Survival analysis

An alternative medical example: an experiment with fabricated data

A Phase IIB clinical trial is conducted to assess the efficacy of a new anti-cholinergic agent in reducing chronic obstructive pulmonary disease (COPD) dyspenea events in smokers. The trial has three TREATMENT arms: treatments 1 and 2 are low and high doses respectively of the drug, whilst treatment 3 is a placebo. We have 100 patient volunteers, all diagnosed with COPD, and they are assigned at random in the proportions 5 low dose, 3 high dose and 2 placebo. Covariates included in the analysis are NUMBER, the average number of cigarettes smoked per day in the month before the screening date and YEARS, the number of years the subject has smoked cigarettes before the trial. The covariates NUMBER and YEARS will be ignored when we obtain the Kaplan-Meier plots, but we will include them in the Cox Regression Analysis.

Patients are followed for 20 weeks and the number of weeks until their first dyspenea event (TTF) is recorded. Patients who have not experienced any dyspenea events at 20 weeks (censored data) are coded one on the bivariate variable RELAPSE and those with a recorded time to relapse are code zero on RELAPSE. Table 14.1 shows the first four cases for each treatment group, and the full dataset can be found on the book website as med.survival.sav.

Table 14.1
The first four cases from each treatment group in the COPD trial (the full dataset can be found as med.survival.sav on the website)

years	number	ttf	relapse[1]	treatment[2]
11	28	8	0	1
9	38	8	0	1
13	29	20	1	1
13	28	16	0	1

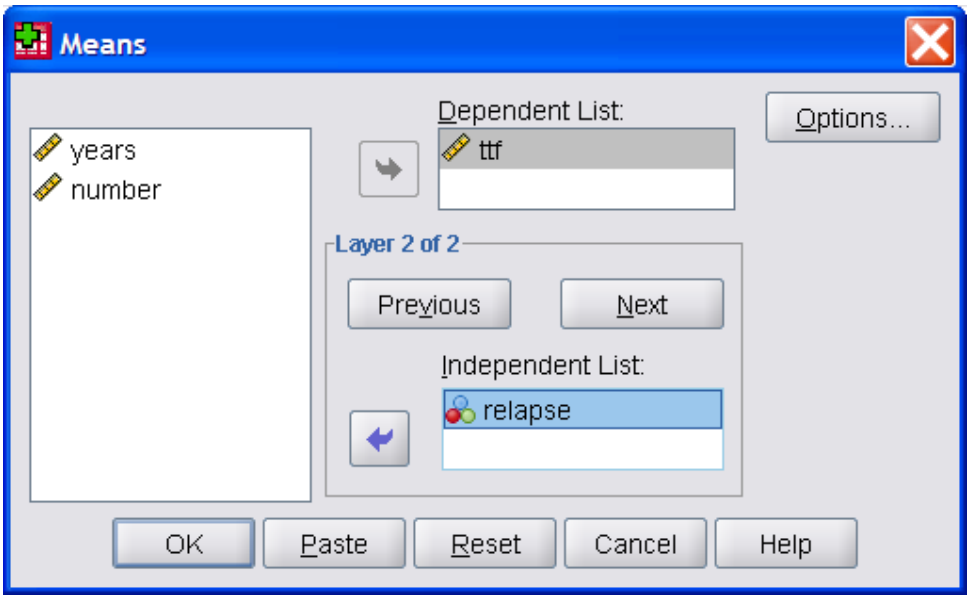
12	30	14	0	2
12	31	5	0	2
12	24	20	1	2
12	29	11	0	2
6	33	4	0	3
10	29	1	0	3
11	30	13	0	3
9	37	4	0	3

[1] relapse: 0 = relapse occurred, 1 = censored (no relapse)
[2] treatment: 1 = low dose, 2 = high dose, 3 = placebo

So, do our data suggest that either treatment is more effective than the placebo? We would expect that if one is better, then participants using it would have longer times to relapse and more censored observations (no dyspenea event by the time the study ends) than those using the others. In fact treatment 2 (high dose) has 14 out of 30 censored cases while treatment 1 (low dose) has 26 out of 50: both close to a half. Only one of those on the placebo is censored, the other 19 all experiencing a dyspenea event within 20 weeks. The uncensored observations have a mean TTF of 10.6 weeks for low dose, 11.4 for high dose and 6.5 for the placebo.

Example box 14.1 about here

You can easily obtain these means by using **Analyze**, **Compare Means**, **Means**. In the dialog box put TTF in the **Dependent List** box and TREATMENT in the **Independent List** box. Click **Next** so that **Layer 2 of 2** replaces **Layer 1 of 1**, and put RELAPSE in the **Independent list** box. Click **OK** to get the table of means. Note that when RELAPSE = 1 (censored cases) the mean is 20 for all treatments, because this is the TTF recorded for censored cases.



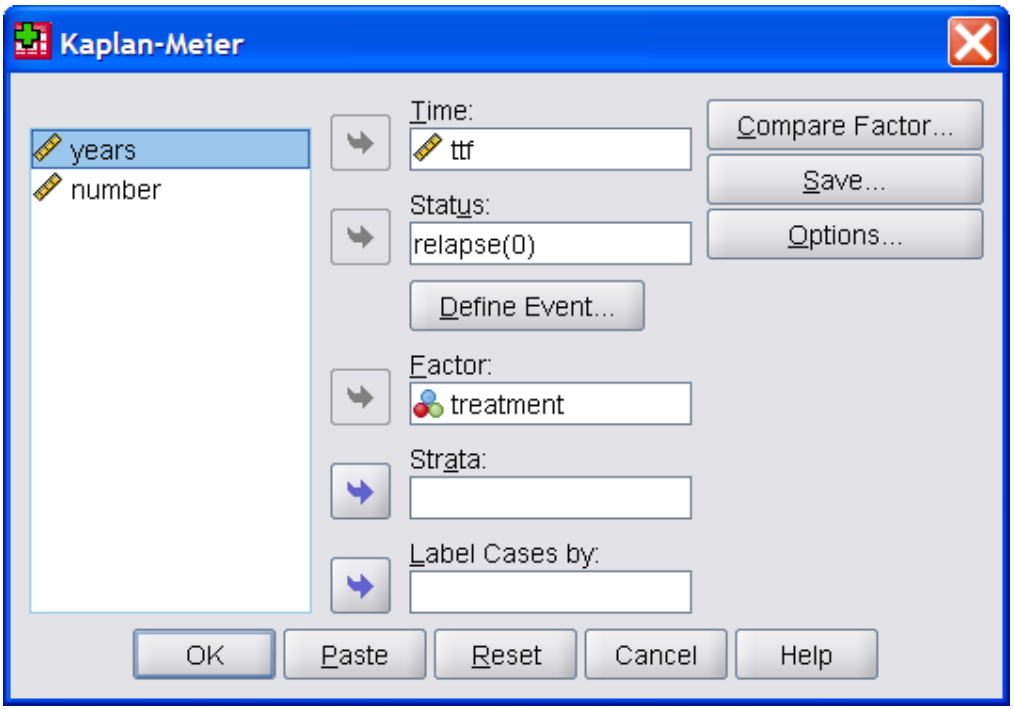
SPSS Dialog Box 14.1. Obtaining treatment group means for uncensored and censored observations

