BIO 244: Unit 9

Cox's Model: Extensions

In this unit we consider several extensions of the usual Cox PH model that are both useful in practice and establish links with variations of the logrank test. Large-sample properties of the extensions will be covered in later units. The extensions considered are:

- extending Cox's model to include strata
- Cox's model for comparing P+1 groups; trend tests
- allowing time-varying covariates
- estimating the underlying cumulative hazard $H_0(t)$
- some words on informative censoring

9.1 Stratified Cox Model:

Suppose first that Z and W are binary covariates.

$$\text{e.g.,} \quad Z = \text{treatment group} = \begin{cases} 0 & \text{group 0} \\ 1 & \text{group 1} \end{cases}$$

$$\text{and} \quad W = \text{gender} = \begin{cases} 0 & \text{female} \\ 1 & \text{male,} \end{cases}$$

and we want to assess the association between Z and survival while controlling for W. One way to do this is to fit Z and W as covariates in Cox's model, and then test that the regression coefficient corresponding to Z is zero. That is, fit the model

$$h(t|Z,W) = h_0(t)e^{\beta_1 Z + \beta_2 W} \ ,$$

and then test the hypothesis that $\beta_1 = 0$. Note that with this model:

$$\text{trmt HR} = \frac{h(t|Z=1,W)}{h(t|Z=0,W)} = e^{\beta_1} \quad \text{and} \quad \text{gender HR} = \frac{h(t|Z,W=1)}{h(t|Z,W=0)} = e^{\beta_2}.$$

Thus, hazard functions of the 4 types of individuals are proportional.

Alternatively, consider the following model:

Females (W=0): $h(t|Z,W=0) = h_0(t)e^{\beta_1 Z}$

Males (W=1): $h(t|Z,W=1) = h_1(t)e^{\beta_1 Z}$ (9.1)

$\longrightarrow$ treatment has same (e^{β_1}) hazard ratio within each level of W, but the underlying hazards for females ($h_0(\cdot)$) and males ($h_1(\cdot)$) are arbitrary.

We can also write this model as:

$$h(t|Z,W) = h_W(t)e^{\beta Z}\ , \quad \text{for} \quad Z = 0,1, \quad W = 0,1.$$

Let's generalize: suppose that:

Z = any covariate (vector or scalar)

W = some categorical covariate with L levels (1,. . ., L),

and consider the model:

$$h(t|Z,W) = h_W(t)e^{\beta Z}.$$

That is,

$$h(t \mid Z, W = l) = h_l(t)e^{\beta Z} \qquad l = 1, \ldots, L.$$

This is called a **stratified Cox proportional hazards model**. The L levels of W can have arbitrary underlying hazards, yet within each the treatment relative risk is e^{β}.

How do we make an inference about β for this model? Consider the following:

For the subset of persons with $W = l$, form the usual PL, say $L_1^{(l)}(\beta)$. Then form an overall partial likelihood function as:

$$L_{\text{strat}}\ (\beta) = \prod_{l=1}^{L}\ L_1^{(l)}\ (\beta).$$

- Note the simple form of this stratified partial likelihood function.

- When Z = binary, the resulting partial likelihood score test of $\beta = 0$ reduces to stratified logrank test (Exercise).

- The model is more general than $h_0(t)e^{\beta_1 Z+\beta_2 W}$ (in what way?).

- We study the asymptotic properties of this approach in a later unit.

9.2 Cox Model for Comparing P+1 Groups:
Suppose that W is a categorical covariate with P+1 levels (say, W=0,1,...,P). One way to compare the survival distributions of the P+1 groups is by a logrank test, yielding a test statistics that has an approximate chi-square dsitribution with P df under the null hypothesis. Alternatively, define the P binary covariates $Z_1, \cdots, Z_P$ by

$$Z_j = 1[W = j] \quad \text{for} \quad j = 1, 2, \cdots, P\ .$$

Consider fitting a Cox model with covariates $Z_1, \cdots, Z_P$, and then testing the hypothesis (say, using the partial likelihood score test) that the P regression coefficients are simultaneously zero; that is, $\beta_1 = \cdots = \beta_P = 0$.

