HW 4

Steph Holm and Shelley Facente

4/11/2019

Dose-response modeling

1. Write out both Cox models that we estimated for the simple linear (Model 0) and restricted quadratic spline (Model 1) coded versions of cigarettes per day in terms of the variables above, and general parameters (e.g. β coefficients).

Simple linear model (Model 0)

$$log[h(t|\boldsymbol{x})] = log[h_0(t)] + age \times \beta_1 + cigsperday \times \beta_2$$

where

$\mathbf{log[h(t|\boldsymbol{x})]}$ *is the log-hazard of death at time* t*, accounting for the covariates in our model.*

$\mathbf{log[h_0(t)]}$ *is the log baseline hazard at time* t *for an individual with the referent value for all of the variables.*

$\boldsymbol{\beta_1}$ *is the coefficient on age in years, meaning that if* e *is raised to the* β_1 *power, this is the hazard ratio comparing the hazard of death for those with a one-year increase in age compared to those one year younger, holding constant the other variables in the model.*

$\boldsymbol{\beta_2}$ *is the coefficient on the number of cigarettes the participant consumed per day, meaning that if* e *is raised to the* β_2 *power, this is the hazard ratio comparing the hazard of death for those with a one-unit increase in cigarettes per day compared to those consuming one fewer cigarette, holding constant the other variables in the model.*

Restricted quadratic spline model (Model 1)

$$log[h(t|\boldsymbol{x})] = log[h_0(t)] + age \times \beta_1 + cigsperday \times \beta_2 + (s^2_{20} - s^2_{60})\beta_3 + (s^2_{40} - s^2_{60})\beta_4$$

where

$\mathbf{log[h(t|\boldsymbol{x})]}$ *is the log-hazard of death at time t, accounting for the covariates in our model.*

$\mathbf{log[h_0(t)]}$ *is the log baseline hazard at time t for an individual with the referent value for all of the variables.*

$\boldsymbol{\beta_1}$ *is the coefficient on age in years, meaning that if e is raised to the β_1 power, this is the hazard ratio comparing the hazard of death for those with a one-year increase in age compared to those one year younger, holding constant the other variables in the model.*

$\boldsymbol{\beta_2}$ *is the coefficient on the number of cigarettes the participant consumed per day, meaning that if e is raised to the β_2 power, this is the hazard ratio comparing the hazard of death for those with a one-unit increase in cigarettes per day compared to those consuming one fewer cigarette, holding constant the other variables in the model.*

$\mathbf{(s^2_{20} - s^2_{60})}$ *is the coefficient on the quadratic spline term for one pack a day minus the uppermost quadratic spline term in the model (three packs a day), which sets a knot in the curve at 20 cigarettes per day, meaning that if e is raised to the β_3 power, this is the hazard ratio comparing the hazard of death for those consuming one pack per day compared to non-smokers, holding constant the other variables in the model. Because the uppermost quadratic spline term has been subtracted here, we are tying down the ends of the model, disallowing erratic behavior in the model due to sparse data. At values below sixty, the s^2_{60} term drops out and the coefficient simplifies to s^2_{20}. Furthermore, at values less than twenty, the entire term drops out.*

$\mathbf{(s^2_{40} - s^2_{60})}$ *is the coefficient on the quadratic spline term for two packs a day minus the uppermost quadratic spline term in the model (three packs a day), which sets a knot in the curve at 40 cigarettes per day, meaning that if e is raised to the β_4 power, this is the hazard ratio comparing the hazard of death for those consuming two packs per day compared to non-smokers, holding constant the other variables in the model. Again, because the uppermost quadratic spline term has been subtracted here, we are tying down the ends of the model, disallowing erratic behavior in the model due to sparse data. At values below sixty, the s^2_{60} term drops out and the coefficient simplifies to s^2_{20}. Furthermore, at values less than forty, the entire term drops out.*

2. Consider the likelihood ratio test that compares the simple linear and restricted quadratic spline models. In words, describe what this is testing. Specifically in terms of the model parameters, (i.e. β coefficients), what are the null and alternative hypotheses that are being tested (use formal notation)? What expression of cigarette smoking would you choose based on this test?

