Survival Analysis: Semiparametric Models

Samiran Sinha

Texas A&M University
sinha@stat.tamu.edu

October 26, 2021

Introduction

- When there is no covariate, or interest is focused on a homogeneous group of subjects, then we can use a nonparametric method of analyzing time-to-event data.

- When there are two or more treatment groups, and each group has *sufficient* number of subjects, then also we can use a nonparametric method of analysis. The advantage of the nonparametric methods is that we do not impose any condition on the behavior of the time-to-event.

- In the presence of several covariates (potential predictors), we consider a parametric method of analysis, and see how the time-to-event is associated with the covariates. In the parametric approaches considered, we modeled the mean of the logarithm of the time-to-event as a linear function of the predictors (AFT model). Specifically, in the AFT model we assume that

  - $\log(T) = X^T \beta + U$
  - The mean of $\log(T)$ was $X^T\beta+$constant
  - We put some distributional assumption on U, that helped us to obtain the density of T and the survival function of T

If we think carefully, in the AFT model we impose the restriction that

- the predictors influence only the mean of $\log(T)$,
- the noise term U is independent of the predictors,
- by assigning a distribution on U, we dictate a particular type of shape to the distribution of T.

- In the nonparametric modelling we do not impose any such restriction. However, we also need to note that in the presence of several predictors, non-parametric modelling is not feasible.

- In this class note we shall talk about a strategy of modelling the effect of a number of explanatory variables on the time-to-event T. This strategy is slightly different from the AFT models and the completely nonparametric approaches discussed previously.

- Alternative to modelling the mean of $\log(T)$ in terms of the predictors, we can model the hazard function in terms of the potential predictors.

- We know that once the hazard function $\lambda(t|X)$ is specified, from there we can obtain the cumulative hazard $\Lambda(t|X)$ and thereby obtain the survival function $S(t|X)$ and the density function $f(t|X)$. In other words,

$$\lambda(t|X) \longrightarrow \Lambda(t|X) \longrightarrow S(t|X) \longrightarrow f(t|X)$$

- Once we know $S(t|X)$ and $f(t|X)$, we can write out the likelihood function that can be used to estimate the model parameters.

Proportional hazard

- In particular, consider this model:

$$\lambda(t|X) = \lambda_0(t) r(X'\beta)$$

- Here $\lambda_0(t) \geq 0$ is called the "baseline" hazard, which describes how the hazard changes with time.

- And $r(X'\beta)$ describes how the hazard changes as a function of the covariates X. Here X does not include any intercept term.

- Cox (1972) proposed $r(X'\beta) = \exp(X'\beta)$, resulting in what became called the Cox Proportional Hazards (CPH) model:

$$\lambda(t|X) = \lambda_0(t)\exp(X'\beta).$$

- In a semiparametric model, the baseline hazard $\lambda_0(t)$ is left unspecified.

Proportional hazard interpretation:

- If $X_i' = (TX_i)$, where TX_i is a binary indicator of treatment group (0 for control, 1 for treatment, say), then the <u>hazard ratio</u> between a treated and a control at time t is:

$$\frac{\lambda(t|(1))}{\lambda(t|(0))} = \exp(\beta),$$

  giving the model a "relative risk"-like interpretation.

- Note also that in the above ratio, the "baseline" hazard λ_0 get canceled.

- Importantly, the proportional hazard assumption implies that the ratio of two hazards for two different set of covariates at any given time is free from the time.

Relating survival and hazard functions

- The cumulative hazard is

$$\Lambda(t|X) = \int_0^t \lambda(u|X)du$$

$$= \int_0^t \lambda_0(u)\exp(X'\beta)du$$

$$= \{\int_0^t \lambda_0(u)du\}\exp(X'\beta)$$

$$= \Lambda_0(t)\exp(X'\beta).$$

Here $\Lambda_0(t)$ is called the baseline cumulative hazard function.

- Let's derive the survival function in this scenario

$$S(t|X) = \exp\{-\Lambda(t|X)\} = \exp\{-\Lambda_0(t)\exp(X'\beta)\}.$$

Relating survival and hazard functions

- The density function is

$$f(t|X) = -\frac{d}{dt}S(t|X)$$

$$= -\frac{d}{dt}\exp\{-\Lambda_0(t)\exp(X'\beta)\}$$

$$= \exp\{-\Lambda_0(t)\exp(X'\beta)\}\frac{d\Lambda_0(t)}{dt}\exp(X'\beta)$$

$$= \exp\{-\Lambda_0(t)\exp(X'\beta)\}\lambda_0(t)\exp(X'\beta)$$

$$= S(t|X)\lambda(t|X).$$

Cox PH model estimation

- For the observed data $(V_i, \Delta_i, X_i), i = 1, \ldots, n$, the likelihood for the Cox PH model is

$$
\begin{aligned}
L(\beta) &= \prod_{i=1}^{n} f^{\Delta_i}(V_i|X_i)\{S(V_i|X_i)\}^{1-\Delta_i} \\
&= \prod_{i=1}^{n} \{\lambda(V_i|X_i)S(V_i|X_i)\}^{\Delta_i}\{S(V_i|X_i)\}^{1-\Delta_i} \\
&= \prod_{i=1}^{n} \{\lambda(V_i|X_i)\}^{\Delta_i} S(V_i|X_i) \\
&= \prod_{i=1}^{n} \{\lambda_0(V_i)\exp(X_i'\beta)\}^{\Delta_i} \exp\{-\Lambda_0(V_i)\exp(X_i'\beta)\}
\end{aligned}
$$

- To estimate β by maximizing $L(\beta)$, one may specify a parametric form for the function $\lambda_0(\cdot)$. Once the functional form of λ_0 is specified, the model becomes a parametric model.

- In a semiparametric model (Cox PH) λ_0 is left unspecified.

Parametric form for $\lambda_0(\cdot)$

- If $\lambda_0(t) = c_0$, a constant, we obtain the exponential model discussed in the previous class notes.

- If $\lambda_0(t) = c_0 t^{c_1}$, a polynomial in t, we obtain the Weibull model discussed in the previous class notes.