These preliminary results suggest that there may not be a lot of difference between the low and high doses but that either may be more effective than the placebo. But can survival analysis give us more detail and a test of significance?

Requesting a Kaplan-Meier survival plot in SPSS

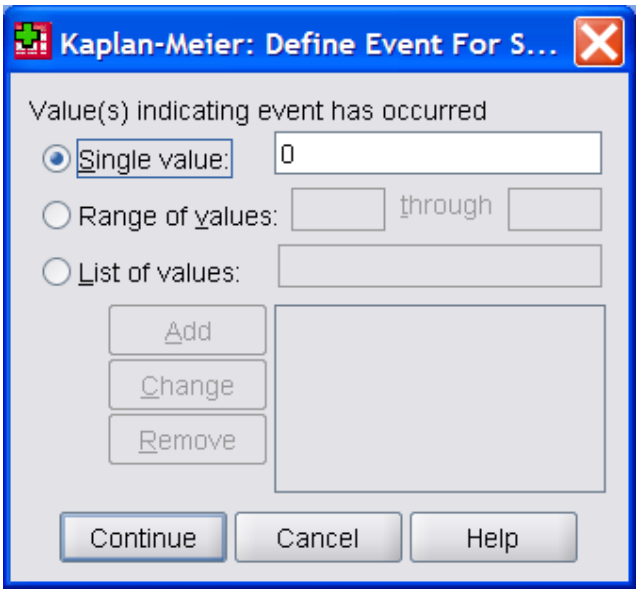
First arrange the datasheet as in Table 14.1. An essential requirement is the variable indicating whether or not the observation is censored. In our example this variable is RELAPSE, and we have used zero for the cases in which relapse is observed to have occurred, and 1 for censored cases. Any coding is possible since SPSS will ask us to state it. The variable giving times to the event must contain a value for censored cases, since if we leave the time blank SPSS will omit those cases from the whole analysis. So these censored cases have the time for which they were observed recorded, always 20 weeks in our example. But you could have a study where people entered at different times and so were observed for different periods up to the end, and then the censored cases would not all have same time recorded. We have three treatment categories. Our other two covariates are quantitative.

With the datasheet set up, we can get the Kaplan-Meier analysis and plots as follows. Choose **Analyze** from the menu bar, then **Survival** and **Kaplan-Meier** to get SPSS Dialog Box 14.2. Use the arrow to put TREATMENT in the **Factor** box and TTF in the **Time** box (this is the variable for the time to the event). In the **Status** box goes the variable to indicate whether the observation was censored, so use the arrow to put RELAPSE in here. There will be a question mark in the brackets following RELAPSE and the **Define Event** button will now be available.



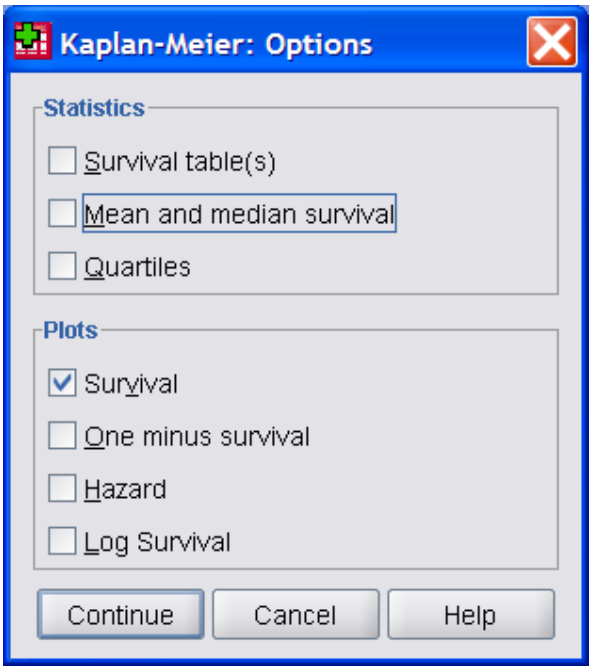
SPSS Dialog Box 14.2. The main dialog box for Kaplan-Meier Survival Analysis

We have to define our code so click **Define Event** to get SPSS Dialog Box 14.3. We have a single value, zero, to indicate that the event, relapse, occurred so type a zero in the box (in our code, 1 indicates censorship, which is when we don't observe the event). Click **Continue** to return to the main dialog box.



SPSS Dialog Box 14.3. Coding for censored observations

Now click the **Options** button to get SPSS Dialog Box 14.4. All we want from this analysis is the survival plot so in the **Plots** group tick **Survival**, and untick all the statistics boxes. Click **Continue** and **OK**.



