Clinical Factors to Investigate Survival Analysis in Cardiovascular Patients

Najada Firza[1,3(✉)], Rossana Mancarella[2], Francesco Domenico d'Ovidio[1], and Dante Mazzitelli[1]

[1] Department of Economics and Finance, University of Bari "Aldo Moro", Largo Abbazia Santa Scolastica, Bari, Italy
najada.firza@uniba.it
[2] Agenzia Regionale per la Tecnologia e l'Innovazione ARTI Puglia, Bari, Italy
[3] Catholic University Our Lady of Good Counsel, Tirana, Albania

Abstract. Cardiovascular diseases are one of the leading causes of mortality worldwide; however, the management of patients with this pathology within hospital facilities has not been uniquely realized.

Through appropriate statistical methods (i.e., survival analysis, Cox regression, and logistic regression), we have tried to illuminate the survival and length of stay of patients with cardiovascular diseases in hospitals. The results obtained classify the clinical factors into risk variables (which reduce survival) and protective variables (which prolong survival) and are linked to the long stay of this particular category of patients.

The aim of this work is to analyze the variables responsible for long stays in patients with cardiovascular diseases to monitor the health of patients and better manage the financial resources of health facilities in relation to this problem.

Keywords: Logistic Regression · Cox Regression · Long Hospital Stay (LOS) · Cardiovascular Diseases (CVD)

1 Introduction

Cardiovascular diseases (CVD) are the leading cause of death in Italy and Europe [1]. The Global Burden of Disease (GBD) study estimates that there were 17.8 million deaths due to CVD [2] worldwide in 2017, which is expected to reach 23.6 million by 2030 [3]. Consequently, CVD is associated with a consistently high number of hospital admissions, a long hospital stay (LOS), and increased mortality rates [4]. All this means that there is an increase in healthcare costs for hospitals [5]. Accordingly, the Italian National Health System dedicates many resources to the prevention, diagnosis, and treatment of these diseases.

Obviously, many variables lead a patient to an LOS, and numerous studies state that an LOS depends on risk factors, socio-demographic variables, the clinical situation at the time of admission, the treatment in hospitals, the development of complications, and so on.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
O. Gervasi et al. (Eds.): ICCSA 2024, LNCS 14813, pp. 29–45, 2024.
https://doi.org/10.1007/978-3-031-64605-8_3

Many studies have investigated the causes associated with a prolonged LOS. These factors have a broad spectrum concerning physiological and clinical aspects and lifestyle: female sex, New York Heart Association (NYHA) functional class [6], low left ventricular ejection fraction (LVEF), concomitant community-acquired infection, arrhythmias, cerebrovascular diseases, dementia, hyponatremia, polytherapy, pressure ulcers, chronic alcohol consumption, peripheral edema, the development of renal failure, the presence of social problems that require special intervention, and concomitant acute medical problems that require specific treatments, just to name a few [7–9].

Regardless of risk factors and the presence of comorbidities, generally, a longer hospitalization is also linked to suboptimal clinical outcomes.

The present research aims to quantify the main factors influencing the length of stay in hospitals, not only to monitor the health risk of these patients, but also to effectively use the resources of facilities. Accordingly, this work considered two representative samples of patients with at least one CVD admitted to hospitals in the Puglia Region, in the last period with available data (2009–2014). We used survival methods (i.e., Cox regression and logistic regression) to explore the length of patients' stays, highlighting factors related to mortality and an LOS. Using more recent data, results could change in some way, but the methodological innovation will remain.

In fact, the ability to identify hospitalized CDA patients at risk of prolonged hospital stays could be assessed by healthcare professionals. A risk stratification for an LOS can help provide good opportunities for patient care by identifying patients who need special attention and certain interventions, such as education and specific therapies [10].

Reducing the LOS of hospitalized patients has been considered a primary strategy for deploying resources efficiently [10]. This study can help improve the treatment of CVD patients while enhancing their clinical outcomes and reducing the costs associated with healthcare.

