Survival analysis URI and CHD

Isabella Chittumuri

3/30/2021

Test questions #1-3, textbook pg. 153

1. Consider a hypothetical 2-year study to investigate the effect of a passive smoking intervention program on the incidence of upper respiratory infection (URI) in newborn infants. The study design involves the random allocation of one of three intervention packages (A, B, C) to all healthy newborn infants in Orange County, North Carolina, during 1985. These infants are followed for 2 years to determine whether or not URI develops. The variables of interest for using a survival analysis on these data are:

T = time (in weeks) until URI is detected or time until censored

s = censorship status (= 1 if URI is detected, = 0 if censored)

PS = passive smoking index of family during the week of birth of the infant

DC = daycare status (= 1 if outside daycare, = 0 if only daycare is in home)

BF = breastfeeding status (= 1 if infant is breastfed, = 0 if infant is not breastfed)

T1 = first dummy variable for intervention status (= 1 if A, = 0 if B, = -1 if C)

T2 = second dummy variable for intervention status (= 1 if B, = 0 if A, = -1 if C)

a. State the Cox PH model that would describe the relationship between intervention package and survival time, controlling for PS, DC, and BF as confounders and effect modifiers. In defining your model, use only two factor product terms involving exposure (i.e., intervention) variables multiplied by control variables in your model.

$$h(t,X) = h_0(t)e^{\beta_1 T_1+\beta_2 T_2+\beta_3 PS+\beta_4 DC+\beta_5 BF+\beta_6(T_1*PS)+\beta_7(T_2*PS)+\beta_8(T_1*DC)+\beta_9(T_2*DC)+\beta_{10}(T_1*BF)+\beta_{11}(T_2*BF)}$$

b. Assuming that the Cox PH model is appropriate, give a formula for the hazard ratio that compares a person in intervention group A with a person in intervention group C, adjusting for PS, DC, and BF, and assuming interaction effects.

$$\begin{aligned} HR &= \frac{h(t,X*)}{h(t,X)} \\ &= \frac{h_0(t)e^{1\beta_1+0\beta_2+\beta_3 PS+\beta_4 DC+\beta_5 BF+\beta_6(1*PS)+\beta_7(0*PS)+\beta_8(1*DC)+\beta_9(0*DC)+\beta_{10}(1*BF)+\beta_{11}(0*BF)}}{h_0(t)e^{-1\beta_1-1\beta_2+\beta_3 PS+\beta_4 DC+\beta_5 BF+\beta_6(-1*PS)+\beta_7(-1*PS)+\beta_8(-1*DC)+\beta_9(-1*DC)+\beta_{10}(-1*BF)+\beta_{11}(-1*BF)}} \\ &= e^{2\beta_1+\beta_2+2\beta_6 PS+\beta_7 PS+2\beta_8 DC+\beta_9 DC+2\beta_{10} BF+\beta_{11} BF} \end{aligned}$$

c. Assuming that the PH model in part 1a is appropriate, describe how you would carry out a chunk test for interaction; i.e., state the null hypothesis, describe the test statistic and give the distribution of the test statistic and its degrees of freedom under the null hypothesis.

Degrees of freedom = number of predictors in $H_0 = 6$

R = reduced model

F = full model

$H_0 : \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = 0$

$H_a : \beta_6 \neq \beta_7 \neq \beta_8 \neq \beta_9 \neq \beta_{10} \neq \beta_{11} \neq 0$ (all of those β_i not equal to 0)

Likelihood ratio test statistic: $-2lnL_R - (-2lnL_F)$, which is approximately χ^2_6 with 6 degrees of freedom (d.f.) under H_0

d. Assuming no interaction effects, how would you test whether packages A, B, and C are equally effective, after controlling for PS, DC, and BF in a Cox PH model without interaction terms (i.e., state the two models being compared, the null hypothesis, the test statistic, and the distribution of the test statistic under the null hypothesis).