SPSS Dialog Box 14.4. Ordering a Kaplan-Meier survival plot

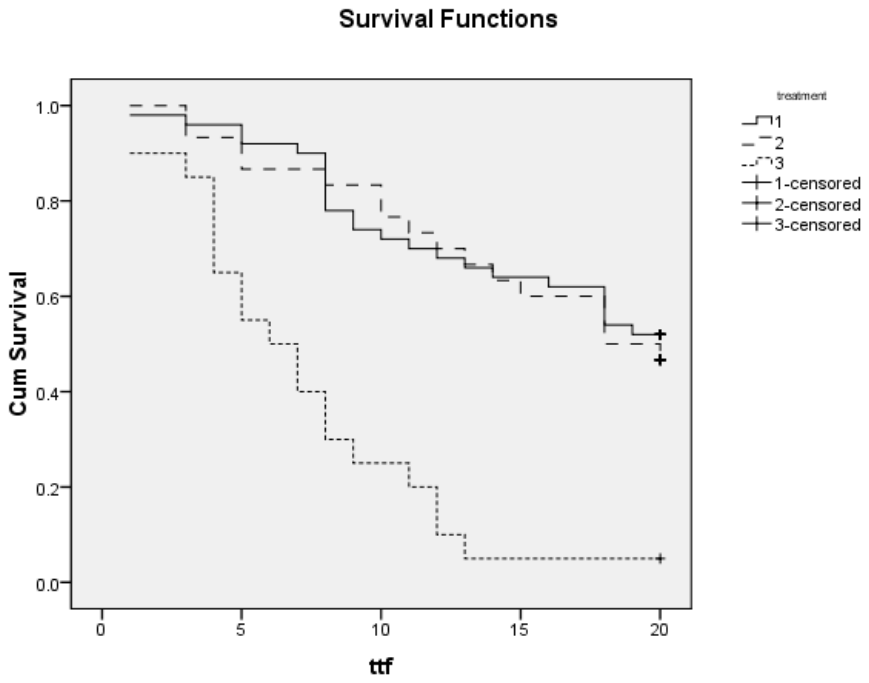
Understanding the Kaplan-Meier plot output

We only requested a plot from this analysis, but the output begins with a table showing the number and percentage of censored observations in each treatment group, as shown in SPSS Output 14.1. We see that just over half of the observations from treatment 1 (low dose) were censored, so just over half of these people had still not experienced a dyspenea event when the study ended. Just under half of the high dose group (treatment 2) were censored but only one of the placebo group had still not experienced a dyspenea event at the end of the study.

The survival plot, also shown in SPSS Output 14.1, shows the proportion of each group who had not experienced a dyspenea event by the end of each week of the study. Two of the placebo group (10%) had already experienced a dyspenea event at the end of week 1, and you can see that 0.9 of that group (treatment 3) are survivors at the end of week 1. Then no one else from that group experienced a dyspenea event until week 3, when one person succumbs (0.05 of the group of 20). Four more of treatment group 3 (0.2 of the group) experienced a dyspenea event by the end of week 4 and you can see the big step down in their survival function. For the high dose group (treatment 2), no one experienced a dyspenea event until week 3, so the survival function for that group stays at 1.0 until week 3, the whole group still survives until then.

Case Processing Summary

			Censored	
treatment	Total N	N of Events	N	Percent
1	50	24	26	52.0%
2	30	16	14	46.7%
3	20	19	1	5.0%
Overall	100	59	41	41.0%



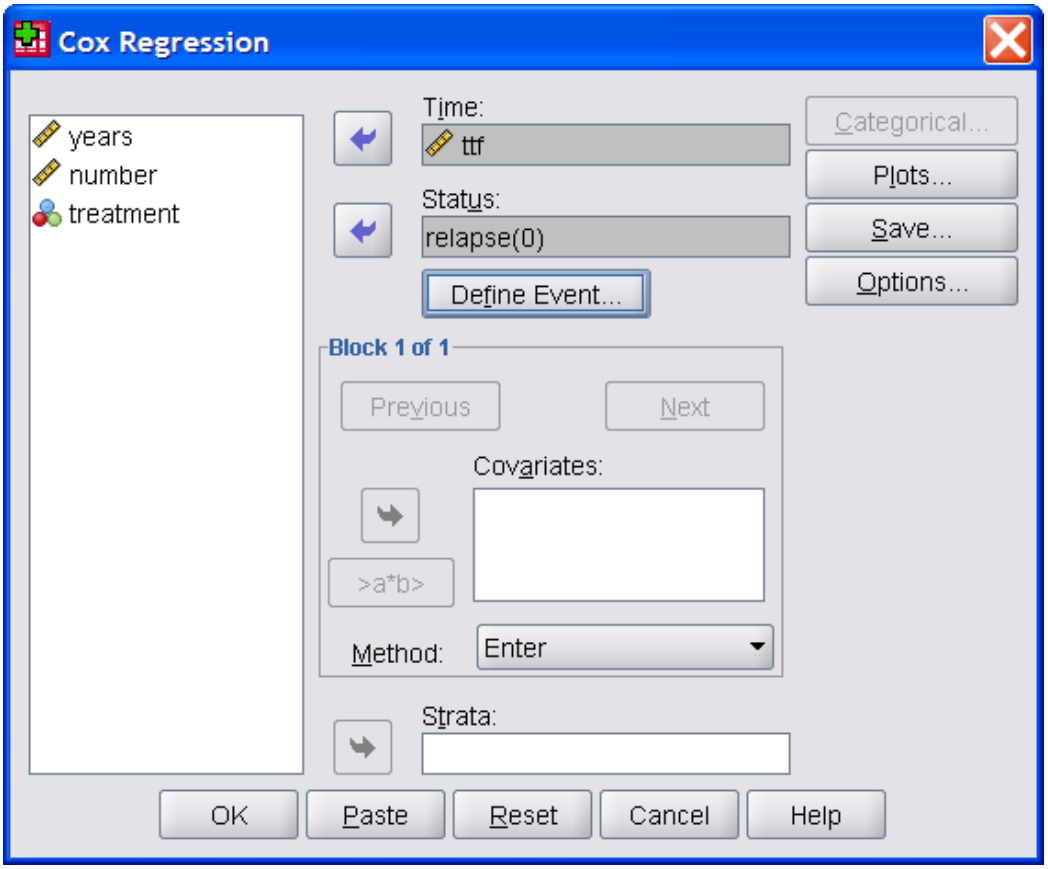
SPSS Output 14.1. Numbers of censored observations and the survival plot from a Kaplan-Meier survival analysis

There are tests of significance for the differences among treatments, and they can be requested by clicking the **Compare Factor** button. Three tests are offered, of which the Log rank test is most commonly used. However we do not discuss these here since we want to allow for the possible effect of our covariates YEARS and NUMBER when comparing treatments. To do this we need the Cox Regression method.

