Contents

Contents

2

1 Modelling Time to event data

1 Modelling Time to event data

1.1 Introduction

Survival (or "time-to-event") analysis is concerned with the length of time that passes before a particular event occurs. It was initially developed for events that can only occur once, such as death, and that is the only type of event we will consider here. However, it can be extended to cover events that can happen repeatedly.

The main issue with time to event studies is censoring. That is, some people may not have had the event yet, so we don't know how long will have elapsed when it happens. We need to have a way to include such subjects in our analysis.

In the following chapter, we will explain what censoring is, then present methods for:

1. Presenting the survival experience of a group

2. Comparing the survival experiences of two groups

3. Measuring the impact of multiple predictors on survival

1.2 Censoring and Survival

If we do not know the exact time that an event occurred, we say that the event time is censored. In survival analysis, we are most commonly interested in the case where the event hasn't happened yet: that individual is still alive. We know that the event will happen, but not the time at which it will happen. This is commonly known as right-censoring: we think of time as running along an axis from left to right, and we know that the event will happen somewhere to the right of where we are now, but not how far to the right.

Right censoring commonly occurs because we wish to analyse our data before every subject in the sample has died. It can also occur if a subject is lost to followup, but were alive when we last had contact with them. The methods presented here can allow for right censoring, provided that the censoring is unrelated to the time to event. I will give an example shortly in which this assumption is unlikely to be true, and a more sophisticated analysis may be called for.

Figure 1.1 shows an example of a study which was planned to run for 18 months. The subjects were recruited during the first 6 months of the study, and followed up for between 12 and 18 months, depending on when they were recruited. The data was analysed as it was 18 months after the study began. In Figure 1.1, filled circles represent the event of interest (subjects 1, 5, 7 and 10) and open circles represent censoring (subjects 2 and 4 were lost to followup, subjects 3, 6, 8 and 9 had not had the event at the end of the study).

However, when analysing the data, we need time for each subject to start when they are recruited into the study, not when the study begins. Hence the data for analysis will look like Figure 1.2. Everyone starts at time 0, and although subjects 3, 6, 7 and 8 were all censored a the same time chronologically, they are censored at different points in their followup since their followup started at different times chronologically.

1.2.1 Other types of censoring

Survival analysis is intended to deal with right censoring. However, there are other types of censoring that require different approaches to deal with them. They are left censoring and interval censoring.

Left censoring happens when you know that the event has taken place, but you do not know when. For example, vertebral fractures are often asymptomatic, and can only be diagnosed
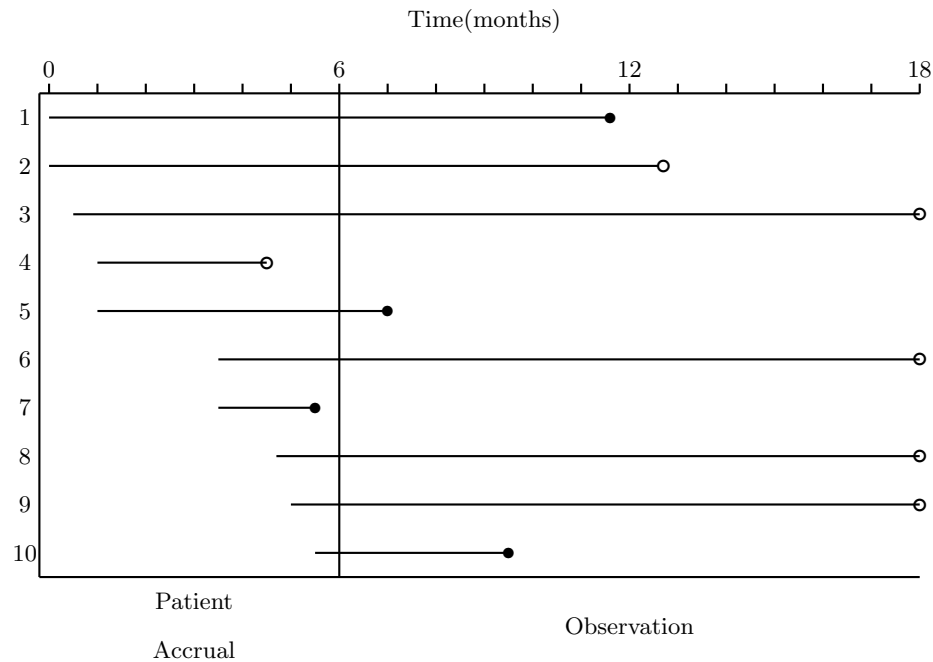
Time(months)

Figure 1.1: Censoring example: chronological time

when a spinal radiograph is taken. If a vertebral fracture is apparent on the radiograph, you know that the fracture happened, but not when it happened.