The two models being compared:

Full model(F): $h(t, X) = h_0(t)e^{\beta_1 T_1 + \beta_2 T_2 + \beta_3 PS + \beta_4 DC + \beta_5 BF}$

Reduced model(R): $h(t, X) = h_0(t)e^{\beta_3 PS + \beta_4 DC + \beta_5 BF}$

$H_0 : \beta_1 = \beta_2 = 0$

$H_a : \beta_1 \neq \beta_2 \neq 0$

Likelihood ratio test statistic: $-2lnL_R - (-2lnL_F)$, which is approximately χ^2_2 with 2 d.f. under H_0

e. For the no-interaction model considered in parts 1c and 1d, give an expression for the estimated survival curves for the effect of intervention A adjusted for PS, DC, and BF. Also, give similar (but different) expressions for the adjusted survival curves for interventions B and C.

Intervention A:

$$\hat{S}(t, X) = \hat{S}_0(t)^{e^{\hat{\beta_1} + \hat{\beta_3}(\bar{PS}) + \hat{\beta_4}(\bar{DC}) + \hat{\beta_5}(\bar{BF})}}$$

Intervention B:

$$\hat{S}(t, X) = \hat{S}_0(t)^{e^{\hat{\beta_2} + \hat{\beta_3}(\bar{PS}) + \hat{\beta_4}(\bar{DC}) + \hat{\beta_5}(\bar{BF})}}$$

Intervention C:

$$\hat{S}(t, X) = \hat{S}_0(t)^{e^{-\hat{\beta_1} - \hat{\beta_2} + \hat{\beta_3}(\bar{PS}) + \hat{\beta_4}(\bar{DC}) + \hat{\beta_5}(\bar{BF})}}$$

2. The data for this question consists of a sample of 50 persons from the 1967–1980 Evans County Study. There are two basic independent variables of interest: AGE and chronic disease status (CHR), where CHR is coded as 0 = none, 1 = chronic disease. A product term of the form AGExCHR is also considered. The dependent variable is time until death, and the event is death. The primary question of interest concerns whether CHR, considered as the exposure variable, is related to survival time, controlling for AGE. The edited output of computer results for this question is given as follows:

a. State the Cox PH model that allows for main effects of CHR and AGE as well as the interaction effect of CHR with AGE.

$$h(t,X) = h_0(t)e^{\beta_1 CHR + \beta_2 AGE + \beta_3 (CHR * AGE)}$$

b. Carry out the test for significant interaction; i.e., state the null hypothesis, the test statistic, and its distribution under the null hypothesis. What are your conclusions about interaction?

$H_0 : \beta_3 = 0$

$H_a : \beta_3 \neq 0$

LR (interaction in model 3): $-2lnL_{model2} - (-2lnL_{model3}) = (-2 * -132.45) - (-2 * 132.35)$

$= 264.9 - 264.7 = .2$

LR of 0.2 is approximately χ^2 with 1 d.f. under H_0, which is not significant. Therefore, we should not include the interaction term in the model.

c. Assuming no interaction, should AGE be controlled? Explain your answer on the basis of con- founding and/or precision considerations. (reference tb page 106)

Model 2, where AGE is controlled:

$HR_{CHR} = e^{.8051} = 2.236$

Model 1, where AGE is not controlled:

$HR_{CHR} = e^{.8595} = 2.362$

The hazard ratios are not significantly different, therefore AGE is not a confounder.

Confidence Interval (CI): $e^{\hat{\beta}_1 \pm 1.96 se_{\beta_1}}$

Model 2 CHR Precision $e^{.8051 \pm 1.96(.3252)} = (1.183, 4.231)$ The width between CI is $4.231 - 1.183 = 3.048$

Model 1 CHR Precision $e^{.8595 \pm 1.96(.3116)} = (1.282, 4.350)$ The width between CI is $4.350 - 1.282 = 3.068$

For both models, the width between CI is relatively the same, with model 2's width being slightly smaller. In conclusion, controlling AGE does not have great effect on the final interval estimates. Therefore, AGE should not be controlled.

d. If, when considering plots of various hazard functions over time, the hazard function for persons with CHR = 1 (has chd) crosses the hazard function for persons with CHR = 0 (does not have chd), what does this indicate about the use of any of the three models provided in the printout? (reference tb page 125)

If hazard functions cross, this indicates that the hazard ratios are not constant over time. Therefore, Cox PH model inappropriate because each model assumes constant HR.

e. Using model 2, give an expression for the estimated survival curve for persons with CHR = 1, adjusted for AGE. Also, give an expression for the estimated survival curve for persons with CHR = 0, adjusted for AGE.