Requesting a Cox Regression Analysis in SPSS

Now we do the Cox Regression and take account of NUMBER and YEARS, which may well influence the results: choose **Analyze** from the menu bar, then **Survival** and **Cox Regression**, to get SPSS Dialog Box 14.5. Use the arrow to put TTF in the **Time** box (this is the variable for the time to the event). In the **Status** box goes the variable to indicate whether the observation was censored, so use the arrow to put RELAPSE in

here. There will be a question mark in the brackets following RELAPSE and the **Define Event** button will now be available, just as in the Kaplan-Meier analysis above.



SPSS Dialog Box 14.5. The main dialog box for Survival Analysis

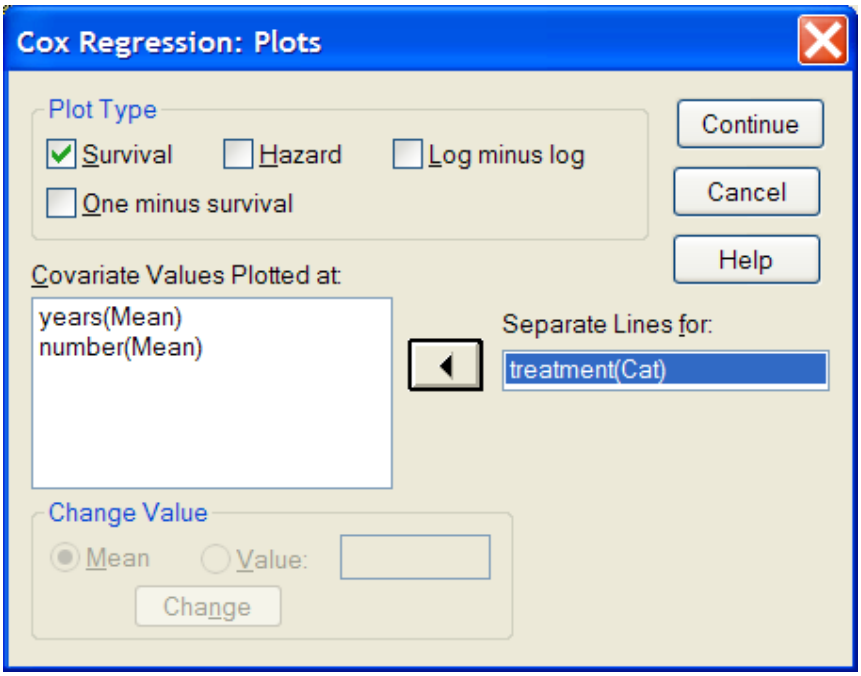
Clicking **Define Event** produces a dialog box just like SPSS Dialog Box 14.3, and once again we have a single value, zero, to indicate that the event, relapse, occurred so type a zero in the box as before. Click **Continue** to return to the main dialog box.

Cox Regression includes treatment factors in the term covariates, so our covariates are YEARS since starting to smoke, NUMBER smoked per day, and TREATMENT, and we are primarily interested in whether the treatments are equally effective or not. The other covariates are there because the effectiveness of treatment may be affected by them. We could just put in all the covariates and see which have a significant effect, but as

we did in multiple regression, we can take a hierarchical approach. We shall enter YEARS and NUMBER as covariates in Block 1, and TREATMENT in Block 2, as we now explain.

First use the arrow to put YEARS and NUMBER in the **Covariates** box. Whenever we have more than one covariate in a block, we can use a stepwise method (see Chapter 4 on Regression) by clicking the arrow next to the **Method** box. Here we have chosen to enter both our covariates for Block 1 together so we left **Method** at the default setting of **Enter**. Now click **Next** so that **Block 2 of 2** replaces **Block 1 of 1**. Now use the arrow again to put TREATMENT in the **Covariates** box. So now the effect of TREATMENT will be examined, having allowed for the effects of YEARS and NUMBER.

We still have to specify which covariates are categorical so click **Categorical** and in the dialog box which appears, put TREATMENT in the **Categorical covariates** box and click **Continue**. We also want to specify a graph, so click the **Plots** button to get SPSS Dialog Box 14.6. Tick **Survival** and put TREATMENT in the **Separate Lines for** box. This will give us survival graphs showing our three treatment groups as separate lines.



SPSS Dialog Box 14.6. Specifying a plot

Click **Continue** and when you return to the main dialog box click the **Save** button. In the dialog box that appears, select **Partial residuals** in the **Diagnostics** group, and click **Continue** again.

Click the **Options** button. In the dialog box that appears select **CI for exp(B)**. (This dialog box also allows you to alter the criteria for entering and removing variables when using a stepwise method, just as in regression.) Click **Continue**, then **OK** to get the analysis.

Understanding the Cox regression output

The first table (which we haven't shown) just gives a case processing summary, showing the number of cases, the number that were censored, and how many with missing data (none of ours). The second table (SPSS Output 14.2) shows the coding used by SPSS for any categorical variables. We only have one, TREATMENT, and we used 1, 2 and 3 to denote low dose, high dose and placebo. The SPSS default is to use

the last category as the reference category, and to assign a code of zero to it. The other categories are defined by dummy variables just as in multiple regression. Because we have three categories, there are two dummy variables. The first takes the value 1 for treatment 1 and zero for treatments 2 and 3, the second takes the value 1 for treatment 2 and zero for treatments 1 and 3, so the three treatments are defined by the pairs of values (1, 0), (0, 1) and (0, 0).

Because treatment 3 is the reference category, treatments 1 and 2 will each be compared with treatment 3 in the analysis. You can change this if you want, and use the first instead of the last category as reference (we show you how in the next section).

