**Computer Fraud and Security**
**ISSN (online): 1873-7056**

# Cox Distributions and Multiphase Distributions

**Jerzy Dorobisz**

Institute of Information Technology and Cyber-security, Faculty of Cybernetics, WAT, 2 Gen. Sylwestra Kaliskiego St., 00-908 Warsaw

email jerzy.dorobisz@wat.edu.pl

**Summary** The aim of this work is to present the relationships between the Cox proportional hazards model and phase-type distributions, particularly Cox distributions. Phase-type distributions form a flexible class of lifetime distributions that can be interpreted as distributions of time to absorption in a Markov chain with a finite number of phase states. This allows for accurate approximations of a wide range of empirical survival time distributions, even in situations where classical parametric distributions (e.g., exponential or Weibull) do not provide a satisfactory fit. The work discusses the basic concepts of survival analysis: survival function, distribution function, density function, and hazard function. Special attention is given to the Cox model, which is a standard tool for analyzing the impact of multiple prognostic factors on survival time. Its assumptions, interpretation of regression coefficients, and the concept of hazard ratios are presented. In the following sections, the class of phase-type distributions is characterized, and it is shown how they can be used to model hazard functions in medical applications (among others.oncology, hepatology) and cybernetics – such as the analysis of the reliability of complex technical systems and computer networks. The theoretical considerations are illustrated by examples drawn from the classic works of Cox and Christensen regarding the survival of patients with cancer and liver diseases. Additionally, a cybernetic example is presented, in which the Cox model and multi-phase distributions are applied to describe the time to failure in an information processing system.

**Keywords:** survival analysis, Cox model, hazard function, multi-phase distributions, Cox distributions, hazard ratio, system reliability.

Introduction

Survival analysis deals with modeling the time until the occurrence of a specific event. This event can include, among others, patient death, disease recurrence, equipment failure, as well as phenomena from the social sciences, such as divorce or graduation. Consequently, various names and interpretations can be found in the literature; however, from a mathematical perspective, random durations are always analyzed.

Cox distributions and multiphase distributions.

Classical survival analysis models are based on the assumption of a specific parametric distribution, such as exponential, Weibull, or log-normal. However, in many applications—especially in medicine—they prove to be too inflexible to accurately describe the observed data. At the same time, advanced regression models are used in practice, allowing for the simultaneous consideration of multiple prognostic factors.

One of the most important tools of this type is the Cox proportional hazards model. Its main advantage is that it does not require the assumption of a specific parametric distribution of survival time—it is a semi-parametric model. The hazard function in the Cox model takes the form.

λ(t,Z)=λ0(t)exp(β1Z1+β2Z2+⋯+βpZp)

where λ0(t) is the so-called baseline hazard, and the vector Z describes the values of the prognostic (explanatory) variables for a given object (e.g., a patient or a component of a technical system). The parameter vector β indicates the direction and strength of the influence of these variables on the hazard function.$\lambda_0(t) Z = (Z_1, \dots, Z_p) \beta$

The Cox model is widely used in oncology and cardiology, in studies of survival after surgical procedures, organ transplants, or heart attacks. In many medical publications, this model is used merely as a ready-made procedure available in statistical packages, without a broader discussion of its assumptions and limitations. One of the goals of this work is therefore to provide an intuition behind the Cox model and to organize the used symbolism.

In survival analysis, three related functions play a key role:— the survival function, describing the probability of surviving at least until time t,— the distribution function, describing the probability of an event occurring by time t,— the density function, which is the derivative of the distribution function. The hazard function is related to them.$S(t) t F(t) t$

$f(t)$

λ(t)=1−F(t)f(t)=S(t)f(t),

which describes the "instantaneous intensity" of the occurrence of an event at time t, given that it has not yet occurred by that time. For the exponential distribution, the hazard function is constant over time, while for other families of distributions, it may increase or decrease, which has significant interpretative implications in medical and technical applications.$t$

The classical Cox model does not impose a specific shape on the baseline hazard function. On one hand, this is an advantage (the lack of necessity to assume a rigid survival time distribution), but on the other hand, it complicates the intuitive interpretation of this function and the construction of parametric families of distributions that could accurately reflect empirical data. In this context, the role of multi-phase distributions arises, particularly Cox distributions, which can be treated as a rich family of lifetime distributions capable of approximating a very wide range of shapes of the hazard function.$\lambda_0(t)S(t)\lambda(t)$

Multi-phase distributions are defined as distributions of time to absorption in a finite Markov chain with one absorbing state and several transient (phase) states. The appropriate selection of the chain structure and transition intensities allows for various shapes of the hazard function—ranging from decreasing to increasing, and even to functions with one or several maxima. Cox distributions represent a special case of multi-phase distributions, where the phases are arranged serially and have exponential distributions.

