

# Chapter 2

## A Survival Analysis Guide in Oncology



Smaranda Belciug

**Abstract** Survival analysis has a crucial role in oncology. These statistical methods are ubiquitous in oncology, helping physicians determine the death risk, the best course of treatment, and even help in discovering new therapies. The aim of this chapter is to provide a guide on some survival analysis statistical methods as well as practical examples on how to apply them in oncology. We begin with some theory regarding survival analysis in general, followed by Kaplan–Meier survival curves, logrank test and hazard ratio to determine the best course of treatment. The chapter ends with the Cox regression. This chapter is a step-by-step guide in performing survival analysis and in interpreting the obtained results.

**Keywords** Survival analysis · Kaplan–Meier curve · Logrank test · Hazard ratio · Cox regression

### 2.1 Introduction

Cancer remains the main health challenge we are facing nowadays. Seems like every day you find out that another person you know has been diagnosed with cancer. After the shock wave, you feel sad for that person, and guilty because you are somehow happy that this did not happen to you or your loved ones. Even if heart diseases still remain the first cause of death worldwide, the death rate caused by them is rapidly falling. Why is that? Because we found out what causes heart diseases: high blood pressure, high weight, high glucose levels, smoking, drinking, etc. Knowing the cause, we have the means to prevent them. Unfortunately, this is not the case for cancer. We still do not know what causes cancer.

Sadly enough, nowadays most people Google first the symptoms, before speaking to a physician. No matter what we are Googling, cancer will appear on the result page. So, we are alarmed, and while waiting for the doctor's appointment we start Googling for that type of cancer. The most searched questions are: what is the 5-year

---

S. Belciug (✉)

Department of Computer Science, Faculty of Sciences, University of Craiova, Craiova, Romania  
e-mail: [sbelciug@inf.ucv.ro](mailto:sbelciug@inf.ucv.ro)

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023  
C. P. Lim et al. (eds.), *Artificial Intelligence and Machine Learning for Healthcare*,  
Intelligent Systems Reference Library 229,  
[https://doi.org/10.1007/978-3-031-11170-9\\_2](https://doi.org/10.1007/978-3-031-11170-9_2)

29

survival rate? What is the overall survival? If there is need for surgery, what is the morbidity? What about the mortality rate? So, the number one outcome of interest is survival—overall, disease-free, recurrent, surgery survival.

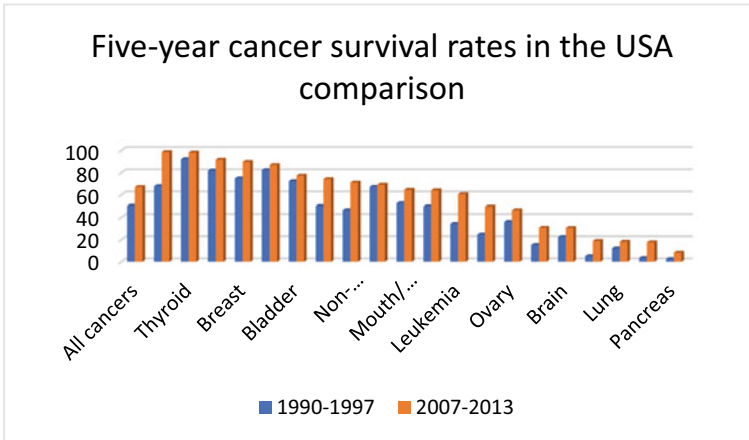
The Global cancer prevalence rose from 0.54 to 0.64% since 1990. For instance for prostate cancer the rates rose from 67.8% to 98.6%, due to better AI prediction, [1]. Even if the cancer rates are rising, the death rates are falling. This can only mean one thing: early diagnosis and/or better novel treatments, hence people have better and longer survival rates. The 5-year survival rates for all cancers have increased from 50.3% to 67% [2–4]. Table 2.1 and Fig. 2.1 present how the 5-year survival rates for different types of cancers have changed from 1970–1977 to 2007–2013 in the USA.

WHO’s global target ( $25 \times 25$ ) is a 25% reduction in deaths from cancer in people aged 30–69 years by 2025 [5]. Cancer survival research is crucial for developing cancer control strategies [6], control measures [7], so that the effectiveness and costs of them to be assessed [8, 9].

Survival analysis concerns the time until a certain event takes place: i.e. the time that passes from the start of chemotherapy until the tumor stops shrinking (the patient stops responding to treatment), the time elapsed from when a cancer surgery is over

**Table 2.1** 5-year cancer survival rates in the USA comparison

Cancer type	1990–1997 (%)	2007–2013 (%)
All-cancers	50.3	67
Prostate	67.8	98.6
Thyroid	92.1	98.2
Melanoma	81.9	91.7
Breast (female)	74.8	89.7
Uterus	82.3	86.9
Bladder	72.3	77.3
Kidney	50.1	74.1
Non-hodgkin lymphoma	46.5	71
Cervix uteri	67.1	69.1
Mouth/throat	52.5	64.5
Colon	49.8	64.1
Leukemia	34.3	60.65
Myeloma	24.6	49.6
Ovary	36	46.5
Stomach	15.2	30.6
Brain	22.4	30.5
Esophagus	5	18.8
Lung	12.2	18.1
Liver	3.4	17.6
Pancreas	2.5	8.2



**Fig. 2.1** 5-year cancer survival rates comparison

until the patients gets out of the ICU (Intensive Care Unit), or the time that passed from the moment the patient started radiotherapy, till she/he passed away.