Categorical Variable Codings[b]

		Frequency	(1)	(2)
treatment[a]	1	50	1	0
	2	30	0	1
	3	20	0	0

a. Indicator Parameter Coding

b. Category variable: treatment

SPSS Output 14.2. Dummy variables for categories defined by SPSS

The analysis begins by ignoring all the covariates (Block 0), just as the Kaplan-Meier analysis does. You could think of the likelihood statistic that is given here as analogous to the total sum of squares in ANOVA or ANCOVA. Subsequently, as covariates are added, the reduction in this statistic is the basis for tests of significance for the covariates. We omit this table but the statistic from it is also shown (note a) below the first table in SPSS Output 14.3 (-2 Log likelihood = 502.587). SPSS Output 14.3 shows the results of entering our Block 1 variables, YEARS and NUMBER. Because

we chose not to use a stepwise method of entering these two, we are reminded that they were both entered together (Method = Enter).

Block 1: Method = Enter

Omnibus Tests of Model Coefficients [a,b]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
464.783	36.583	2	.000	37.804	2	.000	37.804	2	.000

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 502.587

b. Beginning Block Number 1. Method = Enter

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
years	.039	.071	.298	1	.585	1.040	.904	1.196
number	.245	.041	36.296	1	.000	1.277	1.179	1.383

Variables not in the Equation [a]

	Score	df	Sig.
treatment	31.185	2	.000
treatment(1)	11.359	1	.001
treatment(2)	.018	1	.894

a. Residual Chi Square = 31.185 with 2 df Sig. = .000

SPSS Output 14.3. The effect of YEARS and NUMBER on the relapse hazard rate

In the first table of SPSS Output 14.3, look at the column labelled Change from Previous Step. This gives a chi-square of 37.804 on 2 degrees of freedom, a highly significant value. The 2 *df* refer to the two covariates we added at this step, YEARS and NUMBER. The chi-square value is the reduction in the likelihood ratio statistic for this step. You get the Block 0 value (no covariates included) from note a or from the Block 0 table we have not shown (this is 502.587). The Block 1 value (464.783) is in the left column of this table, hence the Chi-square Change From Previous Step is 502.587 – 464.783 = 37.804. So we know that together our two covariates have a

significant effect on the probability of relapse at any time (the hazard rate). The column on the right, Change from Previous Block, just gives the same information since we added both our covariates at the same time.

The next table shows the significance levels separately for the two variables, and we see that YEARS is not significant: the significance observed above is entirely due to NUMBER. (The test statistic used here, which is assumed to follow a normal distribution, is the Wald statistic, is just $(B/SE)^2$, the square of the parameter estimate divided by its standard deviation.) Look at the right hand column of this table: if Exp(B), the hazard ratio, > 1, as here, the hazard rate will increase (so expected time to the event, relapse, will decrease) for increasing values of the covariate. The fact that the hazard ratio for YEARS is so close to one reflects its non-significance. For each unit increase in NUMBER the hazard rate will be multiplied by 1.277. The more our patients smoked before starting the study, the greater the chance of a dyspenea event at any stage.

The last table in SPSS Output 14.3 tells us which covariates are omitted. Only TREATMENT is omitted at this stage, but because it is a categorical variable the two dummy variables that represent it are also listed in the table, together with the significance in each case. We consider this in SPSS Output 14.4, which shows the tables for Block 2, when we entered TREATMENT.

Block 2: Method = Enter

Omnibus Tests of Model Coefficients[a,b]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
441.467	66.371	4	.000	23.316	2	.000	23.316	2	.000

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 502.587

b. Beginning Block Number 2. Method = Enter

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
years	.057	.072	.622	1	.430	1.059	.919	1.220
number	.264	.045	34.738	1	.000	1.302	1.193	1.422
treatment			26.209	2	.000			
treatment(1)	-1.709	.338	25.550	1	.000	.181	.093	.351
treatment(2)	-1.324	.377	12.353	1	.000	.266	.127	.557

Covariate Means and Pattern Values

	Mean	Pattern		
		1	2	3
years	11.600	11.600	11.600	11.600
number	28.880	28.880	28.880	28.880
treatment(1)	.500	1.000	.000	.000
treatment(2)	.300	.000	1.000	.000

SPSS Output 14.4 The effect of TREATMENT on the relapse hazard rate after allowing for YEARS and NUMBER

The first table in SPSS Output 14.4 again shows the chi-square statistic for the Change From the Previous Step (23.316). There are 2 *df* again even though we only added one covariate, because TREATMENT is categorical with three categories and hence two dummy variables are needed to define it. Once again we see that the change from the previous step is significant. The change from the previous block again repeats the same information since the two dummy variables were entered together.

Now look at the second table in SPSS Output 14.4. Here we see that, with all covariates included, YEARS still does not have a significant effect on the hazard rate. NUMBER does have a significant effect with a hazard ratio of 1.302 (the 95%

