REVIEW ARTICLE

Survival Analysis

Part 15 of a Series on Evaluation of Scientific Publications

Isabella Zwiener, Maria Blettner, Gerhard Hommel

SUMMARY

Background: Survival times are often used to compare treatments. Survival data are a special type of data, and therefore have to be analyzed with special methods.

Methods: We illustrate special techniques for analyzing survival times by applying them to a publication on the treatment of patients with brain tumors. The present article is based on textbooks of statistics, a selective review of the literature, and the authors' own experience.

Results: Survival times are analyzed with the Kaplan-Meier method, which yields two measures of interest: survival rates and the median survival time. The log-rank test is used to compare survival times across treatment groups. Cox regression is used in multivariable models. The hazard ratio, a descriptive measure for differences in survival times, is explained.

Conclusion: If survival times are analyzed without the use of special techniques, or if the underlying assumptions are not taken into account, faulty interpretation may result. Readers of scientific publications should know these pitfalls and be able to judge for themselves whether the chosen analytical method is correct.

► Cite this as:
Zwiener I, Blettner M, Hommel G: Survival analysis—part 15 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2011; 108(10): 163–9.
DOI: 10.3238/arztebl.2011.0163

In many areas of medicine, the primary target parameter is the time until an event occurs. Examples include the time from diagnosis of lung cancer to death, the time from fitting dentures to first repair, and the time from the beginning of treatment for urinary incontinence until successful treatment outcome. An "event" may be either success (cure) or failure (death). It is important that both the beginning of the period of time and the time of the event are clearly defined. The time between the two is generally called survival time, even when the event which ends it is not death.

Almost all specialized medical publications include articles in which survival analysis techniques are used. A recent example of this is a trial in patients with brain tumors. Von Hoff et al. (1) investigated 280 children and young people with medulloblastoma in the two-arm, randomized trial HIT '91 (HIT = *Hirntumor* [German for brain tumor]). Patients in arm 1 received chemotherapy before and after radiotherapy ("sandwich" chemotherapy), while patients in arm 2 first received radiotherapy and then chemotherapy (maintenance chemotherapy). The trial investigated whether one of the two types of treatment led to longer patient survival times.

In order to interpret the results and value of such publications correctly, readers should be familiar with the methods used to analyze survival times. This article provides a step-by-step introduction to survival analysis techniques based on the HIT '91 trial and enables readers to understand and interpret them themselves.

The nature of survival time data

For both ethical and financial reasons, clinical trials last for only a limited period of time. In some patients, the expected event, e.g. death or success of treatment, does not occur until after the end of the trial, or even not at all. This means that the only information available on these patients is that no event has yet occurred as of a particular point in time. This is known as censoring. Censoring can also occur when individuals leave a trial. This occurs, for example, when they no longer wish to take part in the trial or die for reasons unrelated to the trial.

In oncology, a distinction is often made between overall survival (the time from diagnosis to death for any reason) and tumor-specific survival (the time from diagnosis to tumor-related death). In tumor-specific survival, patients who die for reasons unrelated to their

Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI) Universitätsmedizin Mainz: Dipl.-Math. Zwiener, Prof. Dr. rer. nat. Blettner, Prof. Dr. rer. nat. Hommel

BOX 1

Typical errors in survival time analysis

1. Evaluation of raw event frequencies

- For each patient, the only thing taken into account is whether or not an event was observed during the trial. When the event occurred and how long patients were observed with no events occurring are not considered.
  ⟶ Problem: comparison of treatments is based only on the frequencies of the observed events. This is usually incorrect, because the length of time until the event occurs is not taken into account. Events will be observed more frequently in patients with long follow-up times than in patients with short follow-up times.
- Example: There has been a standard treatment for brain tumors for the last 10 years. A new treatment was introduced one year ago (maximum follow-up time of patients receiving the new treatment: 1 year). We would like to investigate whether fewer patients die if they receive the new treatment. As many patients die 2 or 3 years after diagnosis, further patients who have received the new treatment will die in the future. Evaluation of raw event frequencies will produce biased results.

