Survival analysis—part 2: Cox proportional hazards model

Salil Vasudeo Deo [1,2] · Vaishali Deo [2] · Varun Sundaram [1,2,3,4]

Abstract

Learning objectives:

1. To understand the log-rank test and limitations of the log-rank test in comparing survival between groups.
2. To understand the fundamental concepts of the proportional hazards assumption.
3. To understand basic steps in the development of the Cox proportional hazards model and reported hazard ratios.
4. To understand how results of a Cox model run using STATA© (a commonly used proprietary statistical software) can be understood and interpreted.

Introduction

As mentioned in the first part of survival analysis, observational studies and randomized clinical trials (RCT) often involve a time to event outcome, where patients are followed up from the start of the study (e.g., after coronary artery bypass grafting) until the occurrence of the outcome of interest (time to event, e.g., time to first myocardial infarction after surgery) or the end of follow-up period [1]. In outcomes research, especially RCTs, a hazard ratio is often estimated from a Cox proportional hazards (CPH) model and is reported as the main measure of therapeutic efficacy. In this review, the authors elaborate on the rationale for the use of CPH model, its important assumptions, limitations and the key aspects related to the inappropriate interpretation of results from CPH models [2].

✉ Varun Sundaram
  vxs173@case.edu

1   Louis Stokes Veteran Affairs Medical Center, Cleveland, OH, USA

2   Case Western Reserve University School of Medicine, Cleveland, OH, USA

3   Harrington Heart and Vascular Institute, University Hospitals Cleveland Medical Center, Cleveland, OH, USA

4   Department of Cardiology, Louis Stokes Cleveland VA Medical Center, Cleveland, OH 44106, USA

Rationale for the Cox proportionate hazard model

In 1972, David Cox developed the proportional hazards model which derives robust estimates of covariate effects using proportional hazards assumption. In this review, we shall illustrate CPH model using an example of an observational study comparing mid-term survival after surgery for stage III lung cancer among males and females. The data for this example is available in the "survival" package in R (The R Foundation for Statistical Computing, Austria). As it is publicly available, institutional board approval was not needed for presenting these results. The first step in the analysis would be to report the observed survival for males and females in our cohort. These survival estimates can be easily calculated by the Kaplan and Meier method. These values can be graphed to present the survival estimates for patients in each patient group (i.e., two survival curves, one each for males and females). Figure 1 presents the survival estimates for females and males in our study of post-surgical patients with stage III lung cancer. This is to be followed by a formal statistical test, a log-rank test, to investigate whether the survival estimates for the two groups are statistically different (reported as p value in Fig. 1). This hypothesis testing is performed at a pre-specified confidence level (most commonly we set it at the 95% confidence level; hence, the importance of a p value < 0.05). The log-rank is a test of the whole survival estimates, rather than of the survivor functions at a particular time [3].

While log-rank test enables effective comparison of the survival in these two groups (females versus males), it has

certain important limitations. (i) Firstly, the log-rank test can only assess the effect of one variable at a time on prognosis. (ii) The log-rank test can be used to investigate the impact of a categorical confounder by looking at survival curves for the main exposure within strata defined by that confounding variable. However, it does not allow us to investigate the simultaneous impact of multiple categorical variables or continuous variables (e.g., age, body mass index, ejection fraction) on survival. In an observational study, it is important to control for multiple potential confounders in the analyses. (iii) The log-rank test can only tell us if there is a statistically significant difference between groups. It cannot provide a hazard rate or hazard ratio. Hence, it cannot quantify this difference. [4]

