Cox Regression

Aim

To describe the association of a time to event outcome and other explanatory variables.

Objectives

By the end of this session you will be able to:

- show the association between the log-rank test and the univariate Cox regression
- obtain and interpret the results of a cox regression analysis
- ensure that the assumptions for the Cox regression are met
- checking proportionality assumption
- report results in a graph or a table
- critically appraise the survival analyses reported in the literature.

A Cox regression model (which is also called a Cox proportional hazards regression) provides an estimate of survival time while adjusting for the effects of other explanatory variables (referred to as covariates in this type of model). For example, in predicting an event such as death, factors such as age of the patient or number of years smoking cigarettes can be included in a Cox regression model. Compared to the Kaplan–Meier method where only categorical variables can be used to predict the event, with the Cox regression analysis a combination of categorical and/or continuous variables can be used to predict survival. In addition, Cox regression models can also manage censored data.

Cox regression is similar to other regression models such as linear regression or logistic regression, in that regression coefficients are generated, interaction between variables can be examined and adjustment for confounding factors can be made. Where linear and logistic regressions models are used respectively to predict scores for a continuous and a binary variable, Cox regression models are used to predict the rate of an event.

A rule of thumb is that Cox models should have a minimum of 10 outcome events per predictor variable determining the maximum number of independent variables.

At any point in time, t, an individual, i, has an instantaneous risk of reaching the endpoint (often known as the hazard, or $\lambda_i(t)$), given that s/he has not reached it up to that point in time. For example, if death is the endpoint, the hazard is the risk of dying at time t. This instantaneous hazard is usually very small and is of limited interest. However, we may want to know whether there are any systematic differences between the hazards, over all time points, of individuals with different characteristics. For example, is the hazard generally reduced in individuals treated with a new therapy compared with those treated with a placebo, when we take into account other factors, such as age or disease severity?

Cox model

In Cox regression model, the hazard function at a given time is predicted by the baseline hazard function $\lambda_0(t)$, which estimates the overall risk of the event where all explanatory variables equal zero. This term is similar to the constant term (i.e. the intercept) in linear regression. Regression coefficients are also generated for the explanatory variables or covariates that are included in the model.
Thus, the Cox regression model can be given by the following equation:

$$\textbf{ln (Hazard at time t)} = \lambda_0(t) + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

or

$$\lambda(t) = \lambda_0(t) * \exp(\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p)$$

where $\lambda(t)$ is the hazard at time t, $\lambda_0(t)$ is an arbitrary baseline hazard (in which we are not interested), x1, . . . , xp are explanatory variables in the model and $\beta_1, \ldots, \beta_p$ are the corresponding coefficients. We obtain estimates, $b_1, \ldots, b_p$, of these parameters using a form of maximum likelihood known as **partial likelihood**. The ln (hazard) or the exponential expression involving the X's ensures that the fitted model will always give estimated hazards that are non-negative. We want such nonnegative estimates because, by definition, the values of any hazard function must range between zero and plus infinity, that is, a hazard is always nonnegative. The exponential of these values (e.g. $\exp\{b_1\} = e^{b}_1$) are the estimated **relative hazards** or **hazard ratios**. For a particular value of x_1, the hazard ratio is the

estimated hazard of disease for (x_1 + 1) relative to the estimated hazard of disease for x_1, while adjusting for all other *x*'s in the equation. The hazard ratio is interpreted in a similar manner to the odds ratio in logistic regression therefore values above one indicate a raised hazard, values below one indicate a decreased hazard and values equal to one indicate that there is no increased or decreased hazard of the endpoint.

A confidence interval can be calculated for the hazard ratio and a significance test performed to assess its departure from 1. The hazard ratio is assumed to be constant over time in this model (i.e. the hazards for the groups to be compared are assumed to be *proportional*). It is important to check this assumption either by using graphical methods or by incorporating an interaction between the covariate and time in the model and ensuring that it is non-significant.

Another important property of the Cox model is that the baseline hazard, $\lambda_0(t)$, is an unspecified function. It is this property that makes the Cox model a semiparametric model. Note that the hazard is sometimes symbolized as *h(t)*.

Hazard ratio

Hazard is defined as the immediate risk of event occurrence. The hazard rate or function is the probability that if the event has not occurred, it will occur in the next time interval, divided by the length of that interval. If the time interval is small, the hazard function represents an instantaneous event rate among participants who have not experienced the event. In Cox regression analysis, the dependent variable is the hazard function at a given time. With a Cox regression analysis, the effect of each covariate is reported as a hazard ratio. The hazard ratio is computed as the proportion of the rate (or function) of the hazard in the two groups. The hazard ratio can be used to estimate the hazard rate in a treatment group compared to the hazard rate in the control group. A hazard ratio of 2 indicates that, at any time point, twice as many patients in the one group experience an event compared with the other group.

The hazard ratio is sometimes used interchangeably to mean a relative risk; however, this interpretation is not correct. The hazard ratio incorporates the change over time, whereas the relative risk can only be computed at single time points, generally at the end of the study.