This work aims to:

1. Reminder of the basic concepts of survival analysis and the interpretation of the hazard function.
2. Discussion of the assumptions and properties of the Cox proportional hazards model, with particular emphasis on its applications in medicine.
3. Presentation of the definitions and selected properties of multi-phase distributions, including Cox distributions, as tools for modeling lifetime.
4. Illustration of the above concepts with examples drawn from the literature (including Cox, Christensen), as well as a cybernetic example concerning the reliability of information processing systems.

The structure of the work is as follows. In the first chapter, basic definitions and relationships between the functions , , and are introduced. The second chapter discusses the Cox proportional hazards model and inference methods based on the likelihood function. The third chapter is dedicated to multi-phase distributions and Cox distributions. The final chapter presents examples of the applications of Cox models and multi-phase distributions in survival analysis and cybernetic applications.$F(t)S(t)f(t)\lambda(t)$

**Cybernetic applications of the Cox model and multi-phase distributions**

From the perspective of cybernetics and complex systems theory, it is important not only to model patient survival but also the times of correct operation of elements of technical systems: servers, network links, control systems, or nodes in critical infrastructure. In these applications, an "event" is most often a failure or a permanent transition of the system to a state of incorrect operation.

The hazard function can then be interpreted as the instantaneous failure intensity: the approximate probability that a failure will occur in a very small time interval around the moment, given that the system has been functioning correctly up to that point. The Cox model allows us to describe how explanatory factors, such as: – system load (e.g., the number of requests handled per second), – operating temperature, – type of algorithm used (e.g., memory management strategy), – level of network disturbances or the number of simultaneous connections, affect this intensity.$\lambda(t)t$

In the case of cybernetic systems, the multi-phase nature of the degradation process is often significant: the system sequentially passes through the states of "operational," "partially degraded," "heavily degraded," until it reaches the state of "failure." Describing such a sequence of transitions naturally leads to multi-phase distributions, where each phase corresponds to a state of a Markov chain, and the time to failure is the time to absorption.

**Practical example: time to failure of a server in a data center**

Let's consider a simple model of a server in a data center that handles user requests. We collect data on the uptime of individual servers from the moment they are started until a failure occurs that requires administrator intervention. For each server, we also record explanatory variables such as: – average CPU load (in percentage), – average operating temperature, – number of simultaneously handled services (e.g., databases, application servers).

We can assume that the time to failure follows a multi-phase distribution: – phase 1: new server, failures are unlikely (the so-called "infant mortality" is negligible), – phase 2: period of stable operation, low probability of failure, – phase 3: wear phase, where the hazard of failure increases.$T$

This process can be described using a Markov chain with states corresponding to the above phases and a absorbing state "failure." The total time spent in the transitional states will have a multi-phase distribution.

Additionally, we can apply the Cox model to investigate the impact of load and temperature on the hazard function:

$$\lambda(t, Z) = \lambda_0(t)\ \exp\ (\beta_1 \text{CPU} + \beta_2 \text{TEMP} + \beta_3 \text{USŁUGI}),$$

where: – average CPU load, – operating temperature, – number of running services.CPUTEMPUSŁUGI

The coefficients can be estimated using the maximum likelihood method. The resulting hazard ratios have a direct cybernetic interpretation, for example: – indicates that an increase in CPU load by one unit (e.g., 10 percentage points) increases the failure intensity by about 10%, holding other factors constant, – indicates that with a temperature increase of 1°C, the risk of failure increases by 20%.$\beta$exp $(\beta_j)$exp $(\beta_1) = 1{,}10$exp $(\beta_2) = 1{,}20$

In this way, the Cox model and multi-phase distributions allow: – to forecast the reliability of servers depending on operating conditions, – to design control policies (e.g., load balancing, cooling) that minimize the hazard of failure, – to simulate the behavior of the entire network of servers as a cybernetic system, where individual elements enter and exit degradation phases.

