



Simone Famularo and Davide Bernasconi

Most of the activities we make in clinical practice have a single, simple aim: fighting the disease to increase survival. Typically, although several medical conditions do not require to face the risk of mortality, the general intellect during the centuries has been captured by the medicine's potential to change the natural history of a disease, tearing a great number of people from a destiny already written. Let us think of cancer, probably the leading cause of death worldwide: in recent years, we made exciting steps forward, changing completely the outcomes for those who are affected. No more than 15 years ago, for example, a diagnosis of metastatic colorectal cancer was a death sentence, while now several therapies and combined approaches are available, reducing sensibly the rate of patients who are condemned. Moreover, the integration of new knowledge derived from molecular medicine, oncology, and surgery is leading us to a new scenario where cancer may become a sort of chronic disease.

Clinical studies play a key role in the continuous development of the treatment of cancer to improve the survival of patients. Thus, a solid knowledge regarding how to collect and analyze survival data is crucial for medical researchers involved in such studies. How can we understand the impact of a treatment in modifying the survival probability of our patients? How can we account for the sequence of events

S. Famularo (✉)

Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy

Department of Hepatobiliary and General Surgery, IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy

D. Bernasconi

Bicocca Bioinformatics Biostatistics and Bioimaging Centre–B4, School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy

e-mail: [Davide.bernasconi@unimib.it](mailto: Davide.bernasconi@unimib.it)

that occurred at different time points? How can we be sure that an eventual survival benefit is intrinsically connected to the treatment, and not to a more benevolent disease, not so much aggressive? In this chapter, we will focus our attention on these topics through some clinical examples that may better explain how to manage time-to-event data.

8.1 Generalities About Time-to-Event Data

Time-to-event reflects the time elapsed from an initial event (e.g. diagnosis of the disease, surgery, start of treatment) to an event of interest: death, cancer recurrence, a second episode of diverticulitis after conservative treatment, and so on. The event of interest should not be the death in any case: in fact, despite the name commonly used to refer to the statistical methodology, all kinds of events that can occur in a predetermined time-span can be considered in the analysis.

Generally speaking, these techniques are applicable both in randomized clinical trials and in cohort studies (typically prospective but also when data are collected retrospectively).

The most important aspect for the analysis of this type of data is the planned minimum follow-up time of patients. When we set a prospective study, for example, we may plan to enroll patients for 1 year (e.g. at the time they undergo surgery) and to subsequently follow them for further 2 years. The very first patients enrolled at the start of the study will be followed up for almost 3 years, while people enrolled at the end of the study will be followed up for at most 2 years before the study will be closed. Even if we think about a homogeneous cohort of patients sharing very similar treatment and baseline characteristics, it is quite obvious that the probability to observe the event of interest will be different among these two types of patients: let us think about cancer relapse.

The first enrolled patient has been treated and discharged at home, and now we have started the follow-up period: as previously mentioned, our study will last 2 years since end of enrollment, so he has 36 months of time to develop our event of interest, the recurrence. Depending on the type of cancer we are studying, this time period may be enough to observe the event. For the last patient enrolled, however, we will have only 24 months before the study ends to observe the recurrence. This time may also be enough to observe the event but it is much shorter than the follow-up time of the first patient. Both patients may be classified as no-recurrence; however, this may not be because of the treatment we have administered, but because we have observed them for too little time: their recurrence might have occurred when the study was already closed. How can we manage this situation where different follow-up times are present? Moreover, how can we account for the fact that for some patients we may not observe the event of interest (and thus the time of event occurrence)? Should we exclude these patients from our analysis? Obviously not: survival analysis methods have been thought specifically to address these issues, allowing us to manage patients observed for different timespans and that may be still event-free at the end of the study. Another issue that may occur is the patients'

drop-out. In fact, a patient may be enrolled in the very early phase of the study, however, for some reason, he may decide to stop his participation in the study before the end of the planned follow-up (he stops to come at the outpatient visits, or he goes to live in another district or country, or simply he changes his mind on the participation at our study). In this latter case, we will have a shorter follow-up, as in the case of those who are enrolled at the end of the study period. For these patients (dropped-out or late enrolment), we surely have an incomplete follow-up. To sum up, we usually have to deal with patients enrolled at different moments and thus with different potential follow-up times who may quit the study before the end. However, as long as the reasons for these differences in the observation time of patients depend only on study logistics and not on the clinical status of patients, we can say the following: for the patients that do not develop the event of interest during the study, we only know that their survival (i.e. event) time is longer than their last observed follow-up: we call these times **censored**.

Box 8.1: What Is Censoring?

Have a look at Fig. 8.1. We have depicted a situation as the last described: in the Y -axis, we have the patients enrolled, while in the X -axis we have the calendar time. Each patient has a different story: #1 has been enrolled at the study start and has been followed up for the duration of the study without observing a recurrence. Patient #2 has been enrolled on the 4th year and followed up until the end without a recurrence. Patients #3 and #5 have been enrolled at different times, but before the study ended they withdrew: #3 moved to another city and preferred to be followed for his disease in the new location, while #5 did not come to the planned visit, and she did not answer anymore to the phone. Patients #4 and #6 experienced the event of interest (recurrence) at different time points. Thus, since we have observed only two events of interest (patients #4 and #6), our outcome (i.e. the time of event occurrence) is known only for these two subjects. The information we have on the other subjects should not be thrown away! In fact, we have a partial knowledge of the outcome also for censored subjects: we know that their event time is higher than their observed follow-up time. This means that these patients may have a recurrence in the future but we will never observe it. This is true both for patients #1 and #2 who arrived at the end of the study period without experiencing a recurrence as well as for patients #3 who dropped out from the study and patient #5 who was lost to follow-up. Survival analysis methods were designed to account for all these issues, thanks to the following crucial assumption: occurrence of censoring is independent of the likelihood of developing the event of interest. This is certainly true for subject #1 and #2 (the study end is obviously independent from patients survival) and for subject #3 (censoring is due to patient migration so it is again independent from survival), while for patient #5 is not granted: we should speculate what is the reason for loss to follow-up (if the reason is related to the patient status, then independence of censoring assumption may not hold).