In this chapter, we are going to present statistical models that are used in survival analysis, along with examples and explanations regarding the obtained results.

## 2.2 Survival Analysis

Survival analysis deals with survival times. In order for you to start such an analysis you need two variables: a numeric metric of time (i.e. number of days, weeks, or months), and a categorical variable that identifies the event (i.e. irresponsive to treatment, death). The merged variables give us a lot of information about whether a subject has entered or left the study, and if and when the subject has met a certain criterion or not.

Unfortunately, in practice things are never that simple, and we might find ourselves in two situations:

- The *start time* cannot be specified. For example, we cannot establish exactly when was the exact onset of the disease. Some cancers progress at a faster pace than others, being more aggressive, but there is no precise method to determine the debut of cancer.
- The *end time* is difficult to be determined. If the end time is given by the time of death, then there is no problem in establishing it, but if a person decides to leave the study, or she/he survives more than the time that was set to be recorded (i.e. 5-year survival rate), things change. This is the case of the *censored survival time*.

In Fig. 2.2 we present an example of how we can record the survival times. The figure is converted into a table, Table 2.2.

The dotted line from Fig. 2.2 gives us information about the study: the patients were recruited the first 5 months of the study. The timeline from month 5 till month 9 represents the follow-up part of the study. The black square signifies that the patient has died, whereas the grey circle means that the patient did not die. To be more specific: patient 1 was recruited at the beginning of the study and stayed in the study till month 3, time of which she/he died. The second patient enrolled in the study from month 0, and did not have any event (i.e. died) during the whole observation period, so she/he represents a censored data. Patient 3 also started from the beginning of the study, but died five months later. The fourth patient enrolled on month 1 and stayed

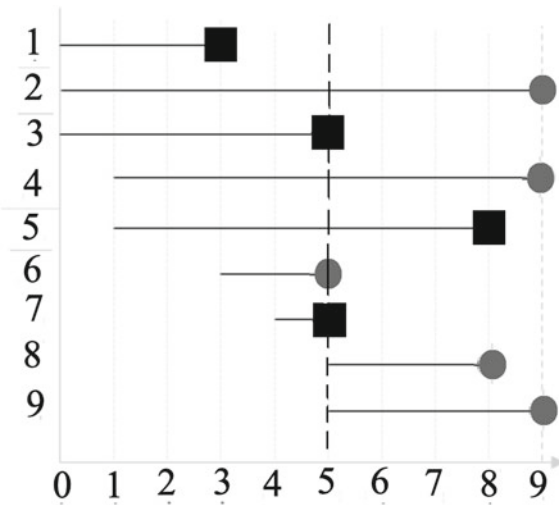
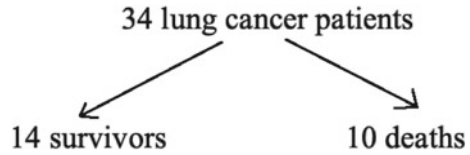


Fig. 2.2 Survival time recordings

Table 2.2 Tabulated survival time recordings

Patient	Start time	Time of death/censored	Death/censored	Survival time
1	0	3	D	3
2	0	9	C	9*
3	0	5	D	5
4	1	9	C	9*
5	1	8	D	7
6	3	5	C	2*
7	4	5	D	1
8	5	8	C	3*
9	5	9	C	4*

**Fig. 2.3** Survival analysis tree structure diagram



until the end of it. The fifth patient enrolled on month 1, and had an event (i.e. death) on month 8. The sixth patient enrolled on month 3, but left the study on month 5. Patient 7 joined the study on month 4 and passed away on month 6. Both patients 8 and 9 joined on month 5, the first leaving the study on month 8 (censored data), while the latter on month 9 (censored).

You can see in Table 2.2 that the censored data are marked with (\*).

In general, we use survival analysis in oncology to review the outcomes of clinical trials, cohort studies, etc. For instance, if we have a cohort of 30 patients who have been diagnosed with lung cancer between 2017 and 2019, they have started chemotherapy and/or immunotherapy and were observed until the end of 2021, we wish to review their *survival time*. So far, we have seen that survival analysis contains a *starting period*, in which the patients are enrolled, followed by the *observation period* or *follow-up*, when the patients are observed. Besides these two stages, there exists another one named the *final period*. In this stage the collected data is analyzed, and conclusions are drawn.