Cox PH model (λ_0 is unspecified) estimation

For the semiparametric model ($\lambda(t|X) = \lambda_0(t)\exp(X'\beta)$), Cox proposed to estimate β by maximizing the "partial likelihood" function

$$L_p(\beta) = \prod_{i=1}^{n}\left\{\frac{\exp(X_i'\beta)}{\sum_{j \in R(V_i)}\exp(X_j'\beta)}\right\}^{\Delta_i},$$

- $R(V_i)$ is the "risk set" at time V_i, comprised of all individuals with survival or censoring times $\geq V_i$;
- using mathematics beyond the scope of this course, it can be shown that $\widehat{\beta}$ obtained by maximizing $L_p(\beta)$ has the same distributional properties as that obtained by maximizing $L(\beta)$;

Cox PH model estimation

To maximize $L_p(\beta)$, we first log transform $L_p(\beta)$

$$\ell_p(\beta) = \sum_{i=1}^{n} \Delta_i \left[X_i'\beta - \log\{ \sum_{j \in R(V_i)} \exp(X_j'\beta)\} \right]$$

then differentiate

$$\frac{\partial}{\partial \beta} \ell_p(\beta) = \sum_{i=1}^{n} \Delta_i \left\{ X_i - \frac{\sum_{j \in R(V_i)} X_j \exp(X_j'\beta)}{\sum_{j \in R(V_i)} \exp(X_j'\beta)} \right\},$$

and we can solve $\frac{\partial}{\partial \beta} \ell_p(\beta) = 0$ by numerical methods, to obtain $\widehat{\beta}$.

Cox PH model estimation continues...

The estimator of the baseline hazard is

$$\widehat{\lambda}_0(t) = \begin{cases} \frac{\Delta_k}{\sum_{j \in R(V_k)} \exp(X'_j \widehat{\beta})} & \text{if } t = V_k \text{ for some } k \\ 0 & \text{otherwise.} \end{cases}$$

The estimator of the cumulative baseline hazard is

$$\widehat{\Lambda}_0(t) = \int_0^t \widehat{\lambda}_0(u) du = \sum_{V_k \leq t} \frac{\Delta_k}{\sum_{j \in R(V_k)} \exp(X'_j \widehat{\beta})}.$$

The estimator of the survival function at time τ is

$$\widehat{S}(\tau | X) = \exp\{-\widehat{\Lambda}_0(\tau) \exp(X^T \widehat{\beta})\}.$$

Cox PH model standard errors

What about standard errors for $\widehat{\beta}$? We can estimate $\text{Var}(\widehat{\beta})$ by $\boldsymbol{I}^{-1}(\widehat{\beta})$, where

$$\boldsymbol{I}(\beta) = -\frac{\partial^2 \ell_p(\beta)}{\partial \beta \partial \beta'}$$

is called the "observed information matrix," and $\boldsymbol{I}(\widehat{\beta})$ is obtained by plugging $\widehat{\beta}$ in for β. Standard errors for $\widehat{\beta}$ are then the square root of the diagonal elements of $\boldsymbol{I}^{-1}(\widehat{\beta})$.

A linear model connection: Information matrix and MLEs

- In the linear regression model, $Y_i = X_i^T \beta + \epsilon_i$, $i = 1, \ldots, n$, with $(\epsilon_1, \ldots, \epsilon_n)^T \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$.

- Then $\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ and $\mathrm{Var}(\widehat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, where

$$
\mathbf{X} = \begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix}, \ \mathbf{Y} = (Y_1, \ldots, Y_n)^T.
$$

A linear model connection: Information matrix and MLEs

- We obtain these results via ML estimation.
- The log-likelihood is:

$$\ell(\beta) = \text{constant} - \frac{1}{2\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{Y} - \boldsymbol{X}\beta)$$

- Then the score function is

$$\frac{\partial}{\partial \beta}\ell(\beta) = -\frac{1}{2\sigma^2}\left(-2\boldsymbol{X}'\boldsymbol{Y} + 2\boldsymbol{X}'\boldsymbol{X}\beta\right)$$

- The Hessian matrix is

$$\frac{\partial^2}{\partial \beta \partial \beta'}\ell(\beta) = -\frac{1}{\sigma^2}(\boldsymbol{X}'\boldsymbol{X}),$$

- The observed information matrix is $\boldsymbol{I} = -\partial^2\ell(\beta)/\partial\beta\partial\beta'$ so $\text{Var}(\widehat{\beta})$ is estimated by $\widehat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

Likelihood ratio tests

- With estimates $\widehat{\beta}$, we can also carry out likelihood ratio tests as usual, but by using the partial likelihood.

- Suppose that there are two explanatory variables, X and Z, and the corresponding regression coefficients are β_1 and β_2, respectively. Let $\beta = (\beta_1^T, \beta_2^T)^T$. We are interested in testing if X has any association with the hazard of the time-to-event. Then $H_0 : \beta_1 = \mathbf{0}$ and $H_a : \beta_1 \neq \mathbf{0}$.

- The test statistic is
$$T = -2\{\log(\mathcal{L}_{p0}) - \log(\mathcal{L}_{pa})\},$$
where $\mathcal{L}_{p0}$ and $\mathcal{L}_{pa}$ are the maximized partial likelihood value under H_0 and H_a.

- When H_0 holds, T approximately follows χ_q^2, where q is the difference in the number of parameters for the unrestricted and null models.

Wald tests

An alternative test is the "Wald" test. Suppose that we are interested in testing the jth component of the β vector. Suppose that $H_0 : \beta_j = \beta_j^*$ versus $H_a : \beta_j \neq \beta_j^*$. Then the test statistic is

$$T = \frac{\widehat{\beta}_j - \beta_j^*}{se(\widehat{\beta}_j)},$$

which approximately follows $N(0, 1)$ under the null hypothesis. Note that this is essentially the t-statistic we use in the linear regression. The p-value is calculated based on the Z distribution, and use $2pr(Z > |T_{\text{obs}}|)$ as the p-value for this two-sided alternative hypothesis. Here T_{obs} denotes the observed value of the test statistic T.

Wald tests

Wald's test can be used in a more general context. Suppose that we are interested in testing $H_0 : A\beta = b$ versus $H_a : A\beta \neq b$. Then the test statistic is

$$T = (A\widehat{\beta} - b)^T \Sigma^{-1} (A\widehat{\beta} - b),$$

where $\Sigma = A\text{Var}(\widehat{\beta})A^T$. Under H_0, T approximately follows χ_q^2 with q being the rank of A.

Application to the lung cancer data

- Consider the Veteran Lung cancer data given in the `survival` package of `R`
  `https://stat.ethz.ch/R-manual/R-devel/library/survival/html/veteran.html`

- The model for the hazard is

$$
\begin{aligned}
\lambda(t|\text{predictors}) = & \ \lambda_0(t)\exp\{\beta_1 age + \beta_2 I(\text{prior therapy} = \text{Yes}) \\
& + \beta_3 I(\text{cell type} = \text{small}) + \beta_4 I(\text{cell type} = \text{adeno}) \\
& + \beta_5 I(\text{cell type} = \text{large})\}
\end{aligned}
$$

Application to the lung cancer data

Code

```
library(survival)
data(veteran)
head(veteran)
 trt celltype time status karno diagtime age prior
1   1 squamous   72      1    60        7  69     0
2   1 squamous  411      1    70        5  64    10
3   1 squamous  228      1    60        3  38     0
4   1 squamous  126      1    60        9  63    10
5   1 squamous  118      1    70       11  65    10
6   1 squamous   10      1    20        5  49     0
out=coxph(Surv(time, status)~age+as.factor(prior)+celltype,
  data=veteran)
```