2 Materials and Methods

2.1 Study Design and Data Collection

The data analysed was collected by the Apulian Regional Health Agency (AreS-Puglia) at the administrative level. The agency continuously monitors data quality. IBM SPSS 22 software was used for the data analyses.

The original database, which contained over 4 million records registered over a period of 6 years, was obtained from the complete registration of hospital discharge cards (HDC) for all patients admitted to Apulian hospitals between 2009 and 2014. The data were confidential and under the responsibility of AreS-Puglia, and were granted to us solely for statistical analysis in aggregate form.

The S*emi-parametric Cox Regression model* and the *Logistic Regression model* are the methods used in this work. These methodologies were chosen by the authors over other methodologies because they are known in the literature to resolve issues regarding survival analysis. (see Sect. 2.3 and 2.4).

For our research, we investigated a sub-population in the original database consisting of patients who were admitted to hospitals in Puglia during the specified time

period and had at least one diagnosis related to cardiovascular diseases (Category IX of the International Classification of Diseases, 9th Revision, codes I00 to I99). The subpopulation in question consists of over 1,300,000 HDC, with almost 1,150,000 being ordinary hospitalizations.

According to Table 1. Distribution of the outcomes of hospitalizations for cardiovascular diagnoses in Puglia, 2009–2014.), over 96% of patients hospitalized for cardiovascular diagnoses in Puglia survived from 2009–2014, despite a slightly increasing mortality trend.

Table 1. Distribution of the outcomes of hospitalizations for cardiovascular diagnoses in Puglia, 2009–2014.

	Observed Data			%		
Year	Surviving	Dead	Total	Surviving	Dead	Total
2009	229161	6496	235657	97,2	2,8	100,0
2010	242011	6563	248574	97,4	2,6	100,0
2011	223147	6810	229957	97,0	3,0	100,0
2012	198833	7302	206135	96,5	3,5	100,0
2013	191884	7242	199126	96,4	3,6	100,0
2014	183684	7583	191267	96,0	4,0	100,0
Total	1268720	41996	1310716	96,8	3,2	100,0

2.2 Differential Sampling

Two representative samples were randomly extracted from the entire population to demonstrate the robustness of the results and provide a concrete example for comparison. Of course, robustness should be confirmed by both clinical relevance and statistical significance of the analyses. Additionally, the age factor divided by class was used in the second sample to investigate possible relationships with LOS and confirm the other factors in the Cox model. We analyzed Sample I, which represents approximately 0.10% of the population (with over 11,000 hospitalizations), and Sample II, which represents approximately 0.6% of the population (with over 6,000 hospitalizations).

Table 2. Distribution of the outcomes of hospitalizations for cardiovascular diagnoses in Puglia in the two samples, 2009–2014., highlights two randomly extracted samples from the dataset of 1,310,716 admissions recorded in public and private hospitals in Puglia from 2009 to 2014. The admissions had at least one diagnosis related to cardiocirculatory diseases.

It is important to note that the dynamics of deaths in Table 1. Distribution of the outcomes of hospitalizations for cardiovascular diagnoses in Puglia, 2009–2014., should not cause alarm. It is necessary to highlight the increasingly frequent recourse, first to a day hospital and then to outpatient services, for the treatment of milder pathological

forms (thus avoiding hospitalization and, consequently, detection through HDC). Below, we must also mention the aging Apulian population, which is more exposed to potentially fatal cardiovascular events.

The relevance of the present research lies in the proportion of hospital deaths in the observed population, which should not be excessive, and the maintenance of this condition in the extracted samples. This ensures good representativeness while maintaining temporal dynamics Table 2. Distribution of the outcomes of hospitalizations for cardiovascular diagnoses in Puglia in the two samples, 2009–2014., confirms these conditions, despite random variations. Therefore, the two determined samples can be used as a reference for analyzing hospitalizations.

Table 2. Distribution of the outcomes of hospitalizations for cardiovascular diagnoses in Puglia in the two samples, 2009–2014.

