

Introduction

The Application of the Cox Proportional Hazard Model and Its Application and Effectiveness

The Cox Proportional Hazard Model, also known as the Cox Regression or Cox PH model is an important statistical tool used in survival analysis. This model was developed in 1972 by a British statistician by the name of Sir David R. Cox to analyze time-to-event data. Survival analysis such as the Cox Regression is a time-to-event analysis, used to investigate the time until the occurrence of an event of interest. Applications of this regression model go far beyond statistical setting and are widely used in medical research and are growing in use in reliability engineering and other outcomes research. In this introduction, we will review Cox Regression, delve deeper into how the Cox Regression has been applied to three different medical research papers, note some of its uses outside of the medical field, and discuss how effective statistical analysis is.

Prior to the publication of the Cox Proportional Hazard model, tests such as the log-rank test were used but limited to observing the effect of one variable at a time [1]. Additionally, the logrank test could not quantify the effect of the variables [1]. The Cox proportional hazards model estimates the hazard as [1]:

$$h(t) = h_n(t) \exp(B_1 X_1 + B_2 X_2 + \dots B_p X_p)$$

h is the hazard at time t , $h_u(t)$ is the baseline hazard, X are the covariates.

The Cox proportional hazard model assumes the hazard rate is a product of a baseline hazard rate, that that baseline hazard rate is the hazard rate when the covariates have no effect on the event. A regression vector is estimated by maximizing a marginal, partial, or maximum likelihood function. The significance of the covariates can be tested with analytical methods (log rank test, or chi-squared test) or graphical methods (cumulative hazard rate plots or residual plots). Insignificant covariates can be removed and a regression vector recalculated [2].

Model Assumptions

Dividing both sides by $h_0(t)$ yields the hazard ratio

$$h(t)/h_0(t) = \exp(B_1 X_1 + B_2 X_2 + \dots B_p X_p).$$

This hazard ratio must remain constant over time to fit the assumptions of the Cox PH model. Babinska, et al. emphasize this in their paper looking at the Cox PH model's use in analyzing the outcomes of patients with acute coronary syndrome [3]. Cox proportional hazard model

also assumes events are independent and that censoring is uninformative [1]. That is that the time to event for one person (or thing) is independent of the time to event for another person (or thing) and people (or things) who do not continue to the end of the study, aka. are censored, have the same risk of the event as those that do continue until the end of the study period.

Reasons the Cox proportional hazards model is popular is that it does not require the baseline hazard, $h_0(t)$ to be known [1].

Babinska, et al. look into the limitations of Cox Regression in predicting mortality rates in patients with acute coronary syndrome (ACS) [3]. Although there have been significant advances in medicine and interventional therapies, cardiovascular disease (CVD) still burdens the medical system. The researchers see the importance of Cox Regression but worry about a potential violation of the proportionality of hazard assumption using a simple Cox regression model [3]. The study at first examined if proportional hazard assumptions are met in various factors including homocysteine and sodium concentrations. These assumptions are important for accurate Cox regression analysis because they ensure that hazard ratios remain constant. The study found there was a violation in some of the cases, such as homocysteine and sodium concentrations. Without verification of these assumptions, the increased risk of death due to ACS might be incorrectly identified. The study looked into one hundred and fifty consecutive patients with acute coronary syndrome (ACS) with a mean age of 65 (range; 21-92 years), 33 (22%) deaths was registered during the follow-up period of 64 months [3]. In the Cox regression model, several factors were identified as risk factors for long-term prognosis in patients with ACS. The factors included smoking, the presence of diabetes and anemia, the duration of coronary artery disease, and abnormal serum concentrations of uric acid, sodium, homocysteine, cystatine C, and NT-proBNP. The study stressed the importance of ensuring that the proportional hazard assumption is met in the Cox regression. Failure to meet the assumption can lead to false positives in the conclusion and incorrect identification of factors leading to ACS-related deaths. To remedy the violation, the study recommends using the Cox stratified regression model or extended regression models with time-dependent variables. The next study investigated the impact of regional medical disparities on complications in patients with hypertension while using the Cox Proportional Hazard model. The study took place in South Korea since it has become the fastest-aging country worldwide, and there is a greater likelihood of an increase in the prevalence of hypertension [4]. As of 2022, Seoul, the capital of South Korea, had 4.8 doctors per 1,000 inhabitants, and except for metropolitan cities (Busan, Daegu), all the other areas had an average of less than 3.2 doctors per 1,000 inhabitants [5]. The disparities between the number of physicians in the more populated regions and the rural areas play an important role in hypertension complications. The study's results aligned with similar studies conducted in China, Romania, and the United States. These studies also reported a correlation between disparities in hypertension treatment and control, particularly in rural areas. The study also suggested that the higher risk of hypertension complications in medically vulnerable regions may be due to disparities in access to healthcare. The study included a cohort of 246,490 participants, with the goal of investigating the interaction between residential regions (vulnerable vs. non-vulnerable) and diagnosis areas (outside vs. inside) in

