

**Ritesh Singh, Keshab
Mukhopadhyay¹**

*Departments of Community Medicine,
and ¹Pharmacology, College of
Medicine and JNM Hospital, Kalyani,
West Bengal, India*

Address for correspondence:

Dr. Ritesh Singh,
Department of Community Medicine,
College of Medicine and JNM Hospital,
Kalyani, West Bengal – 741 235, India.
E-mail: drriteshsingh@yahoo.com

Abstract

Survival analysis in clinical trials: Basics and must know areas

Many clinical trials involve following patients for a long time. The primary event of interest in those studies is death, relapse, adverse drug reaction or development of a new disease. The follow-up time for the study may range from few weeks to many years. A different set of statistical procedures are employed to analyze the data, which involves time to event analysis. It is a very useful tool in clinical research and provides invaluable information about an intervention. This article introduces the researcher to the different tools of survival analysis.

Key words: Cox proportional hazard model, hazard ratio, survival analysis

INTRODUCTION

Clinical trials are conducted to assess the efficacy of new treatment regimens. The major events that the trial subjects suffer are death, development of an adverse reaction, relapse from remission, and development of a new disease entity.^[1] Medical articles dealing with survival analysis often use Cox's proportional hazards regression model. These statistical models take into consideration time until an event of interest occurs and compare the cumulative probability of events over time for two or more cohorts, while adjusting other influential covariates. This article outlines the must know areas of survival analysis and introduces the reader to often-used terms in the survival analysis.

HISTORY OF SURVIVAL ANALYSIS

Survival analysis is a collection of statistical procedures

for data analysis, for which the outcome variable of interest is time until an event occurs. It is the study of time between entry into observation and a subsequent event. The term 'Survival analysis' came into being from initial studies, where the event of interest was death. Now the scope of the survival analysis has become wide. Today scientists are using it for time until onset of disease, time until stock market crash, time until equipment failure, time until earthquake, and so on.^[2] Common events studied are death, disease, relapse, and recovery. Few examples of studies where tools of survival analysis are used are: leukemia patients and time in remission, time to develop a heart disease for normal individuals, elderly population and time until death, and heart transplants and time until death.^[3]

No one is sure of the birth of this statistical procedure. Probably it originated centuries ago, but only after World War II a new era of survival analysis has emerged, being stimulated by an interest in the reliability of military equipment. At the end of the war these newly developed statistical methods, emerging from strict mortality data research to failure time research, quickly spread through the private industry as customers became more demanding of safer, more reliable products.

Access this article online

Quick Response Code:



Website:

www.picronline.org

DOI:

10.4103/2229-3485.86872

LIFE TABLE ANALYSIS

In longitudinal studies it is often of interest to estimate a 'survival' curve for the population. What proportion of the population survive beyond a specified time interval without a particular event happening?^[4] The most straightforward way to describe the survival in a sample is to compute the Life Table. The life table technique is one of the oldest methods for analyzing survival data. The distribution of survival times is divided into a certain number of intervals. For each interval we can then compute the number and proportion of cases or objects that entered the respective interval 'alive,' the number and proportion of cases that failed in the respective interval (number of terminal events, or number of cases that died), and the number of cases that were lost or censored (will be described later) in the respective interval. Based on those numbers and proportions, several additional statistics can be computed, such as, the number of cases at risk, proportion failing, proportion surviving, survival function, hazard rate, and median survival time. This procedure is used for larger samples where the time intervals are large enough to be broken down into smaller units.^[5] By using the life table analysis we can find out the probability of whether a woman who retained an IUD for the first six months will still have it by the end of the twentieth month. Similarly we can find out if Mrs. A, who has retained her IUD until now (the beginning of the eleventh month) and Mrs. B, who has also retained her IUD until now (the beginning of the thirteenth month) will both lose their IUDs within the next six months.

SOME COMMON TERMS USED IN SURVIVAL ANALYSIS

Censoring

Most survival analyses consider a key analytical problem called censoring. It occurs when we have some information about individual survival time, but we do not know the survival time exactly. Three reasons of censoring are: When a person does not experience the event before the study ends, when a person is lost to follow-up during the study period, and when a person withdraws from the study because of death (if death is not the event of the interest) or some other reason like adverse drug reaction. Censoring is of two types, right and left.^[6] We generally encounter right-censored data. Left-censored data can occur when a person's survival time becomes incomplete on the left side of the follow-up period for the person. As an example, we may follow up a patient for any infectious disorder from the time of his or her being tested positive for the infection. We may never know the exact time of exposure to the infectious agent.