	Sample I data						Sample II data				
Year	Surviving %		Dead %		Total	Year	Surviving %		Dead %		Total
2009	1969	96,8	66	3,2	2035	2009	1123	96,6	40	3,4	1163
2010	2040	96,3	78	3,7	2118	2010	1204	97,6	29	2,4	1233
2011	1944	96,9	63	3,1	2007	2011	1112	96,1	45	3,9	1157
2012	1741	96,1	70	3,9	1811	2012	983	95,1	51	4,9	1034
2013	1658	95,9	71	4,1	1729	2013	865	96	36	4	901
2014	1569	95,6	73	4,4	1642	2014	863	95,1	44	4,9	907
Total	10921	96,3	421	3,7	11342	Total	6150	96,2	245	3,8	6395

The aim of this study is to investigate whether specific hospitalisation variables affect the length of stay. This is logical when considering the diseases and required interventions. However, other hospitalisation-related variables may also significantly influence hospitalisation duration. The study analyses the main characteristics of hospital stays, with a focus on the impact of patient mortality.

The analysis does not include day hospital admissions or day surgery, nor does it cover outpatient or emergency room care that does not require admission.

Patient data includes personal information such as gender, age, and residence, as well as the length of hospital stay. Additionally, recorded are characteristic elements of hospitalization, such as the health facility where it occurred, the admission and discharge ward, and any transfers. Clinical variables, such as diagnosis, diagnostic procedures, interventions, and admission outcomes, are documented.

2.3 Semi-parametric Cox Regression Model

The survey methodology is a duration analysis performed through some of the best-known techniques concerning survival analysis. In particular, we used Cox's semiparametric model to identify variables that influence the duration of hospitalization, reducing

any confounding effects. The variable response is the duration of hospitalization, considering deaths that logically determine the interruption of hospitalization for reasons opposite to those desirable (i.e., the discharge of patients for at least a partial recovery).

In the biomedical field, the Cox model is one of the best-known and most important models for survival analysis under the hypothesis of proportional risks over time. It is called "semiparametric" because no assumptions are made about the form of the survival function and, therefore, about the risk function. In this model, however, the influence of explanatory variables (or prognostic factors) on risk, itself, is parameterized. Its theoretical foundations start from the concept of the risk function, whose general estimate (with events of interruption, default, or death, generically indicated with di and not necessarily equispaced in time) is given in a group of n statistical units with r distinct times of interruption according to the following:

$$\widehat{H}_k(t) \cong \sum_{i=i}^{k} \frac{d_i}{n_i} = \sum_{i=1}^{k} \hat{q}_i, \textit{ with } t_k \leq t \leq t_{k+1} \textit{ and } k = 1, 2, \ldots, r,$$

where r < n and n_i equal the number of units that survived until event d_i [11].

The Cox model [12] was created to verify, through a sample, the influence that some characteristics, both quantitative and qualitative (i.e., explanatory variables or "prognostic factors"), possessed by the sampled units can have on survival times.

It defines $h_i(t)$, the hazard function at time t (for $t \geq 0$), for any unit with certain values of prognostic factors and $h_0(t)$ the risk function (baseline function) at time t for any unit having prognostic factors with a zero value. This condition implies that quantitative factors are rescaled so their minimum value is zero and that qualitative factors are defined with zero indicator variables corresponding to basic modalities often assumed to have the least risk [13]. That said, assuming that at any moment the risk function $h_i(t)$ is proportional to the function $h_0(t)$ at a rate of the parameter ψ, the Cox model is expressed by the following formulation:

$$h(t) = \psi h_0(t) = \exp(\beta)\ h_0(t), \ \ \beta \in IR.$$

Notably, $\psi = h(t)/h_0(t)$ represents the default risk at time t of any statistical unit with respect to another unit belonging to the base group, for which it is called the *relative risk*, or *risk ratio*. Of course, if $\psi < 1$ (and, therefore, $\beta = \ln\psi < 0$) the default risk for said unit is smaller than for the unit belonging to the base group. The opposite is true if $\psi > 1$ and $\beta > 0$ [14].