Assumptions of Cox regression model

The Cox regression model is a semi-parametric model and no assumptions are made about the distribution of survival. However, the following assumptions should be met:

- The participants must be independent, that is, each participant appears only once in their group
- The groups must be independent, that is, each participant is only in one group
- All participants are event free when they enrol in the study
- The measurement of time to the event is precise
- The start point and the event are clearly defined
- Participants' survival prospects remain constant, that is, participants enrolled early or late in the study should have the same survival prospects
- The probability of censoring is not related to the probability of the event
- The hazards should be proportional

The proportional hazards assumptions means that the hazard (rate of the event) in one group should be a constant proportion of the hazard in the other study group over all time points. This assumption is important since the hazard ratio estimated by the model is for all time points. To test this assumption, when there are only two groups and no covariates, a simple test is to plot the Kaplan–Meier survival curves of the two groups together. If the curves are proportional and approximately parallel, then the assumption of proportional hazards is satisfied (figure 1). If the curves cross or if curves are not parallel and diverge they indicate that the rate of the event between the two groups is different (e.g. rate for one group increases constantly and the other group only slowly increases), and that the assumption of proportional hazards is not met (figure 2).

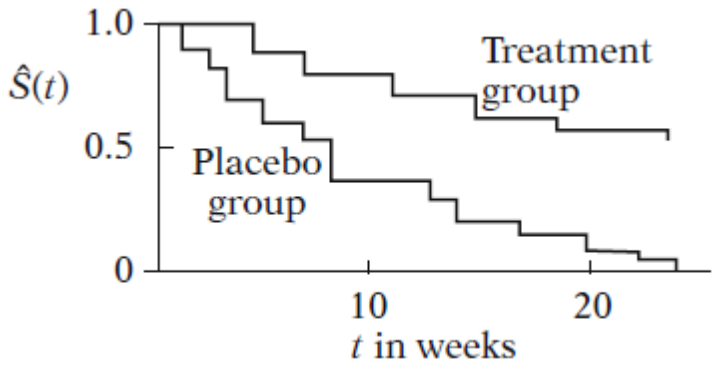


Figure 1: Proportional hazards assumption satisfied (parallel survival curves)

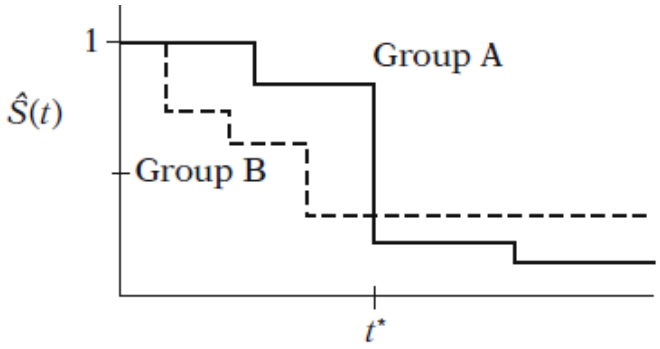


Figure 2: Proportional hazards assumption violated (survival curves cross)

However, with small data sets the error around the survival curve is increased and therefore this test may not be accurate. In addition, these plots are more complex to interpret with multivariable models. More appropriate methods are the log-minus-log plot and examination of the partial residuals. The log-minus-log of the survival function, is the ln(−ln(survival)), versus the survival time. When the hazards are proportional, the curves should be approximately parallel. The residuals when plotted should be horizontal and close to zero (shown later in the chapter) if the hazards are proportional.

In survival analysis, the assumption of proportional hazards can also be checked by assessing whether there is an interaction between time and treatment, as well as the covariates. This is referred to an extended Cox regression model.

Checking the Proportional Hazards Assumption

There are three general approaches for evaluating the proportional hazards (PH) assumption of the Cox model—a graphical procedure, a goodness-of-fit testing procedure, and a procedure that involves the use of time-dependent variables. We now briefly overview each approach, starting with graphical techniques.

Graphical procedure

There are two types of graphical techniques available. The most popular of these involves comparing **estimated –ln(–ln) survivor curves** over different (combinations of) categories of variables being investigated. We will describe such curves in detail in the next section. Parallel curves, say comparing males with females, indicate that the PH assumption is satisfied, as shown in figure 3 for the variable Sex.

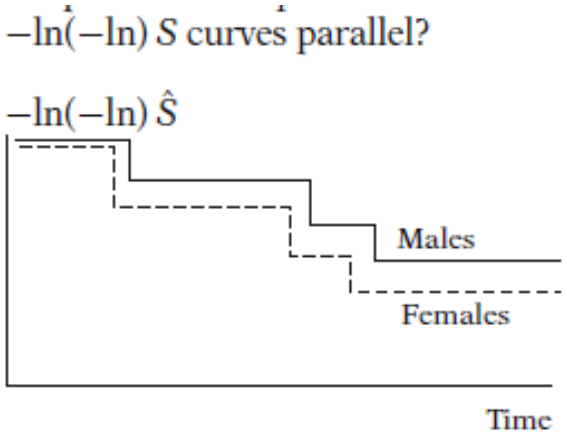


Figure 3: Example of parallel –ln(-ln) S curves indicating that the PH assumption is satisfied

An alternative graphical approach is to compare observed with predicted survivor curves. The observed curves are derived for categories of the variable being assessed, say, Sex, without putting this variable in a PH model. The predicted curves are derived with this variable included in a PH model. If observed and predicted curves are close, then the PH assumption is reasonable as in figure 4.

