A Review of Time-to-Event Data Methods

Thomas Braun

Department of Biostatistics
University of Michigan School of Public Health
BIOSTAT 699: Design and Analysis of Biostatistical Investigations

Introduction to Survival Data

- Commonly used names:
  - Survival data
  - Censored data
  - Time-to-event data
  - Failure time data

- Numerous applications in biomedicine as well as engineering

- Outcome of interest: T, which is the span of time from entry into a study until the occurrence of some event (death, disease recurrence, implant failure, stroke, etc.)

Introduction to Survival Data

- T is a continuous random variable, but is restricted to be positive
  - Assuming normality of T seems implausible
  - Inference about means seems less useful
- More often we are interested in (cumulative) probabilities, i.e.
  - what is the probability that a cancer patient relapses within one year of chemotherapy?
  - does this probability vary by the stage of cancer?

Introduction to Survival Data

- In order to compute cumulative probabilities, we need to estimate the entire distribution of T
- Parametric approaches:
  - Exponential
  - Weibull
  - Log-normal
- Non-parametric approaches:
  - No assumed form for distribution of T
  - Most common approach

Introduction to Survival Data

- There are five functions that characterize a survival distribution:

  (1) Cumulative distribution function (CDF)

$$\begin{aligned} F(t) &= \text{Pr}(T \leq t) \\ &= \text{probability that event occurs by time } t \end{aligned}$$

  (2) Probability density function (PDF)

$$\begin{aligned} f(t) &= \lim_{\Delta \to 0} \frac{F(t+\Delta) - F(t)}{\Delta} \\ &= \text{instantaneous probability that event} \\ & \text{occurs at time } t \end{aligned}$$

Introduction to Survival Data

(3) Survival function

$$
\begin{aligned}
S(t) &= \text{Pr}(T > t) \\
&= 1 - \text{Pr}(T \leq t) \\
&= 1 - F(t) \\
&= \text{probability of no event by time } t
\end{aligned}
$$

(4) Hazard function (force of mortality)

$$
\begin{aligned}
\lambda(t) &= \lim_{\Delta \to 0} \frac{\text{Pr}(t \leq T < t + \Delta \mid T \geq t)}{\Delta} \\
&= f(t)/S(t) \\
&= \text{instantaeous event rate at time } t \\
& \text{given event-free up to time } t
\end{aligned}
$$

Introduction to Survival Data

(5) Cumulative hazard function

$$\Lambda(t) = \int_0^t \lambda(s)ds$$

- If we know any one of $f(t)$, $F(t)$, $S(t)$, $\lambda(t)$ or $\Lambda(t)$, we know the other four
- We are most interested in $S(t)$, which tells us the probability that a subject makes it to time t without the event

Introduction to Survival Data

- An important relationship is

$$\begin{aligned}
-\frac{d}{dt} log\{S(t)\} &= \frac{dS(t)/dt}{S(t)} \\
&= \frac{f(t)}{S(t)} \\
&= \lambda(t)
\end{aligned}$$

  so that

$$\begin{aligned}
log\{S(t)\} &= -\int_0^t \lambda(u)du \\
S(t) &= exp\{-\Lambda(u)du\}
\end{aligned}$$

- Thus, regression methods for survival data focus upon modeling $\lambda(t)$ as a function of covariates and then getting $S(t)$ indirectly from $\lambda(t)$

Introduction to Survival Data

- The major challenge with time-to-event outcomes is that many subjects will be not be followed long enough to observe when the outcome occurred
  - We have missing data!
  - We refer to missing outcomes as being (right) censored
  - For these subjects, we do not have an event time, but have an amount of time followed without the event
- With censoring, we make a crucial assumption that censoring is independent of the times-to-event, conditional on covariates
  - In other words, censored individuals are representative of individuals still under observation at the same time
  - Non-independent censoring can lead to severe biases, but it is difficult in most situations to gauge the magnitude or direction of the biases
    - We need to model the censoring distribution - do we have the right model?