Now, we explicitly consider a variable X relative to a prognostic factor, with determinations x_i taking the value 0 if the unit belongs to the base group. For the generic individual of the sample, the previous formulation becomes as follows:

$$h_i(t) = \psi(x_i)h_0(t) = \exp(\beta x_i)h_0(t), \beta \in \text{IR}.$$

Logically, if the survival of the sampled units is affected by several factors, this model is generalized as follows:

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi})h_0(t), \beta_j \in \text{IR}.$$

Cox's model assumes that:

– The explanatory variables have a multiplier effect on the risk. That is, the link function is exponential. This assumption is generally plausible when the risk ratio is ψ positive;
– The relationship between the risk functions of individuals with different values of prognostic factors is constant between the layers identified by the factors;
– The invariance over time of the values of explanatory variables: If this does not happen, it means that it is no longer legitimate to record the values of prognostic factors (or a set of them) only at the beginning of the study. However, there is also a model for non-time-independent covariates.

Consequently, it is always advisable to check in advance at least for the absence of the stratification effect. That is, survival data should be classified according to the levels of the factors, with all the subgroups (strata) defined based on all the J combinations of the levels of factors so individuals in the same stratum are characterized by the values of the same explanatory variables. In the case of quantitative covariates, a preliminary reduction into classes must be performed.

In each subgroup, the survival function is estimated with the Kaplan–Meier method [11]. The points—t, $ln\ \widehat{H}_j(t)$—are reported in a Cartesian reference, where.

j = 1, 2…, J identifies the subgroup where the analysis takes place.

If the diagrams (one for each value of *j*) obtained by joining the points t, $ln\ \widehat{H}_j(t)$ appear parallel, then the risks of the various groups can be considered perfectly proportional.

Cox's model, however, is robust enough to provide reliable results even when the parallelism of the lines is not perfect. At most, it may be sufficient to verify that they do not intersect, except possibly at the extremes [15].

In our work, we used the Kaplan–Meier method [16] to identify variables that are significant for Cox regression. Subsequently, with the forward stepwise method that was validated through a backward stepwise elimination procedure, the statistically significant variables were validated for Cox regression ($p < 0.05$).

2.4 Logistic Regression Model

In order to compare and confirm the results obtained with the Cox model, a logistic regression analysis was also used in this work.

Generally, in epidemiological studies, it is assumed that the baseline is characterized by healthy subjects.

Accordingly, the "of interest" event is death. The variable Y of the model is defined as a function of a set of predictive, categorical, and/or quantitative variables:

$Y = f(X_1, X_2, \ldots, X_k)$.

In the case of a Bernoullian answer with the average $E(Y) = p$, where p is the probability of success, however, the classical linear relation:

$E(Y) = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k$.

the following transformation is conducted:

$$g(E(Y)) = a + b_1x_1 + \cdots + b_kx_k$$

where the function *g* is given by: $g(p) = log\ \frac{p}{1-p}$ [17].

The probability of success is given by $p = \frac{exp^{a+b_1x_1+\cdots+b_kx_k}}{1+exp^{a+b_1x_1+\cdots+b_kx_k}}$.

As for the interpretation of the coefficients within the equation of the logit model, consider that each coefficient b is the logarithm of the odds ratio conditioned to the value assumed by the predictive x:

$$\frac{p|x+1}{1-p|x+1} = \exp(b)\frac{p|x}{1-p|x}$$

The logit model was first used with variables that were significantly relevant for a correlation with the outcomes of the analysis. Subsequently, with the forward stepwise method, validated through a backward stepwise elimination procedure, the statistically significant variables for logit regression ($p < 0.05$) were selected. The selection criterion for the progressive insertion and eventual removal of predictive variables is based on the test of the difference in the likelihood of subsequent models, while the subsequent interpretation of the model is usually based on the Wald index bj/SE(bj) [18].

3 Results and Discussion

In this paper we considered it mandatory to combine the results paragraph with the discussion paragraph because the multitude and detailed specificity of the results obtained had to be discussed simultaneously so as not to confuse the reader and lead him into misunderstandings about the research carried out.