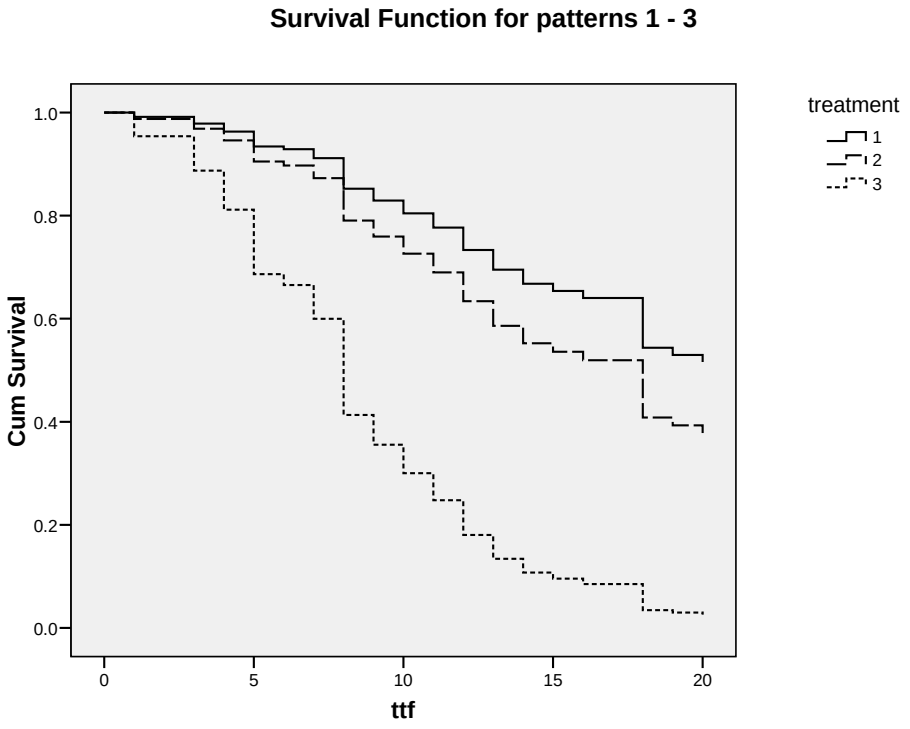
confidence interval for this value is shown as 1.193 to 1.422). TREATMENT is also significant, but more interesting than the overall significance of this covariate is whether either of the treatments gives a significant reduction of the hazard rate compared with the placebo. In fact we see that both do. The low dose (treatment 1) has a significant effect with a hazard ratio of 0.181 (the 95% CI is 0.093 to 0.351), or about 18%, of that on the placebo. Likewise the high dose is significant and reduces the hazard rate to about 27% of that on the placebo. We consider whether the difference between the low and high doses is significant later. For now, look at the last table in SPSS Output 14.4.

The first column gives the mean values for all the covariates. The mean time participants have smoked is 11.6 years, and they have smoked an average of 28.88 cigarettes per day before giving up. The first dummy variable for TREATMENT has a mean of 0.5 because 50 of our 100 participants received treatment 1, so the first dummy variable was coded 1 for them and zero for everyone else. The second dummy variable was coded 1 for the 30 participants on treatment 2 and zero for everyone else, so its mean value is 0.3. The three columns labelled Pattern 1, 2 and 3 repeat the means for YEARS and NUMBER, and give the dummy variable codings for the three treatments, (1, 0), (0, 1) and (0, 0). These are patterns 1 – 3 in the title for the survival graph we requested, which is shown in SPSS Output 14.5.

These survival graphs for the three treatment groups are the expected values for the Cox Regression model. Since they are also affected by the values of the covariates YEARS and NUMBER that we also have included, they are calculated using the mean values of YEARS and NUMBER. On the graphs you can see that the expected proportion

of those who have not relapsed falls more rapidly for the placebo than for either dose of the drug, and is nearly zero by the end of the 20 weeks. This agrees well with our observation that only one patient had still not experienced a dyspenea event. The two drug doses each have expected proportions of around a half not relapsed by the end, close to what we observed by counting the censored observations at the start of our analysis.

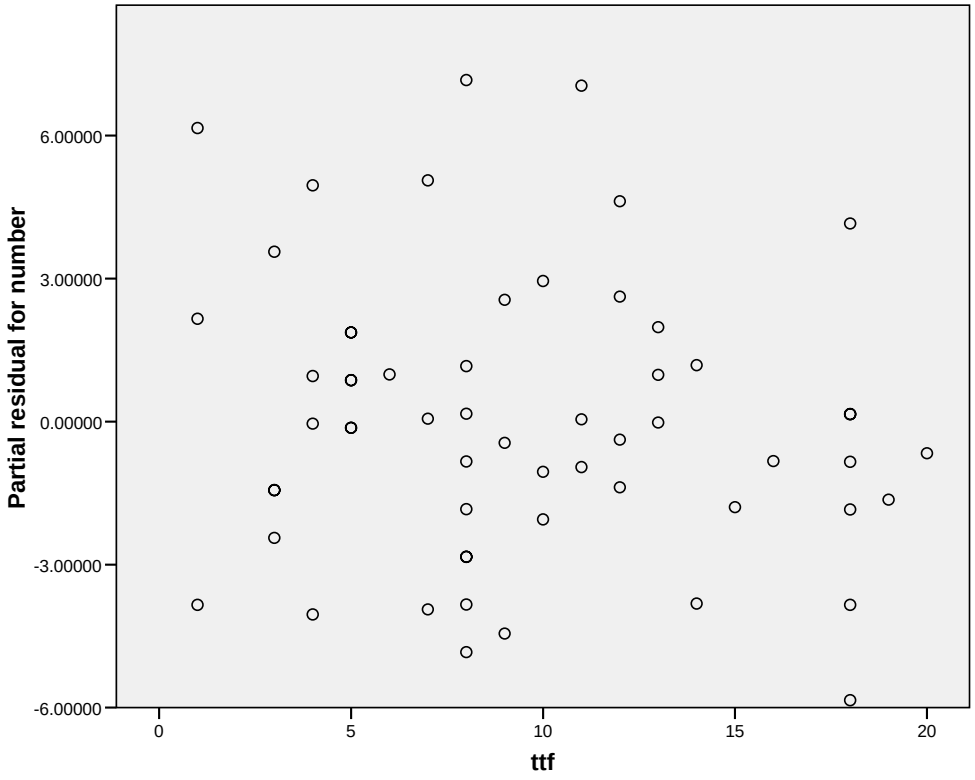
SPSS also provides a single survival graph for all treatment categories together. Since the values of the other covariates affect this, it is also calculated for the mean values of YEARS and NUMBER, and is the one shown in the Introduction as Figure 14.1.



SPSS Output 14.5. Survival functions for each TREATMENT in the COPD trial

Before leaving this stage of the analysis we consider how we might check the proportional hazards assumption. We requested that partial residuals should be saved, and these will be added to the datasheet. The columns in the datasheet are labelled

PR1_1, PR2_1 etc, but if you let the mouse rest over the column names, the more helpful Partial residual for years, Partial residual for number, etc will be shown. Since only NUMBER of our quantitative covariates was significant, we will use it to illustrate our check. From the menu bar, choose **Graphs**, then **Chart builder**. Click **OK** when you get the warning about defining your variables correctly. From the gallery, choose a simple scatterplot and drag and drop it into the graph area. Then drag TTF from the variable list into the x-axis space and PARTIAL RESIDUAL FOR NUMBER into the y-axis space. Click **OK** and get the graph in SPSS Output 14.6.