This test is equivalent to the logrank test for comparing the P+1 groups. To verify this, compute the Px1 partial likelihood score vector of partial derivatives of the log partial likelihood with respect to the P regression coefficients, and then evaluate it at $\beta_1 = \cdots = \beta_P = 0$. The score test is based on the quadratic form resulting from this vector and the observed information matrix evaluated at $\beta_1 = \cdots = \beta_P = 0$. This will be seen to be the same as the logrank test for P+1 groups. By making this "connection" between a likelihood-based test and the logrank test, we can infer that the logrank test for P+1 groups is oriented towards alternatives where the P+1 groups have proportional hazard functions. Another advantage of making this connection is that the Cox model provides a way of generalizing the logrank test to adjust for other covariates.

Trend Test: Next suppose the categories of W (=0,1,2,...,P) are ordinal, such as increasing doses of the same drug, and let Z^* be the scalar covariate defined by:

$$Z^* = c_j \qquad \text{if} \quad W = j,\ j = 0, 1, ..., P\ ,$$

where $c_0 = 0$ and $c_1, \cdots, c_P$ are some constants. Suppose we fit a Cox proportional hazards model with the scalar covariate Z^* and denote the corresponding regression coefficient by β^*. Then it can be shown that the partial likelihood score test of $\beta^* = 0$ is equivalent to the logrank trend test using weights $c_1, \cdots, c_P$. For example, choosing $c_j = j$ gives the most commonly used version of the logrank trend test (and the default value in STATA), and corresponds to a proportional hazards model where the regression coefficient β represents the log hazard ratio between consecutive values of the ordinal covariate. That is,

$$\frac{h(t \mid W = j+1)}{h(t \mid W = j)} = e^{\beta} \qquad \text{for} \quad j = 0, 1, \cdots, P-1\ .$$

This connection shows that the logrank trend test is oriented towards the alternatives given by this proportional hazards model. We return to the choice of c_j in a later unit. However, it is worth noting that another nice feature of this connection between the logrank test and inferences from Cox's model is that we can now easily extend the logrank trend test to adjust for other covariates by simplying adding these to the Cox proportional hazards model.

Let's illustrate some of these methods with an example using the AZT-ddI data set. As before, we focus on the subjects receiving AZT (rx=0) or 500 mg ddI (rx=1), and who have not yet develop AIDS at the time of enrollment (as-ar-a<2). However, to intentionally 'skew' the data, we only consider those ddI patients with a cd4 count of at least 30. This gives a total of 209 AZT patients and 190 ddI patients. Consider the binary covariate W denoting grouped cd4 count:

$$W = \begin{array}{c}\text{Grouped cd4} \\ (\text{“gcd4”})\end{array} = \begin{cases} 0 & \text{cd4} \leq 100 \\ 1 & \text{cd4} > 100. \end{cases}$$

Of the 209 subjects that received AZT, 96 had a cd4 ≤ 100 and 113 had a cd4 > 100. Of the 190 subjects receiving ddI, 67 had a cd4 ≤ 100 and 123 had a cd4 > 100. We consider several different ways of testing the hypothesis of no difference between the AZT and ddI groups. Here, failure is taken to be time until the development of AIDS or death, whichever comes first, represented by the variables Tad and ad. For each method, we present the resulting chi-square statistic (say Q) and p-value.

1. Ordinary logrank test: Q=15.18, P=.0001
2. Cox model with covariate rx; Wald test: Q=14.44, P=.0001
3. Cox model with covariate rx; LR test: Q=15.46, P=.0001
4. Stratified (by W) logrank test: Q=8.46, P=.0036
5. Stratified (by W) Cox model with covariate rx; Wald test: Q=8.20, P=.004
6. Stratified (by W) Cox model with covariate rx: LR test: Q=8.66, P=.0033
7. Cox model with covariates rx and W; Wald test: Q=8.41, P=.004
8. Cox model with covariates rx and W; LR test: Q=8.89, P=.003

Note that tests 1-3, which do not adjust for cd4, are asymptotically equivalent. Tests 4-6 do adjust for cd4 and are asymptotically equivalent. Tests 7-8 also adjust for cd4, but in a different way, and are asymptotically equivalent.