3.1 Use of the Semi-parametric Cox Regression Model in Sample I

To identify the possible explanatory variables to include in the model, a preliminary analysis used the Kaplan–Meier method [16]. The graphic results of this analysis are reported conclusively only for prognostic factors significantly influencing the duration of hospitalization (except deaths). The prognostic factors that produced roughly parallel risk and survival curves (thus implying roughly proportional risks) and that were therefore included in the Cox model are listed below, divided by categories [19].

a. Cardiovascular diagnostic CCS.

96. Heart valve diseases.
98. Essential hypertension.
100. Acute myocardial infarction.
107. Cardiac arrest and ventricular fibrillation.
108. Congestive heart failure, not from hypertension.

b. CCS of cardiovascular interventions and procedures.

44. Coronary artery bypass.
47. Diagnostic cardiac catheterization and coronary arteriography.
48. Listing, reviewing, replacing, or removing cardiac pacemakers or an automatic defibrillator.
54. Other vascular catheterizations, excluding the cardiac category.

c. Nosological classifications (besides cardiovascular diseases).

1. Infectious and parasitic diseases.
2. Neoplasms.
13. Diseases of the musculoskeletal system and connective tissue.
16. Injuries and poisoning.
17. Symptoms, signs, undefined conditions, and other factors influencing the state of health.

d. Operative classifications (other than cardiovascular interventions).

8. Operations of the blood and lymphatic system.
14. Operations of the musculoskeletal system.

All the "non-clinical" characteristics of hospitalizations are not proportional or even time dependent, in some cases [20].

As seen in Table 3. Prognostic factors for the analysis of cardiovascular patients' hospital stays, Cox model.), the model was quite significant when it was obtained by applying the objective variable "duration of hospitalization" with the forward stepwise technique concerning the aforementioned prognostic factors. The probability of exclusion was $\alpha = 0.05$. The maximum loglikelihood test ensures, in fact, an almost infinitesimal probability ($p < 0.0000001$) that the relationships identified in the model are due to chance.

The effects of individual prognostic factors are also all statistically significant, some exceptionally. The variables that increase the probability of "risk" (of short hospitalizations, unfortunately, including deaths) by reducing "survival," that is, the duration of hospitalization, are as follows:

– CCS of cardiac arrest (a "risk" more than 25 times higher than the lack of diagnoses in this classification) [21];
– A diagnosis belonging to the nosological category "Symptoms, signs, undefined conditions, other factors influencing the state of health" (probability 1.8 times greater than the absence of one of these diagnoses);
– Non-cardiac vascular catheterization surgery (1.75 times greater probability than cases without interventions);
– CCS of congestive heart failure, not from hypertension (probability almost 1.6 times greater than its lack).

These conclusions are drawn, of course, from the last column of the table, which depicts the proportionality parameter $\text{Exp}(\beta) = \psi$ (also called the "odds ratio").

Some prognostic factors, presenting negative coefficients, protect patients from the risk of short hospitalizations (or death) and, hence, are statistically connected to long hospitalizations.

Table 3. Prognostic factors for the analysis of cardiovascular patients' hospital stays, Cox model.

Prognostic factors	β	SE	p-values	EXP(β)
Diseases of the heart valves	–1,143	0,286	<0,001	0,319
Essential hypertension	–0,730	0,168	<0,001	0,482
Cardiac arrest and ventricular fibrillation	3,234	0,108	<0,001	25,385
Congestive heart failure, not due to hypertention	0,463	0,125	<0,001	1,589
Diagnostic cardiac catheterization, coronary arteriography	–0,837	0,284	<0,005	0,433
Insertion, review, replacement, removal of cardiac pacemaker or automatic defibrillator	–1,417	0,582	<0,020	0,242
Other vascular catheterization, excluding cardiac cathegory	0,560	0,188	<0,005	1,750
Diseases of the musculoskeletal system and connective tissue	–0,826	0,266	<0,005	0,438
Symptoms, signs, undefined conditions, other factors influencing the state of health	0,594	0,125	<0,001	1,811

Going from the most connected factor to the least regarding the duration of hospitalization, these factors are listed in their order of relevance, as follows:

- The insertion, revision, replacement, or removal of a cardiac pacemaker or automatic defibrillator. (Compared to those who do not undergo interventions of this type, the probability of short hospitalizations is reduced to less than a quarter);
- CCS of heart valve diseases (a third-less probability compared to those without these diseases);
- Interventions of diagnostic cardiac catheterization or coronary arteriography and diagnoses of the nosological category diseases of the musculoskeletal system and connective tissue reduce almost equally the probability of short hospitalizations (43–44% compared to cases without these factors);
- CCS of essential hypertension also significantly reduces (to less than half) the probability of short hospitalizations.