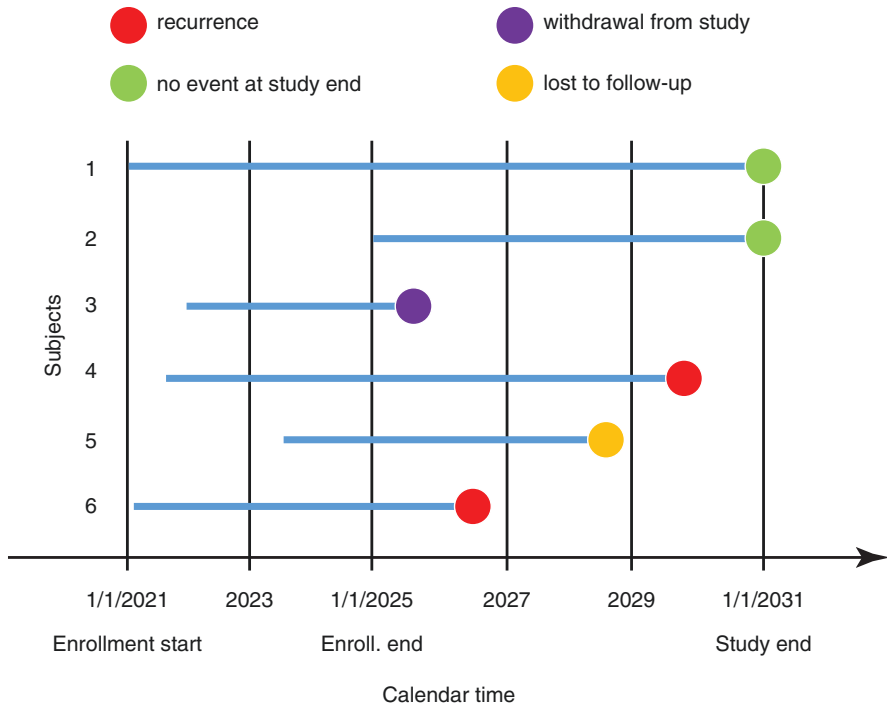


Fig. 8.1 Graphical representation of survival data as they are collected

8.2 The Variables We Need to Make Analysis: Event and Time

Now, we can start to explain how to prepare our data in order to perform a survival analysis. First, for each patient we need a time variable, which should be a continuous variable in which the time (measured in months, days, years, or any other time unit) of observation is expressed. This variable represents the time between the observation start (the first day of a RCT, or the day of surgery, or the day of the diagnosis, depending on the study purpose) and the time of the event or the last time we have notices about the patient. One way to manage this information during data collection is to add two columns in our database, in which we will record the date of the follow-up start, and the date in which the event of interest (e.g. the death) occurred. In case the patient does not

experience the event, we will record the last date we have news about him, for example, the date of the last visit. To increase the accuracy and the effectiveness of our analysis, it may be better to have a complete follow-up for all patients: in this sense, it is strongly recommended to update the follow-up at the same time for each patient (this is why, typically, you can find lots of residents that are busy at the phone, making very strange conversation in which they try to gently understand if the patient who has been treated 10 years ago and then has never been seen anymore is still alive or not). When we are making retrospective studies, this could be difficult and challenging and missing data may occur. Recommended methods to manage the missing data (e.g. multiple imputation) are really too advanced for the purpose of this book. A pragmatic solution is to do all the best to find the data: we know about consultants who required their residents to write letters to the registry offices, generating hatred and frustration which ultimately result in abandoning all ambitions of research in the future.

Once we have the two dates, we can simply calculate the difference, in the time unit we prefer, between the two dates, to obtain a continuous variable measuring the time-span and becoming our time variable.

Second, we need for each patient a categorical variable that indicates whether the observed time just calculated represents the time to an event (e.g. death, recurrence, development of a symptom) or to the last follow-up (i.e. censored observation: the patient did not develop the event of interest during the follow-up period). For the analysis we do not need to distinguish among the possible causes for censoring (e.g. study end, loss to follow-up), provided that the independent censoring assumption holds (see the previous paragraph). Thus, this event indicator variable should always be a dichotomous variable (e.g. dead/alive, recurrence/recurrence-free, yes/no). A little recommendation: the event indicator should be coded as a binary variable assuming value 0 for censored observations and value 1 for observed events. This choice is convenient because it corresponds to the default values in many software which will automatically understand this classification. However, some software (such as STATA) always requires to specify which level of the variable indicates who is censored and which level indicates who has the event.

Now our data will look like those represented in Fig. 8.2, while in Fig. 8.3 we can see our toy dataset ready to be loaded into our favorite software to start the analysis.

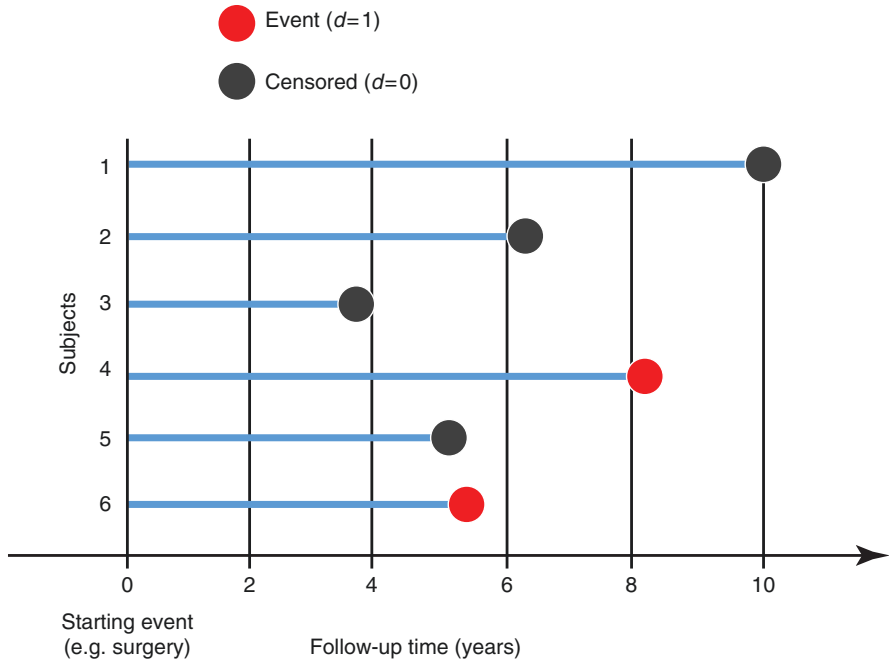


Fig. 8.2 Graphical representation of survival data after they are prepared for being analyzed. The event indicator variable is d and takes value 1 for subjects #4 and #6 who developed an event during follow-up and takes value 0 for the others (censored subjects)