On the other hand, the CPH model enables us to investigate the effects of several continuous and categorical variables on survival, while accounting for possible confounders. Unlike the log-rank test (and other non-parametric models), CPH facilitates quantification of differences in survival distribution between two groups. We do this by estimation of a hazard ratio. The hazard ratio is the ratio of the event rate at any given time in one group (e.g., treatment group) relative to the other (e.g., control group) [5].
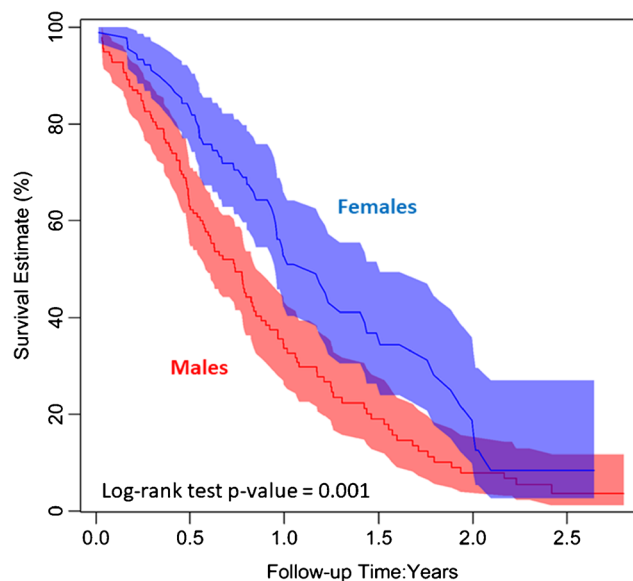

Fig. 1 This is an observational study comparing the all-cause mortality of post-surgical stage III lung cancer patients. The survival estimates of males and females in our cohort were calculated using the Kaplan and Meier estimates. This figure presents the estimated survival of males (red) and females (blue) with their respective 95% confidence intervals (colored bands). The difference in survival between these females and males was compared with the log-rank test (observed p value = 0.001)

What is CPH?

The hazard ratio (HR) is analogous to odds ratio used in multiple logistic regression analysis. It is the ratio of the total number of observed to expected events in two independent comparison groups. In our example of survival outcomes between females and males,

$$\text{HR} = \left(\frac{\Sigma \text{ Observed Events in females in } t}{\Sigma \text{ Expected Events in females in } t} \right) \Big/ \left(\frac{\Sigma \text{ Observed Events in males in } t}{\Sigma \text{ Expected Events in males in } t} \right)$$

Here, the event is death and t is the survival time. With the use of the equation listed above, we have merely examined the association between the type of valve implanted and long-term survival. However, in an observational study where the two groups are not equally balanced with respect to patient characteristics, it is important to measure the impact of confounders. Furthermore, it is often of interest to evaluate the association between several risk factors (both categorical and continuous) and survival time. CPH is one of the most commonly used regression techniques to examine this association while accounting for confounding. The CPH model can be described as follows:

$$h(t) = h0(t)e^{(b1X1+b2X2+\ldots+bpXp)}$$

where h is the expected hazard at time t and $h0(t)$ is the baseline hazard when all predictors $X1, X2\ldots, Xp$ are equal to 0.

Let us assume that in our example, that patient sex is the only predictor variable influencing survival. In this simple model of one predictor variable, the CPH would be $h(t) = h_0(t)e^{b1X1}$, where $X1$ is the sex of the patient. Let us start with the comparison of two participants (one in each group) in terms of the expected hazards; the first patient is female ($X1 = $ Female) and the second patient is male ($X1 = $ Male). The expected hazard for the two patients would be $h(t) = h_0(t)e^{b1\text{Female}}$ and $h(t) = h_0(t)e^{b1\text{Male}}$ respectively. The HR would be the ratio of these two expected hazards, $\text{HR} = (h_0(t)e^{b1\text{Female}})/(h_0(t)e^{b1\text{Male}}) = e^{(b1 * (\text{Female} - \text{Male}))}$.

It is clear from this equation that the time component is cancelled. Hence, the HR does not depend on time t, indicating a proportional hazard over time.

Assumptions for CPH

Like any other statistical model, CPH relies on certain important assumptions:

1. The proportional hazards assumption: In CPH, the hazard ratio is assumed to remain constant throughout the follow-up. In our example, it is reasonable to assume that the hazard for both the groups (females and males) remains same for the entire follow-up. However, this might not be true in all circumstances. For instance, in clinical trials comparing surgical versus medical therapy, as in Coronary-Artery Bypass Surgery in Patients with Left Ventricular Dysfunction (STICH trial), the surgical arm was associated with high mortality immediately after randomization due to procedural risk but conferred lower long-term mortality [6]. In such cases of deviation from proportional hazard assumption, alternative analysis strategies such as accelerated failure time model or a milestone analysis should be considered. [7]

2. Independence of survival times between distinct individuals in the study population. This means that the survival time of one patient does not depend upon the survival time of another. This assumption of independence is a criterion, which is also applied to other statistical methods like linear and logistic regression.