Output

Code

```
 summary(out)
Call:
coxph(formula = Surv(time, status) ~ age + as.factor(prior) +
    celltype, data = veteran)
  n= 137, number of events= 128
                      coef exp(coef) se(coef)     z Pr(>|z|)
age                0.005990  1.006008 0.009367 0.639    0.523
as.factor(prior)10 0.049047  1.050269 0.205806 0.238    0.812
celltypesmallcell  0.999603  2.717202 0.256167 3.902 9.53e-05 ***
celltypeadeno      1.168623  3.217559 0.298658 3.913 9.12e-05 ***
celltypelarge      0.237791  1.268445 0.277956 0.855    0.392
---
```

Output

Code

```
                  exp(coef) exp(-coef) lower .95 upper .95
age                  1.006     0.9940    0.9877     1.025
as.factor(prior)10   1.050     0.9521    0.7016     1.572
celltypesmallcell    2.717     0.3680    1.6446     4.489
celltypeadeno        3.218     0.3108    1.7919     5.778
celltypelarge        1.268     0.7884    0.7357     2.187
Concordance= 0.612  (se = 0.03 )
Rsquare= 0.169   (max possible= 0.999 )
Likelihood ratio test= 25.31  on 5 df,   p=0.0001215
Wald test            = 24.57  on 5 df,   p=0.0001684
Score (logrank) test = 25.99  on 5 df,   p=8.974e-05
```

Output interpretation

- There were 137 observations, and out of them 9 were right censored.

- There are a total of 5 (five) regression parameters.

- The estimate of β_1 is 0.0059 with a standard error of 0.0094. The Wald test statistic for testing $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$, is $T = 0.0059/0.0094 = 0.639$. Since the p-value is 0.523, we fail to reject H_0 and conclude that the data do not provide sufficient evidence that the age has a statistically significant association with the time-to-event in the current model.

- More interpretable quantity is $\exp(\beta_1)$, often referred to as the relative risk of the disease. In other words, $\exp(\beta_1)$ can be interpreted as the risk ratio of the failure for changing age by one year. If the age has no association, then the risk ratio is one. Since the 95% CI for $\exp(\beta_1)$ (0.98, 1.02) includes one, we again conclude that the data do not provide statistical evidence that age has a statistically significant effect on the time-to-event.

Output interpretation

- By default the coxph function returns three test statistics and the corresponding p-values.

- The likelihood ratio (LR) test and the Wald test we have talked about.

- For these test the null hypothesis is $H_0 : \beta = (\beta_1, \ldots, \beta_5) = (0, \ldots, 0)$ and $H_a : \beta = (\beta_1, \ldots, \beta_5) \neq (0, \ldots, 0)$. In words, H_a says that at least one of 5 components of β is non-zero.

- For this data example, the LR and Wald test statistics are 25.31 and 24.57, respectively.

- Concordance denotes the percentage of pairs in the sample, where the observations with the higher risk score will experience the event earlier than the subject with the lower risk score. For the ith subject, by risk score we refer to $X_i' \widehat{\beta}$.

Likelihood ratio test

- Suppose that we are interested in checking if cell type has any effect on the time-to-event.
- The null hypothesis will be $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$ and H_a: at least one of $\beta_2, \beta_3, \beta_4$ is non-zero.

Code

```
out=coxph(Surv(time, status)~age+as.factor(prior)+celltype, data=veteran)
out0=coxph(Surv(time, status)~age+as.factor(prior), data=veteran)
anova(out0, out)
Analysis of Deviance Table
 Cox model: response is  Surv(time, status)
 Model 1: ~ age + as.factor(prior)
 Model 2: ~ age + as.factor(prior) + celltype
   loglik Chisq Df P(>|Chi|)
1 -504.90
2 -492.79 24.22  3 2.248e-05 ***
---
```

Likelihood ratio test

Since the p-value is 2.248e-05, we reject H_0 and conclude that cell type has a statistically significant effect at the 1% level of significance.

Estimation of $\Lambda_0(t)$

Code

```
out2=basehaz(out)
head(out2)
      hazard time
1 0.01307452    1
2 0.01964505    2
3 0.02627565    3
4 0.03297489    4
5 0.05346179    7
6 0.08180175    8

plot(out2[, 2], out2[, 1], type="s",
ylab="Baseline Cumulative Hazard", xlab="Time")
```
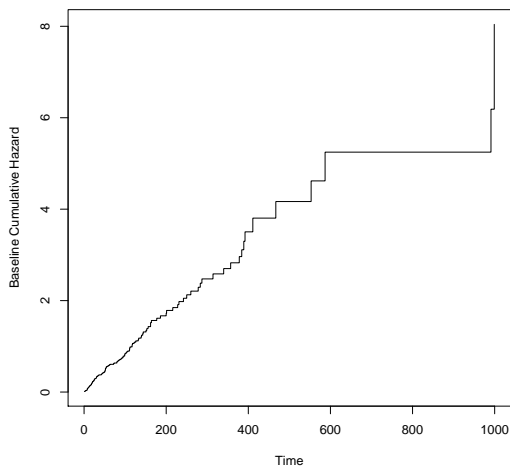
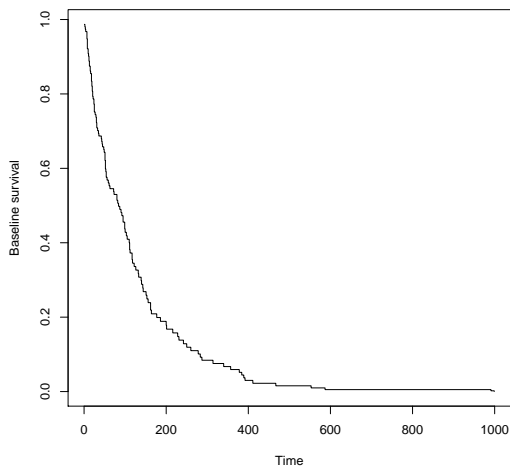# Alternative estimation of $\Lambda_0(t)$

Code

```
out3=survfit(out)
# By taking negative of log transformation of the
# survival probability
plot(out3$time, -log(out3$surv), type="s",
 ylab="Baseline Cumulative Hazard", xlab="Time")
```

Estimated baseline cumulative hazard, $\widehat{\Lambda}_0(t)$

Estimated baseline survival, $\exp\{-\widehat{\Lambda}_0(t)\}$

Code

```
out2=basehaz(out)
index1=findInterval(730, out2$time)
caplambda0=out2$hazard[index1]
```

Prediction

- Suppose that we want to predict the survival probability at time t_* for a subject with covariate X_*. Thus,

$$S(t_*|X_*) = \exp\{-\Lambda_0(t) \exp(X_*^T \beta)\}$$

- The estimator of $S(t_*|X_*)$

$$\widehat{S}(t_*|X_*) = \exp\{-\widehat{\Lambda}_0(t_*) \exp(X_*^T \widehat{\beta})\}$$

- Suppose that we want to estimate the survival probability for $t_* = 730$ days (2 years) for a subject with age 62 years, cell type squamous, and had a prior therapy.