2. Exclusion of censored patients

- Only patients who have suffered an event are included in evaluation. Censored patients (those who have not suffered an event) are excluded from analysis. The time until the event occurs is compared using a t-test.
  ⟶ Problem: censored patients are patients who have not suffered an event at any point during the observation time. This is important information which must not be excluded from analysis.
- Example: 10% of patients who receive treatment A die within one year. 50% of patients who receive treatment B die within one year. If we only take into account *when* the 10% or 50% of the patients died, i.e. only those patients who died within the observation period (one year), then if censored patients are excluded all the patients included in the analysis died and both treatments appear equal. Information on how many patients did not die (i.e. censored patients) must therefore also be taken into account.

3. Censoring time = event time

- All patients are included in evaluation. However, no distinction is made as to whether a patient suffered an event or was censored. Survival time (censored or until an event) is compared using a t-test.
  ⟶ Problem: the results are biased because censorings are interpreted in the same way as events.
- Example: The same example as in point 1. With the standard treatment, there are patient data available for a period of up to 10 years. For the new treatment, data are only available for the first year. If we interpret all censored patients in the same way as patients who have suffered an event, the new treatment appears to be worse than the old treatment, as it has no survival times of more than 1 year (with the standard treatment there are patients with survival times of up to 10 years).

BOX 2

Kaplan–Meier method: an example based on data from 5 children with brain tumors

Patient data are shown in Table 1, in order of observation time. 3 of the 5 patients died.

TABLE 1

Survival times and Kaplan–Meier estimators

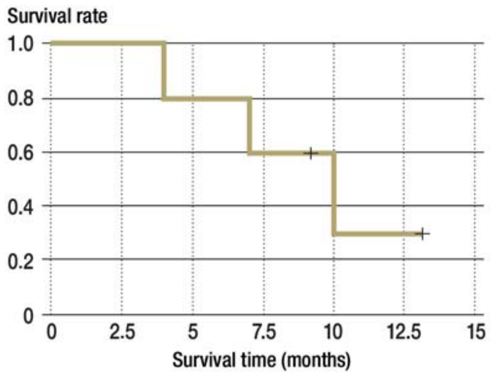
Pa-tient no.	Died	Survival time t_i (months)	n_i	d_i	Kaplan–Meier estimator $S(t_i)$
3	Yes	4	5	1	$\frac{4}{5}$ = 80%
5	Yes	7	4	1	$\frac{4}{5} \times \frac{3}{4}$ = 60%
2	No	9	3	0	
1	Yes	10	2	1	$\frac{4}{5} \times \frac{3}{4} \times \frac{1}{2}$ = 30%
4	No	13	1	0	

t_i: time of event no. i; n_i: no. of patients at risk at time t_i; d_i: no. of patients who have suffered an event by time t_i; $S(t_i)$: Kaplan-Meier estimator of the survival function at time t_i

Calculating the Kaplan–Meier estimator *(Table 1)*

- Month 4: 1 of 5 patients dies → Probability of surviving until at least the end of month 4 = $\frac{4}{5}$ = 80%
- Month 7: 1 of 4 patients dies ($\frac{3}{4}$ survive from month 4 to month 7) → Probability of surviving until at least the end of month 7 = $\frac{4}{5} \times \frac{3}{4}$ = 60% (the overall probability of surviving until at least the end of month 7 is the product of the two previous probabilities)
- Month 9: 1 censored patient (the patient did not suffer an event during the trial) → Only 2 remaining patients at risk but the Kaplan–Meier estimator remains the same (because up to this point in time no further patient has died)
- Month 10: 1 of 2 patients dies → Probability of surviving until at least the end of month 10 = $\frac{4}{5} \times \frac{3}{4} \times \frac{1}{2}$ = 30%
- Month 13: 1 censored patient → No more patients at risk → Kaplan–Meier estimator ends, and with it the Kaplan–Meier curve