This approach combines classical survival analysis with a cybernetic perspective on reliability and control, showing that tools developed in medical statistics can also be effectively used in modeling complex technical and informational systems.

The Cox proportional hazards model is one of the most commonly used models in oncology and, more broadly, in survival analysis. However, in clinical practice, it is often treated as a kind of "black box." Statistical packages offer extensive modules for survival analysis, including implementations of multivariate regression models, but the documentation of these tools often focuses on describing how to run procedures and interpret basic results, with less emphasis on presenting the model assumptions and deriving the formulas used. Consequently, a user may correctly utilize ready-made procedures without having full insight into their theoretical foundations. A formal, theoretical description of the proportional hazards model can be found, among others, in Cox's classic work [1].

This paper aims to discuss the above-mentioned model, including explaining the basic concepts related to it, namely: hazard functions, cumulative distribution function, and survival function.

The Cox proportional hazards model belongs to the class of multivariate regression models used in survival analysis. In its basic form, it assumes the possibility of simultaneously considering multiple explanatory (predictive) variables that may influence the duration of observation. It does not assume that the outcome (survival time) is explained by a single variable, while the influence of other factors is negligible. On the contrary, the Cox model was designed specifically to examine the combined and independent effects of multiple prognostic factors on the risk of an event. Oncological applications are a typical example of situations where it is necessary to simultaneously consider multiple variables (e.g., age, stage of disease, type of treatment, biochemical parameters). In such cases, univariate models are insufficient, which is why complex regression models, including the Cox proportional hazards model, are widely used, allowing for the assessment of the impact of each factor while controlling for the others. [2]

The developed model based on the survival table allows for the analysis of the impact of prognostic factors on survival. The author presents a model using a table based on a single case, which can also be applied to other cases of oncological or cardiological diseases after heart transplantation or myocardial infarctions.

Referring to multivariate analysis, we usually obtain a set of factors with a specific measure of impact on the outcomes for each factor separately, assuming the constancy of the others.

The Cox model is very often used with commercial statistical packages, but usually without a detailed discussion of its principles. The documentation of these packages mainly emphasizes how to run procedures and interpret basic results, with less focus on presenting the theoretical foundations of the model. In oncological research, the Cox proportional hazards model is one of the most commonly used tools for survival analysis, alongside other regression models such as logistic models or log-logit models. It is used in both univariate and multivariate analyses, allowing for the assessment of the impact of multiple prognostic variables on the risk of an event. [3].

Survival analysis usually involves various functions in relation to statistics. Its object of interest is usually the time that has elapsed since a certain event occurred. Such events typically include the death of a patient, equipment failures, committing a crime, divorce, or graduating from school. Considering the diversity of such events, various terminologies are usually encountered.

In contrast to earlier, simplified approaches to survival analysis, the Cox proportional hazards model is not limited to a single predictive variable. From the beginning, it was designed as a multivariate model, allowing for the simultaneous consideration of multiple prognostic variables, even if their individual impact is small. In clinical practice—especially in oncology—it is rare for survival to depend on a single factor. Therefore, it is the ability of the Cox model to analyze the combined and independent effects of multiple variables that makes it widely used in biomedical research. The model does not assume that the sum of the influences of the factors equals zero; on the contrary, it allows for the estimation of relative risk (hazard ratio) for each variable while controlling for the others. [4].

Cox developed this type of model for the survival table (see also Breslow's work), which enables the analysis of the impact of prognostic factors on survival. He illustrated the application of such models for the case of leukemia, but it is also intended for the analysis of survival in many other diseases. It is related to oncological or cardiological diseases after transplantation or myocardial infarctions [5].

Considering the multivariate analysis of the model, we usually obtain a set of specific factors with a specific measure of impact that translates into the outcome of each factor separately when combining separate factors into a whole. This type of model is very often used in commercial statistical projects, which generally do not provide explanations of fixed projects and their data analysis, but merely explain its functioning [6].

In promoting its use in oncology, there are many other multivariate models that complement the data analysis of the discussed algorithm.

Below, the functions of the multivariate Cox model will be discussed. Among the most significant of these, we can include:

- Hazard function;
- Cumulative distribution function;
- Survival function [7].