Survival function

Survival function, $S(t)$ gives the probability that a person survives longer than some specified time t . It gives the probability that the random variable T exceeds the specified time t . The survival function is fundamental to a survival analysis. The survivor function is often expressed as a Kaplan-Meier curve. The name is a misnomer as in the actual data case scenario we get the step functions rather than smooth curves. Vertical drop in a Kaplan-Meier curve indicates an event.

Hazard function

The hazard function $h(t)$ gives the instantaneous potential per unit time for the event to occur, given the individual has survived up to time t . It is the probability of failure in an infinitesimally small time period between y and $y + \Delta y$ given that the subject has survived up till time y . In this sense, the hazard is a measure of risk: The greater the hazard between times y_1 and y_2 , the greater the risk of failure in this time interval. The hazard function has its own importance, as it provides an insight into the conditional failure rates; it may be used to identify a specific model form and it is the vehicle by which mathematical modeling of survival data is carried out.

Hazard ratio

Hazard ratio (HR) is akin to relative risk. It has been used to describe the outcome of therapeutic trials where the question is, to what extent can treatment shorten the duration of an illness.^[7] The hazard ratio is an estimate of the ratio of the hazard rate in the treated versus the control group. For example if there are two groups, group 1 and group 2, $HR = 4.5$ for treatment means that the risk (of relapse) for group 2 is 4.5 times that of group 1. If $HR = 1$ then $Group\ 1\ h(t) = Group\ 2\ h(t)$.

COX PROPORTIONAL HAZARDS MODEL

Clinical trials commonly record the length of time from study entry to a disease endpoint for a treatment and a control group. These data are commonly depicted with a Kaplan-Meier curve, from which the median (time at which, in 50% of cases, an event of interest has occurred) and the mean (average time for the event) can be derived. There are several methods available to analyze time-to-event curves, such as Cox proportional hazards, log-rank, and Wilcoxon two sample tests. The Cox proportional hazards model has been the most widely used because of its applicability to a wide variety of clinical studies.^[8] The Cox model was introduced by Cox, in 1972, for analysis of survival data with and without censoring, for identifying differences in survival due to treatment and prognostic factors (covariates or predictors or independent variables) in clinical trials. The Cox model is a regression

method for survival data. It provides an estimate of the hazard ratio and its confidence interval. Cox regression is considered a 'semi-parametric' procedure because the baseline hazard function, $h_0(t)$, does not have to be specified. There are two assumptions about the Cox proportional hazard model: The hazard ratios of two people are independent of time, and are valid only for time-independent covariates. This means that the hazard functions for any two individuals at any point in time are proportional. In other words, if an individual has a risk of death at some initial point in time that is twice as high as that of another individual, then at all later times the risk of death remains twice as high.

In a survival study, one should ensure that patients are not removed from the study just before they die. Survival studies often recruit patients over a long period of time and so it is also important to verify that other factors remain constant over the period, such as, the way patients are recruited into a study, and the diagnosis of the disease. The Cox model is popular as it is robust, the estimated hazards are always non-negative and the hazard ratio can be calculated.

Logistic regression is applied when the investigators examine the relationship between risk factors and various disease events. The ability to consider the time element of event occurrences by proportional hazards models has meant that logistic regression has played a less important role in the analysis of survival data.^[9] The Cox model is preferred over the logistic model, which ignores survival time and censoring information.^[10] Given a Cox model and the coefficients, we can subsequently estimate the baseline hazard function and the survival curves.

LOG RANK TEST

The log rank test (also known as the Mantel log-rank test, the Cox Mantel log-rank test, or the Mantel-Haenszel test) is a form of Chi-square test.^[11] It calculates a test statistic for testing a null hypothesis that the survival curves are the same for all groups, in other words, to test a null hypothesis where there is no difference between the populations in the probability of an event at any time point. For each time point the observed number of deaths in each group and the number expected if there has been no difference, are calculated. The number of expected is calculated as the proportion of subjects who are at risk at a given time point multiplied by the total number of events at that point. The log rank test is based on the same assumptions as the hazard ratio that the survival probabilities are the same for subjects early and late in the study, and the events happen at the time specified. The test is more likely to detect a difference

between groups when the risk of an event is consistently greater for one group than another. It is unlikely to detect a difference when survival curves cross. Hence it is useful to plot survival curves when analyzing survival data. Under the null hypothesis, the log-rank statistic is approximately chi-square with one degree of freedom. Thus, a *P*-value for the log-rank test is determined from tables of the chi-square distribution.

There are other tests for survival data. One of the important one is the 'Peto test'. It is an alternative to the log-rank test. In contrast to the log-rank test, the Peto test uses a weighted average of the observed minus expected score. It places more emphasis on the information at the beginning of the survival curve where the number at risk is large.