FIGURE 1

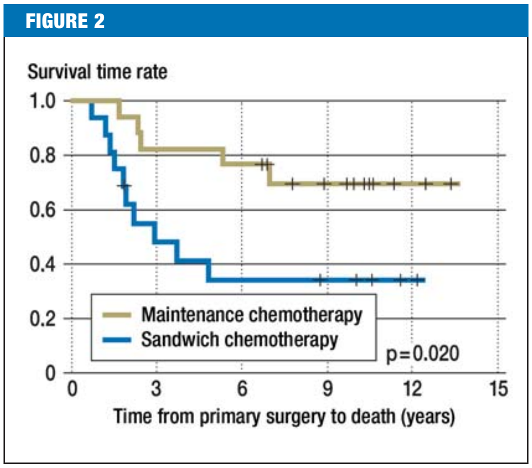


Kaplan–Meier curve *(Figure 1)*

With each death, the Kaplan–Meier curve drops. Censored patients are indicated by a vertical line (shown here in black). The Kaplan–Meier method does take censored patients into account:

- If we assume that patient #2 would have died, only 1 of 5 patients had survived the maximum observation time of 13 months: 20%.
- If we assume that patient #2 would have remained alive, 2 of 5 patients had survived the observation time of 13 months: 40%.
- However, we do not know what happened to the censored patient. The Kaplan–Meier estimator reflects this by estimating the survival rate at 30%.

FIGURE 2



Kaplan–Meier curve for 33 children and adolescents with medulloblastoma and metastasis status M1

Maintenance chemotherapy:	10-year survival rate = 70%, median survival rate cannot be determined
Sandwich chemotherapy:	10-year survival rate = 36%, median survival rate = 2.9 years

(From: von Hoff K., Hinkes B., Gerber N.U., Deinlein F., Mittler U., Urban C. et al.: Long-term outcome and clinical prognostic factors in children with medulloblastoma treated in the prospective randomised multicentre trial HIT '91. EJC 2009; 45: 1209–17 [1]; printed with the kind consent of Elsevier Publishers, Oxford)

tumors are censored because the event "tumor-related death" has not occurred. In more complex evaluations, both events can be investigated in parallel (as competing risks). However, this will not be examined in this article. HIT '91 investigated the time from primary brain tumor operation to death for any reason.

Alongside data from patients with known survival times, data from censored patients must also be included in evaluation. Specific evaluation strategies are needed in order for censored patients' data to be sufficiently reflected in analysis.

If survival time data are not evaluated in this way, the results are generally faulty. The mistakes most commonly made when evaluating survival time data are described in *Box 1*.

When evaluating survival times, it is important to take into account both the time until an event occurs and censored patients. This article describes methods for evaluation and graphical representation of survival time data on the basis of the trial HIT '91. Simple introductions to survival analyses are provided by textbooks by Weiß (2) and Schumacher and Schulgen (3). Textbooks by Collett (4) and by Kalbfleisch and Prentice (5) may be consulted for further reading.

Kaplan–Meier curves

Table 1, *Box 2* shows the survival times of five children with brain tumors. The probability that a patient has survived up to a certain point in time is calculated using the Kaplan–Meier method (6). The survival times can be shown graphically using a Kaplan–Meier curve (also called a survival time curve) (*Figure 1* in *Box 2*). Patients' survival times are plotted on the x-axis, and the probability of survival calculated according to the Kaplan–Meier method is plotted on the y-axis.

Calculation of the probability of survival and graphical representation using a Kaplan–Meier curve are explained step by step in *Box 2*.

Survival rates and median survival time

Survival rates can be determined using the Kaplan–Meier curve. Survival rates indicate the number of patients in whom no event has occurred up to a certain point in time. In the example above, the 1-year survival rate is 30% (*Box 2*). This can be interpreted as follows: one year after diagnosis, we can expect 30% of patients to be still alive. When stating survival rates, it is important to also state the point in time to which it corresponds. When comparing two treatment groups, it is advisable to plot Kaplan–Meier curves for both treatment groups, as these provide more information than survival rates alone.

The mean survival time is very much affected by censorings. Because of this, median values of survival times are always given. The median survival time is the time at which half the patients have suffered an event. The median survival time of the five brain tumor patients is ten months. If the Kaplan–Meier estimator for the whole observation period is more than 50%, the median survival time cannot be determined. In such cases, fewer than half the patients have suffered an event by the end of the observation period.