Fig. 8.3 A screenshot of our toy dataset once ready to be analyzed

	A	B	C
1	id	time	event
2	1	10	0
3	2	6.2	0
4	3	3.7	0
5	4	8.3	1
6	5	4.6	0
7	6	4.7	1

8.3 The Survival Curve and Life Tables: The Kaplan–Meier Method

The main goal of survival analysis is to assess the probability that patients from a certain population can survive (or remain event-free) until some time. We want to compute this probability for every time unit (e.g. every year, month, or day) up to a fairly distant time horizon. Once we have prepared the data as mentioned, we can

Table 8.1 Kaplan–Meier estimation of the survival probability over time in our toy dataset

Time index	Time (years)	Number at risk, N_t	Number of deaths, D_t	Number of censored, C_t	Survival probability $S_{t+1} = S_t \times ((N_{t+1} - D_{t+1})/N_{t+1})$
$t = 0$	0	6	0	0	1 (by definition)
$t = 1$	3.7	6	0	1	$1 = 1 \times (6 - 0)/6$
$t = 2$	4.6	5	0	1	$1 = 1 \times (5 - 0)/5$
$t = 3$	4.7	4	1	0	$0.75 = 1 \times (4 - 1)/4$
$t = 4$	6.2	3	0	1	$0.75 = 0.75 \times (3 - 0)/3$
$t = 5$	8.3	2	1	0	$0.375 = 0.75 \times (2 - 1)/2$
$t = 6$	10	1	0	1	$0.375 = 0.75 \times (1 - 0)/1$

proceed to estimate this quantity. The result will typically be presented as a survival curve, in which the X -axis shows the follow-up time, and the Y -axis the probability to survive (the proportion of people surviving) until that moment.

The estimation of the survival probability used to generate the curve is typically done using the Kaplan–Meier method, which can be described by the following formula:

$$S_{t+1} = S_t \times ((N_{t+1} - D_{t+1}) / N_{t+1})$$

This means that if we know the survival at time t (S_t) we can compute the survival at next time $t + 1$ (S_{t+1}) by multiplying S_t with the probability of surviving in the next time unit $t + 1$. The last one is computed as the proportion of patients NOT died at $t + 1$ (i.e. patients alive at $t + 1$ minus patients died at $t + 1$: $N_{t+1} - D_{t+1}$) over patients alive at $t + 1$ (N_{t+1}). The first value of the survival S_0 is by definition equal to 1 since at time 0 all patients are alive. How does the method account for censored subjects? At time $t + 1$ the number of subjects still alive is obtained by taking patients alive at t (N_t) and subtracting patients who died at t (D_t) but also subjects censored at t (C_t), thus $N_{t+1} = N_t - D_t - C_t$.

With this formula, we can create a table like the following (Table 8.1):

In the first column only times when something happens (i.e. at least one patient died or censored) are reported in increasing order. The number at risk reported in the second column represents the patients that have not yet experienced the event of interest at that time point and that have not yet been censored. This number will decrease as time passes and give us a very important piece of information regarding how reliable the survival estimate is (see Box 8.2 to better understand why this is very important). In the third column, we find the number of deaths: in this example, we are measuring the overall survival but in other examples here we can find the number of patients who experienced the event of interest, at the time it occurred. Then we have a column for censored observations. Finally, for all the relevant time points, the survival probability is computed. As time increases, the value of the survival can only remain the same or decrease and the value is updated only at times when at least one patient has the event (for this reason, the plotted curve has a “stair” shape). To know the estimated survival probability at a certain time (e.g. 5 years), we need to find in the time column the maximum time lower than the time of interest and read the corresponding survival value. For instance, the estimated survival at 5 years is 0.75 (we should look at the row where time = 4.7).

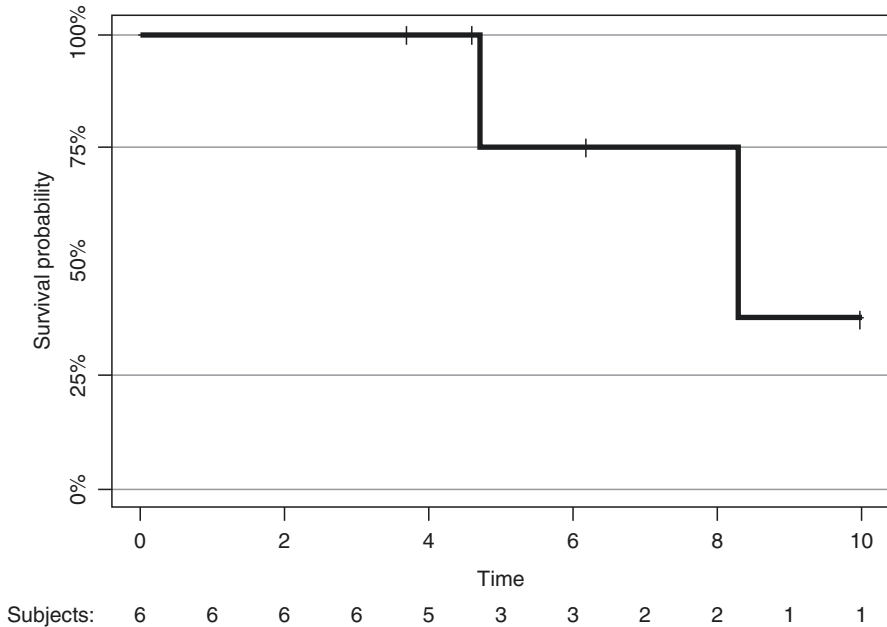


Fig. 8.4 Kaplan–Meier survival curve estimated on toy data

Looking at Fig. 8.4, we can easily understand the survival probability of our cohort at different time points. Finding the intersection among X and Y axis, we can estimate that after 5 years of follow-up, 75% of our patients are survivors. By definition, to estimate the median survival of our cohort, we need to follow the Y -axis at 0.50 and find the intersection on the X -axis: in this example, the median survival time of our cohort is a bit higher than 8 years. As expected, patients are 100% alive at time 0 (the X -axis origin): this obvious assumption conditions the figure of the curve, which is always decreasing to the right, that is the direction in which the time increases. The slower the curve decreases, the higher will be the survival of patients, even after a long time. When the curve decreases sharply, we are facing a disease that is very aggressive, with a high probability of death. When we compare two survival curves, for example, estimated on two groups of patients under two different treatments, the higher one will belong to the treatment with the best prognosis (we will reconsider this theme after).