The first of the above-mentioned functions, also known as the hazard function (risk function), is often interpreted as the intensity of instantaneous mortality or the instantaneous failure rate. In renewal theory, it most commonly describes the failure rate of technical elements. The risk function is expressed in the language of probability calculus as the ratio of the probability density function to the survival function, where a denotes the distribution function:$\lambda(t) f(t) S(t) S(t) = 1 - F(t) F(t)$

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)}.$$

[8].

$$\Lambda(t) = f(t)/[1 - F(t)] \tag{1}$$

The following graph shows the course of the risk function estimated in the multidimensional Cox model. The shape of the graph – approximately linear over time – is significant in assessing whether the assumption of proportional hazards is met in the analyzed model.$\lambda(t)$

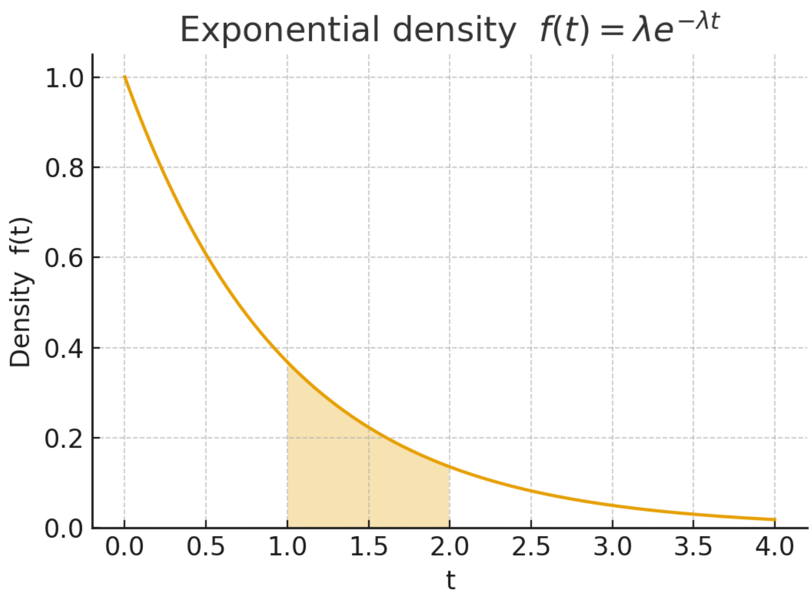


Figure 1. Since the risk function λ(t) is the ratio of the probability density function f(t) to the survival function (1 – F(t)), for the exponential distribution, λ(t) = λ. The probability that t falls within the interval (1/λ, 2/λ) is equal to the area under the curve within the marked boundaries [5]

In survival analysis, the distribution function is usually denoted by the symbol. For the random variable of time, it is defined as$F(t)T$

$$F(t) = P(T \leq t).$$

The function is non-decreasing over time and – in general – approaches the value of 1 at the boundary, when (the distribution function is defined over the domain, and not just up to a certain „maximum" value).$F(t)tt \to +\infty(-\infty, +\infty)$

The survival function is defined as

$$S(t) = P(T > t) = 1 - F(t),$$

It describes the probability that the time to the event will exceed the value.$t$

The probability density function is such that its integral over an interval gives the probability of the variable taking values from that interval:$f(t)[t_1, t_2]T$

$$P(t_1 \leq T \leq t_2) = \int_{t_1}^{t_2} f(t)\, dt.$$

The schematic diagram below illustrates example shapes of the function, and for the case when it is the density of the exponential distribution. [9]:$f(t)F(t)S(t) = 1 - F(t)f(t)$

$$f(t) = \lambda \exp[-\lambda t] \quad (2)$$

In the case of the exponential distribution, the hazard function is constant over time, meaning it does not depend on the value. In medical interpretation, this would mean a constant risk of death independent of the duration of observation. This assumption is practically unrealistic for many cancers, for which the risk of death changes over time (e.g., it may be higher in the first years after diagnosis and then decrease, or vice versa) [10]. Hastings and Peacock [11] provided examples of density functions, survival functions, and hazard functions for many families of distributions used in survival analysis. Graphical methods for studying the shape of the hazard function have been discussed, among others, by Nelson [12]. In turn, the distributions of survival functions and methods for estimating their parameters have been detailed by Gross and Clark, as well as Kalbfleisch and Prentice.$\lambda(t)t$