Interval censoring is a combination of right and left censoring. You know that the event happened within a given time interval, but not exactly when. Continuing with the vertebral fracture example, a patient may have 2 radiographs taken several years apart. There may be a fracture visible on the second radiograph that was not visible on the first. You therefore know the fracture happened some time during the interval between the two radiographs, but not exactly when.

1.3   Describing Survival

1.3.1   Survivor function

Describing survival does not lend itself easily to being reduced to a single number. It is most commonly presented as a graph of the Survival Function $(S(t))$: the probability of surviving to time t, plotted against time.

To see how to calculate the Survival Function, it is easiest to think of time as passing in discrete chunks, which I'll refer to as "ticks". Suppose that r_k subjects reach the kth tick, and f_k of them have the event during that tick. Then the probability of surviving the kth tick, for those that reach it, is $\frac{r_k - f_k}{r_k}$. The probability of surviving to the beginning of the kth tick is $S(k-1)$, so $S(k) = S(k-1) \times \frac{r_k - f_k}{r_k}$.

Note that if no-one has the event, or is censored, at time k, then $S(k) = S(k-1) \times 1$, and hence the Survival Function remains unchanged. So we only need to recalculate the function at times when an event or censoring happens.
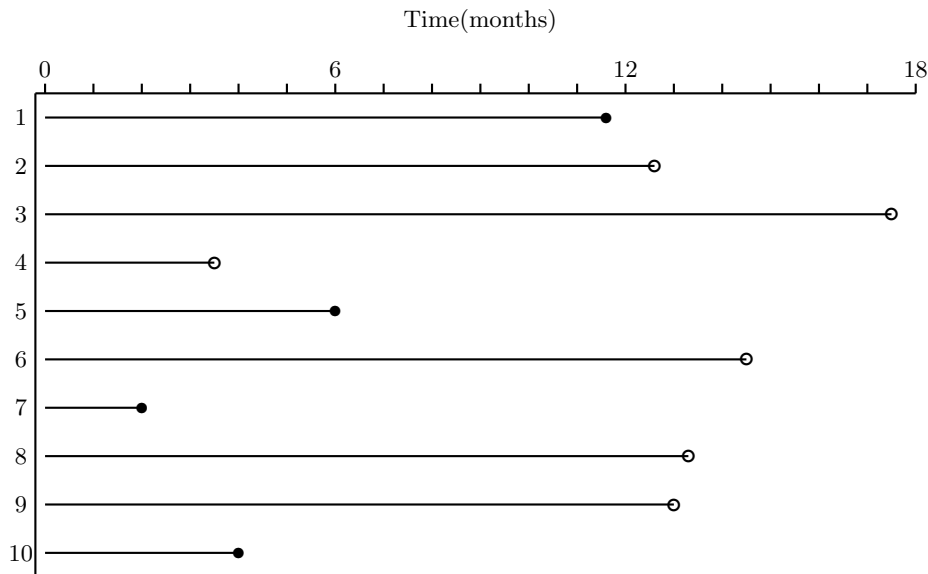
Time(months)

Figure 1.2: Censoring Example: followup time

Survivor Function Example

In an experiment to investigate motion sickness for the U.S. Navy **??**, 21 subjects were put into a cabin on a hydraulic piston, and bounced up and down for 2 hours, or until they vomited, whichever came first. 5 subjects were ill, after 30, 50, 51, 82 and 92 minutes. In addition, 2 subjects insisted on ending the experiment early, after 50 and 66 minutes. It is possible, even likely, that these 2 subjects were likely to vomit sooner than the others had they continued with the experiment, and hence censoring is not independent of survival time in this case. However, for the sake of a simple example, we will proceed as if it were independent: methods for dealing with "informative" censoring are complicated and beyond our scope.

So the first time we need to consider is 30 minutes, when the first subject vomited. At this point, 20 of the 21 subjects were surviving, so $S(30) = \frac{20}{21} = 0.952$. One of the remaining subjects vomited after 50 minutes, so $S(50) = \frac{20}{21} \times \frac{19}{20} = 0.905$. However, the first censoring occured at this time, so the probability of surviving the 51st minute is $\frac{17}{18}$, not $\frac{18}{19}$, and $S(51) = 0.905 \times \frac{17}{18} = 0.855$.