This likelihood ratio test is comparing the performance of two nested models, the simple linear and restricted quadratic spline models. This test is used to determine which model is a better fit for the data. In this case the null and alternative hypotheses being tested are:

$$H_0 : \mathcal{L}(log[h_0(t)] + age \times \beta_1 + cigsperday \times \beta_2) = \mathcal{L}(log[h_0(t)] + age \times \beta_1 + cigsperday \times \beta_2 + (s_{20}^2 - s_{60}^2)\beta_3 + (s_{40}^2 - s_{60}^2)\beta_4)$$

$$H_A : \mathcal{L}(log[h_0(t)] + age \times \beta_1 + cigsperday \times \beta_2) \neq \mathcal{L}(log[h_0(t)] + age \times \beta_1 + cigsperday \times \beta_2 + (s_{20}^2 - s_{60}^2)\beta_3 + (s_{40}^2 - s_{60}^2)\beta_4)$$

Based on the results of this test (p-value = 0.11), we would choose the linear model as the null hypothesis is not rejected and this is a simpler model.

3. Using the results from the spline model, report the hazard ratios and 95% confidence intervals comparing:

a) 30 cigarettes per day vs. 10 cigarettes per day.

Using our restricted quadratic spline model, we find that the hazard ratio for 30 cigarettes per day compared to 10 cigarettes per day is 1.39 (95% CI: 1.14, 1.68).

b) 50 cigarettes per day vs. 10 cigarettes per day.

Using our restricted quadratic spline model, we find that the hazard ratio for 50 cigarettes per day compared to 10 cigarettes per day is 0.98 (95% CI: 0.43, 2.25).

Interaction and Effect Modification

4. Write out the expressions for the following in terms of the model parameters and show the corresponding coefficient vector (k, above) for the contrast of interest. For each of these, write out the two equations for relevant log-hazard functions to help show your work.

a) The HR comparing obese (BMI $\geq$ 30) vs. normal weight (18.5 $\leq$ BMI $<$ 25.0) among females.

For the obese female group:

$$0 \times \beta_1 + 1 \times \beta_2 + 0 \times \beta_3 + 0 \times \beta_4 + 0 \times \beta_5 + 0 \times \beta_6 + 0 \times \beta_7 + 0 \times \beta_8 + 0 \times \beta_9$$

For the normal weight female group:

$$0 \times \beta_1 + 0 \times \beta_2 + 0 \times \beta_3 + 0 \times \beta_4 + 0 \times \beta_5 + 0 \times \beta_6 + 0 \times \beta_7 + 0 \times \beta_8 + 0 \times \beta_9$$

$$\begin{aligned}\log(\text{HR}_{10}) &= (0\times\beta_1 + 1\times\beta_2 + 0\times\beta_3 + 0\times\beta_4 + 0\times\beta_5 + 0\times\beta_6 + 0\times\beta_7 + 0\times\beta_8 + 0\times\beta_9) - \\ &\qquad (0\times\beta_1 + 0\times\beta_2 + 0\times\beta_3 + 0\times\beta_4 + 0\times\beta_5 + 0\times\beta_6 + 0\times\beta_7 + 0\times\beta_8 + 0\times\beta_9) \\ &= \beta_2 - 0 \\ &= \beta_2\end{aligned}$$

Coefficient vector (k) : `(0, 1, 0, 0, 0, 0, 0, 0, 0)`

b) The HR comparing obese (BMI $\geq$ 30) vs. normal weight (18.5 $\leq$ BMI $<$ 25.0) among males.

For the obese male group:

$$0\times\beta_1 + 1\times\beta_2 + 1\times\beta_3 + 0\times\beta_4 + 0\times\beta_5 + 0\times\beta_6 + 0\times\beta_7 + 0\times\beta_8 + 1\times\beta_9$$

For the normal weight male group:

$$0\times\beta_1 + 0\times\beta_2 + 1\times\beta_3 + 0\times\beta_4 + 0\times\beta_5 + 0\times\beta_6 + 0\times\beta_7 + 0\times\beta_8 + 0\times\beta_9$$

$$\begin{aligned}\log(\text{HR}_{11}/\text{HR}_{01}) &= (0\times\beta_1 + 1\times\beta_2 + 1\times\beta_3 + 0\times\beta_4 + 0\times\beta_5 + 0\times\beta_6 + 0\times\beta_7 + 0\times\beta_8 + 1\times\beta_9) - \\ &\qquad (0\times\beta_1 + 0\times\beta_2 + 1\times\beta_3 + 0\times\beta_4 + 0\times\beta_5 + 0\times\beta_6 + 0\times\beta_7 + 0\times\beta_8 + 0\times\beta_9) \\ &= (\beta_2 + \beta_3 + \beta_9) - (\beta_3) \\ &= \beta_2 + \beta_9\end{aligned}$$

Coefficient vector (k): `(0, 1, 0, 0, 0, 0, 0, 0, 1)`

c) The HR comparing obese (BMI $\geq$ 30) males vs. normal-weight (18.5 $\leq$ BMI $<$ 25.0) females.