Estimating the Survival Function $S(t)$

Parametric Approach

- Each subject $i = 1, 2, \ldots n$ has two data points:

$$\begin{array}{rcl} t_i & = & \text{amount of time followed} \\ \delta_i & = & \begin{cases} 0 & \text{if no event at } t_i \text{ (censored)} \\ 1 & \text{if event at } t_i \text{ (not censored)} \end{cases} \end{array}$$

- Suppose we assume event times have a Weibull distribution:

$$\begin{array}{rcl} \lambda(t;\theta,\alpha) & = & \alpha\theta(\theta t)^{\alpha-1} \\ S(t;\theta,\alpha) & = & exp\{-(\theta t)^{\alpha}\} \end{array}$$

- We can estimate θ and α via maximum likelihood; our likelihood is:

$$L(\theta,\alpha|t_1,\ldots,t_n;\delta_1,\ldots,\delta_n) = \prod_i \{f(t_i;\theta,\alpha)^{\delta_i} \times S(t_i;\theta,\alpha)^{1-\delta_i}\}$$

Non-Parametric Approach

- To motivate this approach, we start with a simple set of data:
  - We have four subjects whose cancer returned 10, 13, 14, and 23 weeks, respectively after treatment
  - We have one subject who was cancer-free at 14 weeks and then was lost to follow-up

- The first step is to identify and order the *unique* times among the subjects with an event
  - For our example, there are $J = 4$ values: $t_1 = 10$, $t_2 = 13$, $t_3 = 14$, and $t_4 = 23$
  - These times define $J = 4$ non-overlapping intervals:

$$[10, 13), [13, 14), [14, 23), [23, \infty]$$

  - For each interval, we compute p_j, the probability of surviving through the entire interval, given being in the study (at risk) at the beginning at the interval

Kaplan-Meier Estimate of $S(t)$

- For the j^{th} interval, $j = 1, 2, \ldots J$, we have:

$$\begin{aligned} p_j &= 1 - \frac{\text{\# of events at } t_j}{\text{\# of subjects at risk just prior to } t_j} \\ &= 1 - \frac{d_j}{n_j} \\ &= \frac{n_j - d_j}{n_j} \\ &= \frac{s_j}{n_j} \end{aligned}$$

- Note that if any subject is censored at the same time when an event occurs, we assume the censoring occurs *after* the event

Kaplan-Meier Estimate of $S(t)$

- We assume that each of these intervals is independent of the others

- Thus, the probability of surviving to the end of interval j^* is simply the product of the probability of all intervals prior to and including interval j^*

- For example:

$$\begin{aligned}
Prob(\text{surviving to end of third interval} &= \\
Prob(\text{surviving first interval}) &\times \\
Prob(\text{surviving second interval}) &\times \\
Prob(\text{surviving third interval})&
\end{aligned}$$

Kaplan-Meier Estimate of $S(t)$

- This concept defines the (Kaplan-Meier) KM estimate of $S(t)$:

$$\widehat{S}(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{n_j}\right) = \prod_{j:t_j \leq t} \frac{n_j - d_j}{n_j} = \prod_{j:t_j \leq t} \frac{s_j}{n_j}$$

- In words, the Kaplan-Meier estimate of survival to time t is the product of surviving each interval that occurs before or includes t

- For our example, the estimated survival to $t = 20$ would be the product of surviving the intervals $[10, 13)$, $[13, 14)$, and $[14, 23)$

Kaplan-Meier Estimate of $S(t)$

- For our example, we rewrite the data as:

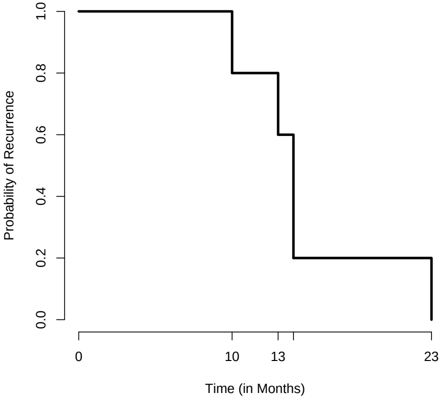
	Event	Number	Number	Number of
Index	Time	at Risk	of Events	Non-Events
j	t_j	n_j	d_j	s_j
1	10	5	1	4
2	13	4	1	3
3	14	3	1	2
4	23	1	1	0