The survival function is tabulated in Table 1.3.1. Such a table is commonly referred to as a life table. The same information can also be presented as a graph, with time running left to right on the x-axis and survival probability on the Y-axis. The graph is drawn as a step function, with a step downwards at each failure time, the size of the step depending on the number of subjects who fail at that time. An example of a survival curve is shown in Figure 1.3. This approach to calculating the survival function is often referred to as the Kaplan-Meier method, since they first proposed it [**?**].

1.3.2   Stata Commands

Stata has a large suite of commands for doing survival analysis, all of which begin with the letters `st`. First, the data needs to be set as survival data using the `stset` command. This takes a single variable containing the failure time. There is also a `failure()` option, which should be

ID	Time	Censored	r_k	f_k	$S(t)$	
1	30	No	**21**	**1**	**20/21**	**= 0.952**
2	50	No	**20**	**1**	**19/20 × S(30)**	**= 0.905**
3	50	Yes	**19**	**0**	**19/19 × S(50)**	**= 0.905**
4	51	No	**18**	**1**	**17/18 × S(50)**	**= 0.855**
5	66	Yes	**17**	**0**	**17/17 × S(51)**	**= 0.855**
6	82	No	**16**	**1**	**15/16 × S(66)**	**= 0.801**
7	92	No	**15**	**1**	**14/15 × S(82)**	**= 0.748**
8	120	Yes	**14**	**0**	**14/14 × S(92)**	**= 0.748**
⋮						
21	120	Yes	**14**	**0**	**14/14 × S(92)**	**= 0.748**

a variable containing the value 1 for failures and 0 for censorings.

Having set the data with `stset`, a life table can be generated with `sts list` and a survival curve produced with `sts graph`. An example of the output from `sts list` is given below: Figure 1.3 was produced with `sts graph`.

```
sts list if group == 1

        failure _d:  fail
  analysis time _t:  time


          Beg.           Net     Survivor      Std.
  Time    Total   Fail   Lost    Function     Error      [95% Conf. Int.]
  -------------------------------------------------------------------------
    30      21     1      0       0.9524      0.0465     0.7072    0.9932
    50      20     1      1       0.9048      0.0641     0.6700    0.9753
    51      18     1      0       0.8545      0.0778     0.6133    0.9507
    66      17     0      1       0.8545      0.0778     0.6133    0.9507
    82      16     1      0       0.8011      0.0894     0.5519    0.9206
    92      15     1      0       0.7477      0.0981     0.4946    0.8868
   120      14     0     14       0.7477      0.0981     0.4946    0.8868
  -------------------------------------------------------------------------
```

1.4 Comparing Survival Between Groups

If we are interested in the survival experience of two (or more) groups, we can draw survival curves for each group separately. However, testing for a difference between the groups is trickier than testing for a difference in prevalence or rate, since each of those can be summarised with a single number for each group. A survival curve is not a single number, so there are a number of ways the two groups can differ: there may be a consistent difference, the groups may be the same at first then separate, they may differ at first then continue parallel, the survival curves may even cross, with one group having better survival initially then being overtaken by the other group.