3. The last assumption is that the censoring is uninformative about the outcome of interest, i.e., it is important that those who have been censored have the same risk of suffering the study end-point as those who continue to be followed. To explain this further, the Cox model holds only if patients that are censored have the same risk of mortality, if they were still included in the study. For example, consider that we are conducting a trial to evaluate the benefit of a medication on 5-year survival with a regular periodic follow-up. Consider that a patient does not come for his next follow-up visit, because he suffers from a side effect of the treatment and hence visits another physician. Then, this patient will be censored from the trial. However, this type of censoring is *not* uninformative. Given his side effects, he may now be at a higher risk of suffering from the end-point specified in our study. But consider another scenario. A patient fails to keep his follow-up appointment because he moved to another city; this, however, is an example of uninformative censoring. We can safely assume that this patient continues to have the same risk of suffering the end-point in the new city, although he is censored from our present study.

Benefits of CPH

The CPH model is very popular among clinical researchers for numerous reasons. It does not need the researcher to specify the function of the baseline hazard. Provided proportional hazards assumptions are met, the results are robust. With results from the CPH model, the coefficients obtained can be

Table 1 The results of the CPH model explained in the text. The table provides the hazard ratios and their 95% confidence interval with p values for each variable included in the model. Hazard ratios are exponentials of the coefficients. When the coefficient is a positive number, then the hazard ratio is greater than unity. Similarly, when the coefficient is a negative number, the hazard ratio will be less than unity. If the hazard ratio > 1, then the risk is higher for the study group versus the control group. The converse is true when the hazard ratio is < 1

Covariate	Coefficient	Hazard ratio	Confidence interval	p value
Age at surgery	0.017	1.01	0.99–1.03	0.06
Female sex	− 0.513	0.59	0.43–0.83	0.001

used to model and predict the expected survival of patients with specific values of covariates included in the model. To understand this, we will again go back to the example dataset of 228 stage III lung cancer patients who underwent surgery. We would like to understand the association of patient sex and age at surgery with all-cause mortality. For this purpose, we will fit a CPH including these two covariates in the model.

From (Table 1), we observe that both variables independently influence all-cause mortality. Keeping sex constant (i.e., comparing only males or only females), a unit increase in age increases the risk of mortality by 1%. However, for patients with the same age, compared to males, females have a 41% reduced risk of mortality. Another benefit of a regression model, like the CPH model, is that it can be used to predict outcome for patients with specific values for covariates included in the model. While understanding the results of regression models, it is important to consider the confidence interval. The range of confidence interval provides an understanding of uncertainty inherent in the analysis.

We provide, below, an outline of steps used to conduct a CPH model. In the supplement, we present an example using STATA© (STATACorp, Station College, TX), a simple yet powerful proprietary statistical software. We have also provided a simplified dataset that readers can download and open in STATA©.

Steps in survival analysis using CPH

1. Create a null hypothesis, e.g., survival time $S(t)$ for females = $S(t)$ of males
2. Derive survival estimates using the Generate Kaplan and Meier method. This method accounts for right censoring observed in the data.
3. Log-rank test to investigate whether the survivor curves for the two groups are statistically different (p value).
4. Check the proportional hazards assumption for each covariate considered for the multi-variable CPH model. Hypothesis testing and plotting residuals from the model

against time are some methods to test the CPH assumptions. For all conventional time to event analyses, independence of observations and non-informative censoring are important assumptions that need to be accepted. There are different techniques available when patients are clustered together in groups, for example, they are operated by the same surgeon, or treated in the same hospital in a multi-institutional study. However, they are not the focus of this paper and will be discussed in future articles.