Estimated survival function for a subject with age 62 years, cell type squamous, and had a prior therapy

Code

```
out=coxph(Surv(time, status)~age+as.factor(prior)+celltype, data=veteran)
 plot(survfit(out, newdata=data.frame(age=62, celltype="squamous",
prior=as.factor(10)) ) , ylab="Estimated survival function", xlab="Time")
```

Estimated survival function for a given covariate value
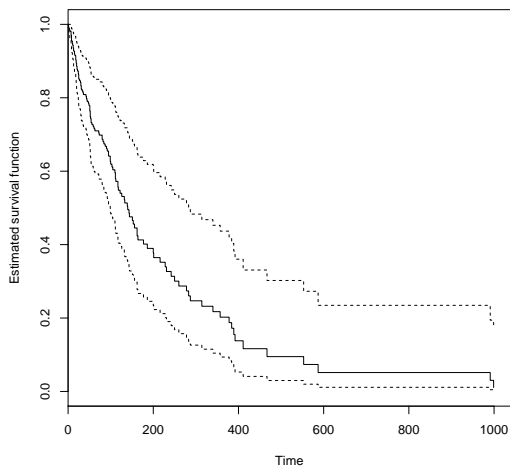
Estimated survival probability at a given time $t = 730$ days and for a given covariate value

Code

```
out=coxph(Surv(time, status)~age+as.factor(prior)+celltype, data=veteran)
out200=survfit(out, newdata=data.frame(age=62, celltype="squamous",
 prior=as.factor(10)) )
index1=findInterval(730, out200$time)
out200$surv[index1] # estimate of S(730|given the covariate value)
c(out200$lower[index1], out200$upper[index1]) # the 95% CI
```

Re-analysis of the veteran lung cancer data

In the previous analysis we treated age as a numeric variable and assumed that its effect on the hazard is in a log-linear form. How about we bin the age into different groups, and assume that the age effect is constant within a group, but varies across the groups. This approach is more general and more nonparametric than assuming a log-linear form of the effect of age. Usually, for many diseases the age effect is not always linear on the log-hazard, and in those cases it is better to use age as a categorical variable. On the other hand, we should avoid creating many categories that will result in highly variable/unreliable estimates specially when the number of observations corresponding to each category of the variable is small.

Code

```
out=coxph(Surv(time, status)~age+as.factor(prior)+celltype, data=veteran)
myage=cut(veteran$age, breaks=c(0, 51,  62, 66,  100), labels=c("A",
 "B", "C", "D"))
  out2=coxph(Surv(time, status)~myage+as.factor(prior)+celltype,
 data=veteran)
extractAIC(out)
[1]    5.0000 995.5898
extractAIC(out2)

[1]    7.0000 994.8146
```

A quick comparison of the two coxph objects

Code

```
 out
Call:
coxph(formula = Surv(time, status) ~ age + as.factor(prior) +
    celltype, data = veteran)

                     coef exp(coef) se(coef)    z       p
age               0.00599   1.00601  0.00937 0.64    0.52
as.factor(prior)10 0.04905   1.05027  0.20581 0.24    0.81
celltypesmallcell 0.99960   2.71720  0.25617 3.90 9.5e-05
celltypeadeno     1.16862   3.21756  0.29866 3.91 9.1e-05
celltypelarge     0.23779   1.26844  0.27796 0.86    0.39

Likelihood ratio test=25.31  on 5 df, p=1e-04
n= 137, number of events= 128
> out2
Call:
coxph(formula = Surv(time, status) ~ myage + as.factor(prior) +
    celltype, data = veteran)

                     coef exp(coef) se(coef)    z       p
myageB            -0.6324    0.5313   0.3524 -1.79 0.07272
myageC            -0.3089    0.7343   0.3350 -0.92 0.35644
myageD             0.4267    1.5322   0.7806  0.55 0.58459
as.factor(prior)10 0.0408    1.0416   0.2058  0.20 0.84300
celltypesmallcell  0.9903    2.6920   0.2568  3.86 0.00012
celltypeadeno      1.0927    2.9824   0.3010  3.63 0.00028
celltypelarge      0.1995    1.2208   0.2790  0.72 0.47454

Likelihood ratio test=30.08  on 7 df, p=9e-05
n= 137, number of events= 128
```

Practical application continues

- If we want to change the reference category of cell type to *adeno*, we may use the following code.

Code

```
myveteran=within(veteran, celltype<-relevel(celltype, ref="adeno"))
out3=coxph(Surv(time, status)~age+prior+celltype, data=myveteran)
```

Practical application continues

- Next look at the pbc data in the survival package of R.

- A description can be found at https: //stat.ethz.ch/R-manual/R-devel/library/survival/html/pbc.html

Code

```
library(survival)
  head(pbc)
 head(pbc)
  id time status trt       age sex ascites hepato spiders edema bili chol
1  1  400      2   1 58.76523   f       1      1       1   1.0 14.5  261
2  2 4500      0   1 56.44627   f       0      1       1   0.0  1.1  302
3  3 1012      2   1 70.07255   m       0      0       0   0.5  1.4  176
4  4 1925      2   1 54.74059   f       0      1       1   0.5  1.8  244
5  5 1504      1   2 38.10541   f       0      1       1   0.0  3.4  279
6  6 2503      2   2 66.25873   f       0      1       0   0.0  0.8  248
  albumin copper alk.phos     ast trig platelet protime stage
1    2.60    156   1718.0 137.95  172      190    12.2     4
2    4.14     54   7394.8 113.52   88      221    10.6     3
3    3.48    210    516.0  96.10   55      151    12.0     4
4    2.54     64   6121.8  60.63   92      183    10.3     4
5    3.53    143    671.0 113.15   72      136    10.9     3
6    3.98     50    944.0  93.00   63       NA    11.0     3
```

Crude or unadjusted model, stage as the only explanatory variable

Code

```
mypbc=pbc[complete.cases(pbc), ]
nstatus=mypbc$status
nstatus[nstatus==1]=0
nstatus=nstatus/2
uout=coxph(Surv(mypbc$time, nstatus)~as.factor(mypbc$stage))
```

Adjusted model, age is included along with stage as an explanatory variable

Code

```
aout=coxph(Surv(mypbc$time, nstatus)~as.factor(mypbc$stage)+mypbc$age)
```