Let us presume that from the 30 lung cancer patients, 6 of them left the study during the observation period. The 6 patients will be excluded from the statistical analysis process. From the remaining 24 patients, 14 survived and 10 passed away. This observation is depicted in a tree structure diagram, as the one presented in Fig. 2.3.

A more thorough tree diagram will include even the patients that left the study. See Fig. 2.4.

We can compute the *death rate* or the *death risk* using the following formula:

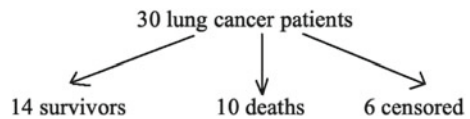
$$deathrate = \frac{number\ of\ deaths}{number\ of\ subjects}$$

In our example the death rate is 0.41, that is 41%.

The *death probability* is computed as:

$$Deathprobability = \frac{D}{N - 0.5 \times L}$$

**Fig. 2.4** Survival analysis tree diagram 2



where  $N$  is the cohort size,  $D$  is the number of deaths, and  $L$  is the number of patients that left the study during the observation period. The death probability in our case is 0.37, that is 37%.

Having computed the death risk, we can compute the *survival probability* as 1—*death probability for that interval*. By plotting the cumulative survival probability, we obtain the *survival curve*. The curve starts at 1 meaning that all patients are alive and approaches 0 as patients start to die. In the following sections we shall discuss more about survival curves, starting with the *Kaplan–Meier survival curves*.

### 2.3 Kaplan–Meier Survival Curve

Kaplan–Meier curves were invented in 1958 by Edward L. Kaplan and Paul Meier, and they can be used if the data is incomplete [10]. They represent the standard for reporting the survival rate of patients, being used in over 70% of the oncology papers [11].

Kaplan–Meier curves use three types of data regarding the patient: the date the patient entered the study, the last date of observation (i.e. the last time the patient was seen alive), and whether the last observation was due to the death of the patient, or because the patient left the study. We can use Kaplan–Meier curves to determine the survival probability of a patient given certain conditions. For example, by recording the survival times of patients that undergo chemotherapy, we can compute the probability of a new patient to survive a certain period of time if she/he undergoes the same protocol.

We denote the survival time with a random variable  $X$ .  $P_n$  is the probability of a patient to survive the  $n$ th day after the last chemotherapy session, conditioned by the fact that she/he survived all the other  $n - 1$  days before that.  $\overline{P}_n$  is the total probability of surviving all the  $n$  days, and we compute it as follows:

$$\overline{P}_n = P_1 \cdot P_2 \cdot \dots \cdot P_n$$

We compute the intermediate survival probabilities using:

$$p_k = p_{k-1} \times \frac{r_k - f_k}{r_k},$$

where  $p_k$  is the probability of surviving  $k$  units of time,  $r_k$  is the number of patients with a death risk at the  $k$  moment, that survived  $k$  units of time, and  $f_k$  the number of deaths reported at the  $k$  moment. If no patient has died the survival rate is 100%. We compute the standard error of the probability of surviving using:

$$SE_{p_k} = p_k \cdot \sqrt{\frac{(1 - p_k)}{r_k}}.$$

If we presume that  $p_k$  is governed by the Normal distribution, than we can compute the 95% interval as follows:

$$(p_k - 1.96 \times SE_{p_k}, p_k + 1.96 \times SE_{p_k})$$

The standard error does not always give accurate approximations if there are extreme values in the data sample. If this is the case, the Greenwood formula is preferred:

$$SE_{p_k} = p_k \cdot \sqrt{\sum_{j=1}^k \frac{f_j}{r_j \cdot (r_j - f_j)}}$$

Let us presume that we have a sample data that contains 16 patients that have been diagnosed with stage IV lung cancer. All the patients have undergone chemotherapy treatment with a certain type of drug, drug A. The patients are monitored 14 months. We start the survival analysis from day 0 (Table 2.3).

Using the survival probability formula, we compute the survival probability at a given time. Table 2.4 presents these calculations.

The corresponding Kaplan–Meier curve is plotted in Fig. 2.5.

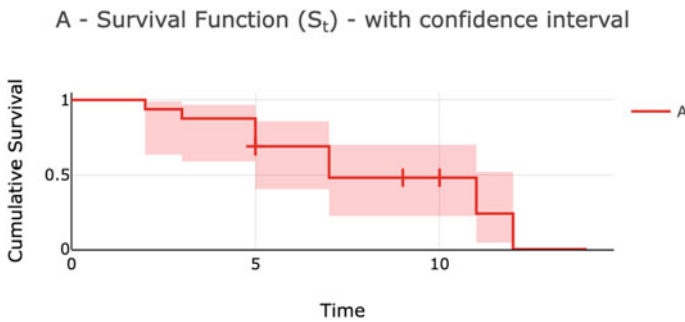
Kaplan–Meier survival curves have a drawback: if we wish to compare two or multiple sample data, we can obtain only a comparison at a certain moment in time, not a global one. To resolve this issue, we can use the logrank test, or hazard ratio.