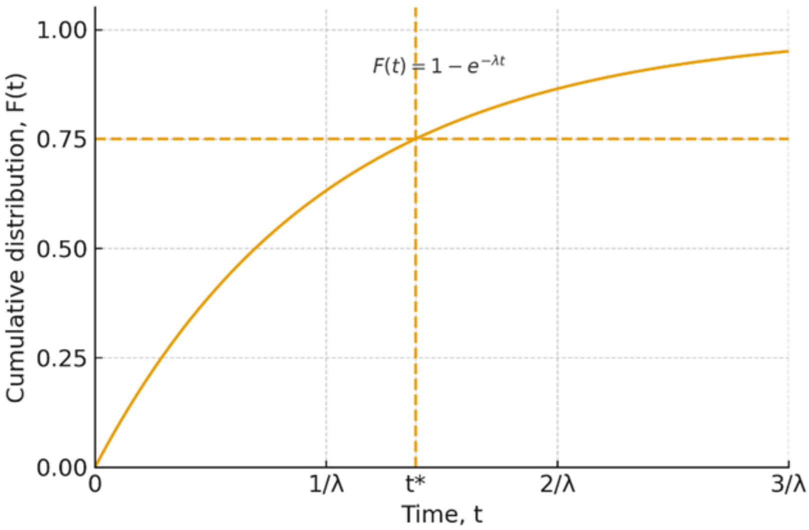


Figure 2. The probability density function f(t) is the first derivative of the probability distribution function F(t). f(t) = d/dt[F(t)] => d/dt{1 – exp[–λt]} = λexp[–λt] As shown in the graph, the probability that t is in the interval (0, t*) is equal to 0.75 [16]

## 1. Assumptions of the Cox proportional hazards model

The above model assumes that the relative risk of death in the two studied groups is generally constant over time, referring to the assumptions of proportionality of risk, and that the relative risk is generally constant concerning fixed parameters, except in cases where appropriate interactions are introduced into the model. This last event is common to all regression models, not just the model described above. Additionally, this model assumes that the effect of independent variables on the hazard function is log-linear. Cox proposed the proportional hazards model in the form:

where:

- $\lambda(t, Z)$denotes the hazard function at time for the vector of independent variables,$tZ = (Z_1, \dots, Z_p)$
- $\lambda_0(t)$is the baseline hazard function (baseline hazard),
- $\beta = (\beta_1, \dots, \beta_p)$β is the vector of regression parameters,
- the product denotes the sum.$Z\beta \sum_{j=1}^{p} Z_j\beta_j$

$$\lambda(t, Z) = \lambda0(t)exp[\beta1Z1 + \beta2Z2 + .. + \beta pZp]: \quad (4)$$

where β is the vector of unknown constant parameters (weights), and λ0(t) is an unknown risk function for Z equal to zero (the so-called baseline risk) [13].

The formulated risk model is not expressed in terms of a distribution such as the Weibull distribution, but is defined by the distribution of the analyzed data. For a given Z, exp[β1Z1 + β2Z2 + .. + βpZp] is a constant value, and thus the risk function λ(t, Z) is a baseline function λ0(t), multiplied by a constant. Consequently, Cox introduced the definition: proportional hazards model. However, it should be noted that a limitation of the Cox model is the necessity of having Z values for each patient. There are methods for estimating the values of missing data, but they require additional assumptions that are generally difficult to justify [14].

Based on the algorithm described above, a case study of breast cancer is presented.

Based on the literature, it is worth noting that a typical example of multivariate results using the aforementioned model concerns locoregional breast cancer documented in Cox's 1972 article [15]. Another reference seems to be information about the statistical package used. Scientific oncology publications rarely, if ever, include a description of the methodology of the Cox model, instead directing interested readers to rather inaccessible statistical publications.

The literature discusses many diverse examples of Cox's applications, but analyses of prognostic factors for locoregional recurrences in breast cancer are of the greatest interest. Identifying a subgroup of patients with a favorable prognosis is very important due to the high incidence, as well as the poor prognosis of patients after locoregional recurrence following mastectomy [16].

The publication presents, among other things, algorithms indicating that some prognostic factors that were significant in univariate analysis are not significant in multivariate analysis, likely being a substitute factor for another variable, meaning strongly correlated with it [17].

Table 1. Selected results from the publication by Willner et al., presenting univariate and multivariate analyses of prognostic factors affecting survival after recurrence. NI = not significant, pT – degree of primary advancement T, PD = primary diagnosis, DN = diagnosis of recurrence, * indicates exclusion from multivariate analysis due to a large number of missing data.