DISCUSSION

Survival analysis is a very good tool when a researcher takes into account the time till an event occurs and the censored data. There are some common mistakes performed by researchers when applying tools of survival analysis for their research.^[12] The first being, only data related to an event of interest occurring is reported. The *time* of the event is not mentioned. How long patients were observed with no events occurring is not considered. It is evident that events would be observed more frequently in patients with longer follow-up times than in patients with a short follow-up. Evaluation of raw event frequencies without mention of time will produce biased results. Similarly, when we get biased results, no distinction is made as to whether a patient suffered an event or was censored. The third error is not including the censored data in the analysis. If we take a specific proportion of events from both the groups, without taking into account the censoring, a different method of statistics should be employed, and not the survival analysis technique.

CONCLUSION

There are three primary goals of survival analysis, to estimate and interpret survival and / or hazard functions from the survival data; to compare survival and / or hazard functions, and to assess the relationship of explanatory variables to survival time. Survival analysis provides a great tool for analyzing the time to an event type of data, which is very common in any clinical trial. Researchers are not using it frequently because they are not confident in the theory of its application and its interpretation. There are books available that provide the basic knowledge on survival analysis. They should not make common mistakes while applying these tools to their data.

REFERENCES

1. Lee HP. On clinical trials and survival analysis. *Singapore Med J* 1982;23:164-7.
2. Smith T, Smith B. Survival analysis and the application of Cox's proportional hazards modeling using SAS. *Statistics, Data Analysis, and Data Mining*.
3. Survival analysis. A self-learning text. Kleinbaum DG, editor. USA: Springer; 2005.
4. Booth JG, Hirschl TA. Life Table analysis using weighted survey data. June 2005. Available from URL: <http://bscb.cornell.edu/~booth/papers/lifetable.pdf>. [Last accessed on 2011 Sep 06].
5. Ives M, Funk R, Dennis M. Survival Analysis/Life Tables. Available from URL: http://www.chestnut.org/li/downloads/training_memos/survival_analysis.pdf. [Last accessed on 2011 Sep 06].
6. Prinja S, Gupta N, Varma R. Censoring in clinical trials: Review of survival analysis techniques. *Indian J Community Med* 2010;35: 217-21.
7. Spruance SL, Reid JE, Grace M, Samore M. Hazard ratio in clinical trials. *Antimicrob Agents Chemother* 2004;48:2787-92.
8. Cox DR, Oakes D. Analysis of survival data. London, England: Chapman and Hall; 2001.
9. Abbott RD. Logistic regression in survival analysis. *Am J Epidemiol* 1985;121:465-71.
10. Seminar in Statistics: Survival Analysis Presentation 3: The Cox proportional hazard model and its characteristics. In: Fabsic P, Evgeny V, Zemmer K, editors. Zurich; 2011.
11. An Introduction to Survival Analysis. Mark Stevenson. EpiCentre, IVABS, Massey University. June 4, 2009.
12. Zwiener I, Blettner M, Hommel G. Survival analysis. *Dtsch Arztebl Int* 2011;108:163-9.

How to cite this article: Singh R, Mukhopadhyay K. Survival analysis in clinical trials: Basics and must know areas. *Perspect Clin Res* 2011;2:145-8.

Source of Support: Nil. **Conflict of Interest:** None declared.

Author Help: Reference checking facility

The manuscript system (www.journalonweb.com) allows the authors to check and verify the accuracy and style of references. The tool checks the references with PubMed as per a predefined style. Authors are encouraged to use this facility, before submitting articles to the journal.

- The style as well as bibliographic elements should be 100% accurate, to help get the references verified from the system. Even a single spelling error or addition of issue number/month of publication will lead to an error when verifying the reference.
- Example of a correct style
Sheahan P, O'leary G, Lee G, Fitzgibbon J. Cystic cervical metastases: Incidence and diagnosis using fine needle aspiration biopsy. *Otolaryngol Head Neck Surg* 2002;127:294-8.
- Only the references from journals indexed in PubMed will be checked.
- Enter each reference in new line, without a serial number.
- Add up to a maximum of 15 references at a time.
- If the reference is correct for its bibliographic elements and punctuations, it will be shown as CORRECT and a link to the correct article in PubMed will be given.
- If any of the bibliographic elements are missing, incorrect or extra (such as issue number), it will be shown as INCORRECT and link to possible articles in PubMed will be given.

Copyright of Perspectives in Clinical Research is the property of Medknow Publications & Media Pvt. Ltd. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.