**Table 2.3** Life table for lung cancer patients that underwent chemotherapy with drug A

No. of subjects	Survival time (months)	Event/Censored
1	2	1
2	3	1
3	5	1
4	5	1
5	5	1
6	5	0
7	7	1
8	7	1
9	7	1
10	9	0
11	10	0
12	10	0
13	11	1
14	11	1
15	12	1
16	12	1

**Table 2.4** Survival probabilities, standard error and confidence interval

Time	Events	Survival rate	$p_k$	$SE$	Confidence interval
0	0	1	0	0	(1, 1)
2	1	0.9375	0.9375	0.0605	(0.6323, 0.991)
3	1	0.875	0.9333	0.0826	(0.586, 0.9672)
5	3	0.6875	0.7857	0.1159	(0.4046, 0.8563)
7	3	0.4812	0.7	0.1285	(0.2241, 0.6993)
9	0	0.4812	1	0.1285	(0.2241, 0.6993)
10	0	0.4812	1	0.1285	(0.2241, 0.6993)
11	2	0.2406	0.5	0.1364	(0.04473, 0.5204)
12	2	0	0	0	(0, 0)



**Fig. 2.5** Kaplan–Meier survival curve

## 2.4 The Logrank Test

The logrank test is a non-parametric test that uses the null hypothesis  $H_0$ : “there is no difference between the two groups”. We perform this test by dividing the time scale according to observed events (i.e. deaths), while ignoring the censored data. For each interval we compute the observed number of deaths and the expected one, summing them up.

If we have two groups of patients, each group receives a certain chemotherapy drug. We will divide the survival time in time periods. Each period ends with one or multiple deaths. For each death unit, and each patient group, we compute the number of patients that are at death risk. Let us denote with  $r_1$  the number of patients with death risk for sample group 1, and with  $r_2$  the number of patients with death risk for sample group 2. Next, we compute the number of observed deaths for each group,  $f_1$  and  $f_2$ . With this information we proceed on building the following table, Table 2.5.

The expected number of deaths for each group is computed using the following formula:



**Table 2.5** Death and survival data regarding the two sample groups of patients

	Group 1	Group 2
Deaths	$f_1$	$f_2$
Survivals	$r_1 - f_1$	$r_2 - f_2$
Total	$r_1$	$r_2$

$$e_i = \frac{r_i \cdot f}{r}, i = 1, 2.$$

Next, we must sum up the observed values,  $O_i$ , as well as the expected values,  $E_i$ :

$$O_i = \sum_j f_{ji}, \quad i = 1, 2,$$

$$E_i = \sum_j e_{ji}, \quad i = 1, 2.$$

The logrank statistics is computed as follows:

$$X^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}.$$

For multiple groups we will use:

$$T = \sum_{j=1}^n \sum_{i=1}^m \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

To verify whether we accept or reject the null hypothesis, we will use the  $O_1 + O_2 = E_1 + E_2$  as control equality. The statistic value is compared with a  $\chi^2$  distribution with  $(n - 1)(m - 1)$  degrees of freedom.  $n$  represents the number of groups, whereas  $m$  represents the number of time intervals [12, 13].

Let us exemplify how the logrank test works. Let us presume that we are conducting a clinical trial, with two types of immunotherapy drugs,  $A$  and  $B$ . For 14 months we have monitored the two groups of patients diagnosed with IV grade lung cancer. The first group contains the 16 patients from the Kaplan–Meier section, whereas the second group contains 12 patients. Table 2.6 presents the data regarding the two sample groups.

Using the survival probability formula, we compute the survival probability at a given time. Table 2.7 presents these calculations.

The corresponding Kaplan–Meier curve is plotted in Fig. 2.6.

For a better comparison we shall plot both curves in the same plot. Figure 2.7 present this plot.

Using the logrank equations we can build the following table, Table 2.8:

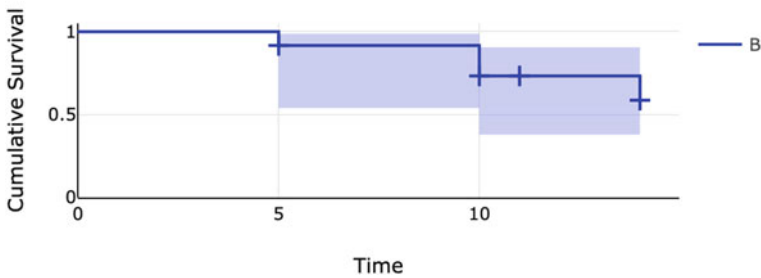
**Table 2.6** Life table for lung cancer patients that underwent chemotherapy with drug *B*