Given the imbalance between the AZT and ddI groups, one would tend to place greater trust in tests 4-8. Of these, tests 4-6 make fewer assumptions. However, in this example, one gets similar results with tests 4-6 as with tests 7-8.

Now let's consider the covariate W* which represents grouped age, with W*=0,1,2 if the patient's age is 30 or less, 31-40, or over 40, respectively. Also, define W1*=1[W*=1] and W2*=1[W*=2]. Consider testing the hypothesis that grouped age is not associated with time until AIDS or death in the AZT-ddI dataset, but without creating the imbalance by leaving out some patients randomized to ddI. Below we report the resulting chi-square statistics and p-values.

1. Logrank test for W*: Q (2 df)=4.16, P=.125
2. Cox model with covariates W1*, W2*: (LR test) Q (2 df)=4.69, P=.0957
3. Logrank trend test for W*: Q (1 df)=3.04, P=.081
4. Cox model with covariate W*; Wald test: Q (1 df)=3.03, P=.082
5. Cox model with covariate W*, LR test: Q (1 df)=3.07, P=.080

Note that tests 1 and 2 are asymptotically equivalent, and that tests 3-5 are asymptotically equivalent. Tests 3-5 can be expected to be more powerful if there is a trend in the age-response relationship. They are directed towards alternatives where there is a trend. We return to the relative properties of these types of tests later in the course.

9.3 Time-Varying (or Time-Dependent) Covariates

In some settings, we want to assess association between a time-varying covariate, say $Z(t)$, and survival:

Example 9.1 $Z(t) =$ air pollution level at time t

$T =$ time until asthma attack

$$h(t|z(t)) = h_0(t)e^{\beta \cdot Z(t)}$$

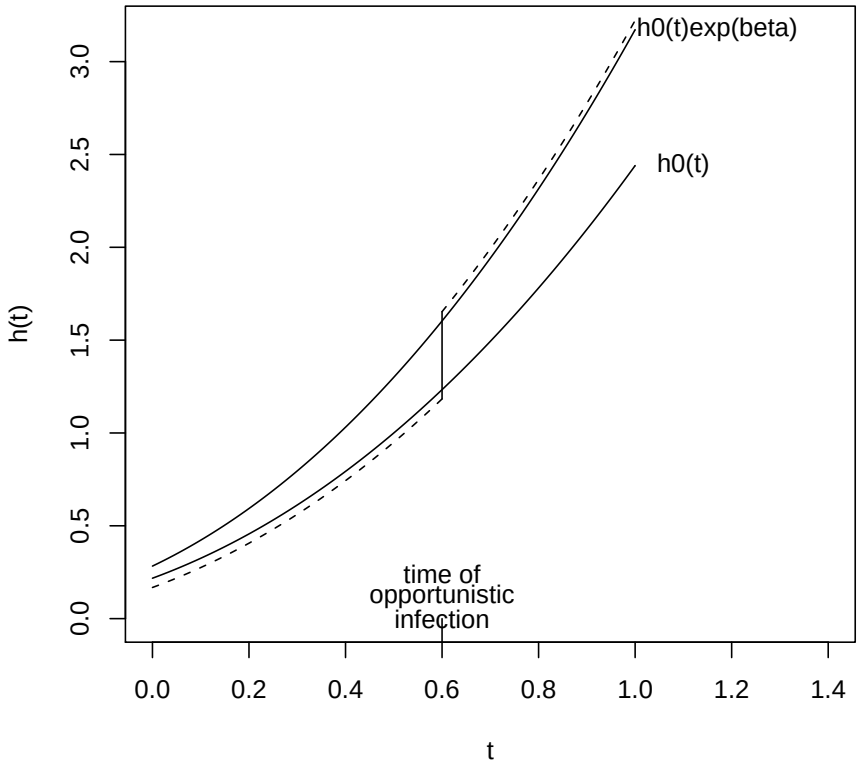
$\longrightarrow$ on a given day (t), the hazard function for an asthma attack is

$$\underbrace{h_0(t)}_{\text{baseline hazard}} \cdot e^{\beta \cdot (\text{pollution level on that day})}.$$