If the coefficient estimate for the treatment (or the main exposure variable) for the adjusted and unadjusted models are *different* then we say age has a confounding effect, and a measure of change is

$$100\frac{(\widehat{\theta} - \widehat{\beta}_1)}{\widehat{\beta}_1}$$

- $\widehat{\theta}$: the estimated coefficient for treatment in uout (unadjusted model)
- $\widehat{\beta}_1$: the estimated coefficient for treatment in aout (adjusted model)

Results

Code

```
uout
Call:
coxph(formula = Surv(mypbc$time, nstatus) ~ as.factor(mypbc$stage))

                        coef exp(coef) se(coef)    z      p
as.factor(mypbc$stage)2 1.34      3.81     1.04 1.29 0.1966
as.factor(mypbc$stage)3 1.93      6.87     1.01 1.90 0.0571
as.factor(mypbc$stage)4 2.81     16.63     1.01 2.78 0.0054

Likelihood ratio test=43.16  on 3 df, p=2e-09
n= 276, number of events= 111
aout
Call:
coxph(formula = Surv(mypbc$time, nstatus) ~ as.factor(mypbc$stage) +
    mypbc$age)

                          coef exp(coef) se(coef)    z       p
as.factor(mypbc$stage)2 1.23784   3.44816  1.03563 1.20 0.23199
as.factor(mypbc$stage)3 1.83148   6.24310  1.01288 1.81 0.07058
as.factor(mypbc$stage)4 2.57977  13.19405  1.01229 2.55 0.01082
mypbc$age               0.03513   1.03576  0.00981 3.58 0.00034

Likelihood ratio test=55.98  on 4 df, p=2e-11
n= 276, number of events= 111
```

Adjusted model, age is included along with stage as an explanatory variable

For this example, the percentage of change is no more than 10%. So, the confounding effect is not worth mentioning.

Effect modifier

If the effect of an exposure on the outcome varies across groups defined by a third variable, then we say the third variable is an effect modifier. Usually, in statistics, one way of detecting effect modification is to check the presence of a statistically significant interaction term.

Code

```
aout
Call:
coxph(formula = Surv(mypbc$time, nstatus) ~ as.factor(mypbc$stage) +
    mypbc$age + as.factor(mypbc$stage) * mypbc$age)

                                   coef exp(coef) se(coef)     z    p
as.factor(mypbc$stage)2          2.9395   18.9070   5.4020  0.54 0.59
as.factor(mypbc$stage)3          2.7014   14.9004   5.2859  0.51 0.61
as.factor(mypbc$stage)4          3.1816   24.0842   5.2725  0.60 0.55
mypbc$age                        0.0521    1.0535   0.1013  0.51 0.61
as.factor(mypbc$stage)2:mypbc$age -0.0339    0.9667   0.1050 -0.32 0.75
as.factor(mypbc$stage)3:mypbc$age -0.0175    0.9827   0.1026 -0.17 0.86
as.factor(mypbc$stage)4:mypbc$age -0.0126    0.9875   0.1022 -0.12 0.90

Likelihood ratio test=56.49  on 7 df, p=8e-10
n= 276, number of events= 111
```

Effect modifier

One purpose of identifying effect modifier to check if there is any high risk group. If there is really an effect modifier, then that should be properly taken into account in the analysis to accurately estimate the effect of the exposure. If effect modification is suspected, it should also be taken into account in the design stage of the study.

Many covariates: Stepwise variable selection

We shall use the stepwise variable selection procedure (mixture of 'forward' and 'backward') to find the best model. The 'variable list' contains relevant covariates and some of their interaction terms (or moderators). Use the stepwise variable selection with the AIC criteria (you may also use the BIC criteria).

Code

```
data(cancer)
newcancer=cancer[complete.cases(cancer), ]
newcen=newcancer$status-1
out=coxph(Surv(time, newcen)~age+wt.loss+pat.karno+ph.karno+sex+ph.ecog+mea
data=newcancer)
> out
Call:
coxph(formula = Surv(time, newcen) ~ age + wt.loss + pat.karno +
    ph.karno + sex + ph.ecog + meal.cal, data = newcancer)

               coef   exp(coef)   se(coef)      z       p
age        1.080e-02   1.011e+00  1.160e-02   0.931  0.35168
wt.loss   -1.420e-02   9.859e-01  7.766e-03  -1.828  0.06748
pat.karno -1.207e-02   9.880e-01  8.116e-03  -1.488  0.13685
ph.karno   2.244e-02   1.023e+00  1.123e-02   1.998  0.04575
```

Final output of the step function

Code

```
step(out)
Call:
coxph(formula = Surv(time, newcen) ~ wt.loss + pat.karno + ph.karno +
    sex + ph.ecog, data = newcancer)

              coef exp(coef)  se(coef)      z       p
wt.loss   -0.014494  0.985611  0.007693 -1.884 0.05957
pat.karno -0.012401  0.987675  0.007978 -1.554 0.12008
ph.karno   0.020366  1.020575  0.011080  1.838 0.06604
sex       -0.558190  0.572244  0.199202 -2.802 0.00508
ph.ecog    0.742983  2.102197  0.227604  3.264 0.00110

Likelihood ratio test=27.28  on 5 df, p=5.028e-05
n= 167, number of events= 120
```

Checking the proportional hazards (PH) assumption

- Consider a single binary covariate X (1 for treatment, 0 for control, say).
- The Cox model is

$$\lambda(t|X) = \lambda_0(t) \exp(X\beta)$$

- The key assumption is that the effect of the covariate does not depend on time

$$\frac{\lambda(t|1)}{\lambda(t|0)} = \exp(\beta),$$

a constant in time.

- How to check whether this is a reasonable assumption?

Checking the PH assumption

Recall that $S(t|X) = \exp\{-\Lambda(t|X)\}$, where

$$\Lambda(t|X) = \int_0^t \lambda(u|X)du = \Lambda_0(t)\exp(X\beta)$$

We can compute a nonparametric estimate of $\widehat{S}(t|X)$ for each covariate group using the Kaplan-Meier (KM) method. In above scenario, we would compute two KM curves: $\widehat{S}_1(t)$ for $X = 1$ and $\widehat{S}_0(t)$ for $X = 0$.

Checking the proportional hazards assumption:

If the PH assumption holds, then we can write

$$\widehat{S}_1(t) \approx \exp\{-\Lambda(t|1)\}$$

and

$$\widehat{S}_0(t) \approx \exp\{-\Lambda(t|0)\},$$

and consequently,

$$\log\left[-\log\left\{\widehat{S}_1(t)\right\}\right] \approx \log\{\Lambda(t|1)\} = \log\{\Lambda_0(t)\} + \beta$$

and

$$\log\left[-\log\left\{\widehat{S}_0(t)\right\}\right] \approx \log\{\Lambda(t|0)\} = \log\{\Lambda_0(t)\}.$$

What we see is that if the PH model assumption holds, the cloglog transformation of the KM survival functions yields two parallel curves. That means the vertical distance between the curves is constant. So, we can plot , $\log[-\log\{\widehat{S}_1(t)\}]$ and $\log[-\log\{\widehat{S}_0(t)\}]$ against t, and verify if they are separated by an approximately constant amount.