- Using this table, we can compute the KM estimate:

For t in	$\widehat{S}(t)$
$[0, 10)$	1
$[10, 13)$	$1 \times 4/5 = 0.8$
$[13, 14)$	$1 \times 4/5 \times 3/4 = 0.6$
$[14, 23)$	$1 \times 4/5 \times 3/4 \times 2/3 = 0.4$
$[23, \infty)$	$1 \times 4/5 \times 3/4 \times 2/3 \times 0/1 = 0.0$

Kaplan-Meier Plot

- These estimates of survival are displayed in a Kaplan-Meier plot:



Inference with KM Estimates

- The variance for $\widehat{S}(t)$ was derived by Greenwood:

$$\widehat{Var}\{\widehat{S}(t)\} = \widehat{S}(t)^2 \left\{ \sum_{t_j : t_j \leq t} \frac{d_j}{n_j(n_j - d_j)} \right\}$$

- However, the standard 95% CI for $S(t)$

$$\widehat{S}(t) \pm 1.96 \sqrt{\widehat{Var}\{\widehat{S}(t)\}}$$

  is not restricted to have values inside $[0, 1]$.

Inference with KM Estimates

- An alternative approach:
  - Transform $S(t)$ to $\log[-\log\{S(t)\}]$, which ranges over $(-\infty, \infty)$
  - Compute CI for $\log[-\log\{S(t)\}]$
  - Transform back to find CI for $S(t)$
    - This confidence interval will be guaranteed to lie in $[0, 1]$
- This is similar to what we do when finding a confidence interval for an odds ratio or rate ratio:
  - We first found a CI for the log-odds and then transformed back

Inference with KM Estimates

- The variance of $\log[-\log\{\widehat{S}(t)\}]$ is

$$\widehat{Var}\left(\log[-\log\{\widehat{S}(t)\}]\right) = \frac{1}{\log \widehat{S}(t)}\left\{\sum_{t_j:t_j\leq t}\frac{d_j}{n_j(n_j-d_j)}\right\}$$

- A 95% CI for $\log[-\log S(t)]$ is

$$\log[-\log\{\widehat{S}(t)\}] \pm 1.96\sqrt{\widehat{Var}\left(\log[-\log\{\widehat{S}(t)\}]\right)}$$

- Thus, a 95% CI for $S(t)$ is

$$\widehat{S}(t) \times \exp\left\{1.96\sqrt{\widehat{Var}\left(\log[-\log\{\widehat{S}(t)\}]\right)}\right\}$$

Inference with KM Estimates

- For the example, we have:

Time	$\widehat{S}(t)$	$SE\{\widehat{S}(t)\}$	95% Conf Int	
			Lower Bound	Upper Bound
10	0.80	0.18	0.52	1.00
13	0.60	0.22	0.29	1.00
14	0.40	0.22	0.14	1.00
23	0.00	n/a	n/a	n/a

Estimating the Hazard Function

- Recall our definition of the hazard:

$$\lambda(t) = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathsf{Pr}(t \leq T \leq t + \Delta \mid T \geq t).$$

- For t in $[t_j, t_{j+1})$, we estimate $\lambda(t)$ as

$$\widehat{\lambda}(t) = \frac{d_j / n_j}{t_{j+1} - t_j},$$

  which takes the probability of an event in the whole interval (d_j / n_j) and derives the probability of an event per unit time

Estimating the Hazard Function

- Using our data:

Index	Event Time	Number at Risk	Number of Events	Number of Non-Events
j	t_j	n_j	d_j	s_j
1	10	5	1	4
2	13	4	1	3
3	14	3	1	2
4	23	1	1	0

we estimate the hazard to be:

For t in	$\widehat{\lambda}(t)$
$[0, 10)$	0
$[10, 13)$	$\frac{1/5}{13-10} = 1/15$
$[13, 14)$	$\frac{1/4}{14-13} = 1/4$
$[14, 23)$	$\frac{1/3}{23-14} = 1/27$
$[23, \infty)$	n/a