Example 9.2: T = time to death,

$$Z(t) \;=\; \begin{cases} 0 & \text{if no opportunistic infection (OI) by time } t \\ 1 & \text{if an opportunistic infection occurs on/before } t \end{cases}$$

$$(9.2) \qquad h(t|Z(t)) \;=\; h_0(t)e^{\beta Z(t)} \;=\; \begin{cases} h_0(t) & \text{if } Z(t) = 0 \\ h_0(t)e^{\beta} & \text{if } Z(t) = 1 \end{cases}$$



$\longrightarrow$ for a particular patient, his or her hazard function begins as $h_0(t)$ but "jumps" to $h_0(t)e^{\beta}$ at the random time he or she has an opportunistic infection

$\longrightarrow$ different patients will "jump" at different times, but all will begin with $h_0(t)$ and, when they jump, jump to $h_0(t)e^{\beta}$.

Example 9.3

$$Z_1 = \begin{matrix}\text{treatment} \\ \text{group}\end{matrix} = \begin{cases} 0 & \text{treatment group 0} \\ 1 & \text{treatment group 1} \end{cases}$$

$$Z_2(t) \overset{\text{def}}{=} Z_1 \cdot \ln t$$

$$h\left(t \mid Z_1,\ Z_2(t)\right) = h_0\left(t\right) e^{\beta_1 Z_1 + \beta_2 Z_2(t)}$$

$$= \begin{cases} h_0(t) & \text{treatment group 0} \\ h_0(t) e^{\beta_1 + \beta_2 \cdot \ln t} & \text{treatment group 1.} \end{cases}$$

$$\text{Treatment hazard ratio} = e^{\beta_1 + \beta_2 \cdot \ln t}.$$

$\longrightarrow$ Note that $Z_2(t)$ is not really a new covariate but a way to allow for a nonproportional hazards relationship between the 2 levels of Z_1.

NOTE:

- $\beta_2 = 0 \quad \leftrightarrow \quad \dfrac{h(t|\text{treatment 1})}{h(t|\text{treatment 0})} = \text{ind. of } t \text{ (i.e., PH)}$
- $\beta_1 = 0 \mid \beta_2 = 0 \iff$ no treatment difference, given PH
- $\beta_1 = \beta_2 = 0 \iff$ no treatment difference.

Inference with time-dependent Cox proportional hazards model: Computationally, inference for time-dependent covariates proceeds just as with fixed covariates, except that at each risk set, we evaluate the value of the time-dependent covariate for a subject at that time point. Properties of time-dependent Cox models can be more complex, as we will see later.

9.4 Estimating $H_0(\cdot)$:

- Until now, and in most applied settings, the focus is on making inferences about the regression coefficient β, as this describes association between Z and T.

- However, we sometimes are also interested in estimating $H_0(t)$. Why?
  - Learn about its shape (e.g., does $h_0(t) \uparrow$ with t?)
  - Prediction:

$$S(t|Z) = e^{-e^{\beta Z} H_0(t)}.$$

    To estimate this, also need an estimate of

$$H_0(t) = \int_0^t h_0(u) \ du.$$

  - Model testing (more later).

No unique correct way to estimate $h_0(t)$.

Most popular (proposed by Breslow):

$$\widehat{H}_0(t) = \sum_{\tau_j \leq t} \widehat{\Delta H}_0(\tau_j),$$

where

$$\widehat{\Delta H}_0(\tau_j) = \frac{d_j}{\sum_{l \epsilon R_j} e^{\widehat{\beta} Z_l}}.$$

NOTE: This is a discrete estimator.

NOTE: When $\widehat{\beta} = 0$ it reduces to $d_j / Y(\tau_j)$, as in the Nelson-Aalen estimator of $H_0(\cdot)$.

Heuristic Justification: Consider risk set, R_j, at τ_j.