No. of subjects	Survival time (months)	Event/Censored
1	14	1
2	14	0
3	5	1
4	5	0
5	10	1
6	10	1
7	10	0
8	11	0
9	11	0
10	14	0
11	14	0
12	14	0

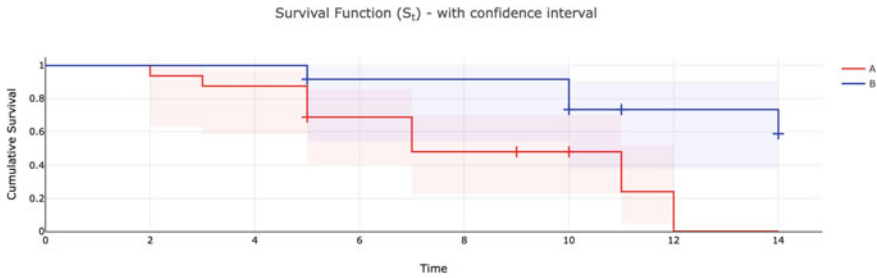
**Table 2.7** Survival probabilities, standard error and confidence interval

Time	Events	Survival rate	$p_k$	SE	Confidence interval
0	0	1	0	0	(1,1)
5	1	0.9167	0.9167	0.0605	(0.6323, 0.991)
10	2	0.7333	0.8	0.0826	(0.586, 0.9672)
11	0	0.7333	1	0.1159	(0.4046, 0.8563)
14	1	0.5867	0.8	0.1285	(0.2241, 0.6993)

B - Survival Function ( $S_t$ ) - with confidence interval



**Fig. 2.6** Kaplan–Meier survival curve second group

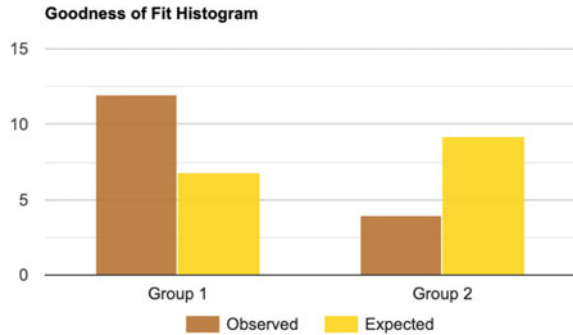


**Fig. 2.7** Kaplan–Meier curves for both sample patients

**Table 2.8** Goodness of fit table

Group	Observed frequency	Expected frequencies
Group 1	12	6.829
Group 2	4	9.170

**Fig. 2.8** Goodness of fit histogram



The Goodness of Fit histogram is presented in Fig. 2.8.

The test statistic  $\chi^2$  equals 6.829, while the  $p$ -level equals 0.0089. This means that we will reject the null hypothesis, implying that there are significant differences between the survival rates of the patients that use drug *A*, versus drug *B*.

## 2.5 The Hazard Ratio

Besides the information regarding the difference between two groups of observations, we might be interested in seeing how truly different the two groups are. In this matter, we cannot use the logrank test, but we can apply the hazard ratio. Technically, we will measure the relative survival between the two groups by comparing the observed and expected numbers [14–19]. The hazard ratio is computed using the following formula:

$$R = \frac{O_1/E_1}{O_2/E_2}$$

Returning to our example, we obtain the following results (Fig. 2.9; Table 2.9).

The obtained results show that the estimated relative risk of dying when undergoing immunotherapy with drug A is 3.1839 of the estimated relative risk of dying when undergoing chemotherapy with drug B. More specifically, if the hazard ratio equals 1, it means that there is no difference in survival rates / event rate over time between the two sample groups. If the hazard ratio is greater than 1, just like in our example, then the risk of having an event is greater in the group that uses drug A versus the group that uses drug B. Please note that the hazard ratio indicates an increase of hazard when using drug A, which is an increase in the rate of the event, not the chances of it happening.