We can also provide information about the censoring: a proper figure, in fact, should report the presence of patients censored at each time point. This is usually visualized by a sign on the curve (in our example, a small vertical line is present when there is a censored case). Remember that the aim of your graphic representation is not to hide data, but to summarize the highest quantity of information and to make it easy to be understood by other physicians. Depending on the sample size, the survival curve can be more of a staircase rather than a proper curve: the higher the number of patients, the more the survival line becomes similar to a smooth curve.

Another consideration should be made about the direction of the curve which can provide a description of the natural history of a disease. For example, if the survival curve tends to flatten after a certain time point, this suggests that patients who are still event-free at that time point are not anymore at risk of developing the event. In contrast, if the curve falls down to zero this means that after the time the curve reaches the horizontal axis no patient survives.

Be careful about extrapolating the results of a curve beyond a certain time point! The Kaplan–Meier curve could be artificially projected up to very far time points that we do not really observe. For example, one could be tempted to draw the survival curve until 15 years, even if patients in the cohort were observed for a maximum of 5 years. This is very speculative and should be avoided. A good practice is to report the median follow-up time of the study (together with inter-quartile range), allowing readers to know which survival times have been really observed (see Box 8.2 to know how to calculate appropriately the median follow-up time of your study).

Box 8.2: Patient-At-Risk and Median Follow-Up Time: How to Interpret

We now want to focus our attention on two important aspects both connected to the correct interpretation of a survival curve: the number of patients at risk in the right tail of the curve and the calculation of the median follow-up time.

All statistical estimators are subject to some variability which reflects our uncertainty on the point estimate we calculate and this variability depends also on the inverse of the sample size: the higher is the number of subjects on which the estimate is calculated, the lower is our uncertainty around that value. This is true also for Kaplan–Meier curves with one additional caveat: the sample size decreases through time. The number of patients at risk is in fact eroded for two reasons: patients who die (or have the event of interest) and patients who are censored. As a consequence, even if our initial sample size is particularly high, as we move towards the right tail of the curve the number of patients at risk becomes smaller and smaller. This means that also the precision of our Kaplan–Meier survival estimates decreases through time and, from a certain time onwards, may become unreliable because its updated value could be based on a very small number of patients still at risk. In Fig. 8.5 we show a Kaplan–Meier curve together with 95% confidence interval (shaded area). As you can see, the amplitude of the confidence interval increases through time suggesting that the estimate becomes less and less precise. At time 4.5 we have only 10 patients still at risk: the curve is thus updated based on the mortality observed on that restricted group of patients. Reading survival probability estimates on the right tail of the curve must be considered with caution.

Providing a summary measure of follow-up time is always requested when reporting the results of a cohort study. A typical measure is the median

which can be easily calculated directly on the observed follow-up time of patients. However, with survival times we may get in trouble even for this simple task. In fact, if our end-point is death or if our observation ends as patients experience the event of interest, simply calculating the median of the observed times would lead to an underestimation of the follow-up time. The survival time of patients who died is obviously shorter than their potential follow-up time (especially for those who died early). The question is: for how long would we have followed patients if they did not die? How can we obtain a more accurate estimate of this quantity? Surprisingly, the solution is again Kaplan–Meier but... reversed! In this case, censored times are those we are really interested in, while survival times represent a lower limit for the true follow-up time of a patient; thus, we can simply estimate the curve considering censored observation as events and death as censorings (Fig. 8.6). The time where this curve reaches 50% is about 4.2 and this is the best estimate of the median follow-up time we can get. On the same data, if we calculated the simple median of the observed times, we would get 1.55, a clear underestimation.

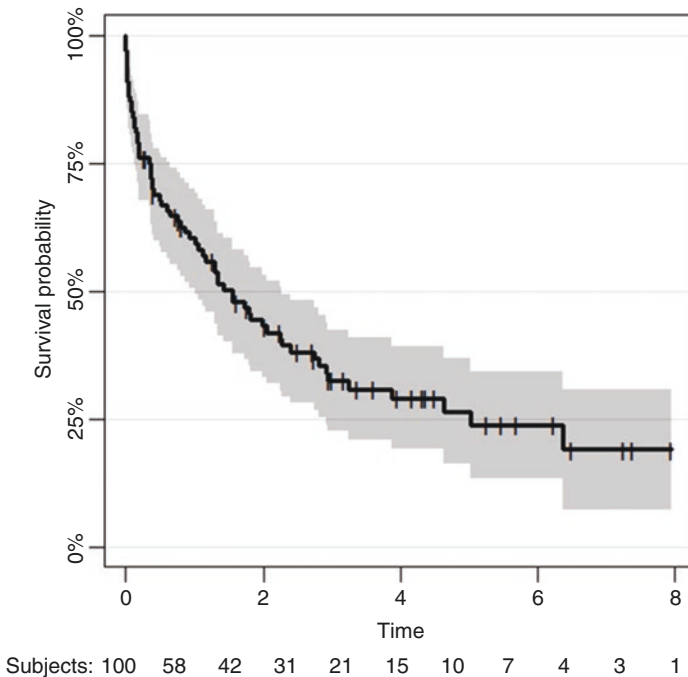
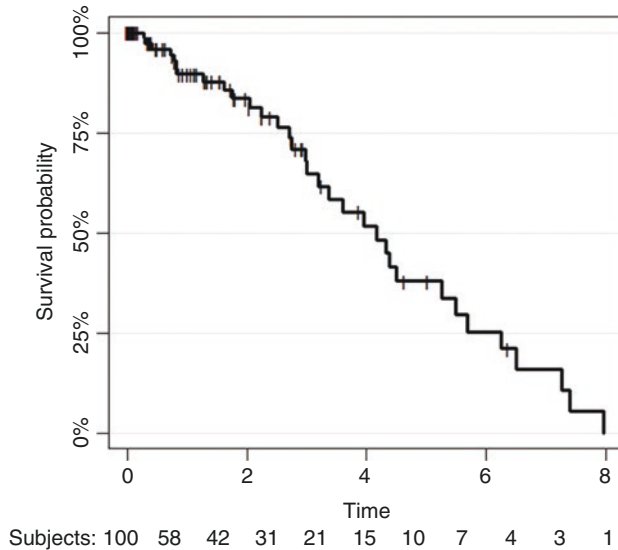