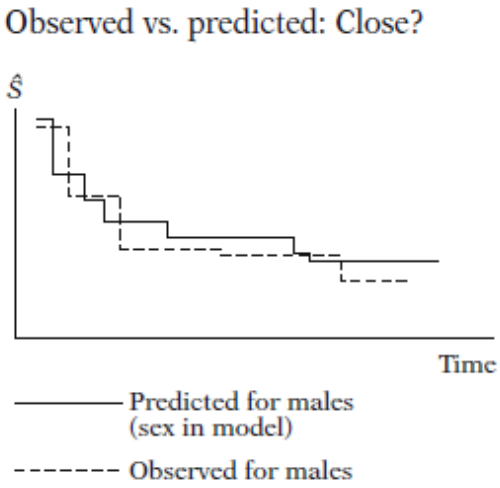


Figure 4: Example of observed versus predicted curves that are close indicating that the PH assumption is satisfied

Goodness-of-fit test

A second approach for assessing the PH assumption involves goodness-of-fit (GOF) tests. This approach provides large sample Z or chi-square statistics which can be computed for each variable in the model, adjusted for the other variables in the model. A number of different tests have been proposed in the literature. The most common is the test of Harrel and Lee (1986), a variation of a test originally proposed by Schoenfeld (1982) and based on the residuals defined by Schoenfeld, now called the Schoenfeld residuals. A p-value derived from a standard normal statistic is also given for each variable. This p-value is used for evaluating the PH assumption for that variable. A nonsignificant (i.e., large) p-value, say greater than 0.10, suggest that the PH assumption is reasonable, whereas a small p-value, say less than 0.05, suggests that the variable being tested does not satisfy this assumption.

Time-dependent covariates

When time-dependent variables are used to assess the PH assumption for a time-independent variable, the Cox model is extended to contain **product** (i.e., interaction) **terms** involving the time independent variable being assessed and some function of time. For example, if the PH assumption is being assessed for Sex, a Cox model might be extended to include the variable "Sex*t" in addition to Sex. If the coefficient of the product term turns out to be significant, we can conclude that the PH assumption is violated for Sex.

$$\ln(\text{Hazard of } y=1 \text{ at time } t) = \lambda_0(t) + \beta_1 SEX + \beta_2 SEX*t$$

If $\beta_2 \neq 0 \Rightarrow$ PH assumption violated ($P<0.05$).

The GOF approach provides a single test statistic for each variable being assessed. This approach is neither as subjective as the graphical approach nor as cumbersome computationally as the time-dependent variable approach. Nevertheless, a GOF test may be too "global" in that it may not detect specific departures from the PH assumption that may be observed from the other two approaches.

It is recommended that the researcher use at least two of these approaches when assessing the PH assumption.

Example

MAC prophylaxis study was a Phase III randomized, double-blind study comparing rifabutin, clarithromycin, and the combination of rifabutin and clarithromycin for prevention of MAC disease in HIV-infected patients with CD4 lymphocyte counts less than 100 cells/mm3. A total of 1177 patients were accrued over an 8-month period and randomized in equal proportions to the three treatment arms. Patients were followed for a median of 19 months (the study closed in August 1995, prior to the advent of widespread use of protease inhibitors).

The primary endpoint for this study was the development of MAC (mycobacterium avium complex) disease, which is one of the most common opportunistic infections occurring in AIDS patients and is associated with significant mortality. Secondary endpoints of the trial included survival and drug toxicity resulting in permanent discontinuation of study drugs. In the following example we will work with the survival as an outcome.

Question: Are the CD4 cell count, the use of antiretroviral therapy or the type of antibiotic use independent predictors of survival?

Variables: Outcome variable=death (binary event)

Explanatory variables: CD4 cell count (continuous), gender (binary, two levels: 0=male, 1=female), antibiotic use (categorical, three levels: 1=rifabutin, 2=clarithromycin, 3=combination of rifabutin+clarithromycin)

Simple Cox model:

First, we will fit the univariate models.

Example 1: Continuous explanatory variable: CD4 cell count and survival

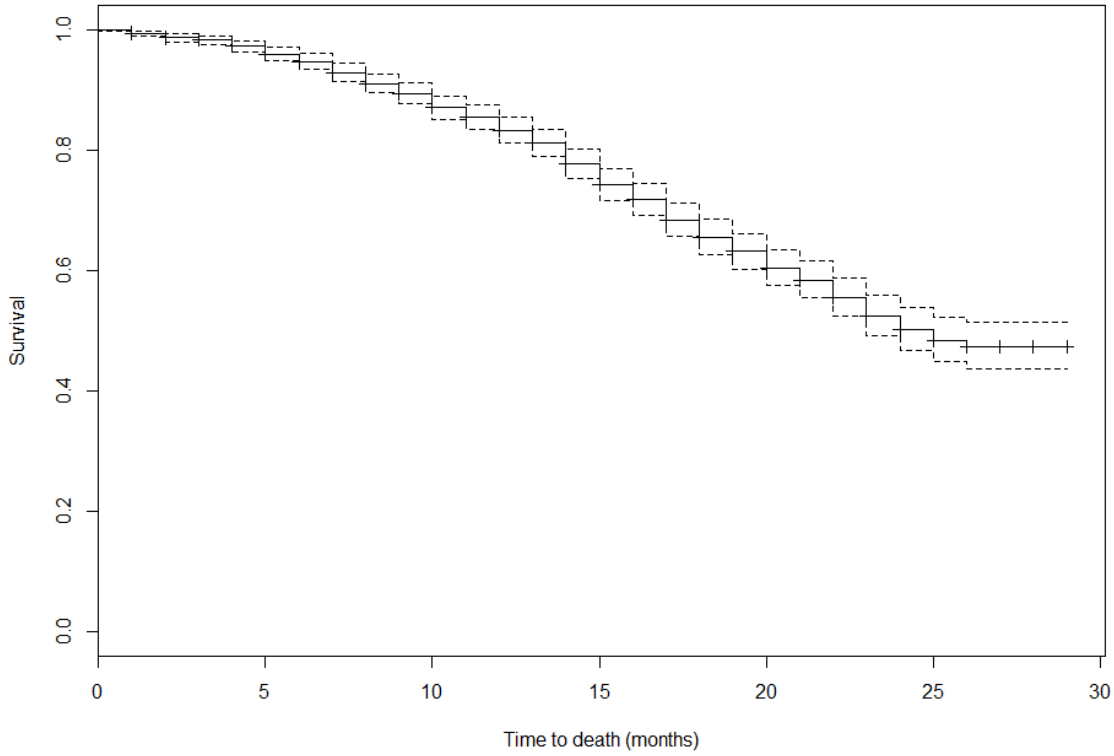


Figure: Kaplan-Meier curve of CD4, the dotted lines are the 95% confidence intervals