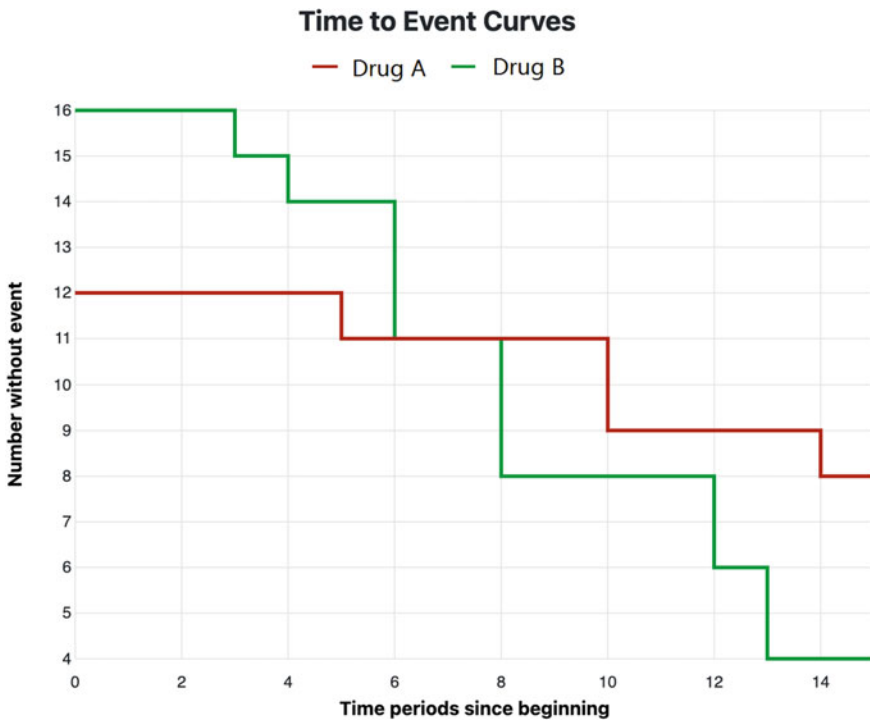


Fig. 2.9 Hazard ratio—Time to event curves

Table 2.9 Hazard ratio results

Hazard	95% Confidence interval	$\chi^2$	p-value
3.1839	(1.194, 8.487)	4.494	0.034

### 2.5.1 Cox Regression Model

The chapter ends with the *Cox's proportional hazard regression model* or *Cox regression*, which creates a *survival function* that gives us a certain event's probability (i.e. death, irrespective to treatment) to happen at a particular time  $t$ . Having previously observed and recorded data, we can build the model, and afterwards use it to make predictions on new patients. Cox regression can analyze multiple factors. Some may think that instead of using the Cox regression, we might be able to use the multilinear regression. This is not possible due to the following:

- in general, samples that contain survival times have exponential or Weibull distributions, and multiple linear regression cannot be applied unless the sample data is governed by the Gaussian distribution.
- Survival times contain censored data.

When using the Cox proportional hazard regression method, we need to compute the *survival function* and the *hazard function*. We compute the survival function as it follows:

$$S(t) = \{T > t\},$$

where  $t$  is the time, and  $T$  is the time remaining till the patient's death. Hence, we can write the *lifetime* distribution as:

$$F(t) = 1 - S(t).$$

We compute the number of deaths per time unit as:  $f(t) = \frac{d}{dt}F(t)$ . The *hazard function* is:

$$\lambda(t) = P\{t < T < t + dt\} = \frac{f(t)dt}{S(t)} = -\frac{S'(t)dt}{S(t)}.$$

Practically, by computing the hazard function we find the patient's death risk within the timeframe  $dt$ , when previously given  $T$  time left to live. The Cox regression model presumes that variables within the hazard function are independent and have a constant effect over the time of the survival, and each of them can be a predictor or covariance:

$$h(t; Z_1, Z_2, \dots, Z_k) = h_0(t) \cdot \exp(b_1Z_1 + b_2Z_2 + \dots + b_kZ_k).$$

The function can be afterwards transformed into:

$$\ln\left[\frac{h(t; Z_1, Z_2, \dots, Z_k)}{h_0(t)}\right] = b_1Z_1 + b_2Z_2 + \dots + b_kZ_k.$$

The  $h_0(t)$  is the *underlying hazard function* and represents the hazard when all the variables equal 0. Two assumptions must be fulfilled [20–22]:

- The hazard and the independent variables have a log-linear relationship;
- The *hypothesis of proportionality*: the relationship between the underlying hazard function and log-linear function of covariates exists.

Let us see how the Cox regression works on another fictional example. Our sample data contains 17 patients diagnosed with lung cancer. The dataset contains four attributes, four predictor variables (time, age, number of affected lymph nodes, number of months that have passed since the surgery), and the categorical variable (survival). The data is presented in Table 2.10.

First, we were interested in plotting the Kaplan–Meier curve to see the survival after the oncological lung surgery. Figure 2.10 show the curve together with 95% confidence interval.

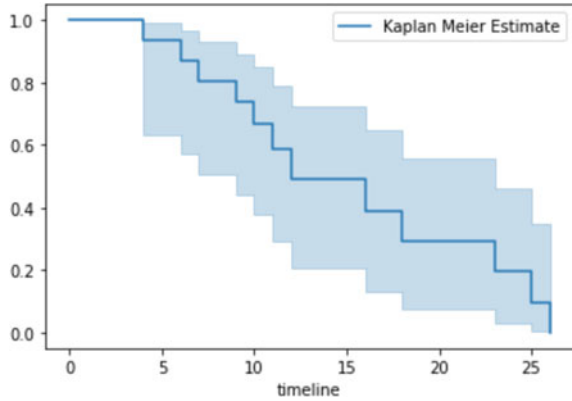
Next, we built two cohorts of patients. The first had no cancerous lymph node detected, the other had more than one. We have plotted the survival curve for both groups using Kaplan–Meier (Fig. 2.11).