There are a number of different tests that can be used to compare groups, a the power of each test depends on the type of difference that exists. For instance, the logrank test is typically most powerful against consistent differences between the groups, whilst the modified Wilcoxon test is most powerful against early differences. For this reason, it is essential to decide which test to use *before* looking at the data, otherwise you risk a false positive test result.
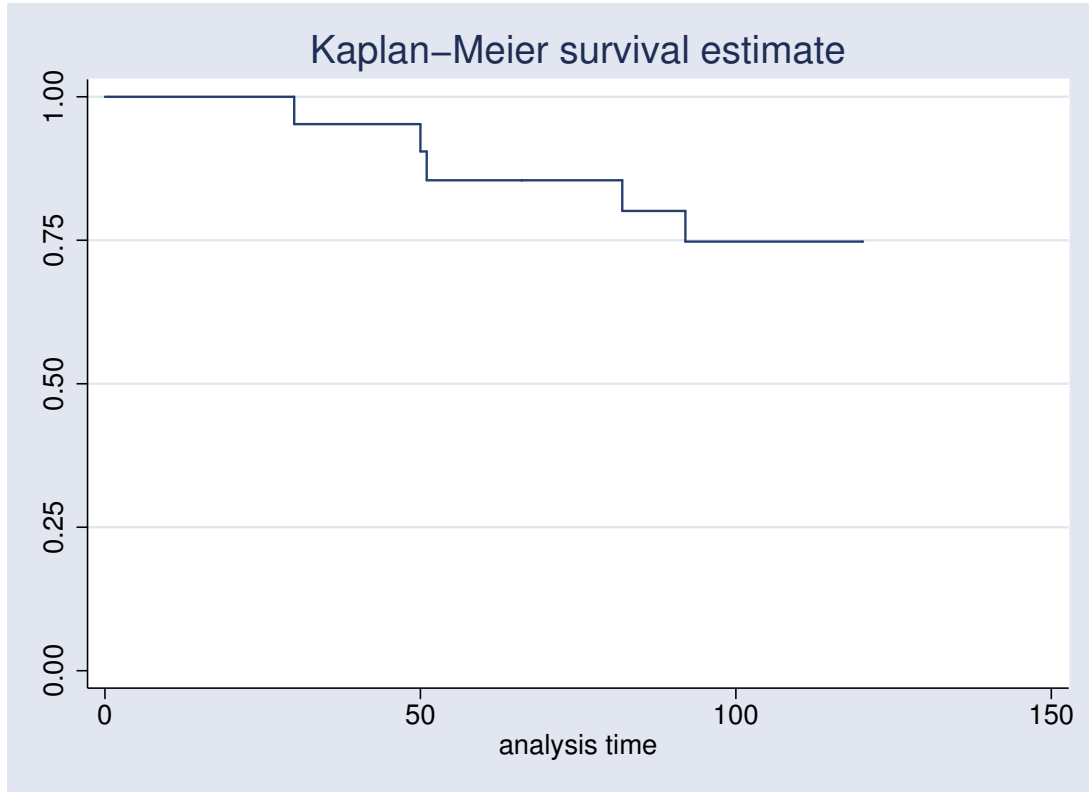
Figure 1.3: Kaplan Meier Survival Curve for Motion Sickness

1.5   Modelling Survival

Normally, modelling provides more information than hypothesis testing, and survival analysis is no exception. If you want to compare two groups while adjusting for potential confounders, or you are interested in the effects of different predictors on survival, modelling is the way to go.

1.5.1   The hazard function

Rather than modelling the survival function, we model the hazard function. The hazard function $h(t)$ is the probability of having the event of interest at time t, assuming that you have reached time t. For example, in the motion sickness study, $h(30) = \frac{1}{21}$ and $h(50) = \frac{1}{20}$. The hazard function is modelled as a function of both time t and the predictors we are interested in $\mathbf{x}$:

$$h(t, x) = h_0(t) \times \Psi(\mathbf{x})$$

The baseline hazard, $h_0(t)$, may be estimated parametrically or non-parametrically. Since $\Psi(\mathbf{x})$ is always parametric, if $h_0(t)$ is also parametric, the model is parametric, if not the model is semi-parametric. In medical research, survival analysis almost always uses a semi-parametric model, and that is all that we will consider here.

1.5.2   Cox Regression

The most common form of semi-parametric regression is Cox regression. Here, $h_0(t)$ is piecewise constant, and $\Psi(\mathbf{x})$ takes the form $e^{\beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k}$. Note that the linear predictor does not

contain an intercept, since the role of the intercept is played by $h_0(t)$. In a Cox regression model, if x_i increases by 1, the hazard is multiplied by e^{β_i}, which is referred to as the *hazard ratio*.

The stata command for performing Cox regression is `stcox`. You need to have `stset` your data before you can use it. By default, the command will produce hazard ratios, but there is a `nohr` option which will give coefficients instead.