See above for expressions for obese males and normal-weight females.

$$\begin{aligned}\log(\text{HR}_{11}) &= (0\times\beta_1 + 1\times\beta_2 + 1\times\beta_3 + 0\times\beta_4 + 0\times\beta_5 + 0\times\beta_6 + 0\times\beta_7 + 0\times\beta_8 + 1\times\beta_9) - \\ &\qquad (0\times\beta_1 + 0\times\beta_2 + 0\times\beta_3 + 0\times\beta_4 + 0\times\beta_5 + 0\times\beta_6 + 0\times\beta_7 + 0\times\beta_8 + 0\times\beta_9) \\ &= (\beta_2 + \beta_3 + \beta_9) - 0 \\ &= \beta_2 + \beta_3 + \beta_9\end{aligned}$$

Coefficient vector: `(0, 1, 1, 0, 0, 0, 0, 0, 1)`

5. Write out the expression for the RERI for the obesity (BMI $\geq$ 30)-sex interaction in terms of the hazard in each exposure group. Define/interpret each term in the equation and express each term using coefficients from the above regression model.

The RERI for the obesity-sex interaction is:

$$\begin{aligned}\text{RERI}_{\text{HR}} &= \text{HR}_{11} - \text{HR}_{10} - \text{HR}_{01} + 1 \\ &= e^{(\beta_2+\beta_3+\beta_9)} - e^{\beta_2} - e^{\beta_3} + 1\end{aligned}$$

where

$$\text{HR}_{11} = 2.68 = \textit{the hazard ratio for obese males compared to the referent group},$$
$$\text{HR}_{10} = 1.8 = \textit{the hazard ratio for obese females compared to the referent group},$$
$$\text{HR}_{01} = 1.84 = \textit{the hazard ratio for normal-weight males compared to the referent group, and}$$
$$\text{the final term} = 1.0 = \textit{the hazard ratio for normal-weight females (the referent group)}$$

All of these hazard ratios are referring to the hazard of stroke based on BMI and sex.

6. Complete Table 1, being sure to include point estimates and confidence intervals where appropriate, as well as the relevant parts of the footnotes.

Table 1. Adjusted hazards ratio estimates and 95% CI of the association between baseline BMI status and mortality. The Framingham Cohort Study, 1948-1972, Framingham, MA.

BMI	Female	Male (single ref)	Male (stratum-specific)
18.5 - 24.9	1 (ref)	1.46 (1.02, 2.08)	1 (ref)
25.0 - 29.9	1.14 (0.82, 1.59)	1.55 (1.13, 2.13)	1.07 (0.76, 1.49)
$\geq$ 30.0	1.8 (1.24, 2.61)	2.68 (1.7, 4.22)	1.84 (1.15, 2.93)

All hazard ratios adjusted for age and education.

Effect modification on additive scale (95% CI):
RERI male vs. female/25.0-29.9 vs. 18.5-24.9: -0.05 (-0.66, 0.57);
RERI male vs. female / $\geq$ 30.0 vs. 18.5-24.9: 0.42 (-0.81, 1.64).

Effect modification on multiplicative scale (95% CI):
Male vs. female / 25.0-29.9 vs. 18.5-24.9: 0.93 (0.58, 1.49);
Male vs. female / $\geq$ 30.0 vs. 19.5-24.9: 1.02 (0.56, 1.85);
Likelihood ratio test p-value: 0.937.

7. What are your conclusions regarding effect modification on the additive and multiplicative scales? Your answer should include individual interpretations of interaction on each scale.

We conclude that given the wide confidence intervals, there is no convincing evidence of effect measure modification - either positive or negative - on the additive scales for either overweight or obese males and females. In both cases the lower and upper bound of the CIs comfortably contain the null (zero).

The same is true on the multiplicative scales for both overweight and obese males and females. Again in both cases the lower and upper bound of the CIs for tests of interaction comfortably contain the null (in this case, one). This is further supported by the p-value on the likelihood ratio test (which compared our model with interactions included to that without interaction terms), which is 0.937 - quite high.

8. What assumptions regarding confounding have you made in assessing effect modification? If you were instead interested in the causal interaction between BMI and these variables would you have to consider any other relationships? If so, what?

In assessing effect modification, we have assumed that the only additional variables that confound our primary exposure (BMI) are age and education, which we controlled for in our model. If we were interested in the ***causal*** *interaction between BMI and these variables, however, we would also need to identify whether age and education confound the relationship between stroke and our secondary exposure (sex), as well as whether any other confounders exist for that relationship.*