Log-rank test

In HIT '91, the survival times of patients from the two treatment groups were compared according to their metastasis statuses. Kaplan–Meier curves can be used for descriptive comparison of the two treatment groups' survival times for patients with metastasis status M1 *(Figure 2)*. The standard method, the log-rank test, was used for statistical comparison of survival times. The log-rank test examines whether there is a difference between two groups' survival times. This involves not only a specific point in time, such as the 6-month survival rate, but also the whole observation period. To put it more simply, we might say that Kaplan–Meier curves are compared with each other.

An extended form of the log-rank test can be used to compare three or more groups, e.g. to compare the survival times of patients with metastasis status M0 versus M1 versus M2/3. This means examining whether survival times are longer or shorter in at least one group than in the other groups.

In HIT ’91, the p-value of the log-rank test used to compare the treatment groups is 0.020. The difference between survival times is significant, with a significance level α = 5%. The group represented by the top curve is the group with the longest survival times. In this example, it is the group receiving maintenance chemotherapy. Patients who receive maintenance chemotherapy live longer than patients who receive sandwich chemotherapy.

Hazard and hazard ratio

Essentially, hazard is the instantaneous death rate for a particular group of patients. The hazard ratio is a quotient of hazards of two groups and states how much higher the death rate is in one group than in the other group. The hazard ratio is a descriptive measure used to compare the survival times of two different groups of patients. It should be interpreted as a relative risk (for relative risks see Ressing et al. [7]) and is described in more detail in *Box 3*. If the hazard ratio is 2.3 for patients with metastasis as compared to patients with no metastasis, the risk of death of patients with metastasis is 2.3 times as high as that of patients with no metastasis (in other words 130% higher).

BOX 3

Hazard and hazard ratio

Hazard h(t)

The risk of suffering an event at exactly time t is called the *hazard h(t)* and can be understood as the instantaneous risk of death. This risk may change over time, and is therefore dependent on time t. For example, if we consider the time from administration of medication to the occurrence of a particular adverse effect, e.g. nausea, the risk of nausea (the hazard) directly after administration of medication is higher than the risk of nausea one day or one week later.

Hazard ratio

When comparing two groups, the hazard functions $h_1(t)$ and $h_2(t)$ can be determined for the groups. The hazard ratio is the quotient of the two hazard functions:

$$\text{Hazard ratio} = \frac{h_2(t)}{h_1(t)}$$

The hazard ratio is a measure of how high the risk of an event is in group 2 in comparison to group 1. Group 1 is therefore considered to be the reference group. The following holds true:

- Hazard ratio >1 → Risk of event in group 2 higher than in group 1
- Hazard ratio <1 → Risk of event in group 2 lower than in group 1
- Hazard ratio ≈ 1 → Risk of event approximately equal in both groups

Cox regression

The simultaneous effects of several variables on survival time can also be investigated. The parameters examined in the HIT ‘91 study include the following:

- Treatment
- Sex
- Degree of resection
- Metastasis status.

The effect on survival time of age at operation, a continuous variable, should also be examined. Cox regression (8) can be used in both cases. Cox regression can also be used to obtain an estimator of the effect size. This estimator takes the form of the hazard ratio.

Underlying assumptions

Cox regression is based on the assumption that the hazard ratio remains constant over time (it is therefore also known as proportional hazards regression). This is true provided that the risk of an event (the hazard) of group 2 is proportional to that of group 1 (assumption of proportional hazard). Although the risk of an event (hazard) may vary over time, the variations over time must be the same in both groups. This assumption is not always justified, but can be approximately assessed using Kaplan–Meier curves. If the hazard in one of the two groups exceeds the hazard in the other permanently and to the same extent, the assumption of proportional hazard is valid. Represented graphically, this is the case when the Kaplan–Meier curves do not cross. If they do cross, it is not the case. Parmar and Machin (9) describe how to test the assumption of proportional hazard. The log-rank test is also based on the assumption of proportional hazard.

An example of a situation in which this assumption does not hold is the following: The survival times of patients who have undergone an operation need to be compared to those of patients who received radiotherapy instead of surgery. The risk of death is high immediately after surgery and then drops. In patients who receive radiotherapy, the risk of death at the beginning of treatment is low, but it may rise over time if radiotherapy is insufficiently effective. This means that the two death rates are not proportional to each other.