Model

We have the following Cox model equation:

ln (Hazard at time t)=$\lambda_0(t)+\beta_1$CD4

The results when fitting a Cox regression to the data are the following

```
Call:
coxph(formula = Surv(dthtime, dthstat) ~ cd4, data = mac)
        coef exp(coef) se(coef)     z       p
cd4 -0.01160   0.98847  0.00156 -7.45 9.4e-14

Likelihood ratio test=67  on 1 df, p=3.33e-16
n= 1177, number of events= 514
```

Interpretation

We notice that the estimated coefficient ($\hat{\beta}_1 = -0.012$) of CD4 is negative. $\hat{\beta}_1$ is the estimated change in the log hazard of death for one cell/ µL increase in CD4. Here, we notice a negative association between CD4 and death.

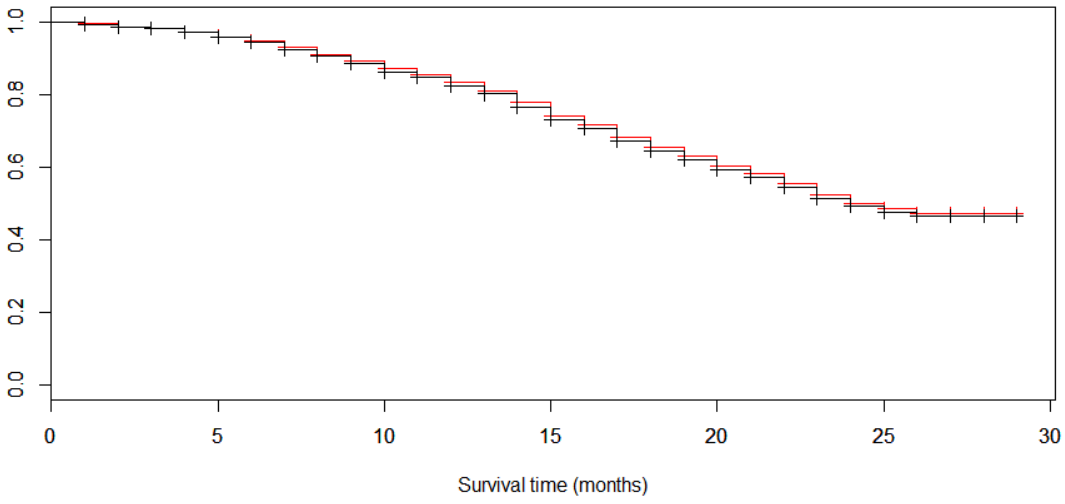
To convert these values to hazard ratio (HR) we just need to take the exponential value of log hazard. So, the HR for $\hat{\beta}_1$ is $e^{-0.012} = 0.988$. This means that the risk of death is reduced by about (0.98-1=-0.02) 2 % as the patient's CD4 cell count increase by one cell/µL. The related p-value is <0.001 which means that this is a statistically significant result. The corresponding 95% confidence interval is (0.985, 0.991) does not include 1, which also indicates that the result is significant.

In case we would like to express the HR for every 100 cells/ µL increase in patient's CD4 cell count we just need to multiply the b coefficient with 100 and then exponentiation, (β_1*100 = -0.012*100= -1.2) , so the HR = $e^{-1.2} = 0.301$ which is equivalent with raising the hazard ratio on the power of 100. So, $0.988^{100} = 0.300$, which means that the risk of death is reduced by about (1-0.30=-0.70) 70% for every 100 cells/ µL increase in the CD4 cell counts.

Proportionality assumption using the observed vs. predicted graphic method:

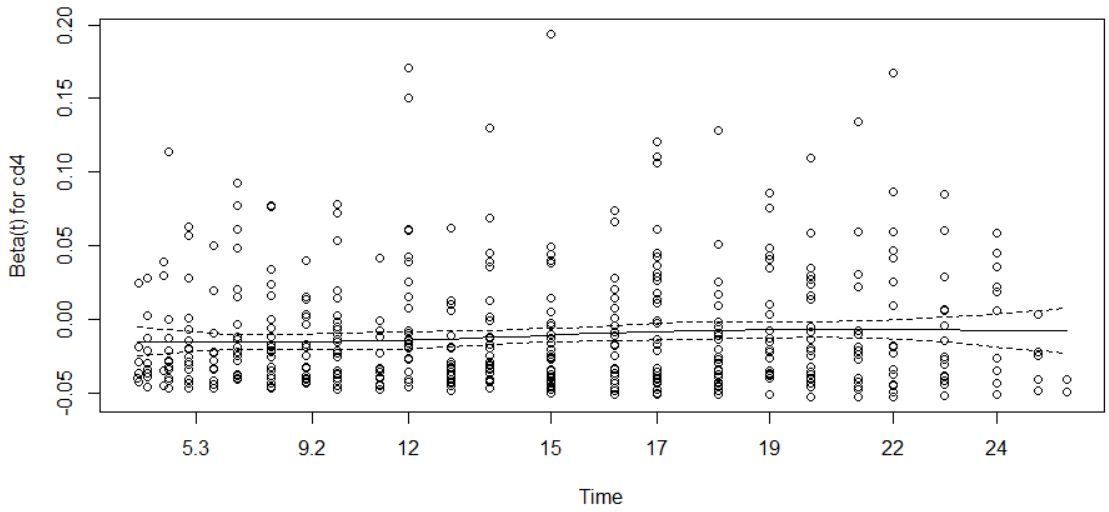
Red curve: Predicted model

Black curve: Observed model



Proportionality assumption using the GOF test:

```
       rho chisq      p
cd4 0.0823  4.34 0.0372
```