Table X. Prognostic factors – univariate and multivariate analysis

| Prognostic factor | Univariate analysis | Multivariate analysis |
|---|---|---|
| pT (T1–2 vs T3–4) | P < 0.001 | P < 0.01 |
| Differentiation grade G (G1–2 vs G3–4) | P < 0.01 | * |
| Vascular invasion | P < 0.01 | NS |
| Tumor necrosis | P < 0.001 | P < 0.01 |
| Age PD (< / > 50 years) | NS | NS |
| Axillary lymph node status PD | P < 0.001 | P < 0.05 |
| Post-mastectomy chemotherapy | P < 0.01 | NS |
| Site of recurrence | P < 0.01 | P < 0.001 |
| – Chest wall | | |
| – Axillary lymph nodes | | |
| – Supraclavicular region | | |
| – Multiple | | |

<table>
<tr><td>| Time to recurrence<br>(< / > 12 months)</td><td>| P < 0.001</td><td>| P < 0.01</td></tr>
<tr><td>| Age DN (< / > 50 years)</td><td>| NS</td><td>| P < 0.05</td></tr>
<tr><td>| Follow-up time for site of recurrence</td><td>| P < 0.001</td><td>| P < 0.05</td></tr>
</table>

Source: [18]

The above analysis indicates that due to the significant number of missing data in the multivariate model, the interpretation of the results should begin with univariate analysis, especially at the stage of disease stage recognition. The differences between the results of univariate and multivariate analyses provide information about the relationships between prognostic variables and which of them retain independent significance after simultaneously considering the others. The relationship between the probability of survival and the risk of death described by the hazard function is crucial here. The dependence between the survival function and the risk function in the Cox model is illustrated in the figure below.$S(t)\lambda(t)$

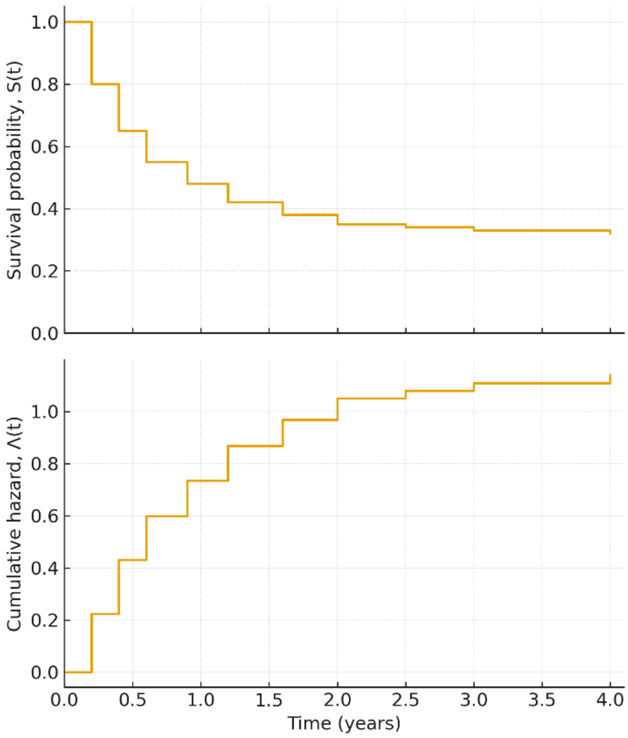


Figure 3. If we denote the probability of survival as a function of time by , and the cumulative risk function by , then the relationship between these functions can be expressed as$tS(t)\Lambda(t)$

$$\Lambda(t) = -\log S(t) \text{ oraz równoważnie } S(t) = \exp\left(-\Lambda(t)\right).$$

Figure 3 shows an example of the survival function and its corresponding cumulative risk function. One of the problems in interpreting the Cox model is the large number of symbols used in the literature, which are not always consistently recorded across different publications. [21]$S(t) = 1 - F(t)\Lambda(t)$[1]

For illustration purposes, the results of calculations based on Kaplan-Meier survival tables for 30 patients from the group described by Christensen et al. are presented. The figure shows the estimated survival function (the Kaplan-Meier curve) and its corresponding cumulative risk function. The graph illustrates how the shape of the survival curve reflects the increasing risk of an event over time.$XS(t)\Lambda(t)$

**2. Likelihood ratio test**

The Cox model for Christensen's data, which includes three variables: ALB (albumin), log10 LGB (bilirubin), and ALC (alcoholism), takes the form:

[1] Rysunki wszystkich struktur krystalicznych zostały wykonane w programie Vesta 3.

$$\lambda(t, Z) = \lambda 0(t)\exp[\beta 1Z1 + \beta 2Z2 + \beta pZ3] \quad (4)$$

where: Z1 = ALB, Z2 = LGB, and Z3 = ALC

In the model fitting process, a total of seven possibilities can be considered (Table 1). Three models include one variable each, three include two variables each, and one includes all three variables, resulting in a total of seven possible models.