Checking the PH assumption

- In general, with more than 2 comparison groups, or with continuous covariates, the same idea can be applied to get a rough feel for whether the PH model is appropriate.

- With continuous covariates, we can bin the covariates to create artificial categorical variables and groups.

- For other model checking tools, see Hosmer and Lemeshow (2000).

- If PH is not a reasonable assumption, consider parametric models (Reference: Klein & Moeschberger, 2003).

Example, the veteran lung cancer data

Code

```
out=coxph(Surv(time, status)~celltype, data=veteran)
> out
Call:
coxph(formula = Surv(time, status) ~ celltype, data = veteran)

                  coef exp(coef) se(coef)    z       p
celltypesmallcell 1.001    2.722    0.254 3.95 7.8e-05
celltypeadeno     1.148    3.151    0.293 3.92 8.9e-05
celltypelarge     0.230    1.259    0.277 0.83    0.41

Likelihood ratio test=24.85   on 3 df, p=2e-05
n= 137, number of events= 128
```

Example, the veteran lung cancer data

Code

```
data1=veteran[veteran$celltype=="squamous", ]
data2=veteran[veteran$celltype=="smallcell", ]
data3=veteran[veteran$celltype=="adeno", ]
data4=veteran[veteran$celltype=="large", ]


out1=survfit(Surv(time, status)~1, data=data1)
out2=survfit(Surv(time, status)~1, data=data2)
out3=survfit(Surv(time, status)~1, data=data3)
out4=survfit(Surv(time, status)~1, data=data4)
```

Example, the veteran lung cancer data

Code

```
pdf("fig4_surv_part3.pdf")
plot(out1$time, log(-log(out1$surv)), type="s", ylim=c(-3.3, 1.2),
xlim=c(1, 999), ylab="", xlab="Time", lwd="2", col="red")
par(new=T); plot(out2$time, log(-log(out2$surv)), type="s", ylim=c(-3.3, 1.2),
 xlim=c(1, 999), axes=F, lwd=2, col="blue",
 ylab="", xlab=" ")
par(new=T); plot(out3$time, log(-log(out3$surv)), type="s", ylim=c(-3.3, 1.2),
xlim=c(1, 999), axes=F, lwd=2, col="purple", ylab="", xlab=" ")
par(new=T); plot(out4$time, log(-log(out4$surv)), type="s", ylim=c(-3.3, 1.2),
xlim=c(1, 999), axes=F, lwd=2, col="brown", ylab="", xlab=" ")
dev.off()
```
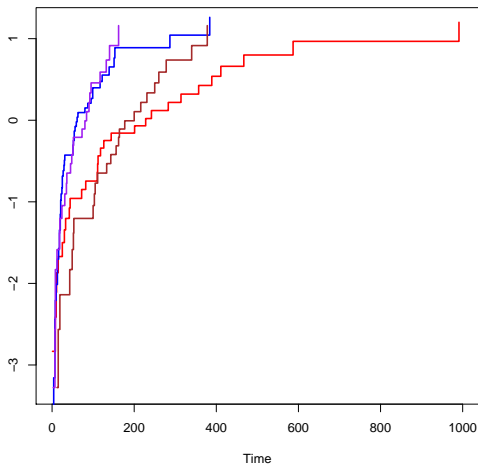
Comments on the figure

- The red and brown curves (squamous and large cell type) are crossing each other, so they cannot be treated as parallel. We call these two curves to form group 1.

- The blue and purple curves (small and adeno cell type) are crossing each other, so they cannot be treated as parallel. We call these two curves to form group 2.

- Although these two groups, 1 and 2, look the same in the early time, they seem not to cross each other over the time period where most of the subjects failed.

A formal test

The above checking is via a visual inspection. A formal test can be conducted via the method proposed in Grambsch & Therneau (1994), *Proportional hazards tests and diagnostics based on weighted residuals, Biometrika, 81, 515–526*.

Code

```
fit <- coxph(Surv(time, status) ~ celltype, data=veteran)
temp <- cox.zph(fit, transform="log")
print(temp)
          chisq df      p
celltype   9.69  3 0.021
GLOBAL     9.69  3 0.021
```

Based on the result of the Global test, we reject H_0 : PH assumption holds at the 5% level.

The cox.zph function will test proportionality of all the predictors in the model by creating interactions with time using the transformation of time specified in the transform option. The default transformation is the KM survival function of time. In this example we are testing proportionality by looking at the interactions with log(time). The last row contains the global test for all the interactions tested at once. A p-value less than 0.05 indicates a violation of the proportionality assumption.

Background information for cox.zph

Allowing the effect of the jth covariate to change over time:

$$\beta_j(t) = \beta_j + \gamma_j g_j(t),$$

where $g_j(t)$ is a specified function of time [e.g., $g(t) = \ln(t)$], in which case

$$\beta_j(t) = \beta_j + \gamma_j \ln(t).$$

The form of the linear predictor is therefore

$$\beta_j x_j + \gamma_j x_j \ln(t).$$

We can test $H_0 : \gamma_j = 0$ via partial likelihood ratio test or Wald test obtained when time varying interaction $x_j \ln(t)$ is added to the proportional hazards model. The cox.zph function also produces the plot of the estimated regression coefficient $\widehat{\beta}_j(t)$ against time. If it is seen than the plot of $\widehat{\beta}_j$ againist time is not much changing over time, then the proportionality assumption is holding otherwise not.

Censoring mechanism: sensitivity analysis

- Liu (2017) consider A randomized, double-blind, parallel group, low-dose active controlled study comparing the safety and efficacy of two doses (400 mg/day versus 50 mg/day) of study medication used as monotherapy for the treatment of newly diagnosed or recurrent epilepsy.

- A total of 487 subjects were enrolled; of those, 16 withdrew during the open treatment phase. Of the 471 subjects randomized, 470 had at least 1 study visit after randomization and were included in the intent-to-treat analysis. Primary efficacy analysis was based on a survival analysis of the difference between TPM 400 and TPM 50 with respect to time to first partial onset seizures or generalized seizures during the double-blind phase (excluding taper). Kaplan-Meier (referred to as KM) estimates were calculated for time to first seizure. Statistical significance of the treatment effect was tested by the log-rank test.

The next table lists the completion/withdrawal status along with p-value of efficacy results for original observed data. The first subject's randomization occurred at 19NOV1999; and afterwards eligible patients were continuously randomized until 15AUG2001. There are 470 subjects (TPM 50=234 and TPM 400=236), with 90 (38%) and 49 (21%) events occurred in the TPM 50 and TPM 400, respectively. Comparison of the KM survival curves of time to first seizure favored TPM 400 over TPM 50 (p-value=0.0002; 2-sided log-rank test). When the trial ended at 26FEB2002, there were 217 (TPM 50=105, TPM 400=112) remained event-free at the time of study termination, which were considered as being administratively censored since censoring was caused by trial operation and thus was also considered as non-informative censoring.