For $l \epsilon R_j$, conditional probability of failing at τ_j (given not failed before) is

$$h_0(\tau_j) e^{\beta Z_l}.$$

The "average" probability of failing at τ_j for the $Y(\tau_j)$ persons at risk is

$$\frac{1}{Y(\tau_j)} \sum_{l \epsilon R_j} h_0(\tau_j) e^{\beta Z_l} = h_0(\tau_j) \cdot \frac{\sum_{l \epsilon R_j} e^{\beta Z_l}}{Y(\tau_j)}.$$

Observed proportion failing $= d_j / Y(\tau_j)$. Equating these and equating $h_0(\tau_j)$ to $\widehat{\Delta H}_0(\tau_j)$ gives:

$$\widehat{\Delta H}_0(\tau_j) \approx \; d_j \Big/ \sum_{l \epsilon R_j} e^{\beta Z_l} \; .$$

Example: $\quad n = 5$

$$\left.\begin{array}{rl} i & : \; 1 \;\; 2 \;\; 3 \;\; 4 \;\; 5 \\ U_i & : \; 2 \;\; 8 \;\; 13 \;\; 5 \;\; 1 \\ \delta_i & : \; 1 \;\; 1 \;\; 1 \;\; 0 \;\; 1 \\ Z_i & : \; 0 \;\; 0 \;\; 1 \;\; 1 \;\; 1 \end{array}\right\} \qquad \begin{array}{ll} \tau_1 = 1 & R_1 = \{1,2,3,4,5\} \\ \tau_2 = 2 & R_2 = \{1,2,3,4\} \\ \tau_3 = 8 & R_3 = \{2,3\} \\ \tau_4 = 13 & R_4 = \{3\} \end{array}$$

$$\widehat{\beta} = -.834$$

$$e^{\widehat{\beta} Z} \; = \; \begin{cases} 1 & z = 0 \\ .43 & z = 1 \end{cases} \qquad\qquad d_1 = d_2 = d_3 = d_4 = 1.$$

Thus,

$$\begin{aligned} \widehat{\Delta H}_0(\tau_1) \; &= \; \frac{1}{2(1) + 3(.43)} \; = \; .3 \\ \widehat{\Delta H}_0(\tau_2) \; &= \; \frac{1}{2(1) + 2(.43)} \; = \; .35 \\ \widehat{\Delta H}_0(\tau_3) \; &= \; \frac{1}{1 + .43} \; = \; .70 \\ \widehat{\Delta H}_0(\tau_4) \; &= \; \frac{1}{.43} \; = \; 2.33 \end{aligned}$$

$$\Longrightarrow \qquad \widehat{H}_0(t) \;=\; \begin{cases} 0 & t < 1 \\ .3 & 1 \le t < 2 \\ .65 & 2 \le t < 8 \\ 1.35 & 8 \le t < 13 \\ 3.68 & 13 \le t. \end{cases}$$

NOTE: $\widehat{H}_0(t)$, and thus $\widehat{S}(t|Z)$, is a step function.

9.5 Informative censoring

In the presence of informative censoring, the standard methods to fit a Cox proportional hazards model can be biased. There exist methods that lead to consistent, asymptotically normal estimation of the parameters in a Cox proportional hazards model in the presence of informative censoring. However, these methods are not routinely used. The methods are described in e.g. Robins (1993), Robins and Rotnitzky (1992) and Robins and Finkelstein (2000). The key element of these methods is in the use of time-dependent surrogate marker data. Their main assumption is that given the surrogate marker data up to a certain time point and the baseline covariates Z, censoring in the next small time interval is independent of the survival time T. This is called Missing At Random (MAR). In the case of non-informative censoring, these methods can be more efficient than the standard methods covered in this and the following units. We may come back to that later in the course.

Exercises

1. Verify that the stratified logrank test statistic essentially arises as a score test from Cox's stratified PL.

2. Suppose you wished to modify (9.1) on page 2 to allow the treatment effect to be different among men than among women.

   - Suggest a model to do this.
   - In your model, what hypothesis corresponds to no treatment effect?
   - What hypothesis corresponds to an equal treatment HR in men and women?