Both approaches agree on the validation of the proportionality assumption: the survival curves are parallel, the p-value is far from significant and the Schoenfeld residuals contain very few outliers.

Linearity assumption

Another assumption that must be satisfied when fitting a continuous variable is that the relationship with the ln(hazard) must be linear. There is no easy visual tool (like a scatterplot) to help with this assessment but an empirical approach can be used by categorizing the continuous explanatory variable into groups, e.g. quartiles, and see if ln(hazard) is increasing or decreasing linearly in consecutive ordinal groups.

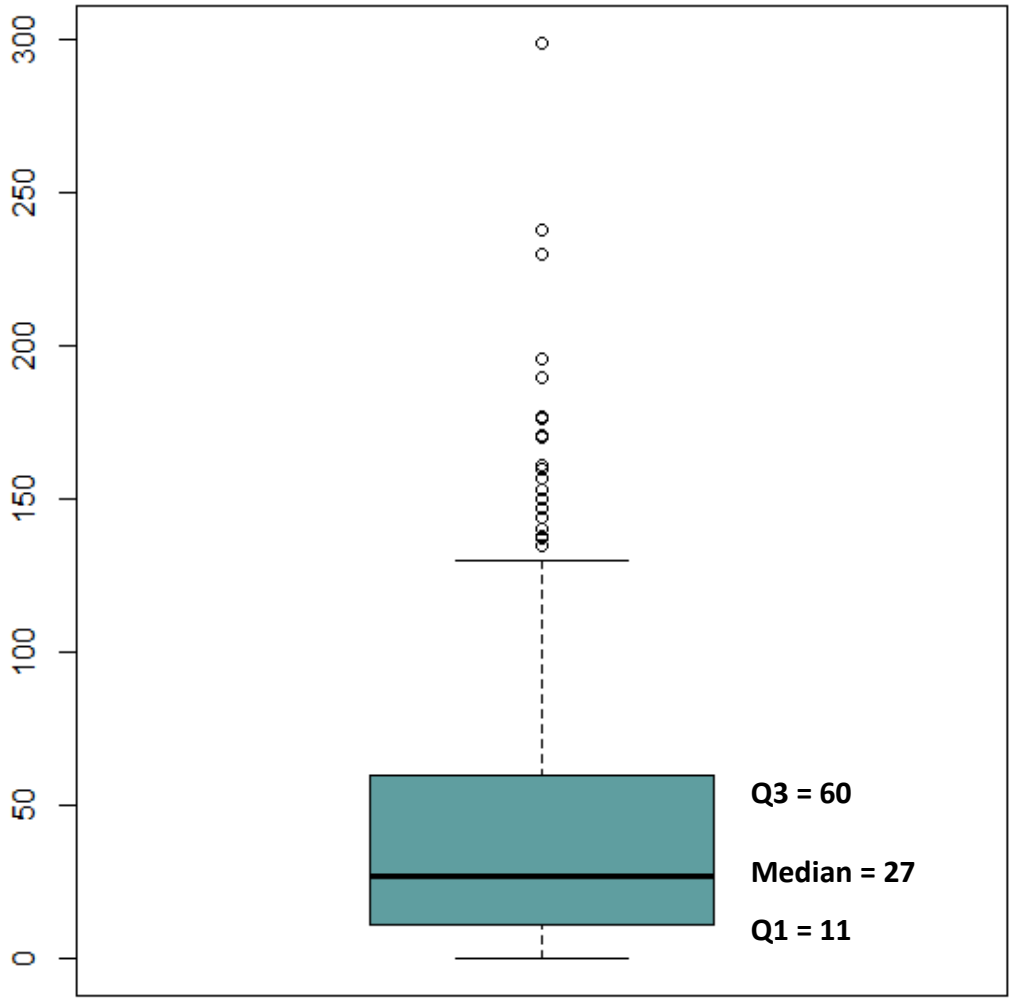


Figure: Boxplot of CD4 cell counts

According to the above boxplot the quartiles that divides CD4 into 4 equal parts are Q1 =11, Q2=27 and Q3=60 cells /µL. Next a Cox model will be fitted with CD4 cell count categorized into 4 quartiles and with quartile 1 as the reference group.

```
Call:
coxph(formula = Surv(dthtime.m, dthstat) ~ varQ, data = mac)


                coef exp(coef) se(coef)     z       p
varQ(11,27]  -0.399     0.671    0.114 -3.50 0.00047
varQ(27,60]  -0.685     0.504    0.119 -5.74 9.6e-09
varQ(60,299] -1.041     0.353    0.132 -7.90 2.8e-15

Likelihood ratio test=74.4  on 3 df, p=4.44e-16
n= 1177, number of events= 514
```

ln(Hazard of death at time t) =$\boldsymbol{\lambda_0(t)-0.399Q_2 - 0.685Q_3 – 1.041Q_4}$

Linearity assumption is satisfied since the coefficients decrease with higher quartiles.