If Kaplan–Meier curves are used for patients with metastasis status M1 from HIT ’91 *(Figure 2)*, we can see that maintenance chemotherapy performs uniformly better. This means there is no evidence against the assumption of proportional hazards.

As with linear regression, there are also several possible methods for variable selection in Cox regression (see Schneider et al. [10]).

Example of Cox regression

In HIT ’91, three variables demonstrated an effect on overall survival *(Table 2)*:

- Treatment (binary)
- Metastasis status on diagnosis (categorial)
- Age on diagnosis (continuous).

TABLE 2

Results of Cox regression for overall survival in 280 children with medulloblastoma

	n	HR	95% CI	p
Metastasis status on diagnosis				
M0*	114			0.001
M1	33	2.11	1.13–3.94	
M2/3	40	3.06	1.76–5.33	
Unknown	93	1.54	0.94–2.52	
Treatment				
Maintenance chemotherapy*	127			0.006
Sandwich chemotherapy	153	1.76	1.17–2.67	
Age on diagnosis (years)	**280**	**0.93**	**0.88–0.98**	**0.005**

n: no. of cases; HR: hazard ratio; 95% CI: 95% confidence interval for hazard ratio; p: p value of likelihood ratio test; * Control group for categorial variables. (From: von Hoff K., Hinkes B., Gerber N.U., Deinlein F., Mittler U., Urban C. et al.: Long-term outcome and clinical prognostic factors in children with medulloblastoma treated in the prospective randomised multicentre trial HIT'91. EJC 2009; 45: 1209–17; printed with the kind consent of Elsevier Publishers, Oxford)

The reference group for the variable treatment consists of patients receiving maintenance chemotherapy. A hazard ratio of 1.76 can be interpreted as follows: The risk of death of children receiving sandwich chemotherapy is 1.76 times as high as that of children receiving maintenance chemotherapy.

There are four possible metastasis statuses:

- M0
- M1
- M2/3
- "Unknown" (Patients with unknown status are those in whom it was not clear whether their status was M0 or M1.)

The reference group used for comparison consists of patients with metastasis status M0. The risk of death in each of the three groups M1, M2/3 and "unknown" is compared with that of the control group, M0. So, three hazard ratios are calculated. The risk of death of children with status M1 is 2.11 times as high as that of children with status M0 (hazard ratio = 2.11); in other words, their risk is 111% higher. The risk of death of children with status M2/3 is 3.06 times as high as that of a child with status M0. The risk of death of patients whose metastasis status is unknown is 1.54 times as high as that of children with status M0. In addition to the hazard ratio, the confidence interval (11) must also be taken into account. The reference value here is "1" (meaning no effect).

With a continuous variable, the hazard ratio indicates the change in the risk of death if the parameter in question rises by one unit, for example if the patient is one year older on diagnosis. For every additional year of patient age on diagnosis, the risk of death falls by 7% (hazard ratio 0.93). Note that the unit chosen for the explanatory variable (in this case age on diagnosis in years, see Schneider et al. [10]) is retained when measures are interpreted.

Other important issues

Time-dependent variables

All the variables examined so far have been known at the beginning of survival time. For example, HIT '91 investigated whether metastases which were present at the time of brain tumor surgery affected survival. To investigate a variable that is still unknown at the beginning of survival time or that changes over time, time-dependent Cox regression must be used. For example, if we wish to know whether diabetes patients' cumulative dose of insulin affects the length of time until a cardiovascular event occurs, we cannot stipulate the cumulative dose as a known quantity at the outset. Patients who survive longer will generally receive a higher total dose. However, this high cumulative dose is not the cause of longer survival. To allow for this, the cumulative dose must be included in Cox regression as a time-dependent variable. Time-dependent Cox regression is a highly complex procedure. It is described at length in Collett's textbook (4).

Patients at risk

The term "patients at risk" refers to patients who are still alive at a particular point in time. The number of patients at risk, which varies over time, is often integrated into the Kaplan–Meier curve (under the time axis). As there are fewer patients at risk on the right-hand edge of the Kaplan–Meier curve (some have already died or been censored), this information allows us to determine how reliable the Kaplan–Meier estimate still is at the right-hand edge. The fewer the patients at risk, the higher the confidence interval of the Kaplan–Meier estimator.