3. Consider the second example of a time-varying covariate (page 8). How might you modify the model (9.2) to allow the size of the jump to depend on the time until the opportunistic infection occurs?

4. Compute and plot $\widehat{S}(t|Z=0)$ and $\widehat{S}(t|Z=1)$ for the example on pages 11–12.

5. Consider a set of censored survival data from a randomized clinical trial for which we have information on 2 covariates: Z_1=treatment group (=0 or 1) and Z_2= disease stage at the time of randomization (=1, 2, or 3). We wish to compare the treatment groups. Three possible tests of treatment group equality are (i) the ordinary logrank test; (ii) the stratified (by disease stage) logrank

test; and (iii) the Wald test of $\beta_1 = 0$ resulting from fitting a PH model with covariates Z_1 and Z_2.

For each test, briefly indicate the types of situations where you might expect it to have better power than the other 2 tests.

6. Consider the same data setting as in the preceding problem, and let $h(t \mid Z_1, Z_2)$ denote the hazard function for someone with covariate values Z_1 and Z_2.

   (a) Propose a model for $h(t \mid Z_1, Z_2)$ which allows the magnitude of the treatment effect to vary with disease stage but which, for given Z_1, is unrestrictive about the hazards for levels of Z_2.

   (b) Given your model, describe a test for the hypothesis that the magnitude of the treatment effect does not vary with disease stage. (You need not give details.)

7. Suppose you are studying elderly married couples and wish to know whether health outcomes in a wife might affect the mortality risk of her husband. Suppose your data set arises from the prospective follow-up of couples from the age of 65 onwards who are generally healthy at the start of follow-up. For any given couple, let T denote the time (from start of follow-up) until the husband dies, and suppose T can be right- censored, so that you really observe the pair (U, δ) instead of T. During follow-up the wife could be hospitalized and/or die, and we wish to assess whether and how either of these events might be associated with T. Suppose that $Z_1(\cdot)$ denotes a covariate process for which $Z_1(t)$ equals zero if the wife is not hospitalized before time t and equals 1 if she is hospitalized on or before time t, where time is defined from the start of follow-up. Similarly, suppose that $Z_2(\cdot)$ denotes the process for which $Z_2(t)$ equals zero if the wife is alive

at time t and equals 1 if she dies on/before time t. Thus, at any time point t, the wife will be in one of 4 states: alive with no prior hospitalization, alive with a prior hospitalization, dead without a prior hospitalization, or dead with a prior hospitalization.

(a) Propose a model that allows the husband's hazard of death to depend on the wife's hospitalization/survival status. Clearly indicate how your assumed hazard function at an arbitrary time t depends on the history of the covariate processes $Z_1(\cdot)$ and $Z_2(\cdot)$ up to time t.

(b) In your proposed model, how would you express the situation that a wife's hospitalization is not associated with the husband's risk of death, but her death is associated with the husband's risk of death?

(c) Ignore, for the moment, the wife's hospitalization status and consider only the possible association between her death (that is, $Z_2(\cdot)$) and the husband's risk of death. Using 2x2 contingency table methods, as we did when introducing the logrank test, propose a statistical testing procedure (a statistic and a criteria for computing a p- value on the basis of this statistic) for assessing the association between $Z_2(\cdot)$ and T. Do not try to formally justify the test, but say in words why you think it would be reasonable.

(d) Next suppose that you again wanted to assess the association between a wife's death and the husband's risk of death, but that now you wanted to adjust for whether the wife was or was not hospitalized (as this might also be associated with the husband's failure). Describe how you might modify the approach in part (c) to do this.

References

The following references can be found on
http://www.hsph.harvard.edu/causal/publications.htm.

Robins, J.M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. Proceedings of the Biopharmaceutical Section, American Statistical Association: 24-33.

Robins, J.M. and Finkelstein, D.M. (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics 56: 779-788.

Robins, J.M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In: AIDS Epidemiology - Methodological Issues: 297-331. Eds: Jewell N, Dietz K, Farewell V. Birkhauser: Boston, MA.