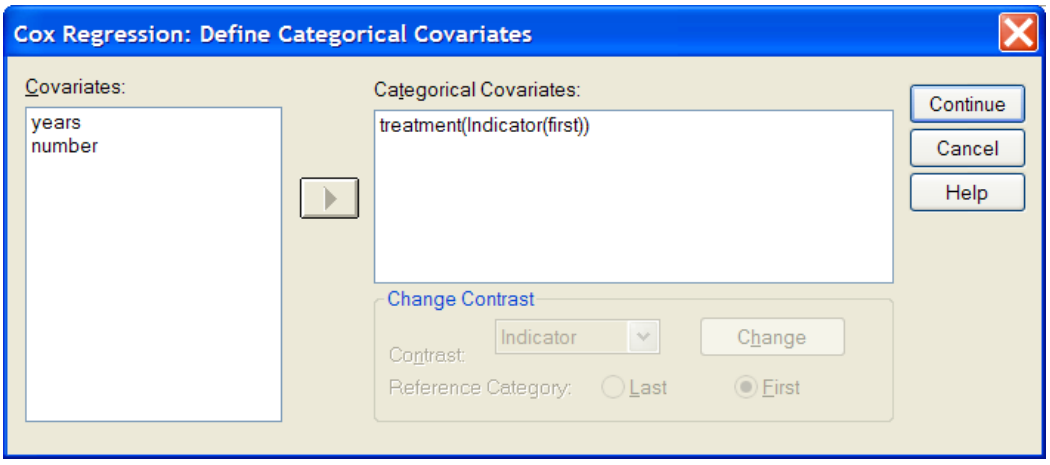
SPSS Output 14.6. A check on the proportional hazards assumption

If the effect of NUMBER on the hazard rate is constant over time, there should be no pattern in this graph of residuals. In fact it looks as if there is a slight downward trend from left to right. If you fit a simple linear regression (**Regression**, then **Linear**, TTF as **Independent** and PARTIAL RESIDUAL FOR NUMBER as **Dependent**) you find that the

negative correlation you see in the graph is very small and does not approach significance. So our proportional hazards assumption seems reasonable here.

Changing the reference category

One way to find out whether the difference between drug doses is significant is to repeat the analysis above but using the first instead of the last of our treatment categories as the reference category. This will mean the effect of the high dose and of the placebo will be compared with the low dose. After clicking the **Categorical** button and putting TREATMENT in the **Categorical Covariates** box, click the **First** radio button at the bottom. You also have to click the **Change** button to make this happen. The dialog box now looks like SPSS Dialog Box 14.7.



SPSS Dialog Box 14.7. Changing the reference category

Now repeat the analysis exactly as before. The Categorical Variable Codings table (like the one in SPSS Output 14.1) now shows that the dummy variables code treatments 1, 2 and 3 as (0, 0), (1, 0) and (0, 1) respectively. The middle table in SPSS Output 14.4 is replaced by the one in SPSS Output 14.7. Here we see that, while the effects of YEARS and NUMBER are the same as before, only the second dummy TREATMENT variable has a significant effect. The first TREATMENT dummy compares

treatment 2 (high dose) to the new reference category, treatment 1 (low dose), and the significance of this (p = 0.252) is well above 0.05. However, the second dummy compares treatment 3 (placebo) to treatment 1 (low dose), and here the effect is highly significant, just as we found when we made the comparison in reverse above. The hazard ratio, 5.521, is just the reciprocal of 0.181 (i.e. 1/0.181), as we would expect since, if the effect of treatment 1 compared to treatment 3 is to multiply the hazard rate by 0.181 as we found above, then the effect of treatment 3 compared to treatment 1 should be the inverse of the multiple. So, as we suspected when we looked at our data at the start, both low and high drug doses reduce the hazard rate compared with the placebo, and there is no significant difference between their effects.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
years	.057	.072	.622	1	.430	1.059	.919	1.220
number	.264	.045	34.738	1	.000	1.302	1.193	1.422
treatment			26.209	2	.000			
treatment(1)	.385	.336	1.312	1	.252	1.469	.761	2.839
treatment(2)	1.709	.338	25.550	1	.000	5.521	2.847	10.710

SPSS Output 14.7. Comparing therapies by changing the reference category

Using a stepwise method

Since one of our covariates turned out not to have a significant effect on the hazard rate, we may want to repeat our analysis omitting this variable (YEARS). However, had we used a stepwise method to enter the two variables in the first block, we could have saved ourselves this step. Here we repeat the first analysis, though keeping the first category of TREATMENT as reference, but this time when we put YEARS and NUMBER in the **Covariates** box (SPSS Dialog Box 14.4), click the arrow by the **Method** box and choose **Forward LR** from the list. (LR stands for likelihood ratio.) SPSS Output 14.8 shows the tables from the Block 1 steps.

Omnibus Tests of Model Coefficients[b,c]

Step	-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
		Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
1[a]	465.082	36.522	1	.000	37.505	1	.000	37.505	1	.000

a. Variable(s) Entered at Step Number 1: number

b. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 502.587

c. Beginning Block Number 1. Method = Forward Stepwise (Likelihood Ratio)

Variables in the Equation

		B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
								Lower	Upper
Step 1	number	.238	.039	38.267	1	.000	1.269	1.177	1.368

Variables not in the Equation[a]

		Score	df	Sig.
Step 1	years	.298	1	.585
	treatment	30.488	2	.000
	treatment(1)	.001	1	.974
	treatment(2)	29.244	1	.000

a. Residual Chi Square = 31.100 with 3 df Sig. = .000

Model if Term Removed

Term Removed		Loss Chi-square	df	Sig.
Step 1	number	37.505	1	.000

SPSS Output 14.8. Block 1 tables with a stepwise method

In the notes below the first table we see that at the first step NUMBER was added (note a). As before, the Likelihood ratio with no covariates is also shown (note b). The chi-square value for the Change from Previous Step is 37.505 (37.804 when we added YEARS and NUMBER together, SPSS Output 14.3), with only 1 *df* since only one covariate was added at this step. The Change from Previous Block is just the same since this is the first step in this block.