Cox Regression Example

In the motion sickness study in Section 1.3.1, less than a third of the subjects vomited. The experimenters wanted to see if increasing the amplitude and frequency of movement induced vomiting more quickly. They recruited a furthe 28 subjects and subjected them to this new motion. They then compared the time to vomiting between the two groups. The survival curves are shown in Figure 1.4, the output from the `stcox` command is below.
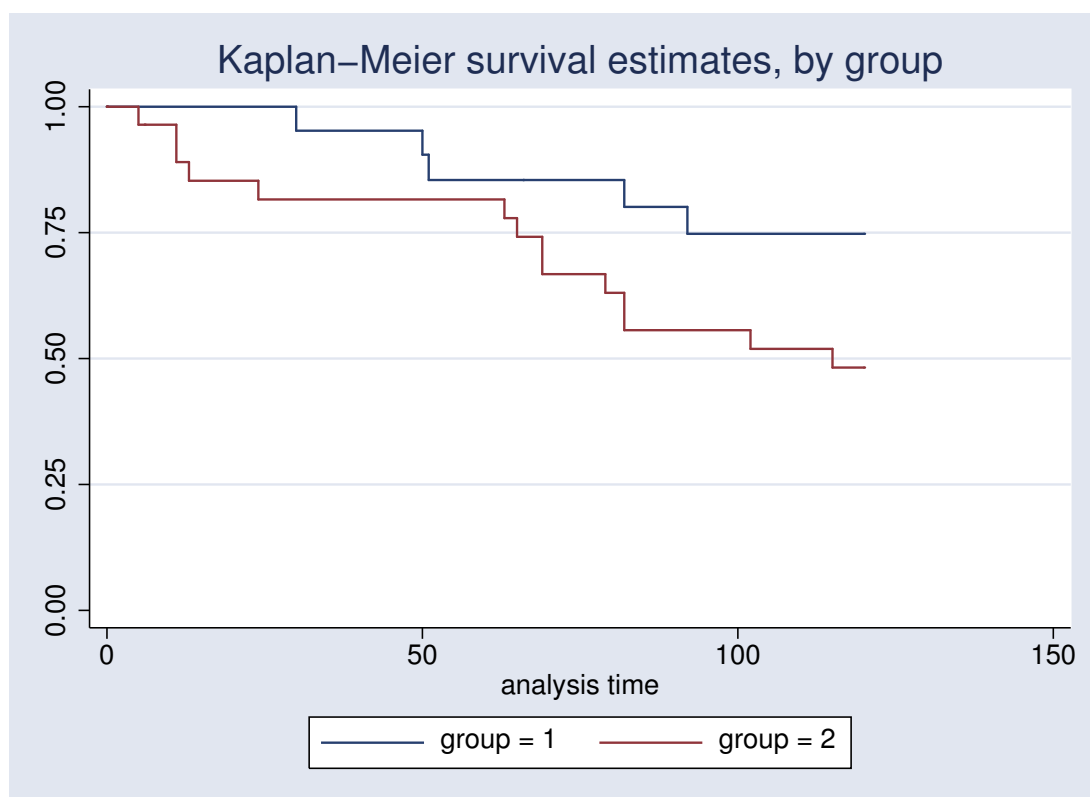

Figure 1.4: Comparing Two Groups in Motion Sickness Study

```
. stcox i.group

Cox regression -- Breslow method for ties

No. of subjects =              49                 Number of obs   =         49
No. of failures =              19
Time at risk    =            4457
                                                  LR chi2(1)      =       3.32
Log likelihood  =   -67.296458                    Prob > chi2     =     0.0685


------------------------------------------------------------------------------
        _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------+------------------------------------------------------------------
   2.group |    2.45073    1.277744     1.72   0.086     .8820678    6.809087
------------------------------------------------------------------------------
```

Figure suggests that subjects did vomit sooner under the new conditions: the survival curve for group 2 is always lower than the survival curve for group one. However, the *p*-value for the comparison is greater than 0.05, so the difference between the groups is not statistically significant. The hazard ratio for group 2 of 2.45 suggests that this group does vomit sooner, but the confidence interval of (0.88, 6.81) contains the null value of 1, so we can't be confident that the true population value of the hazard ratio is greater than 1.

1.5.3   Proportional Hazards Assumption

Cox regression assumes that the hazard ratio is constant over time, in other words that the hazard functions for 2 sets of predictor values are in a constant proportion over time. This is a strong assumption and should be tested. It can be formally tested with a hypothesis test, or informally tested graphically.