Estimating the Cumulative Hazard Function

- Recall our definition of the CHF:

$$\Lambda(t) = \int_0^t \lambda(s)ds$$

- Thus, we define our estimate of $\Lambda(t)$ to be

$$\begin{aligned}\widehat{\Lambda}(t) &= \sum_{t_j:t_j\leq t} \widehat{\lambda}(t_j)(t_{j+1}-t_j) \\ &= \sum_{t_j:t_j\leq t} \frac{d_j}{n_j(t_{j+1}-t_j)}(t_{j+1}-t_j) \\ &= \sum_{t_j:t_j\leq t} \frac{d_j}{n_j}\end{aligned}$$

This is known as the Nelson-Aalen cumulative hazard estimator

Using Nelson-Aalen to Estimate $S(t)$

- Recall that $S(t) = e^{-\Lambda(t)}$
- Using the Nelson-Aalen estimate of $\Lambda(t)$, we have another estimate (Breslow) of survival:

$$\widehat{S}(t) = exp\left\{ -\sum_{t_j : t_j \leq t} \frac{d_j}{n_j} \right\}$$

- Compare this to the Kaplan-Meier estimate:

$$\widehat{S}(t) = \prod_{j : t_j \leq t} \frac{s_j}{n_j} = \prod_{j : t_j \leq t} \left(1 - \frac{d_j}{n_j} \right)$$

Using Nelson-Aalen to Estimate $S(t)$

- Both of these estimates are quite similar because

$$exp(-x) \approx (1-x)$$

for “small” values of x

- Thus we have

$$\begin{aligned} \widehat{S}(t) &= exp\left\{-\sum_{t_j:t_j\leq t}\frac{d_j}{n_j}\right\} = \prod_{j:t_j\leq t} exp\left\{-\frac{d_j}{n_j}\right\} \\ &\approx \prod_{j:t_j\leq t}\left(1-\frac{d_j}{n_j}\right) = \prod_{j:t_j\leq t}\frac{s_j}{n_j} \end{aligned}$$

Comparing Survival Functions

Comparing Two Survival Functions

- Suppose we have follow-up data from two groups
- Group 1 consists of n_1 subjects; Group 2 consists of n_2 subjects
- For subject $i = 1, \ldots, n_1$ in Group 1, we have:

$$\begin{aligned} Y_i &= \text{length of follow-up for subject } i \\ \delta_i &= \begin{cases} 1 & \text{if had event at } Y_i \\ 0 & \text{if censored at } Y_i \end{cases} \end{aligned}$$

Comparing Two Survival Functions

- For subject $j = 1, \ldots, n_2$ in Group 2, we have:

$$\begin{aligned} Y_j &= \text{length of follow-up for subject } j \\ \delta_j &= \begin{cases} 1 & \text{if had event at } Y_j \\ 0 & \text{if censored at } Y_j \end{cases} \end{aligned}$$

- Our goal is to compare $S_1(t)$ to $S_2(t)$ and to formally test

$$H_0 : S_1(t) = S_2(t)$$

versus

$$H_a : S_1(t) \neq S_2(t)$$

$S_1(t)$ and $S_2(t)$ at a Specific t

- We are only interested in comparing survival probabilities at a single time (e.g. $t = 5$ years)
- We look at $\widehat{S}_1(t) - \widehat{S}_2(t)$ at that single time, divide by a standard error estimate and compare to 1.96
- This approach does not allow us to compare whether the two *entire* survival curves are the same

Comparing $S_1(t)$ and $S_2(t)$ at Every t

- Recall that if $S_1(t) \neq S_2(t)$, then $\Lambda_1(t) \neq \Lambda_2(t)$, which implies $\lambda_1(t) \neq \lambda_2(t)$
  - Comparing survival functions is the same as comparing hazard functions
- Thus, our hypotheses are equivalently

$$H_0 : \lambda_1(t) = \lambda_2(t)$$

versus

$$H_a : \lambda_1(t) \neq \lambda_2(t)$$

- This alternative hypothesis is too vague to be tested

Proportional Hazards Model

- We choose to use the specific alternative hypothesis:

  $$H_a : \lambda_1(t) = c\lambda_2(t),$$

  which is known as the assumption of proportional hazards

- The alternative hypothesis states that the two hazard functions are not equal and that

  $$\frac{\lambda_1(t)}{\lambda_2(t)} \equiv c$$

  at every time t, i.e. the hazard functions of the two groups are proportional to each other