In the second table we see that NUMBER does have a significant effect. The default settings will add a covariate if the significance level is less than 0.05. You can change

the default probabilities for entry and removal in the stepwise procedure by clicking the **Options** button in the main dialog box. Next comes a table showing the significance levels for variables not currently included. Then we have a table showing the effect of removing NUMBER. As it is the only variable included at this step, it isn't surprising that removing it will increase the chi-square by exactly the amount that entering the variable reduced it. However, if we had just added a variable to several already there, then the entry of the new variable might possibly make redundant one that had been added at a previous step. So, as in regression, when a new variable is entered, all those currently included will be checked to see if they could now be removed. The default setting for removal is a significance level that exceeds 0.10.

In Block 2, as before we see TREATMENT added. The table giving the significance levels and Exp(B) values and confidence intervals for the final version, which includes just NUMBER and TREATMENT, is shown in SPSS Output 14.9. This is similar to the one we had in SPSS Output 14.7, when YEARS was included, and it confirms that there is not a significant difference ($p = 0.198$) between treatments 1 and 2 (low and high drug doses). Each extra cigarette smoked per day increases the hazard rate by a factor of about 1.3, and the placebo increases the hazard rate by a factor of about 5.4 compared with the low drug dose.

Variables in the Equation

							95.0% CI for Exp(B)	
	B	SE	Wald	df	Sig.	Exp(B)	Lower	Upper
number	.257	.043	34.950	1	.000	1.293	1.188	1.408
treatment			25.796	2	.000			
treatment(1)	.429	.333	1.659	1	.198	1.535	.800	2.949
treatment(2)	1.693	.337	25.243	1	.000	5.435	2.808	10.520

SPSS Output 14.9. The effects of number of cigarettes smoked per day and treatment group on the hazard, with patches (treatment 1) as the reference category.

Other ways to compare treatment effects

By using treatment 3 as reference category, we were able to compare each of our two drug doses with the placebo. Or, using treatment 1 as reference category, we could compare the high with the low dose, and the placebo with the low dose.

There is another approach to comparing treatment effects, which resembles the use of contrasts explained near the end of our description of One-way ANOVA. In that section we showed how to conduct post hoc comparisons of the five levels (ADJECTIVE, IMAGERY, INTENTIONAL, COUNTING and RHYMING) of the IV DEPTH on the DV RECALL. In particular we compared the average of the last three levels with the average of the first two levels of the variable DEPTH. We considered the difference between the two averages,

$$\frac{1}{3}\text{adjective} + \frac{1}{3}\text{imagery} + \frac{1}{3}\text{intentional} - \frac{1}{2}\text{counting} - \frac{1}{2}\text{rhyming.}$$

The coefficients for the three categories must sum to zero and to avoid fractions we used small whole numbers in the same proportions (2, 2, 2, -3, -3).

Using this method, we can consider the difference between the two therapies, high dose – low dose, and then the difference between the placebo and the average of the two drug doses,

$$\text{placebo} - \frac{1}{2}\text{low dose} - \frac{1}{2}\text{high dose.}$$

In Survival Analysis, SPSS offers a short list of contrasts to choose from, but we can't make up our own, and the coefficients are chosen by SPSS. We will now show you how to use this facility.

Return to the main dialog box with TTF in the **Time** box and RELAPSE(0) in the **Status** box. Once again we will enter YEARS and NUMBER into Block 1 of **Covariates** and select the stepwise method **Forward LR**. Then TREATMENT is entered in Block 2 of **Covariates** as before. Now click the **Categorical** button to see SPSS Dialog Box 14.7 again. Move TREATMENT into the **Categorical Covariates** box as before, then click the arrow on the indicator button below **Change Contrast** to get the list of contrasts on offer. Select **Difference** from the list and click **Change**. Click **Continue** and **OK** to repeat the analysis.

This time the categorical variable codings table, giving the dummy variables to define the categories, is as shown in the first SPSS Output 14.10 table. The first dummy variable assigns values (-1/2, 1/2, 0) to treatments 1, 2 and 3, and the second dummy variable assigns values (-1/3, -1/3, 2/3). Once again the three treatments are defined by two dummy variables, this time with values (-1/2, -1/3), (1/2, -1/3) and (0, 2/3). However the first dummy variable contrasts treatments 1 and 2, and the second contrasts treatment 3 with the average of the first two. SPSS has used fractions, but we only want to know whether the contrasts differ significantly from zero, so any multiple of (hypnosis – patches) will do, as will any multiple of (cold turkey - $\frac{1}{2}$ patches - $\frac{1}{2}$ hypnosis).

Categorical Variable Codings[b]

		Frequency	(1)	(2)
treatment[a]	1	50	-.500	-.333
	2	30	.500	-.333
	3	20	0	.667

a. Difference Parameter Coding

b. Category variable: treatment

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
number	.257	.043	34.950	1	.000	1.293	1.188	1.408
treatment			25.796	2	.000			
treatment(1)	.429	.333	1.659	1	.198	1.535	.800	2.949
treatment(2)	1.478	.312	22.487	1	.000	4.386	2.381	8.081

SPSS Output 14.10. Dummy variables defined for the Difference contrast, and the tests of significance

The results appear in the second table of SPSS Output 14.10, which corresponds to SPSS Output 14.9 from our previous analysis. You can see that the first dummy treatment variable does not approach significance ($p = 0.198$), so the difference between hypnosis and patches is not significant. The *p* value is the same as that in SPSS Output 14.9, where we made the same comparison using treatment 1 as the reference category and comparing each of treatments 2 and 3 to it. The second dummy variable for treatment does show a significant effect, but the Wald statistic is not the same as the one in SPSS Output 14.9 because this time the placebo is compared with the average of the low and high drug doses instead of just with the low dose as reference category. Likewise the hazard ratio also differs for the second dummy treatment variable.

There are several other available contrasts. To see a list of definitions, click the **Help** button in SPSS Dialog Box 14.7. When using any of these, check the table of Categorical Variable Codings to make sure it is what you want.