Fig. 8.5 Kaplan–Meier survival curve with 95% confidence interval (shaded area) on a made-up dataset

Fig. 8.6 Reverse Kaplan–Meier applied to the same data of Fig. 8.5. Events and censored observations are flipped over



Example Box: Part 1

DISCLOSURE: The following data are completely created, and the results obtained do not depict a real scenario. The example is completely invented, and the conclusion does not want to suggest anything. The comparisons are made reliable to better highlight the management a clinician should employ to make survival analysis; however, the setting and all the data are not taken from reality.

We are surgical oncologists, and after a few years we have started a laparoscopic program to treat HCC patients. We would like to know if the long-term survival of those patients treated by laparoscopy is longer, similar, or shorter than the classical open approach. For this purpose, we decide to set up a retrospective study, to compare the overall survival among the two surgical approaches.

We know from the literature and the guidelines that the overall survival for liver tumors is conditioned not only by the surgical technique but also by the tumor burden (number and size of the nodules), the comorbidities and the age of patients, their underlying liver function, and some histological characteristics: since the retrospective nature of the study, we want to collect all this information to adjust the risk and be sure about the treatment effect on survival. We start to create a data-sheet in Excel where we collect all the data we need to make our analysis. Likewise, we decide to collect all the following variables: patient's ID, age (continuous variable, years), sex (categorical variables, levels: male and female), number of nodules (continuous variable), size of the nodules (continuous variable, cm), presence of

cirrhosis (categorical variable, levels: yes and no), MELD score (continuous variable), HBV or HCV infections (categorical variables, levels: yes or not), type of procedure executed (categorical variable, levels: open, laparoscopy), date of the procedure, presence of microvascular invasion (categorical variable, levels: yes and no), and satellitosis (categorical variable, levels: yes and no). To simplify the software job, we transform all the categorical variables in no = 0 and yes = 1, female = 0 and male = 1. After this, we also need to collect the variables we need for survival analysis: the event status (alive or dead) and the follow-up time. Since our primary end-point is overall survival, we start checking in our hospital management software all the last visits of our patients, to know if our enrolled patients are still alive or not at the present date. In case we did not visit the patient recently, we decide to call by phone number to speak directly with the patients or the parents to know about their follow-up. We want to know their status at the present day, so we create a column in which we insert the date of the last contact (if the patient is alive) or the death date. Although we want to do the maximum to find the most updated news about the patients' follow-up, for a few of them we won't be able to have any news after a certain period: to don't lose patients, we decide to insert the date of the last available contact, with the event status at that time. Those patients will be considered as censored, but they will still contribute to our analysis, although differently if compared with patients with completed follow-up.

To create our time variable (called OS in this study), we simply make a subtraction between the date of surgery and the date of last follow-up or the date of death (see Fig. 8.7).

After this screening, between 2008 and 2020 we have enrolled 464 patients treated by surgery in our center for HCC. Of them, 65 (14.0%) have been treated by laparoscopy. On a first look, we know that 301 patients (64.8%) died during the follow-up period, 294 in the open surgery group, and 7 in the laparoscopy one. Now we would like to know if this difference is significant and if there is an overall survival advantage with one technique or not.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ID	Age	Sex	Cirrhosis	MELD	HCV	HBV	N_nodules	Size_nodules	date_of_surgery	Laparoscopy	microvascular_invasion	satellitosis	death_event	date_of_last_contact/date_of_death	OS
2	2005033951	57	0	1	14	0	0	1	2	05/07/2005	0	0	0	1	03/12/2012	89.04605263
3	2005040387	53	0	1	6	0	0	1	1.6	05/07/2005	0	1	0	0	12/08/2021	138.0592105
4	2005047733	44	1	1	6	0	1	1	2	09/09/2005	1	0	0	0	12/08/2021	136.9078947
5	2005051308	58	1	1	7	1	1	1	3.3	23/09/2005	0	1	0	1	11/07/2014	105.4934211
6	2005060873	56	0	1	11	1	0	1	1.5	06/12/2006	1	0	0	1	25/04/2008	16.64473684
7	2007013647	71	0	0	5	1	0	1	2	26/01/2007	1	0	0	0	12/08/2021	119.3092105
8	2007029838	64	0	1	7	1	0	1	2	10/05/2007	0	0	0	0	12/08/2021	89.80263108
9	2007034838	47	0	1	6	0	1	1	1.9	08/05/2007	0	0	0	0	12/08/2021	71.34868421
10	2008010513	58	0	1	7	0	1	1	1.5	07/01/2008	1	1	1	1	01/06/2010	28.81578947
11	2008017967	56	0	1	11	0	1	1	2	26/02/2008	1	0	0	1	24/04/2012	49.96710526
12	2008022038	56	1	1	6	1	0	1	1.8	18/03/2008	1	0	1	0	12/08/2021	105.5921053
13	2008030665	75	0	0	6	0	0	1	1.6	20/05/2008	0	0	0	0	12/08/2021	103.8197368
14	2008042122	62	0	1	11	0	0	1	1.8	08/08/2008	0	1	1	1	22/09/2016	97.59868421
15																
16																

Fig. 8.7 An example of a dataset managed with one of the most popular software. All the variables appear as number (and no free text), with only one data per each cell

8.4 Comparing Survival Curves: The Log-Rank Test

During our clinical research, most of the time we are more interested in comparing the effect of two (or more) treatments on survival, rather than knowing the survival of the whole cohort. The Kaplan–Meier estimator can simply be applied separately to groups defined by a categorical variable to evaluate the survival probability (as well as the median survival) observed in each group (provided that enough patients are present in all groups). For example, in a clinical trial with a survival outcome, we might be interested in comparing survival between participants receiving a new drug as compared to a placebo (or standard therapy). In an observational study, we might be interested in comparing survival between men and women, or between persons with and without a particular risk factor (e.g. hypertension or diabetes). However, rather than simply looking at the curves observed in the samples, one might also be interested in assessing the association of treatment or another variable with survival using a statistical hypothesis test.