Proportional Hazards Model

- Thus, we can rewrite our hypotheses again as

  $$H_0 : c = 1$$

  versus

  $$H_a : c \neq 1,$$

  where c is the ratio of the two hazards (hazard ratio)

- Notice the similarity between these hypotheses and those used with binary outcomes

  $$H_0 : OR = 1$$

  versus

  $$H_a : OR \neq 1,$$

  where OR is a ratio of odds rather than hazards

Proportional Hazards Model

- Note that the assumption of proportional hazards does not mean the survival functions are proportional to each other, i.e.

$$\lambda_1(t) = c\lambda_2(t) \quad \Rightarrow \quad S_1(t) = \{S_2(t)\}^c$$

- For example, if one hazard is 50% of the other hazard, one survival function is the square root of the other survival function

- The important property of proportional hazards is that the resulting survival functions *never* intersect each other, i.e. one survival curve is *always* higher than the other.

Two-sample Log-rank Test

- To test a difference in hazard (survival) functions, we use a log-rank test
- The construction of the test statistic is based upon a series of 2×2 tables like those used in a chi-squared test of association
- We first pool both groups together, then identify and order the unique event times (those of non-censored subjects)
    - We label these times $t_1 < t_2 < \cdots < t_J$

Two-sample Log-rank Test

- At each time t_j $(j = 1, \ldots, J)$, we create a 2×2 table

	Group		
	1	2	
# events at t_j	d_{1j}	d_{2j}	d_j
# at risk beyond t_j	s_{1j}	s_{2j}	s_j
Total	n_{1j}	n_{2j}	n_j

- If the event rate (hazard) is the same in the two groups:
  - The total number of events at t_j (d_j) should be divided equally among the two groups in relation to the number at risk at t_j in each group (n_{1j} and n_{2j})
    - Allocate $d_j(n_{1j}/n_j)$ events to Group 1
    - Allocate $d_j(n_{1j}/n_j)$ events to Group 1

Two-sample Log-rank Test

- We then focus upon one of the groups (we'll use Group 1) and express the values in terms we used with the chi-squared test of association:
  - Observed events in Group 1 at t_j:

$$O_j = d_{1j}$$

  - Expected events in Group 1 if H_0 is true:

$$E_j = \frac{n_{1j} d_j}{n_j}$$

- The squared difference $(O_j - E_j)^2$ tells us how valid the null hypothesis is at t_j
  - The bigger the difference, the more likely H_0 is false

Two-sample Log-rank Test

- Forming the typical Pearson statistic for t_j:

$$\frac{(O_j - E_j)^2}{E_j},$$

  we combine the results for all event times $t_1 < t_2 < \cdots < t_J$ into a single statistic:

$$\chi^2_L = \sum_{j=1}^{J} \frac{(O_j - E_j)^2}{E_j}$$

- However, we have one problem:
  - This formula assumes the number of events at each t_j are independent of each other (i.e. the 2×2 tables are independent)

Two-sample Log-rank Test

- Although slight, there is some dependence between the 2×2 tables, as the number of events at one time point restricts how many events can happen later
  - Thus, the statistic just shown is only an *approximation* for the actual statistic

- The log-rank statistic used by all statistical packages is

$$\chi_L^2 = \frac{[\sum_{j=1}^{J}(O_j - E_j)]^2}{\sum_{j=1}^{J} V_j},$$

where

$$V_j = \frac{n_{1j} n_{2j} d_j s_j}{n_j^2 (n_j - 1)}$$

Two-sample Log-rank Test

- Under H_0 (no difference in survival functions), χ_L has approximately a chi-squared distribution with 1 df
    - Thus a value of $\chi_L \geq 4$ is evidence to reject H_0 (gives p-value less than 0.05)
- These concepts can be extended to comparisons of G groups ($G > 2$)
    - In general, χ_L has a chi-squared distribution with $(G - 1)$ df

Two-sample Log-rank Test

- Suppose we have the following data:

Group 1	Y	4	10	15	16
	δ	1	0	1	0
Group 2	Y	7	11	19	22
	δ	1	0	0	1

- For this data, we have:

$$\begin{aligned}\sum_{j=1}^{J}(O_j - E_j) &= \left(1 - \frac{4\times 1}{8}\right) + \left(0 - \frac{3\times 1}{7}\right) \\ &\quad + \left(1 - \frac{2\times 1}{4}\right) + \left(0 - \frac{0\times 1}{1}\right) \\ &= 0.5 - 0.429 + 0.5 \\ &= 0.571\end{aligned}$$

Two-sample Log-rank Test

- The denominator is computed as:

$$
\begin{aligned}
\sum_{j=1}^{J} V_j &= \frac{4 \times 4 \times 1 \times 7}{8 \times 8 \times 7} + \frac{3 \times 4 \times 1 \times 6}{7 \times 7 \times 6} + \frac{2 \times 2 \times 1 \times 3}{4 \times 4 \times 3} \\
&= 0.25 + 0.; 2449 + 0.25 \\
&= 0.7449
\end{aligned}
$$

- Therefore, we find:

$$
\chi^2_L = \frac{[\sum_{j=1}^{J} (O_j - E_j)]^2}{\sum_{j=1}^{J} V_j} = \frac{0.571^2}{0.7449} = 0.438,
$$

yielding a p-value of 0.51 based on a $\chi^2_{(1)}$ distribution

Visually Displaying Results

- This is an example of an excellent Kaplan-Meier plot:

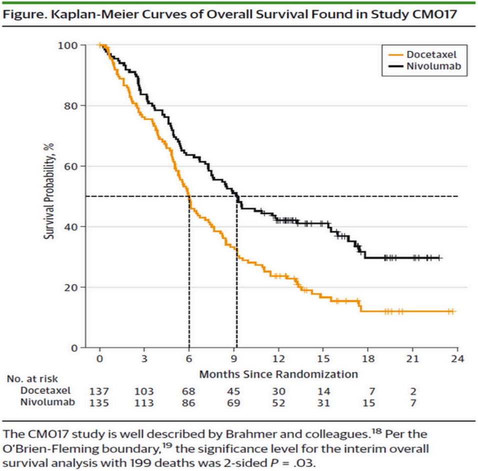


Figure. Kaplan-Meier Curves of Overall Survival Found in Study CMO17

No. at risk	0	3	6	9	12	15	18	21
Docetaxel	137	103	68	45	30	14	7	2
Nivolumab	135	113	86	69	52	31	15	7

The CMO17 study is well described by Brahmer and colleagues.[18] Per the O'Brien-Fleming boundary,[19] the significance level for the interim overall survival analysis with 199 deaths was 2-sided *P* = .03.

Assessing Proportional Hazards

- We can visually assess whether proportional hazards is a reasonable assumption
- If we plot $\log\{-\log \widehat{S}_1(t)\}$ and $\log\{-\log \widehat{S}_2(t)\}$ against $\log(t)$ for each group and proportional hazards holds, the two lines should be roughly parallel to each other
- Why?

$$\begin{aligned}
\lambda_1(t) = c\lambda_0(t) \quad &\Rightarrow \quad \Lambda_1(t) = c\Lambda_0(t) \\
&\Rightarrow \quad -\log S_1(t) = c[-\log S_0(t)] \\
&\Rightarrow \quad \log\{-\log S_1(t)\} \\
&\qquad\quad = \log c + \log\{-\log S_0(t)\}
\end{aligned}$$

Non-proportional Hazards

- If the hazards are not proportional, we use a weighted log-rank statistic:

  (1) Wilcoxon Test

  $$\chi^2_W = \frac{[\sum_{j=1}^{J} n_j(O_j - E_j)]^2}{\sum_{j=1}^{J} n_j^2 V_j}$$

  - This test gives more weight to early event times (when n_j is big) and less weight to late event times (when n_j is small)

  (2) Generalized Wilcoxon (GW) Test

  $$\chi^2_{GW} = \frac{[\sum_{j=1}^{J} w_j(O_j - E_j)]^2}{\sum_{j=1}^{J} w_j^2 V_j}$$