5. If proportional hazard assumption is met, CPH model could be employed to investigate the effects of multiple continuous and categorical variables on the time to event end-point. We can account for possible confounding and quantify differences in survival between the two groups, i.e., by the estimation of hazard ratio. If the proportional hazards assumption is not met, then other extensions of the Cox model are available to account for this. While reading a journal article, we would recommend readers to observe if authors have specified CPH testing in their methods section. The supplement section may provide plots or results of CPH tests for each covariate included in the model. Results obtained from the CPH model are naturally valid only if the data fulfils CPH tests.

Conclusion

Time to event outcomes are commonly used in cardiology and cardiothoracic surgery literature. The CPH model is the most widely used multivariate statistical model for survival analysis. Understanding the rationale and assumptions behind CPH model are important when using the Cox model in time to event analyses. The Cox model provides hazard ratios for variables included in the model. These hazard ratios can be easily understood by clinicians and they aid decision- making. The ability of the model to provide results easily understood by non-statisticians has likely led to the widespread use of this model in medical literature.

Further reading

Applied Survival Analysis: Regression Modeling of Time-to-Event Data

Author(s): David W. Hosmer, Stanley Lemeshow, Susanne May

First published: 26 February 2008 Print ISBN: 9780471754992 |Online ISBN: 9780470258019 |DOI:https://doi.org/10.1002/9780470258019 Copyright © 2008 John Wiley & Sons, Inc. All rights reserved. Book Series:Wiley Series in Probability and Statistics

This is a reference book on survival analysis. It provides detailed information regarding all aspects of analyzing time-to-event data.

Applied Survival Analysis using R.

Author: Dirk F Moore

Print ISBN: 978-3-319-31,243-9 |Online ISBN: 978-3-319-31245-3 |DOI: 10.1007/978-3-319-31245-3 Copyright © 2016, Springer International Publishing Inc. All rights reserved. Book Series: Statistics for Life Sciences, Medicine, Health Sciences.

This book covers methods of survival analysis with solved examples using R. R (The R Foundation for Statistical Computing) is an open-source versatile statistical language. While it has some learning curve, researchers can use R for advanced statistical computing.

An Introduction to Survival Analysis using STATA.

Authors: Mario Cleves, William W. Gould, and Yulia V. Marchenko.

ISBN-13: 978-1-59718-174-7| Publisher: Stata Press, Copyright: 2016.

This is an excellent practical manual for applied researchers that want to start analyzing time-to-event data using STATA© (The STATACorp, College Station, Texas). STATA© is a proprietary statistical software. It also has many user-written commands that allow researchers to apply advanced statistical methods. However, beginners have access to a Graphical user interface (GUI) that helps to learn the code easily.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This is a review article and does not contain confidential patient results. The example used in the paper is from a publicly available source; hence, the study is exempt from institutional board approval.

Informed consent The paper is a review paper. Hence, there is no need for informed consent. The data used as an example is publicly available.

Disclaimer This material is the result of work supported with services and facilities made available at the Louis Stokes Cleveland VA Medical Center. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

References

1. Tolles J, Lewis RJ. Time- to -event analysis. JAMA. 2016;315: 1046–7.

2. Pocock SJ. The simplest statistical test: how to check for a difference between treatments. BMJ. 2006;332:1256–8.

3. Schober P, Vetter TR. Survival analysis and interpretation of time-to-event data: the Tortoise and the Hare. Anesth Analg. 2018;127:792–8.

4. Bland JM, Altman DG. The logrank test. BMJ. 2004;328:1073.

5. Katz MH, Hauck WW. Proportional hazards (Cox) regression. J Gen Intern Med. 1993;8:702–11.

6. Velazquez EJ, Lee KL, Jones RH, et al. Coronary-artery bypass surgery in patients with ischemic cardiomyopathy. N Engl J Med. 2016;374:1511–20.

7. Gregson J, Sharples L, Stone GW, Burman CF, Öhrn F, Pocock S. Nonproportional hazards for time-to-event outcomes in clinical trials. JACC review topic of the week. J Am Coll Cardiol. 2019;74:2102–12.