Parameter estimation and model significance testing are based on the concept of the likelihood function. For a general statistical model, where we observe independent times and event indicators (1 – event occurred, 0 – observation censored), the likelihood function for the parameter vector can be expressed as$T_i\delta_i\theta$

$$L(\theta) = \prod_{i=1}^{n} \quad [f(t_i;\theta)]^{\delta_i}[S(t_i;\theta)]^{1-\delta_i},$$

where denotes the density of time to event, and a – the survival function. Parameters are estimated such that the function (or its logarithm) takes on a maximum value – this is the maximum likelihood method.$f(t;\theta)S(t;\theta)\theta L(\theta)$

In the Cox proportional hazards model, the parameters of interest are the regression coefficients. In practice, the so-called partial likelihood function is used, constructed based on the ordered event times, and then estimators that maximize it are determined. The overall significance of the model can be assessed using a statistic based on the likelihood ratio$\boldsymbol{\beta}L(\boldsymbol{\beta})\widehat{\boldsymbol{\beta}}L(\boldsymbol{\beta})$

$$\Lambda = -2\log\frac{L(\mathbf{0})}{L(\widehat{\boldsymbol{\beta}})},$$

where is the value of the likelihood function for the model without the covariate effect (e.g., ), and a – for the model with the estimated coefficients. A large value indicates a significant impact of at least one explanatory variable on survival time.$L(\mathbf{0})\beta_1 = \beta_2 = \beta_3 = 0L(\widehat{\boldsymbol{\beta}})\Lambda$

For larger values of L(b) or smaller values of the credibility ratio L(0)/L(b), the model explains the observed data better [19]. The significance of each model can be statistically tested using the $\chi^2$ distribution statistics provided below, where the degrees of freedom equal the number of factors in the model [20].

The results of Christensen for seven Cox models, which represent a typical spectrum of data obtained in computer printouts for the Cox model, although not always presented in this arrangement, are shown in Table 1.

$$\chi 2 = -2 \text{ x } \log e[L(0)/L(b)] = -2 \text{ x } [\log eL(0) - \log eL(b)] = 2 \text{ x } [\log eL(b) - \log eL(0)] \quad (5)$$

## 3. Standard normal deviation Z, Wald test, and relative significance of variables

In Table 1, Z represents the standard normal deviation, which is equal to b/SE(b). The significance of the regression coefficients b can be estimated by comparing the values of $Z^2$ with the $\chi^2$ distribution for DF = 1, which is sometimes referred to as the Wald test [21]. If $Z > 1.96$, then b is statistically different from zero at the significance level $\alpha = 0.05$. However, if $Z < 1.96$ for any coefficient b, it does not necessarily mean that the given factor has no effect on the prediction. It may also mean that its effect is too small to be detected in a study with a given number of patients. The relative significance of the variables is determined by the numerical value of Z. The larger the value of Z, the greater the significance of the variable in the model. As shown in Table I for model 7, the significance of the variables decreases in the following order: ALB, LGB, ALC, where ALC is statistically insignificant.

## 4. Relative risk: hazard ratio

Using the values of the regression coefficients b, one can estimate the relative risk, e.g., the incidence in the exposed population compared to the incidence in the unexposed population.

Relative risk is a useful statistic, for example, when analyzing the risk of leukemia or solid tumor formation in a population that survived the atomic bombings in Hiroshima and Nagasaki, depending on the age of exposure and the dose received.

A relative risk value of 1.1 indicates a 10% increase in risk compared to the risk for the reference group. In Cox analysis, relative risks are hazard ratios assigned to different levels of a given factor, while all other factors remain unchanged.

When the analysis is not too complicated, as in the case of Christensen's data with three factors, both selection methods – forward and backward – lead to one final model. However, in the case of a more complex analysis with many variables, these methods may lead to different models [22].