relation to complications in patients. The participants were divided into four groups — vulnerable and outside region, vulnerable and inside region, non-vulnerable and outside region, and non-vulnerable and inside region. Multivariate Cox regression analysis was fitted. The interaction term between the residential region (vulnerable vs. non-vulnerable) and the diagnosis area (outside vs. inside) was significant in the adjusted model (p -value = 0.005) [4]. The Cox regression showed that individuals living in vulnerable and outside regions had the highest risk of complications. This analysis was supported by the hazard ratio (HR) in which vulnerable and outside regions had the highest ratio of 1.156. The vulnerable and outside region group had the highest rate of complications for cardiovascular and cerebrovascular diseases when compared to

the reference group. However, for kidney disease, the non-vulnerable and outside region group had the highest rate of complications. In this third article, researchers are using Cox regression in addition to Artificial Neural Network (ANN) to postoperative mortality after hip fracture surgery. Postoperative mortality is any death, regardless of cause thirty-days after surgery. As the population is aging more hip fracture has become a public health issue [6]. The substantial associated medical costs in the United Kingdom (UK) and Taiwan have increased for these procedures. This is why understanding and predicting the underlying causes of postoperative mortality after hip surgeries is of great importance. Traditional parameter models have not been reliable enough, ANN and Cox regression models are used due to the fact they are the most common models used in the healthcare industry for predicting postoperative mortality. The ANN model captures nonlinear interactions among risk factors, making it favorable in medical decision-making and mortality predictions. Cox proportional hazard (PH) model is standard in survival analysis but has a limitation in predicting longitudinal survival. The purpose of the study is to use ANN and Cox models to identify influential predictors of postoperative mortality after hip fracture and conduct a global sensitivity analysis to assess important predictors. In total, researchers analyzed 10,534 hip fracture procedures. During this period, 71.2% hip fracture patients were referred to lower-level medical institutions for rehabilitation after surgery, and 28.8% patients continued treatment at the same medical institution [7]. The mean age of patients was 68.3 (SD 14.6) years, with females representing 57.6% of the patients [7]. The training data set was 7374 cases, the testing data set of 1580 cases and the same number of 1580 for the validation dataset. The study found that the artificial neural networks (ANN) model outperformed the Cox regression model in predicting postoperative mortality after hip fracture surgery. Comparisons of performance indices in the testing dataset also showed that the ANN model significantly outperformed the Cox model in sensitivity (0.96 vs. 0.92, $p < 0.001$), specificity (0.76 vs. 0.64, $p < 0.001$), PPV (0.88 vs. 0.78, $p < 0.001$), NPV (0.84 vs. 0.77, $p < 0.001$), accuracy (0.93 vs. 0.90, $p < 0.001$), and AUROC (0.93 vs. 0.88, $p < 0.001$) [6]. In addition, what makes this study unique is the fact they are the first to utilize a nationwide population-based dataset for both training and testing an ANN to forecast outcomes of hip fracture surgery. As often, in previous studies data often originates from a single medical center. While the Cox PH model is widely used in medicine, it is also used in other fields as well. Previously reliability engineering used homogeneous Poisson processes and renew process [2]. However, these only look at the variable of time-to-the-event,

ignoring all other variables. In the real world, hazard rate of the failure of an item is influenced by the conditions under which it operates, these conditions, or covariates, are what the Cox proportional hazard model takes into account. Proportional hazard models have been used to model component failure in nuclear plant, ship sonar, aircraft engines, train breaks, safety valves, aircraft doors, power cables, among other areas [2]. This paper also recounts situations that may require covariates that depend on time: wear on a device, growth of a crack [2]. If a time dependent covariate is introduced the model becomes a nonproportional hazards model and there are additional methods beyond the basic Cox PH model for handling this.

In summary, the first study stresses the importance of meeting the proportional hazard assumption in Cox regression analysis, particularly in the context of ACS patients. The importance of not violating the assumption and the need for appropriate modeling techniques to address violations for accurate and reliable prognosis needs to be of the utmost importance when using Cox regression analysis. The second article shed light on regional medical disparities in hypertension complications. The third article highlights the effectiveness of combining the Cox regression model with Artificial neural network (ANN). As the Cox regression model was outperformed by the ANN model when predicting postoperative mortality after hip fracture surgery. It brings to light the importance of equitable healthcare access and the introduction of policies aimed at reducing regional healthcare disparities for better management of patients with hypertension. Lastly the fourth paper reviewed here expands the usefulness of the Cox regression model beyond the medical field to the field of systems failure and reliability engineering.