Example 2: Binary explanatory variable: Gender and survival

In Cox regression the HR of a binary explanatory variable indicates the risk of dying within one category, compared to the risk of dying within the other category which is used as a reference group.

The explanatory variable: Gender (SEX).

$$SEX = \begin{cases} 0, & \text{male} \\ 1, & female \end{cases}$$

We consider male patients (SEX=0) as the reference group.

Since the explanatory variable is categorical this association could be investigated with the log-rank test that we discussed in the introductory course.

Survival plot (Kaplan-Meier curves) with gender

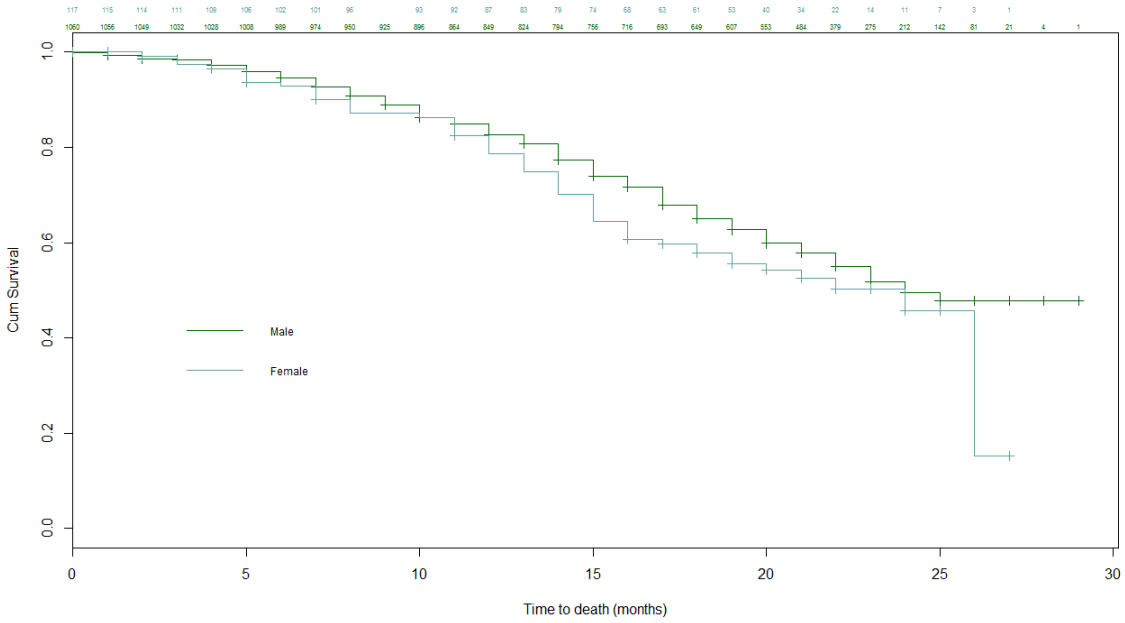


Figure: Kaplan-Meier curves by gender

```
Call:
survdiff(formula = Surv(dthtime, dthstat) ~ sex, data = mac)

         N Observed Expected (O-E)^2/E (O-E)^2/V
sex=0 1060      461    471.2      0.22      2.65
sex=1  117       53     42.8      2.42      2.65

 Chisq= 2.7  on 1 degrees of freedom, p= 0.103
```

According to this analysis, survival was not significantly different in males and females (Log-rank test p=0.103). However, male patients seemed to have a bit better survival than females by looking in the Kaplan-Meier curves (male curve above the female curve). By fitting a Cox model we get the following:

Model

We have the following Cox model equation:

$\lambda(t) = \lambda_0(t) * \exp(\beta 1 * SEX)$

The results when fitting a Cox model to the data are the following

```
Call:
coxph(formula = Surv(dthtime, dthstat) ~ sex, data = mac)
  n= 1177, number of events= 514

      coef exp(coef) se(coef)     z Pr(>|z|)
sex 0.2361    1.2663   0.1453 1.625    0.104

    exp(coef) exp(-coef) lower .95 upper .95
sex     1.266     0.7897    0.9525     1.684

Concordance= 0.51  (se = 0.007 )
Rsquare= 0.002   (max possible= 0.997 )
Likelihood ratio test= 2.48  on 1 df,   p=0.1152
Wald test            = 2.64  on 1 df,   p=0.1042
Score (logrank) test = 2.65  on 1 df,   p=0.1034
```

Interpretation

We notice that coefficient $\hat{\beta}_1 = 0.236$ is positive, so the risk of death is positively associated with gender, meaning that the risk of death is higher in females (codes as 1) than males (codes as 0, reference category).

The HR of coefficient $\hat{\beta}_1$ is $e^{0.236} = 1.266$ meaning that the risk of death is almost 1.3 times higher in female than in male patients. The related p-value is 0.104 which means that this is not a statistically significant result. The corresponding 95% confidence interval is (0.952, 1.683) does contain 1, which also indicates that the result is not significant. Therefore, both analyses, log-rank test and Cox regression yielded similar results (p=0.103 and p=0.104, respectively), that the survival did not differ significantly between female and male patients. Note that a log-rank test will always end up with the same results as simple Cox regression with a categorical explanatory variable.

Example 3: Categorical explanatory variable with more than two categories: Antibiotic use and survival

The explanatory variable here has three categories so we need to create dummy variables for each of these categories. With dummy coding we recode the original categorical variable into a set of binary variables that have values of one or zero meaning whether or not the original variable has that particular category value respectively.