The easiest way to test it graphically is with the command `stcoxkm`. This produces the observed survival curve, along with the expected survival curve based on the Cox regression model. If the two curves are similar, the proportional hazards assumption is probably corrrect. Figure 1.5 shows this plot for the motion sickness data.
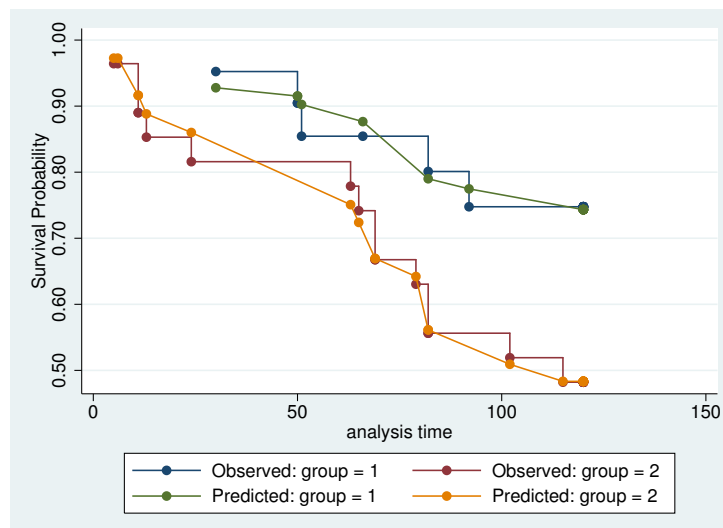

Figure 1.5: Expected and Observed Kaplan Meier Plots for Motion Sickness Data

As you can see, the observed and expected values are very similar for both groups, suggesting

that the proportional hazards assumption holds. This can be confirmed by running `estat phtest` after `stcox`. The output below shows that the *p*-value is greater than 0.05, so we cannot reject the null hypothesis of proportional hazards.

```
. estat phtest

    Test of proportional hazards assumption
    -----------------------------------------------
               |    chi2      df       Prob>chi2
    -----------+-----------------------------------
    global test |    0.03      1          0.8585
    -----------------------------------------------
```

Allowing for non-proportional hazards

What can we do if the proportional hazards assumption does not hold ? This would mean that the hazard ratio for one or more variables changes over time, so we can split the data by time and fit two or more different hazard ratios for the same variable at different times.

This is easiest to see with an example: this is taken from the Stata help for `stcoxkm`. It involves comparing two treatments. The observed and expected Kaplan-Meier curves are shown in Figure 1.6. The observed and expected curves for both treatments are quite distinct from each other: for the standard treatment, the observed curve starts above the predicted curve, crosses it at about 10 months and then remains below it. The reverse is true for Drug B, the observed curve starts below the predicted curve, crosses it at about 12 months and then remains above it. This suggests that the hazard ratio for Drug B is changing over time. This is confirmed by running `estat phtest` after `stcox`: the test is highly significant ($p = 0.0013$).
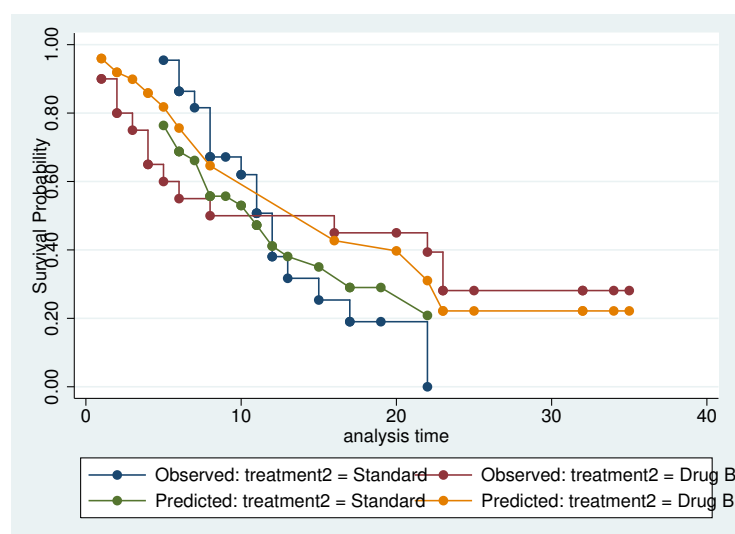

Figure 1.6: Example of Observed and Predicted KM-curves when Hazards are not Proportional

1 Modelling Time to event data

```
. stcox i.treatment2


------------------------------------------------------------------------------
        _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------+------------------------------------------------------------------
 treatment2 |  .7462828    .3001652    -0.73   0.467     .3392646    1.641604
------------------------------------------------------------------------------

. estat phtest

    Test of proportional hazards assumption

    Time:  Time
    ----------------------------------------------------------------
              |                 chi2       df       Prob>chi2
    ----------+-----------------------------------------------------
    global test |               10.28        1         0.0013
    ----------------------------------------------------------------
```