We have applied the Cox regression model having as event the Survival attribute, and as duration the Time attribute. The obtained results are in Table 2.11. The summary statistic table indicates the significance of the covariates in predicting the Survival risk. The large confidence interval indicates that the sample data is small.

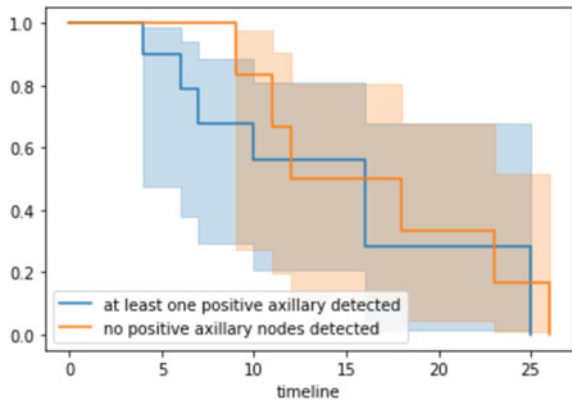
**Table 2.10** Fictional lung cancer patient dataset

# Patient	Time	Age	No lymph nodes	Months from surgery	Survival
0	12	57	1	6	1
1	18	65	3	7	1
2	24	63	0	12	1
3	24	56	2	16	1
4	11	62	4	10	0
5	26	75	0	23	1
6	16	48	0	11	1
7	19	51	0	9	1
8	5	66	1	5	0
9	6	63	2	4	1
10	4	68	4	2	0
11	15	66	5	10	0
12	20	65	0	18	1
13	28	59	1	25	1
14	12	70	1	11	0
15	12	72	2	10	1
16	28	65	0	26	1

**Fig. 2.10** Kaplan–Meier curve and 95% confidence interval for survival after oncological lung surgery



**Fig. 2.11** Kaplan–Meier curve and 95% confidence interval for two lung patient cohorts



The  $p$ -level shows us that the number of months that have passed since the oncological surgery is significant, while the others are not. The hazard ratio for this attribute is 0.71 showing a strong relationship between the number of months that have passed since the surgery and decreased risk of death. Notice that the hazard ration for Age is 1.01, which suggests only a 1% increase for the higher age group. Technically:

- Hazard ratio = 1: no effect
- Hazard ratio < 1: reduction in the hazard
- Hazard ratio > 1: increase in hazard.

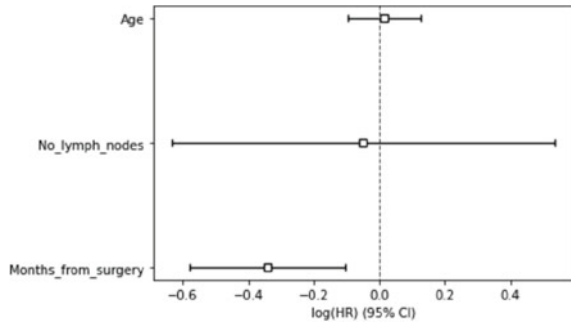
Let us see now which attributes affect the most from the following plot (Fig. 2.12):

From Fig. 2.12, we can clearly see that the number of months that have passed since the surgery is indeed significant, while the others are not. As a final note, we have plotted the survival probabilities for different persons in our dataset. From the graph (Fig. 2.13), we can see that patient 13 has the highest chances of survival, whereas patient 8 has the lowest.

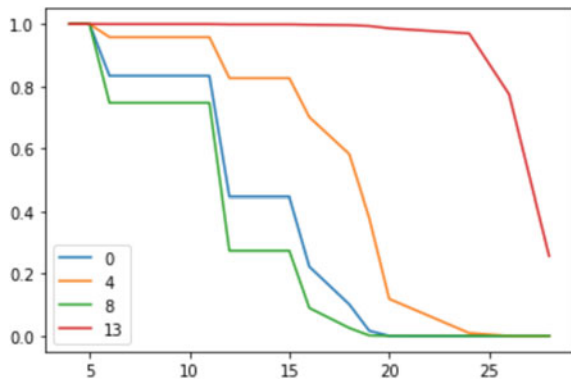
**Table 2.11** Results of Cox hazard regression

	Age	Number of lymph nodes	Months from surgery
Coef.	0.01	-0.05	-0.34
Exp(coef.)	1.01	0.95	0.71
SE(coef.)	0.06	0.30	0.12
Coef. lower 95%	-0.10	-0.63	-0.58
Coef. upper 95%	0.13	0.53	-0.10
Exp. (coef.) lower 95%	0.91	0.53	0.56
Exp. (coef.) upper 95%	1.13	1.71	0.90
Z	0.26	-0.16	-2.82
p-level	0.80	0.87	<0.005
-log2(p)	0.33	0.20	7.70

**Fig. 2.12** Significant attributes



**Fig. 2.13** Survival probabilities for patients: 0, 4, 8, and 13





## 2.6 Conclusions

This chapter provides a survival analysis guide with applications in oncology. Survival analysis represents an important part of cancer research. It can be applied to determine the survival rate of patients, to determine which treatment protocol is more efficient, or to establish whether new therapies are indeed better than the old ones. Using survival analysis in clinical trials we can move forward in providing the best care for cancer patients.