Facing this issue could seem very simple, after all the knowledge acquired in the previous chapters of this book. In fact, once we know the total number of deaths for each treatment, we could imagine simply making a Chi-square test to compare the proportion of events among the groups. Another approach that we could regard as feasible after reading this book may be to perform a T test to compare the mean of time-to-event among groups. Both the ideas are wrong. Let us think again to the first paragraph of this chapter: when we face survival data, we have several issues: patients are observed for different periods, the event of interest can be observed during the follow-up time or occur later, and finally patients may not have completed the predetermined follow-up time. In a few words: we need to account for censoring! Censored observations carry information that we do not want to lose. The log-rank test was designed to tackle these issues, providing us a tool that accounts for the occurrence of events in time and for censoring. There are other types of procedures we can employ to test the hypothesis of equal survival among groups; however, in this chapter we will discuss only the log-rank test, which is undoubtedly the most popular in clinical applications.

The log-rank test considers the null hypothesis (H_0) of equal survival between two or more independent populations. In other words, the test helps to judge whether the survival curves we observe on the sample groups are compatible with the possibility that the “true” survival curves (i.e. the curves in the whole populations of interest from which samples are drawn) are identical (overlapping) between groups or not. The log-rank test is actually a particular kind of stratified chi-square test as it compares observed vs. expected numbers of events at each time point over the follow-up period (the stratification variable is time). Analogously to the other statistical tests, we simply obtain a p -value which should be compared with the chosen level of significance to assess whether the treatment groups are significantly different or not in terms of survival. If we are comparing more than two treatments, it may be useful to make several pairwise log-rank test, which means comparing the treatments one-by-one: this will provide us a better explanation of where the differences are, because, for example, two treatments could have similar survival probability,

but different from a third treatment. Some correction of the p -value (e.g. Bonferroni or more advanced methods) should then be applied to account for the multiple testing problem (similarly to what is done, for example, by ANOVA post-hoc tests).

Box 8.3: OS, DFS, RFS: Defining the Outcomes

A clear definition of the end-point of interest is fundamental, and this is why in the methodology section of research papers, it is mandatory to specify this aspect. Here we provide for you some standard definition of the most popular end-points in surgical oncology, for your convenience.

- Overall survival (OS): The time from the date of treatment (surgery, drug delivery, etc.) to the date of any cause of death.
- Disease free survival (DFS): The time from the date of treatment to the date of the first event among death for any cause or recurrence of the tumor. Consequently, this is a combined end-point that considers together both types of “failures”, deaths and recurrence. This is also sometimes called recurrence-free survival (RFS).
- Time to recurrence (TTR): The time from the date of treatment to the date of recurrence after the treatment. In this case, patients who died during the follow-up are censored at the date of death or, perhaps more properly, death is considered as a “competing risk” (this would require ad hoc analytical methods rather than Kaplan–Meier and log-rank test).

However, be aware that these terms can be used to refer to slightly different end-points. So, besides names you may want to use for your end-points, it is highly recommended to always report each event of interest you included in the analysis.

Example Box: Part 2

Now we have prepared our data, and we want to compare survival among the two surgical groups. We launch our statistical software (we can use, for example, *R*, version 4.0.6) and we upload the database. First, we calculate the median follow-up time, using the reverse Kaplan–Meier method: in our cohort, the median FU was 69.93 months (IQR 39.44–110.10).

Then, we launch a Kaplan–Meier. Here below is the life table obtained (we kept only the rows of 12, 36, and 60 months).

Laparoscopy = No						
Time	N. risk	N. event	Survival	Std. err	Lower 95% CI	Upper 95% CI
12	322	70	0.823	0.0192	0.786	0.861
36	211	83	0.599	0.0252	0.552	0.651
60	124	47	0.448	0.0270	0.398	0.504

Laparoscopy = Yes						
Time	N. risk	N. event	Survival	Std. err	Lower 95% CI	Upper 95% CI
12	51	2	0.965	0.0241	0.919	1.000
36	14	5	0.822	0.0646	0.705	0.959
60	5	0	0.822	0.0646	0.705	0.959

Then we create a survival curve, where we also insert the p -value obtained by the log-rank test (when you create these figures, it is always recommendable to add the p -value).

If we focus on the survival curve, we can immediately understand the survival differences between the two treatments. At the bottom left, we visualize the p -value obtained by the log-rank test. As commonly a $p < 0.05$ is accepted for significance, here the two treatments are significantly different, and laparoscopy (the light blue line) is superior to open surgery (the red line) in terms of overall survival. In our sample, we can conclude that the two treatments are different observing the curves, which showed a large spread between each other. In fact, the spread among the two curves indicates the effect of treatment on survival in our sample, allowing to visualize how “large” is the difference (Fig. 8.8). The figure can give us other important information. The little crosses on the two survival curves are the censored patients: the density on the curve of the crosses gives us information about how many patients we have followed for all the period. Another important information derives from the table “number at risk”: in fact, patients in the laparoscopic group are few, and approximately after 30 months of observation, the survival curve becomes flattened, with only some crosses on it. This means that no death event has been recorded in that group after that time, and the reduction of the number of patients at risk is conditioned by the censoring. When there is so little data, we should carefully evaluate the meaning of those results at least at that specific time points: it is unrealistic that those who were treated by laparoscopy stop to die at a certain time point. Probably, enlarging the sample size, and completing the follow-up, we will note other death events that could better represent the real survival tendency of these patients. Thus, always carefully consider the table of number at risk, because it can give you important information on how realistic the survival prediction is, particularly far in time.