We are including all the categories to the Cox regression model except one which is going to be used as the reference group.

In Cox regression, the HRs of a dummy variable is the risk of death within that category of the explanatory variable compared to the risk of death within the reference category.

The explanatory variable: Antibiotic use (ANTIBIOTIC).

$$ANTIBIOTIC = \begin{cases} 1, & rifabutin \\ 2, & clarithromycin \\ 3, & combination\ of\ both \end{cases}$$

We consider the combination of both antibiotics as the (ANTIBIOTIC=3) as the reference group.

Again since the explanatory variable is categorical this association could be investigated with the log-rank test that we discussed in the introductory course.

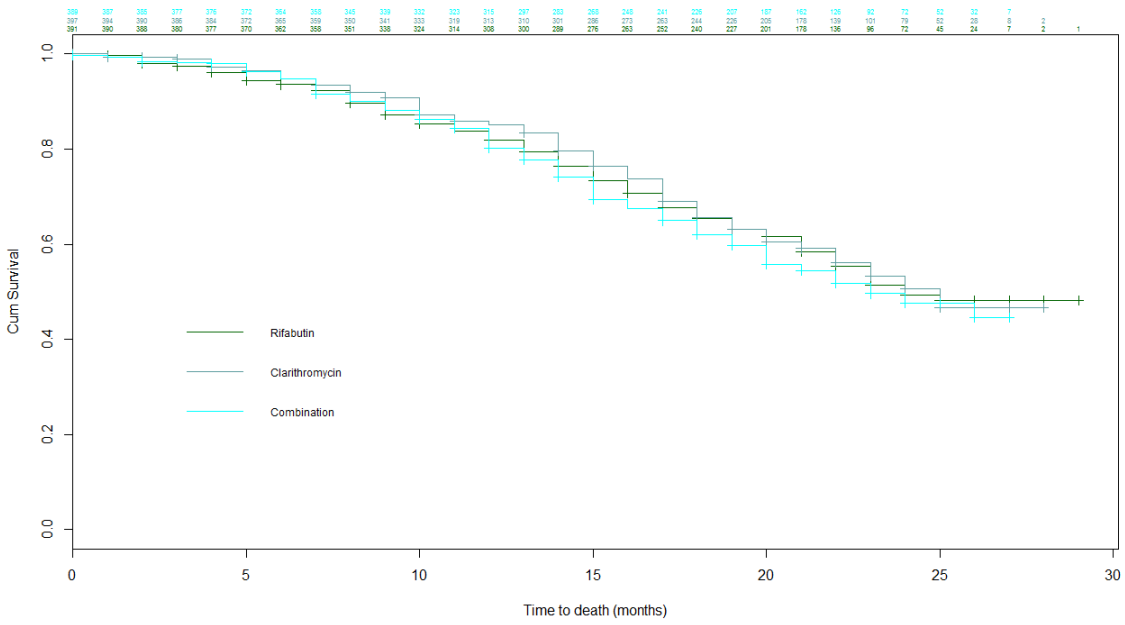


Figure: Kaplan-Meier curves by antibiotic therapy

```
Call:
survdiff(formula = Surv(dthtime, dthstat) ~ antibiotic, data = mac)

                N Observed Expected (O-E)^2/E (O-E)^2/V
antibiotic=1 391      168      171    0.0681     0.102
antibiotic=2 397      167      175    0.3829     0.582
antibiotic=3 389      179      167    0.8050     1.196

 Chisq= 1.3  on 2 degrees of freedom, p= 0.533
```

Model

We have the following Cox model equation:

$\lambda(t) = \lambda_0(t) * \exp(\beta 1 * \text{ANTIBIOTIC})$

```
Call:
coxph(formula = Surv(dthtime.m, dthstat) ~ antibiotic, data = mac)

  n= 1177, number of events= 514

                coef exp(coef) se(coef)      z Pr(>|z|)
antibiotic1 -0.08922   0.91465  0.10744 -0.830    0.406
antibiotic2 -0.11904   0.88777  0.10760 -1.106    0.269

            exp(coef) exp(-coef) lower .95 upper .95
antibiotic1    0.9146      1.093     0.741     1.129
antibiotic2    0.8878      1.126     0.719     1.096

Concordance= 0.518  (se = 0.013 )
Rsquare= 0.001   (max possible= 0.997 )
Likelihood ratio test= 1.33  on 2 df,   p=0.5151
Wald test            = 1.34  on 2 df,   p=0.5122
Score (logrank) test = 1.34  on 2 df,   p=0.5119
```

Interpretation

We notice that both coefficients for the antibiotic use (-0.089 for “rifabutin” and -0.119 for “clarythromycin”) are negative, so risk of death is negatively associated with each of these two categories.

The HR of the coefficient for rifabutin users is $e^{-0.089} = 0.914$ meaning that the risk of death is (0.914 - 1 = -0.086) approximately 8% lower in patients receiving rifabutin compared to patients receiving the combination therapy. Similarly, the HR of the coefficient for clarithromycin users is $e^{-0.119} = 0.887$ meaning that the risk of death is (0.887 - 1 = -0.113) approximately 11% lower in patients receiving clarithromycin compared to patients receiving the combination therapy. The related p-values are 0.406 and 0.269 for rifabutin and clarithromycin users respectively, meaning that the results are non-significant. The corresponding 95% confidence intervals for rifabutin and clarithromycin users are (0.741, 1.129) and (0.719, 1.096) respectively. Both confidence intervals contain 1 which also implies the non-significance of the results.