Category	Sub-category	TPN 50 N=234	TPM 400 N=236	Total N=470	
Event	Seizure	90 (38)	49 (21)	139 (30)	p-value=0.0002
Informative censoring	Withdrawal due to adverse event	13(6)	40 (17)	53(11)	
	Withdrawal due to subject choice	9(4)	13(6)	22(5)	
Non-informative censoring	Administrative censoring	105 (45)	112(47)	217(46)	
	Withdrawal due to lost to follow-up	9(4)	10(4)	19(4)	
	Withdrawal due to other reason	8(3)	12(5)	20(4)	

The proportions of withdrawals due to lost to follow-up and other reason were almost the identical between high and low dose levels (that is: non-differential between two treatment groups), which hinted the claim of non-informative nature for these two kinds of withdrawals. However, at the time of study termination, in the TPM 50 group, 6% (N=13) of subjects had early withdrawal due to adverse event and 4% (N=9) of subjects due to subject choice while having 17% (N=40) of withdrawals due to adverse event and 6% (N=13) of withdrawals due to subject choice in the TPM 400 group. These two types of withdrawals are differential between two treatment groups. Combining these two types of withdrawals together, dis-proportionality in early withdrawal rates between two groups (TPM 400=23% vs. TPM 50=10%) makes people believe that these withdrawals might have informative censoring with being informative with respect to treatment assignment, resulting in violating of non-informative censoring assumption in application of logrank test.

To address this issue, one proposal from the US FDA (Food and Drug Administration) reviewer then was to impute informative censoring subjects and treat them as they have had an event occurred at the time of early withdrawal. The number of events then becomes 112 (48%) in the TPM 50 group and 102 (43%) in the TPM 400 group, resulting in a big decrease in the difference in event proportion between two groups (5% in difference: TPM 50=48% vs. TPM 400=43%) in this naive data as compared with original data (17% in difference: TPM 50=38% vs. TPM 400=21%). More importantly, p-value of log-rank test from the naive data becomes 0.3859, which fails to support the claim of superiority of TPM 400 over TPM 50 in preventing time to first seizure in the double-blind phase. The naive data are very artificial and incorrect because we only know that subjects who were informatively censored at their withdrawal time but with no knowledge on whether or when event occurred afterwards. Surely for them, there was no event occurring at their date of early withdrawal. From this perspective, the naive data can be viewed as the 'worst-case- scenario' imputation of the original data. One question to ask next is: what else imputations could possibly depict intermediate scenarios?

Here we list alternative scenarios, and for each scenario we can consider the log-rank test. Next, we investigate how the p-values are changing across the cases. If the p-values do not change much then that is the best news. If the p-values change substantially, then we report the p-value under the independence assumption with caution, and must state clearly the assumption and need to report the results are sensitive towards the assumption.

- The 'best-case-scenario' is assuming that these subjects were administratively censored. That means for these subjects, the event did not happen until 26FEB2002.

- Imputing the missing values of these censored subjects using some models. Then carry out the analysis (log-rank test). Let $f_{t,c,\omega_j}(\cdot, \cdot | \theta_j)$ be the joint distribution of the failure and censoring time in group j with the known association (correlation) parameter ω_j and unknown parameter θ_j. We fit this model to the data, and estimate θ_j. Then use the model to generate the failure time conditional on the fact that it is larger than the observed censoring time. Once this imputed data are generated, we can carry out the log-rank test, and record the p-value.

Sample size for clinical trials

- Suppose that a number of subjects randomly assigned to two arms (groups), treatment and control. Suppose that X is the binary indicator for the treatment.

- Assume that the hazard of the time-to-event T follows the PH model, that means $\lambda(t|X) = \lambda_0(t) \exp(\theta X)$, where the regression parameter θ is called the log-hazard ratio and $\exp(\theta) = \lambda(t|\text{treatment})/\lambda(t|\text{control})$ is called the risk ratio.

- In a two-arm randomized trial, for given probability of Type-I and II error, α and β, the required number of events, the total in two trials, is

$$m = \frac{(Z_{\alpha/2} + Z_{\beta})^2}{\theta^2 \pi (1 - \pi)},$$

- If clinicians think the treatment provides 25% reduction in the rate of the event, then $\exp(\theta) = 0.75$, so $\theta = \log(0.75)$
- π : proportion of subjects allocated to the placebo, for equal allocation trial set $\pi = 0.5$
- α : the level of significance usually $\alpha = 0.05$
- $1 - \beta$: power of the test, usually $\beta = 0.20$ for 80% power
- Page 340 of the Applied Survival Analysis by Hosmer et al.

Sample size calculation

- This is an ideal scenario where all subjects are recruited at time zero, and all of them are followed-up until the event occurs. In reality that does not happen.

- In practice, subjects are recruited over a specified period, we call it accrual period. Then the subjects are followed for an additional f period of time.

- In practice a fraction of subjects experience the event of interest during the follow-up period, and some will not experience the event of interest during the follow-up (they are right censored). To take into account this censoring we divide the number of events by the overall probability of event by the end of the follow-up period.

- Thus the required number of subjects in the trial is

$$n = \frac{m}{\text{pr}(T \leq a + f)},$$

where $\text{pr}(T \leq a + f)$ is the probability of the event by the end of the accrual period a and then follow-up period f.

Sample size calculation
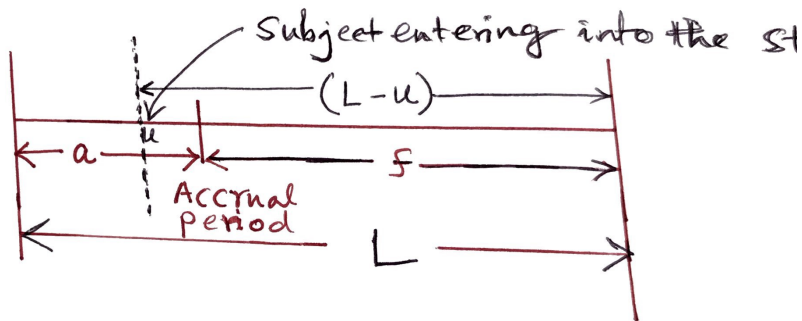
a: accrual period, f: is the follow-up period after the accrual period a, L: total period including accrual and follow-up

u : time when a random subject enters in the study, $L - u$ is the potential follow-up time for the subject

Sample size calculation continues

- Let $\overline{F}$ denotes the CDF of T combining both groups. Note that T is the time to failure measured from the recruitment.