Number of events

In order for results to be reliable, the number of events must be high enough. (N.B.: This does not mean the number of patients.) For each variable investigated using multivariable Cox regression, there must be at least ten events (12). If there is a small number of events, only a few explanatory variables can be investigated simultaneously. In HIT '91 there were 101 cases of death. This means that a maximum of ten variables can be included in Cox regression.

Sample size planning

A sample size calculation can be made for both the log-rank test and Cox regression. In addition to the significance level and the power to be achieved, we also need an estimated survival rate for each group to be compared or the estimated hazard ratio for a continuous explanatory variable (3). Sample size calculation also takes into account the recruitment and follow-up time.

Censoring

If censored patients are distributed differently in each of two treatment groups that are to be compared, biased estimators may result. The degree of completeness of follow-up in each treatment group should therefore be reported (see Clark et al. [13]).

Summary

As survival time data contain censorings, they must always be evaluated using the Kaplan–Meier method and the log-rank test. Analysis based on frequencies of events often produces faulty results. All doctors should understand Kaplan–Meier curves, the log-rank test and the results of Cox regression, as they must be able to explain them to patients (e.g. when choosing a treatment option: whether to treat a brain tumor with sandwich or maintenance chemotherapy).

Multivariable analyses can be performed using Cox regression. Results can be interpreted using hazard ratios and confidence intervals. Unfortunately, the underlying assumptions of Cox regression are not always taken into account (e.g. proportional hazards, time-dependent variables), and many published analyses are therefore faulty. Readers of scientific publications should know these pitfalls and be able to judge for themselves whether the chosen analytical method is correct.

Conflict of interest statement
The authors declare that no conflict of interest exists according to the guidelines of the International Committee of Medical Journal Editors.

Manuscript received on 1 June 2010, revised version accepted on 12 October 2010.

Translated from the original German by Caroline Devitt, MA.

REFERENCES

1. von Hoff K, Hinkes B, Gerber NU, Deinlein F, Mittler U, Urban C, et al.: Long-term outcome and clinical prognostic factors in children with medulloblastoma treated in the prospective randomised multicentre trial HIT´91. EJC 2009; 45: 1209–17.
2. Weiß C: Basiswissen Medizinische Statistik. 5th revised edition. Heidelberg: Springer Medizin Verlag 2010.
3. Schumacher M, Schulgen G: Methodik klinischer Studien. 3rd edition. Berlin, Heidelberg, New York: Springer 2008.
4. Collett D: Modelling survival data in medical research. 2nd edition. London: Chapman and Hall 2003.
5. Kalbfleisch JD, Prentice R: The statistical analysis of failure time data. 2nd edition. New York: Wiley, 2002.
6. Kaplan EL, Meier P: Nonparametric estimation from incomplete observations. JASA 1985; 53: 457–81.
7. Ressing M, Blettner M, Klug SJ: Data analysis of epidemiological studies—part 11 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2010; 107(11): 187–92.
8. Cox DR: Regression models and life tables (with discussion). Journal of the Royal Statistical Society (Series B) 1972; 74: 187–200.
9. Parmar MK, Machin D: Survival analysis: a practical approach. Cambridge: John Wiley and Sons 1995.
10. Schneider A, Hommel G, Blettner M: Linear regression analysis—part 14 of a series on evaluation of scientific publications Dtsch Arztebl Int 2010; 107(44): 776–82.
11. du Prel JB, Hommel G, Röhrig B, Blettner M: Confidence interval or p-value?—part 4 of a series on evaluation of scientific publications Dtsch Arztebl Int 2009; 106(19): 335–9.
12. Peduzzi P, Concato J, Feinstein AR, Holford TR: Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and Precision of regression estimates. Journal of Clinical Epidemiology 1995; 48: 1503–10.
13. Clark TG, Altman DG, De Stavola BL: Quantification of the completeness of follow-up. Lancet 2002; 359: 1309–10.

Corresponding author
Prof. Dr. rer. nat. Maria Blettner
Institut für Medizinische Biometrie (IMBEI)
Johannes Gutenberg-Universität
Obere Zahlbacher Str. 69
55131 Mainz, Germany