In this chapter we have discussed the theory behind Kaplan–Meier survival curves, logrank test, hazard ration, and Cox regression, as well as practical examples. We hope that this chapter will provide data scientists and oncologists a better understanding of the survival analysis process.

**Acknowledgements** The author would like to thank the reviewers for taking their time to review this chapter, and for their important comments that led to a better version of this work.

## References

1. Wulczyn, E., et al.: Predicting prostate cancer specific mortality with artificial intelligence-based Gleason Grading. *Commun. Med.* **1**(10). <https://doi.org/10.1038/s43856021-00005-3>
2. <https://ourworldindata.org/cancer>
3. Jemal, A., Ward, E.M., Johnson, C.J., Cronin, K.A., Ma, J., Ryerson, A.B., Mariotto, A., Lake, A.J., Wilson, R., Sherman, R.L., Anderson, R.N., Hensley, S.J., Kohler, B.A., Penberthy, L., Feuer, E.J., Weir, H.K.: Annual report to the nation of the status of cancer, 1975–2014, feature survival, *JNCI: J. Natl. Cancer Inst.* **109**, 9 (2017). <https://doi.org/10.1093/jnci/djx030>
4. Allemani, C., Weirm H.K., Carreira, H., Harewood, R., Spika, D., Wang, X.S., Bannon, F., Ahn, J.V., Johnson, C.J., Bonaventure, A., Gragera, R.M., Stiller, C., Azevedo e Silva, G., Chen, W.Q., Ogunbiyi, O., Rachet, B., Soeberg, M., You, H., Coleman, M.P.: Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet* **385**(9972), 977–1010 (2015)
5. WHO (2012). Decisions and list of resolutions of the 65th World Health Assembly: prevention and control of noncommunicable diseases: follow-up to the high-level meeting of the United Nations General Assembly on the prevention and control of non-communicable diseases (A65/DIV/3). World Health Organization, Geneva
6. Coleman, M.P., Forman, D., Byrnt, H., et al., The ICBP Module 1 Working Group: Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* **377**, 127–138 (2011)
7. Department of Health: Improving Outcomes: a Strategy for Cancer. Department of Health, London (2011)
8. Rachet, B., Maringe, C., Nur, U., et al.: Population-based cancer survival trends in England and Wales up to 2007: an assessment of the NHS cancer plan for England. *Lancet Oncol.* **10**, 351–369 (2009)
9. Rachet, B., Ellis, L., Maringe, C., et al.: Socioeconomic inequalities in cancer survival in England after the NHS Cancer Plan. *Br. J. Cancer* **103**, 446–453 (2010)
10. Kaplan, E.L., Meier, P.: Nonparametric estimations for incomplete observations. *J. Am. Stat. Assoc.* **53**(282), 457–481 (1958)

11. Stalpers, L.J., Kaplan, E.L.: Edward L. Kaplan and the Kaplan-Meier survival curve. *BSHM Bull.* **33**, 109–135 (2018)
12. Bland, J.M., Altman, D.G.: The logrank test. *BMJ* **328**(74447), 1073 (2004)
13. Gorunescu, F., Belciug, S.: *Incursiune in Biostatistica*, Editura Albastra – Grupul Microinformatica (2014)
14. Altman, D.G.: *Practical Statistics for Medical Research*. Chapman and Hall (1991)
15. Belciug, S.: *Artificial Intelligence in Cancer: Diagnostic to Tailored Treatment*. Elsevier (2020)
16. Georgiev, G.Z.: One-tailed versus two-tailed tests of significance in A/B testing, <https://blog.analytics-toolkit.com/2017/one-tailed-two-tailed-tests-significance-ab-testing/> (2017)
17. Stare, J., Maucoort-Boulch, D.: Odds ratio, hazard ratio and relative risk. *Metodoloskizvezki* **13**(1), 59–67 (2016)
18. Sashegyi, A., Ferry, D.: On the interpretation of the hazard ratio and communication of survival benefit. *Oncologist* **22**(4), 484–486 (2017)
19. Spruance, S.L., Reid, J.E., Grace, M., Samore, M.: Hazard ratio in clinical trials. *Antimicrob. Agents Chemother.* **48**(8), 2787–2792 (2004)
20. Cox, D.R.: Regression models and life tables. *J. Roy. Stat. Soc. B* **34**, 187–202 (1972)
21. Andersen, P., Gill, R.: Cox's regression model for counting processes, a large sample study. *Ann. Stat.* **10**, 1100–1120 (1982)
22. Harrell, F.E., Jr.: *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer (2001)