- If subjects are recruited uniformly over the accrural period interval $(0, a)$, then the probability of observing a failure $(0, L)$ is

$$\text{pr}(T \leq L) = \int_0^a \frac{du}{a} \times \overline{F}(L - u).$$

  The first term of the integral: chance of recruitment over $[u, u + du)$, the second term of the integrand: probability of observing failure over $(0, L - u)$. Since the recruitment can happen over $(0, a)$, the final result is obtained as an integration.

- Using Simpson's rule the above integration is approximated by $\text{pr}(T \leq L) = (a/2)(1/3)(1/a)\{\overline{F}(L) + \overline{F}(L - a) + 4\overline{F}(L - a/2)\}$. For this numerical approximation $(0, a)$ is divided into two parts, $(0, a/2)$ and $(a/2, a)$ with the width of each interval $a/2$, and this width appears as the first term of the Simpson's approximation.

- In terms of survival probability,

$$\text{pr}(T \le a + f) = 1 - \frac{1}{6}\{\overline{S}(f) + 4\overline{S}(0.5a + f) + \overline{S}(a + f)\},$$

where

$$\overline{S}(t) = 1 - \overline{F}(t) = \pi S_0(t) + (1 - \pi)S_1(t),$$

S_0 and S_1 are the estimated survival probability for the placebo and treatment groups, respectively, from the pilot study, and

$$S_1(t) = \{S_0(t)\}^{\exp(\theta)}.$$

- If π^* is the percentage of subjects lost to follow-up during the follow-up period (or withdrawal due to non-compliance), then the required sample size will be $n^* = n/(1 - \pi^*)$.

Example: Applied Survival Analysis by Hosmer, Lemeshow and May, 2008, p. 342.

- Suppose that for a hypothetical example, clinicians think that the treatment provides 50% reduction in the rate of the event, that means $\exp(\theta) = 0.5$, so $\theta = \log(0.5)$

- The accrual period $a = 2$ years, and the minimum follow-up after the accrual $f = 3$ years. Then the maximum follow-up $L = 5$ years. Let $\widehat{S}_0(3) = 0.5$, $\widehat{S}_0(4) = 0.6$, $\widehat{S}_0(5) = 0.7$. With $\theta = 0.5$, $\widehat{S}_1(3) = \widehat{S}_0(3)^{0.5} = 0.5^{0.5} = 0.71$, $\widehat{S}_1(4) = \widehat{S}_0(4)^{0.5} = 0.6^{0.5} = 0.77$, $\widehat{S}_1(5) = \widehat{S}_0(5) = 0.7^{0.5} = 0.84$.

- For equal allocation, $\overline{S}(3) = 0.605$, $\overline{S}(4) = 0.685$, $\overline{S}(5) = 0.77$. Hence, $\text{pr}(T \leq a + f) = 1 - (0.605 + 0.77 + 4 * 0.685) = 0.314$.

- For $\theta = \log(0.5)$, $\alpha = 0.05$, $\beta = 0.2$, $m = (1.96 + 0.84)^2/(\log(0.5)^2 * 0.25) = 65.2$

- $n = m/\text{pr}(T \leq a + f) = 65.2/0.314 = 207.87$.

- For a 20% drop-out/withdrawl (excluding the administrative censoring), the required number of total subjects for both arms is $207.87/0.80 = 259.82 \approx 260$

Alternative way of computing the sample size (following A Tsiatis's class notes)

- Let $r(t)$ be the accrual rate, the rate at which subjects are recruited in the study. Mathematically,

$$r(t) = \lim_{\eta \to 0} \frac{\text{expected number of subjects recruited in } [t, t + \eta)}{\eta}.$$

- Assuming the accrual rate is same for both groups, the expected number of failures for groups 1 and 2 are $D_0 = \int_0^a \{r(u)/2\} F_0(L - u) du$ and $D_1 = \int_0^a \{r(u)/2\} F_1(L - u) du$.

- Under the equal allocation scheme,

$$m = 4 \frac{(Z_{\alpha/2} + Z_\beta)^2}{\theta^2} = D_0 + D_1.$$

Alternative way of computing the sample size

- Suppose that in a hypothetical study, r_0 patients are recruited every year, and they are randomized to one of the two groups. Thus $r(t) = r_0/2$ for every group when t is measured in years. Let the survival time for each group follow exponential distribution with parameters λ_0 and λ_1, respectively. Then

$$
\begin{aligned}
D_0 &= \int_0^a \{r_0/2\}\left[1 - \exp\{-\lambda_0(L - u)\}\right] du \\
&= \frac{r_0}{2}\left[a - \frac{1}{\lambda_0}\left\{\exp\{-\lambda_0(L - a)\} - \exp(-\lambda_0 L)\right\}\right]
\end{aligned}
$$

- Let the median survival time for the control and treatment group are (these are expectations) 4 years and 6 years respectively. The corresponding parameters are $\lambda_0 = \log(2)/4 = 0.173$ and $\lambda_1 = \log(2)/6 = 0.115$. The log-hazard ratio $\theta = \log(4/6) = \log(2/3)$. With $\alpha = 0.05$ and $\beta = 0.2$, we need $m = 190$. Now, the question is what should be the choice of L and a for a given r_0.

Alternative way of computing the sample size

- Let $r_0 = 100$. Then

$$
\begin{aligned}
190 \;=\; & \frac{100}{2}\left[a - \frac{1}{0.173}\left\{\exp\{-0.173(L-a)\} - \exp(-0.173L)\right\}\right] \\
& + \frac{100}{2}\left[a - \frac{1}{0.115}\left\{\exp\{-0.115(L-a)\} - \exp(-0.115L)\right\}\right].
\end{aligned}
$$

- Clearly, there are many solutions for a and L. Let us consider a few options. 1) Set $a = 2$ years, and then find L. 2) Set $a = 3$ years, and find L. 3) Set $a = L/2$ and find L, 4) Set $a = L$ and find L.

- In clinical trials sample sizes are calculated for various possible combinations of the parameters, a, L, θ, β, r_0.

References

Cox, DR. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.

Klein, JP & Moeschberger, ML. (2003). *SURVIVAL ANALYSIS Techniques for Censored and Truncated Data*, Springer: New York.

Hosmer, DW & Lemeshow, S. (2000). *Applied logistic regression*, 2nd edn. John Wiley & Sons, New York.

Hosmer, DW., Lemeshow, S. and May, S. (2008). *Applied Survival Analysis*. 2nd edn. John Wiley & Sons, New York.

Lemeshow, S. & Hosmer, DW. (1982). A review of goodness of fit statistics for use in the development of logistic regression models. *American Journal of Epidemiology*, **115**, 92–106.

Liu Y. (2017). Sensitivity analyses for informative censoring in survival data: A trial example. *J Biopharm Statistics*, **27**, 595–610.