For CHR = 1: $\hat{S}(t,X) = \hat{S}_0(t)^{e^{0.8051 + 0.0856(A\bar{G}E)}}$

For CHR = 0: $\hat{S}(t,X) = \hat{S}_0(t)^{e^{0.0856(A\bar{G}E)}}$

f. What is your overall conclusion about the effect of CHR on survival time based on the computer results provided from this study?

Based on the best model, which is model 1, the effect of CHR on survival time is mild due to the fact that it has a HR of 2.362 and a precision (CI width) of 3.068. Using this model, CHR also has a high significant p-value of 0.0058.

3. The data for this question contain remission times of 42 multiple leukemia patients in a clinical trial of a new treatment. The variables in the dataset are given below:

Variable 1: survival time (in weeks)

Variable 2: status (1 = in remission, 0 = relapse) Variable 3: sex (1 = female, 0 = male)

Variable 4: log WBC

Variable 5: Rx status (1 = placebo, 0 = treatment)

a. Use the above computer results to carry out a chunk test to evaluate whether the two interaction terms in model 1 are significant. What are your conclusions?

Full model: $h(t,X) = h_0(t)e^{\beta_1 Rx+\beta_2 Sex+\beta_3 logWBC+\beta_4(Rx*Sex)+\beta_5(Rx*logWBC)}$

Reduced model: $h(t,X) = h_0(t)e^{\beta_1 Rx+\beta_2 Sex+\beta_3 logWBC}$

$H_0 : \beta_4 = \beta_5 = 0$

$H_a : \beta_4 \neq \beta_5 \neq 0$

LR: $= -2lnL_R - (-2lnL_F) = (-2 * -72.109) - (-2 * 96.515) = 144.218 - 139.03 = 5.19$

Chunk test (also know as LR) of 5.19 is approximately χ^2 with 2 d.f. under H_0, which is not significant at alpha level 0.05. However it is significant at alpha level 0.01. Therefore, the chunk test results in a mild significance of the two interactions.

b. Evaluate whether you would prefer model 1 or model 2. Explain your answer.

We can see from the given table that the interaction term Rx*log WBC has a high p-value, suggesting low significance to the model. The LR test in part (a) verifies this suggestion. Therefore, this interaction term should be omitted from the model and as a result model 2 is the better model.

c. Using model 2, give an expression for the hazard ratio for the effect of the Rx variable adjusted for SEX and log WBC.

$$h(t,X) = h_0(t)e^{0.405(Rx)-1.070(SEX)+1.610(logWBC)+2.013(Rx*Sex)}$$

$$HR = \frac{h(t,Rx=1)}{h(t,Rx=0)} = \frac{h_0(t)e^{0.405(1)-1.070(SEX)+1.610(logWBC)+2.013(1*Sex)}}{h_0(t)e^{0.405(0)-1.070(SEX)+1.610(logWBC)+2.013(0*Sex)}} = e^{0.405+2.013(Sex)}$$

d. Using your answer in part 3c, compute the hazard ratio for the effect of Rx for males and for females separately.

Males (Sex=0): $\hat{HR} = e^{0.405+2.013(0)} = 1.499$

Females (sex=1): $\hat{HR} = e^{0.405+2.013(1)} = 11.223$

e. By considering the potential confounding of log WBC, determine which of models 2 and 3 you prefer. Explain.

Controlling for SEX=female:

$$Model_2 : \hat{HR} = e^{0.405+2.013(1)} = 11.223 Model_3 : \hat{HR} = e^{0.587+1.906(1)} = 2.493$$

The hazard ratios between model 2 and model 3 have a difference of 8.73. This is a meaningful difference therefore log WBC is a likely a confounder.

Log WBC also has a high significant p-value of .000.

Another reason why we want log WBC in the model is because it is well-known prognostic indicator of survival for leukemia patients.

In conclusion, we prefer model 2 because it includes log WBC which is a confounding and has a high significant p-value.

f. Of the models provided which model do you consider to be best? Explain.

Of the models provided, we consider model 2 the best because it includes the confounder log WBC as well as Rx*SEX with a p-value of 0.23.