Multiple Cox model

In building the Cox regression model, as in multiple linear regression, there are a number of different methods for including covariates in the model. Such as automated stepwise methods (forward or backward selection method) or hierarchically. The inclusion or removal of variables is based on the corresponding statistics calculated. As with multiple linear regressions, it is important that both the clinical and statistical significance of variables be considered in building a parsimonious model. However, we do not recommend 'parsimonious' models that only include predictors that are statistically significant at P < 0.05 or even stricter criteria, because the potential for residual confounding in such models is substantial.

Usually a univariate variable screening with P<0.20, is used for building the multivariable model, in order to protect against residual confounding.

Both the backward elimination method and the forward method resulted in the same multivariable model.

Backward elimination: the last model is the final

```
Start:  AIC=6672.64
Surv(dthtime.m, dthstat) ~ age + sex + cd4 + karnof + antiret +
    antibiotic

             Df    AIC
- antibiotic  2 6669.3
<none>          6672.6
- antiret     1 6674.5
- sex         1 6674.7
- age         1 6688.2
- karnof      1 6721.5
- cd4         1 6724.9

Step:  AIC=6669.27

Surv(dthtime.m, dthstat) ~ age + sex + cd4 + karnof + antiret
          Df    AIC
<none>       6669.3
- antiret  1 6671.2
- sex      1 6671.5
- age      1 6685.2
- karnof   1 6719.0
- cd4      1 6721.3
```

Forward selection: the last model is the final

```
Start:  AIC=6812.49
Surv(dthtime.m, dthstat) ~ 1
             Df    AIC
+ karnof      1 6737.4
+ cd4         1 6747.4
+ age         1 6792.2
+ antiret     1 6810.2
+ sex         1 6812.2
<none>          6812.5
+ antibiotic  2 6815.2

Step:  AIC=6737.38
Surv(dthtime.m, dthstat) ~ karnof
             Df    AIC
+ cd4         1 6689.6
+ age         1 6724.7
+ antiret     1 6734.5
+ sex         1 6737.0
<none>          6737.4
+ antibiotic  2 6740.7

Step:  AIC=6689.6
Surv(dthtime.m, dthstat) ~ karnof + cd4
             Df    AIC
+ age         1 6672.9
+ antiret     1 6686.5
+ sex         1 6689.0
<none>          6689.6
+ antibiotic  2 6692.4

Step:  AIC=6672.91
Surv(dthtime.m, dthstat) ~ karnof + cd4 + age
             Df    AIC
+ sex         1 6671.2
+ antiret     1 6671.5
<none>          6672.9
+ antibiotic  2 6676.1

Step:  AIC=6671.2
Surv(dthtime.m, dthstat) ~ karnof + cd4 + age + sex
             Df    AIC
+ antiret     1 6669.3
<none>          6671.2
+ antibiotic  2 6674.5

Step:  AIC=6669.27
Surv(dthtime.m, dthstat) ~ karnof + cd4 + age + sex + antiret
             Df    AIC
<none>          6669.3
+ antibiotic  2 6672.6
```

Reporting the results of Cox regression

Table: Cox regression results for explanatory variables of mortality

		Univariate			Multivariable		
		HR	95%CI	p-value	HR	95%CI	p-value
Age in years		1.024	1.014, 1.034	<0.001	1.021	1.012, 1.031	<0.001
Gender							
	Male	1					
	Female	1.253	0.942, 1.665	0.12	1.367	1.027, 1.820	0.032
CD4 cell count/μL		0.988	0.985, 0.992	<0.001	0.990	0.987, 0.993	<0.001
Karnofsky score		0.955	0.946, 0.965	<0.001	0.963	0.953, 0.973	<0.001
Injection drug use							
	Previous/Current	1.108	0.872, 1.409	0.4			
	Never	1					
Antiretroviral experience							
	Previous/Current	0.813	0.671, 0.987	0.0362	0.818	0.673, 0.995	0.044
	Never/Unknown	1					
Antibiotic treatment							
	Rifabutin	0.916	0.742, 1.131	0.417			
	Clarithromycin	0.971	0.783, 1.202	0.285			
	Combination	1					

*HR: hazard ratio, CI: confidence intervals

According the above results the independent variables of mortality where age, gender CD4 cell count, Karnofsky score and antiretroviral therapy. The interpretation of the variables is similar to the simple Cox regression, e.g. for variable gender we can say that the risk of death was 1.3 times higher in women than men (p=0.032), adjusted for all the other variables in the model. All the other variables can be explained similarly. Below follows a table that presents the results inaccurately.

Table: Cox regression results for explanatory variables of mortality wrongly presented

		Univariate			Multivariable		
		coef	SE	p-value	coef	SE	p-value
Age in years		0.0237	0.0049	0.0000	0.0212	0.0049	0.0000
Gender							
	Male						
	Female	0.2252	0.1453	0.121	0.3127	0.1460	0.0322
CD4 cell count/µL		-0.0116	0.0015	0.0000	-0.0105	0.0016	0.0000
Karnofsky score		-0.0454	0.0051	0.000	-0.0377	0.0051	0.0000
Injection drug use							
	Previous/Current	0.1028	0.1223	0.4			
	Never						
Antiretroviral experience							
	Previous/Current	-0.2064	0.0985	0.036	-0.2004	0.0996	0.0442
	Never/Unknown						
Antibiotic treatment							
	Rifabutin	-0.087	0.107	0.417			
	Clarithromycin	-0.115	0.108	0.285			
	Combination	1					