## Literature


1. Cox D.R., Regression models and life tables, Journal of the Royal Statistical Society, 34B, 1972, s. 187-220.
2. Breslow N.E., Analysis of survival data under the proportional hazards model, International Statistical Review, 443, 1975, s. 45-58.
3. Langlands A.O., Pocock S.J., Kerr G.R. et al., Long-term survival of patients with breast cancer: a study of the curability of the disease, British Medical Journal, 2, 1979, s. 1247-51.
4. *Gore S.M., Pocock S.J., The statistical modelling of survival in breast cancer. Lecture at the Institute of Statisticians Conference, Statistics in Medicine, Cambridge 1981, s. 70-75*
5. Breslow N.E., Day N.E. (eds.), Statistical Methods in Cancer Research. 1. The Analysis of Case-Control Studies, IARC Scientific Publications, Lyon: International Agency for Research on Cancer, 1980, s. 20-22.
6. Breslow N.E., Day N.E. (eds.), Statistical Methods in Cancer Research. 1. The Analysis of Case-Control Studies, IARC Scientific Publications, Lyon: International Agency for Research on Cancer, 1980, s. 33-35.
7. Gore S.M., Pocock S.J., The statistical modelling of survival in breast cancer. Lecture at the Institute of Statisticians Conference, Statistics in Medicine, Cambridge 1981, s. 66-69
8. Gore S.M., Pocock S.J., The statistical modelling of survival in breast cancer. Lecture at the Institute of Statisticians Conference, Statistics in Medicine, Cambridge 1981, s. 70-77.
9. Grochala W., Superconductivity: small steps towards the “grand unification”, Journal of Molecular Modeling, 11, 2005, s. 323-329.
10. *Christensen E., Multivariate survival analysis using Cox’s regression model, Hepatology, 7, 1987, s. 1346-58.*
11. Norwegian Multicentre Study Group, Timolol-induced reduction in mortality and reinfarction, The New England Journal of Medicine, 304, 1981, pp. 801-7.
12. *Nelson W., Theory and application of hazard plotting for censored failure data, Technometrics, 14, 1972, pp. 945-66.*
13. Romiszewski J., Grochala W., Stolarczyk L., Pressure-induced transformations of AgIIF2—towards an ‘infinite layer’ d9 material, Journal of Physics: Condensed Matter, 19, 2007, pp. 116206–116218.
14. Lucier G., Shen C., Casteel W.J. Jr., Chacon L., Bartlett N., Journal of Fluorine Chemistry, 72, 1995, pp. 157–163. Grochala W., Egdell R.G., Edwards P.P., Mazej Z., Žemva B., On the Covalency of Silver–Fluorine Bonds in Compounds of Silver(I), Silver(II) and Silver(III), ChemPhysChem., 4, 2003, pp. 997-1001.
15. Malinowski P., Mazej Z., Grochala W., Probing Reactivity of the Potent AgF2 Oxidizer. Part 2: Inorganic Compounds, Zeitschrift für anorganische und allgemeine Chemie, 634, 2008, pp. 2608-2616.
16. Gao L., Xue Y.Y., Chen F., Xiong Q., Meng R.L., Ramirez D., Chu C.W., Eggert J.H., Mao H.K., Superconductivity up to 164 K in HgBa2Cam−1CumO2m+2+δ (m=1, 2, and 3) under quasihydrostatic pressures, Physical Review B., 50, 1994, pp. 4260-4263.
17. Fisher P., Roult G., Schwarzenbach D., Title of the text, Journal of Physics and Chemistry of Solids, 32, 1971, pp. 1641-1647.
18. Kalbfleisch J.D., Prentice R.L., The Statistical Analysis of Failure Time Data, John Wiley, New York 1980.
19. Grochala W., Porch A., Edwards P.P., Meissner–Ochsenfeld superconducting anomalies in the Be–Ag–F system, Solid State Communications, 130, 2004, pp. 137-142.
20. Romiszewski J., Grochala W., Stolarczyk L., Pressure-induced transformations of AgIIF2—towards an ‘infinite layer’ d9 material, Journal of Physics: Condensed Matter, 19, 2007, pp. 116206–116218.
21. Momma K., Izumi F., VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data, Journal of Applied Crystallography, 44, 2011, pp. 1272-1276.
22. *Nelson W., Theory and application of hazard plotting for censored failure data, Technometrics, 14, 1972, pp. 945-66.*