Since the curves seem to cross at 10 months, we can divide the data at that point. There is a command **stsplit** which will do that for us. The option **at()** takes a list of times at which to split the data, and the command needs to be given a variable which will contain the time of the start of that period after the command has been run. After splitting the data, we create two new variables: one corresponding to **treatment2** in the first 10 months, the other corresponding to **treatment2** after 10 months. The code to create and fit these variables is given below, along with the output from **stcox** and **estat phtest**.

```
stsplit period, at(10)
gen t1 = treatment2*(period == 0)
gen t2 = treatment2*(period == 10)

. stcox t1 t2


------------------------------------------------------------------------------
        _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------+------------------------------------------------------------------
        t1 |  1.836938    .8737408     1.28   0.201     .7231357    4.666262
        t2 |  .1020612    .0853529    -2.73   0.006     .0198156    .5256703
------------------------------------------------------------------------------

. estat phtest

    Test of proportional hazards assumption

    Time:  Time
    ----------------------------------------------------------------
              |                 chi2       df       Prob>chi2
    ----------+-----------------------------------------------------
    global test |                1.34        2         0.5114
    ----------------------------------------------------------------
```

The hazard ratio is greater than 1 (but not significantly so) for the first 10 months, and significantly less than one after this). This suggests that the treatment is no better than standard treatment at first, but for those who survive at least 10 months, it is then markedly better.

2 Modelling Time to event data: Practical

2.1   Practical 10: Survival Analysis

Datasets

The datasets that you will use in this practical can be accessed via http from within stata. However, the directory in which they are residing has a very long name, so you can save yourself some typing if you create a global macro for this directory. You can do this by entering

```
global basedir http://personalpages.manchester.ac.uk/staff/mark.lunt
global datadir $basedir/stats/10_survival/data
```

(In theory, the global variable `datadir` could have been set with a single command, but fitting the necessary command on the page would have been tricky. Far easier to use two separate commands as shown above). If you wish to run the practical on a computer without internet access, you would need to:

1. Obtain copies of the necessary datasets

2. Place them in a directory on your computer

3. Define the global macro `$datadir` to point to this directory.

2.1.1   Life tables and Survival Curves

This section uses the dataset `"leukaemia"`.

1.1   First, set up the data for survival analysis. The time variable is `weeks`, the number of weeks to relapse. The outcome variable is `relapse`, which is 1 if the subject had a relapse at that time and 0 if they did not. Hence the command to set the data up for survival analysis is `stset weeks, fail(relapse)`

1.2   Obtain a life table for the subjects on Drug A with the command `sts list if treatment1 == 1`. What is the median survival in this group (at what time does the survivor function reach 0.5) ?

1.3   How many subjects were lost to followup in this treatment arm ?

1.4   Obtain a life table for the subjects on standard treatment with the command `sts list if treatment1 == 0`. What is the median survival in this group ?

1.5   How many subjects were lost to followup in this treatment arm ?

1.6   Do the answers to your previous questions suggest that Drug A is better, worse, or the same as standard treatment ?

1.7   Produce a Kaplan-Meier curve for each of the treatments with the command `sts graph, by(treatment1)`. Does this confirm your answer to the previous question ?

1.8   Add a horizontal line to the graph by adding the option `yline(0.5)` to the previous command. This line represents half of the group surviving and half having a relapse: the point where it crosses the two survival curves should give you the median survival times you calculated in earlier questions.

1.9   Add the option `lost` to the previous command. This will show how many subjects were censored at each time point. How many subjects were lost to followup in the two treatment arms ? Does this agree with the results you got from `sts list` ?

1.10   Add the option `gwood` to the previous command to obtain confidence bands for the survival curve ? (The odd name for this option is because the formulae used to calculate the confidence bands were developed by a Major Greenwood). Why do the confidence bands get wider over time ?

1.11   Perform a logrank test to compare the survival on Drug A to that on standard treatment, with the command `sts test treatment1`. Is the difference between Drug A and standard treatment statistically significant ?

1.12   Would have had the same answer to the previous question if you had used a Wilcoxon test in place of a logrank test ? (You can do this by adding the option `wilcoxon` to the previous command.)

2.1.2   Cox Regression

1.13   Have a look at the survival curves by white blood cell count using `sts graph, by(wbc3cat)`. Does the white blood cell count affect survival ?

1.14   Do a cross-tabulation of `treatment1` against `wbc3cat` with `tab wbc3cat treatment1, co` Are the proportions of subjects in each of the white blood cell counts categories the same in the two treatment arms ?

1.15   Given that proportion of subjects in the "High" cell count group is greater in the standard treatment arm than in the Drug A arm, would you expect this to have increased or decreased survival in this arm of the trial ?