  - The weights can be chosen to emphasize a particular time or range of times and is most powerful under certain situations

- Note that if $w_j = 1$ for all j, GW test reduces to the log-rank test

Regression Models for Comparing Survival Functions

Proportional Hazards Regression (Non-parametric)

- **Outcome variable**: (T_i, δ_i) $(i = 1, \ldots, n)$, where T_i is the observed survival time for the ith individual, and

$$\delta_i = \begin{cases} 1 & \text{event observed} \\ 0 & \text{event censored} \end{cases}$$

- **Covariates**: $X_{1i}, X_{2i}, \ldots, X_{pi}$ for the ith subject $(i = 1, \ldots, n)$
- **Idea**: Model the hazard function of the event at a particular time t as a function of covariates

Proportional Hazards Regression

- We use the typical regression model

$$\begin{aligned}\log \lambda(t) &= \log \lambda_0(t) + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \\ &= \log \lambda_0(t) + \sum_{j=1}^{p} \beta_j X_j\end{aligned}$$

- However, the model is more commonly written as

$$\begin{aligned}\lambda(t) &= \lambda_0(t) e^{\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p} \\ &= \lambda_0(t) e^{\sum_{j=1}^{p} \beta_j X_j}\end{aligned}$$

and is called a Cox regression model

Proportional Hazards Regression

- In the Cox regression model $\lambda(t) = \lambda_0(t)e^{\sum_{j=1}^{p} \beta_j X_j}$:
  - $\lambda_0(t)$ is known as the baseline hazard, which measures the risk of an event for the reference group, i.e. subjects with all covariates equal to zero
  - We do not make any assumption on the actual functional form of $\lambda_0(t)$
  - There is no intercept term β_0 in the exponent
    - Our baseline hazard serves as our intercept (on the log scale)
  - The hazard ratio is $\lambda(t)/\lambda_0(t) = exp\left\{\sum_{j=1}^{p} \beta_j X_j\right\}$
    - A covariate works to proportionally increase the hazard in reference to the baseline hazard
    - This proportion is constant over time, meaning the hazard ratio is constant over time

Proportional Hazards Regression

- Deriving the interpretation of β_k:
  - We compare two randomly chosen individuals
    - One subject has covariate values $(X_1, \ldots, X_k, \ldots, X_p)$
    - One subject has covariate values $(X'_1, \ldots, X'_k + 1, \ldots, X'_p)$
  - Then we have:

$$\frac{\lambda(t \mid X_1, \ldots, X_k + 1, \ldots, X_p)}{\lambda(t \mid X_1, \ldots, X_k, \ldots, X_p)} =$$

$$\frac{\lambda_0(t) exp^{\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k (X_k + 1) + \cdots + \beta_p X_p}}{\lambda_0(t) e^{\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k (X_k) + \cdots + \beta_p X_p}} = e^{\beta_k}$$

  - Again, the association fof X_k with the time-to-event is assumed to be constant over time

Proportional Hazards Regression: Example

- We have data from a 10-year double-blinded trial in 312 patients with primary cirrhosis of the liver (PBC) who were randomized to either the drug D-penicillamine (DPCA) or placebo
- Available covariates: age, albumin, bilirubin, edema, prothrombin time
- Our current example focuses on a model with variables age, edema and drug (1: DPCA, 0: placebo), in which

$$\text{edema} = \begin{cases} 0 & \text{no edema and no diuretic therapy} \\ 1/2 & \text{edema present w/o or resolved by diuretics} \\ 1 & \text{edema despite diuretic therapy} \end{cases}$$

Partial Likelihood

- Recall that Kaplan-Meier first divided the data by the unique event times
  - Used product of interval probabilities to estimate the distribution of event times
- Estimation of regression parameters β in Cox regression is done in a similar way
- Cox referred to his approach as “partial likelihood” in 1975; the rigorous theory came in 1981 (Tsiatis) and 1982 (Anderson & Gill)
  - Because the baseline hazard is not specified and is nuisance, we would like to remove it when estimating β
  - We condition the risk of each subject with an event (based on their covariate) on the total risk all subjects