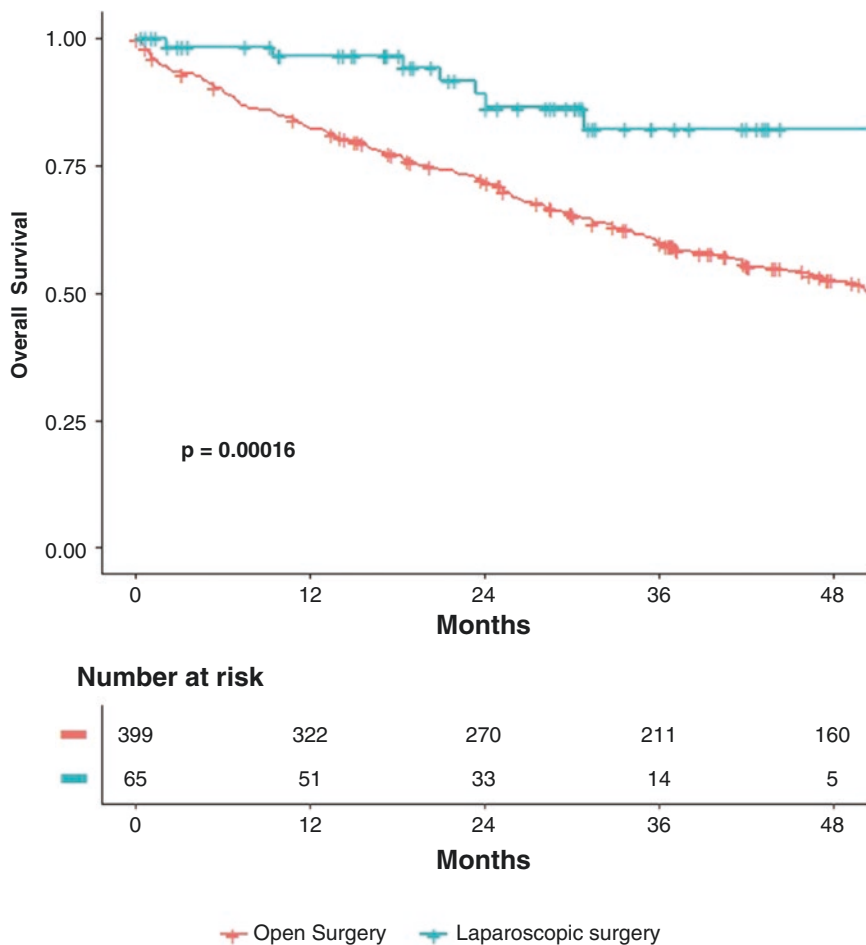


Fig. 8.8 Kaplan–Meier curves and log-rank test p -value of overall survival between surgical techniques

8.5 A Regression Model to Assess the Association of Multiple Predictors with a Survival Outcome: The Cox “Proportional Hazards” Model

The comparison of survival curves between two (or more) treatments is a sort of univariate survival analysis, in which we assess the association of a single risk factor (the treatment) with the outcome. However, we may desire to study several factors simultaneously, as when we perform linear or logistic regressions. We need a regression model, suitable for time-to-event data, that allows us to assess independently the impact of a risk factor in the occurrence of the event of interest in time and consequently to assess if the effect we may have recognized with the Kaplan–Meier

method is real or is driven by some other factors (possible confounders). However, directly modeling the survival function is statistically challenging (although some possible solutions have been proposed). It turns out that it is much more convenient to focus on another quantity of time called “hazard rate.” This can be viewed as an “instantaneous velocity” of the event occurrence at each time or, in other words, as the risk for a patient alive at a certain time to develop the event in the next instant.

One of the most popular regression methods in survival analysis is the Cox proportional hazards model. It is composed of two parts which are multiplied together: one is the “baseline hazard,” the hazard of patients with reference level of all covariates; the other is the effect of each covariate on the baseline hazard, showing how the hazard modifies when covariates change. The first part can be difficult to estimate properly (although suitable estimators have been proposed) but what we really care about is the second part: Sir David Cox invented a method to estimate the effect of covariates without taking care of the baseline hazard (that is why this method is still so popular!). In the end, in a Cox model, the measure of effect is the hazard ratio (HR), which tells us how many times we have to multiply the baseline hazard to obtain the hazard of another level of a covariate. For example, if the HR between treatment A and B is 2, this means that patients treated with A develop the event two times faster than patients treated with B. We can also say that the HR between B and A is 0.5, meaning that the velocity of occurrence of the event is halved for patients treated with B with respect to those treated with A. People tend to interpret HR as a risk ratio (similarly to what happens for the odds ratio). This is not totally correct as we should always bear in mind that the hazard is not a simple risk but a sort of “risk in time.” However, as with risk ratio and odds ratio, an HR approaching 1 suggests no effect of that covariate, an $HR > 1$ means that the covariate is probably a risk factor and an $HR < 1$ indicates a protective role.

Another analogy with risk ratio and odds ratio is that also HR is usually shown with its (95%) confidence interval and possibly the p -value. A statistically significant effect is considered when the confidence interval does not include 1 (the null value) or when the p -value is lower than a nominal level (typically 5%). Beyond statistical significance it is always important to look also at the clinical relevance of the estimated effect, especially in observational studies where no a priori sample size calculations were made.

Have a look at Table 8.2:

Table 8.2 Cox regression analysis

	HR	95% CI		p
		Lower limit	Upper limit	
Age ≥ 75 (versus <75)	1.012	0.990	1.034	0.282
Charlson comorbidity index (per unit)	1.040	0.905	1.195	0.581
Presence of cirrhosis (versus not)	1.966	1.161	3.328	0.012
Number of nodules (per unit)	1.151	0.815	1.627	0.424
Microvascular invasion (versus not)	1.830	1.202	2.786	0.005
Post-operative complication (versus not)	1.221	0.736	2.023	0.440
Laparoscopy (vs. open approach)	0.754	0.654	0.987	0.043
Post-op liver complication (versus not)	2.524	1.344	4.739	0.004