1.16   White blood cell count is a potential confounder, so we need to adjust for it. First, we will perform an unadjusted Cox regression to obtain the hazard ratio before adjusting. This is done with the command `stcox treatment1`. What is the hazard ratio for Drug A, and its 95% confidence interval ?

1.17   Now obtain the adjusted hazard ratio with the command `stcox treatment1 i.wbc3cat`. What is the adjusted hazard ratio and its 95% confidence interval ?

1.18   How did the confounding by white blood cell count affect the apparent effect of Drug A ? Is this what you expected from the earlier questions ?

1.19   Now we need to test the proportional hazards assumption. First for treatment: produce a plot of the observed and predicted Kaplan Meier plots with `stcoxkm, by(treatment1)`. Are the observed and predicted curves close to each other ?

1.20   Now we can test the same assumption for the effect of white blood cell count, with `stcoxkm, by(wbc3cat)`. Are the observed and predicted curves close to each other ?

1.21   To obtain a formal test, we need to store the scaled and unscaled Schoenfeld residuals by running the command `stcox treatment1 i.wbc3cat, sca(sca*) sch(sch*)` Now enter the command `stphtest` to get an overall test of proportionality. Is the regression model valid ?

1.22   Use the command `stphtest, detail` to obtain tests of proportionality for each individual variable. Is there any evidence of non-proportional hazards ?

2.1.3   Non-Proportional Hazards

1.23   There is a second drug used in this trial, stored in `treatment2`.   Compare the survival curves for Drug B and standard treatment with the command `sts graph, by(treatment2)` How does the survival on Drug B compare to that on standard treatment during the first 10 weeks ?

1.24   How does the survival on Drug B compare to that on standard treatment after the first 10 weeks ?

1.25   Superimpose the predicted survival curves from the Cox regression model with `stcoxkm, by(treatment2)`. How do the predicted and observed curves differ ?

1.26   Perform a Cox regression of `treatment2` and `wbc3cat` with `stcox treatment2 i.wbc3cat` Does Drug B have a significant effect on survival ?

1.27   To test the proportional hazards assumption, we need to store the Schoenfeld residuals again. First drop the residual from the previous model with `drop sca* sch*` Then rerun the `stcox` command with the options `sca(sca*) sch(sch*)`. Perform the overall test with `stphtest`: is the model appropriate ?

1.28   Test the proportional hazards assumption for each variable separately with `stphtest, detail`. Which variable does not satisfy the assumption ?

1.29   The Kaplan-Meier curves suggest that Drug B has a negative effect on survival initially, then becomes positive. So we will test for different effects before and after 10 weeks. First produce a life-table with `sts list`

1.30   To be able to split the data, you need to have an id for each subject. We can do this with `generate id = _n`. Now each observation has its record number as an identifier.

1.31   Now, for each subject followed for more than 10 weeks, we will split the data into 2 observations, one for the time up to 10 weeks and one for the time after. First, we must include the id information in the stset command with `stset weeks, fail(relapse) id(id)` Then we can split the data with `stsplit split_time, at(10)`

1.32   Check that the life-table remains unchanged by entering the command `sts list` Is it the same as before ?

1.33   Examine the data with `list id weeks relapse split_time _t0 _t`. You should see that for subjects who were followed up for less than ten weeks, there is still a single record. However, for those followed up for more than 10 weeks, there are two records, one with `split_time == 0`, the other with `split_time == 10`. The start of the interval is given by `_t0`, the end by `_t`

1.34   Now we can generate separate treatment variables for the treatment effect before and after 10 weeks. The commands to use are `gen t1 = treatment2*(split_time == 0)` and `gen t2 = treatment2*(split_time == 10)`.

1.35   Now fit the Cox regression model with both `t1` and `t2` as predictors with the command `stcox t1 t2 i.wbc3cat` What is the hazard ratio for t1, with its 95% confidence interval ?

1.36   What is the hazard ratio for `t2` ?

1.37   Do these hazard ratios confirm what you were expecting ?

1.38   Drop the residuals from the previous model with `drop sca* sch*` then create new residuals with `stcox t1 t2 i.wbc3cat, sca(sca*) sch(sch*)` Now test the proportional hazards assumption with `stphtest`. Is the model now appropriate ?

1.39   Test the proportional hazards assumptions for each of the variables separately with `stphtest, detail`. Do any of the predictors show non-proportionality ?

Bibliography