of smoking. Adjusting for age shows the direct hazard of smoking.

(Q6.c)

The 95% confidence interval is calculated from the estimated coefficient, then exponentiated, and not from the exponentiated coefficient directly. This means we can't expect the confidence interval to be symmetric around the estimated hazard ratio, since they are not calculated from each other.