Partial Likelihood

- Thus, instead of the exact likelihood, we attempt to maximize the partial likelihood:

$$PL(\boldsymbol{\beta};\boldsymbol{X}) = \prod_{t_k=t_1}^{t_d} \frac{exp\{\boldsymbol{X}(t_k)\boldsymbol{\beta}\}}{\sum_{i\in R_k} exp(\boldsymbol{X_i}\boldsymbol{\beta})}$$

in which

$$\begin{aligned} t_1,\ldots,t_d &= \text{unique event times} \\ \boldsymbol{X}(t_k) &= \text{covariate vector for subject with event at } t_k \\ R_k &= \text{group of subjects at risk for event from } (t_{k-1},t_k] \end{aligned}$$

Proportional Hazards Regression: Example

- The resulting fitted model from SAS (code not shown) is:

$$\lambda(t) = \lambda_0(t) exp\{0.035\text{age} + 2.23\text{edema} - 0.11\text{drug}\}$$

- $e^{-0.11} = 0.89$ is the HR comparing DPCA to placebo adjusted for age and edema status
- The estimated HR comparing a patient with age $= 50$, edema $= 0.5$, drug $= 1$ to a patient with age $= 40$, edema $= 0$ and drug $= 1$

$$\begin{aligned} &= e^{0.035(50-40)+2.23(0.5-0)-0.11(1-1)} \\ &= 4.33 \end{aligned}$$

Assessing Proportional Hazards for One Covariate

- We have assumed that a one-unit change in a covariate X leads to a shift in the log-hazard, i.e. hazards are proportional

- If X is categorical, then we plot $\log\{-\log \widehat{S}(t)\}$ against $\log(t)$ for each value of X
  - Proportional hazards holds if the curves are roughly parallel to each other

- If X is continuous, then divide subjects into (four, five?) equally sized groups and fit model using categorical X
  - Use same process as above to assess proportional hazards for X

- Assessing PH gets harder with multiple covariates

Estimating the Baseline Survival Function

- Recall that our model is

$$\begin{aligned}\lambda(t) &= \lambda_0(t)e^{\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p}\\ &= \lambda_0(t)e^{\sum_{j=1}^{p}\beta_j X_j}\end{aligned}$$

  and we have made no attempt to estimate $\lambda_0(t)$

- Thus, we can compare the risk of two subjects relative to each other (hazard ratio)

- We cannot estimate the individual risks (hazard) of each subject

- There are methods for estimating $\lambda_0(t)$; most statistical packages are programmed with these methods

Time-Varying Covariates

- Cox regression allows using covariates that change over time:

$$\lambda(t) = \lambda_0(t) exp\left\{\sum_{j=1}^{p} \beta_j X_j(t)\right\}$$

- But, this impacts our estimation:

$$PL(\boldsymbol{\beta}; \boldsymbol{X}) = \prod_{t_k=t_1}^{t_d} \frac{exp\{\boldsymbol{X}(t_k)\boldsymbol{\beta}\}}{\sum_{i \in R_k} exp(\boldsymbol{X_i}(\boldsymbol{t_k})\boldsymbol{\beta})}$$

  - At each event time t_k, we need covariate values for every subject in the risk set R_k, not just the subject with the event
  - If missing, some suggest using $X_i(t_k)$ as the covariate value measured closest in time to t_k (if reasonably close in time to t_k)

Accelerated Failure Time Model

- A contemporary alternative to Cox regression is a parametric regression model known as the Accelerated Failure Time (AFT) model
  - We assume a baseline distribution $\mathcal{F}_0$ for event times
  - For a subject with covariates $\boldsymbol{X}$, their observed event time is $T = e^{\boldsymbol{\beta X}} T_0$, where $T_0 \sim \mathcal{F}_0$
  - This is a simple parametric regression model

    $$log(T) = \boldsymbol{\beta X} + \epsilon,$$

    with $\epsilon \sim \mathcal{F}_0$
  - $\mathcal{F}_0$ is often parametric, i.e. Normal or Logistic, with mean=0 and variance=σ^2
  - $\mathcal{F}_0$ can be non-parametric - much harder problem to solve

Residual Diagnostics with Cox Regression

- See supplementary slides based on lecture by P. Breheny