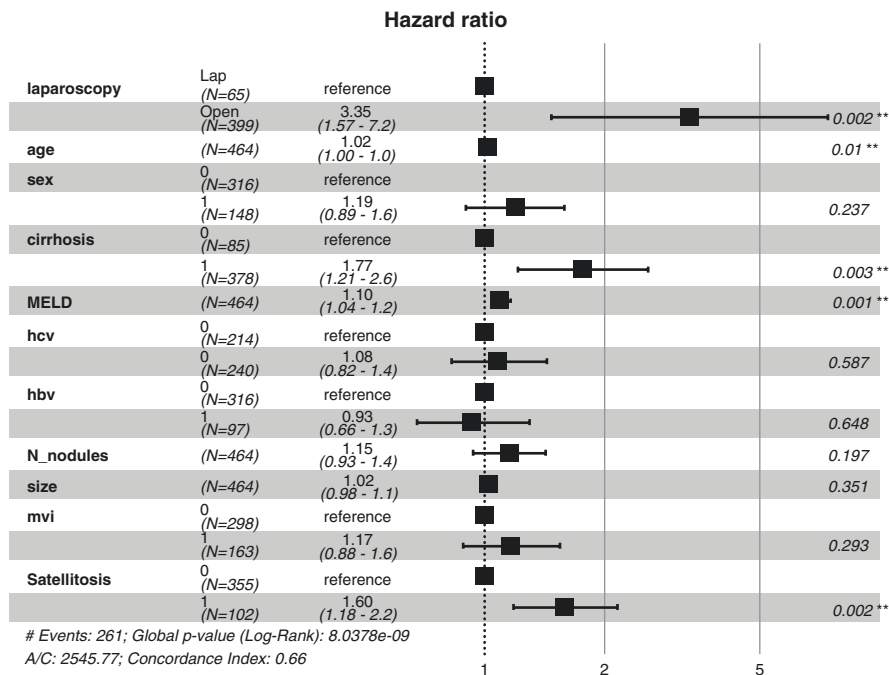


Fig. 8.9 Forest plot depicting the results of a multivariate Cox regression as per our example

This is how to report the results of a Cox regression in a table. You will notice that each variable is followed by brackets, which contain the reference value against which the level of interest of each covariate was compared. It is very important to always declare what we are comparing! The width of the CI is a measure of precision of our estimates: the sharper the CI is, the more precise the obtained estimate is. As already said, the HR may take values from 1 to infinity for risk factors or may take a decimal value between 0 and 1 for protective factors.

To make things clearer, just have a look at one of the variables, e.g. post-op liver complication: the HR is 2.524, 95% CI: 1.344–4.739, $p = 0.004$.

In this example, we expect that experiencing a post-operative liver complication increases the hazard of death by 2.54-fold when compared with patients who did not experience such complications. The result is statistically significant because 1 is not included in the confidence interval, as evident also by the p -value that is <0.05 . In case we are analyzing a continuous variable, the interpretation of the HR is slightly different: the HR represents the multiplicative factor of the hazard for a 1 unit increase in our covariate. In our example, the number of tumor nodules has an HR of 1.151: this means that we expect that the hazard increases by a factor of 1.151 for each additional nodule (the result is not statistically significant). Remember also that, as with other regression tools, we should also explore the assumption of linearity in the effect of continuous covariates (we cannot hereby explain this issue in detail).

To calculate the “excess of hazard” caused by a factor, we may subtract 1 to the HR (HR-1) and then multiply the result per 100. For example, when microvascular invasion is present, we expect the hazard of patients without this problem to increase by 83% (thus, to almost double). In case the HR is below 1, we need to remember to invert the subtraction (1-HR). For example, laparoscopy is associated with a hazard decrease of 25% with respect to patients treated with open surgery.

Another consideration should be made about the number of variables we can insert in a Cox regression. This is a very tricky issue; however, a very simple rule to take home is the so-called one to ten rule: to create a reliable model with valid parameter estimates, we can add an explanatory variable for every ten events that occurred in our cohort. So, if we are investigating the risk of mortality and in our cohort we observe 56 events, we may simultaneously include no more than five to six covariates in our model.

A final consideration should be made to clarify one important limit of the Cox regression. This technique relies on a crucial assumption: the hazard proportionality assumption. Namely, Cox model assumes that the hazards in levels of each covariate (for example, between treatment A and B) are proportional over time, which implies that the effect of a risk factor should be constant over time. So, if HR of A vs. B is 2 we are assuming that patients treated with A die twice more quickly than those treated with B at the beginning of the follow-up, as well as after some time, till the end of follow-up. This assumption is not always tenable as some treatments may have an early efficacy which is lost during time or may show only a late efficacy. We can verify this assumption by several tools, using statistical tests, or graphically. One of the most popular methods is based on scaled Schoenfeld residuals. Otherwise, we can try to figure out whether the assumption may hold directly by looking at the survival curves obtained with the Kaplan–Meier method: if the curves of the two treatments cross each other, then the assumption is definitively violated. This suggests that the hazards are not proportional over time, and the Cox regression is not appropriate: some adjustment must be made to account for non-proportionality. One simple approach is to run a stratified Cox model for the variable for which the assumption is violated (again, refer to more advanced references for details).

Example Box: Part 3—Estimating the Association of Many Variables with Mortality

To complete our survival analysis, we would be sure that the survival advantage we have recorded for laparoscopic patients is not linked to other factors that could have justified the significant difference observed. For example, the laparoscopic advantage may be driven by the fact that patients submitted to laparoscopy presented themselves with a more favorable disease, more little, with a reduced number of nodules, less patient’s comorbidities, younger age, or whatever other medical reason that could modify the risk of mortality.

Moreover, now we know that a laparoscopic approach may increase survival, but we would like to quantify how much the risk is modified when compared with the open technique.

In this condition, we definitely need to perform a multiple Cox Regression analysis. There are several ways to decide which variables should be inserted in the multivariate model, but this is not the point of this paragraph. We will run a model with all the confounders we have at disposal in our dataset. Just note that the total number of deaths is 301, so following the “one to ten” rule of thumbs we can insert in the model up to 30 variables: such a great number of confounders can be explored with this cohort!

As a result of the Cox regression (Fig. 8.9), we can now estimate that performing an open approach independently increases the hazard of mortality by 235% (HR 3.35, 95% CI: 1.5–7.2, p : 0.002) when compared to the laparoscopic one, fixing all the other confounders we have investigated. This is a strong confirmation of the effect of the treatment because now we could be sure that, at least for all the variables investigated (remember that, particularly in the retrospective studies, there could be always other confounders that we did not record that could justify the risk variation), there is an independent and significant survival difference linked to the treatment we are investigating. There are also other factors that, alone and independently, modify the hazard of mortality according to our analysis: the age (an increase of mortality by 2% per each year), the presence of cirrhosis, the MELD score (10% of increase per each point of MELD), and the presence of satellitosis. Now we can conclude satisfactorily our survival analysis and we can discuss the results we have measured!