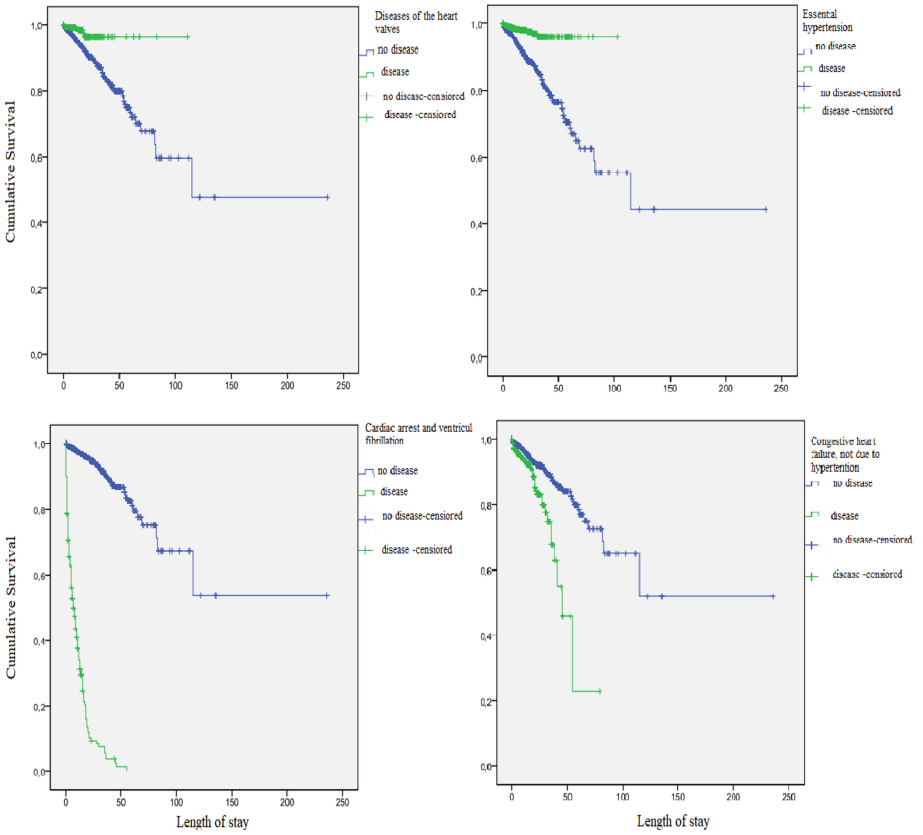

Fig. 1. Prognostic factors for the cardiovascular patients' hospital stays using Kaplan-Meier estimators.

We now observe the effects of notable prognostic factors on hospitalization times using graphs derived from the Kaplan–Meier estimators (Fig. 1) used to identify variables entitled to be included in the Cox model [22].

First, hospitalizations for diseases belonging to the CCS "Heart Valve Diseases" last up to 100 days, with high survival and a very concentrated distribution of deaths in the first month.

Approximately the same observations can be made concerning the pathologies belonging to the CCS (Fig. 1) "Essential Hypertension" and (Fig. 2) the interventions of "Diagnostic Cardiac Catheterization, Coronary Arteriography" (whose hospitalizations do not exceed 50–60 days). The dynamics related to the other prognostic factors are completely opposite, particularly for hospitalizations concerning the CCS (Fig. 1) of "Cardiac Arrest and Ventricular Fibrillation" (the duration curve of which is very steep, with many deaths concentrated in the first 10 days of hospitalization but never exceeding 55 days).

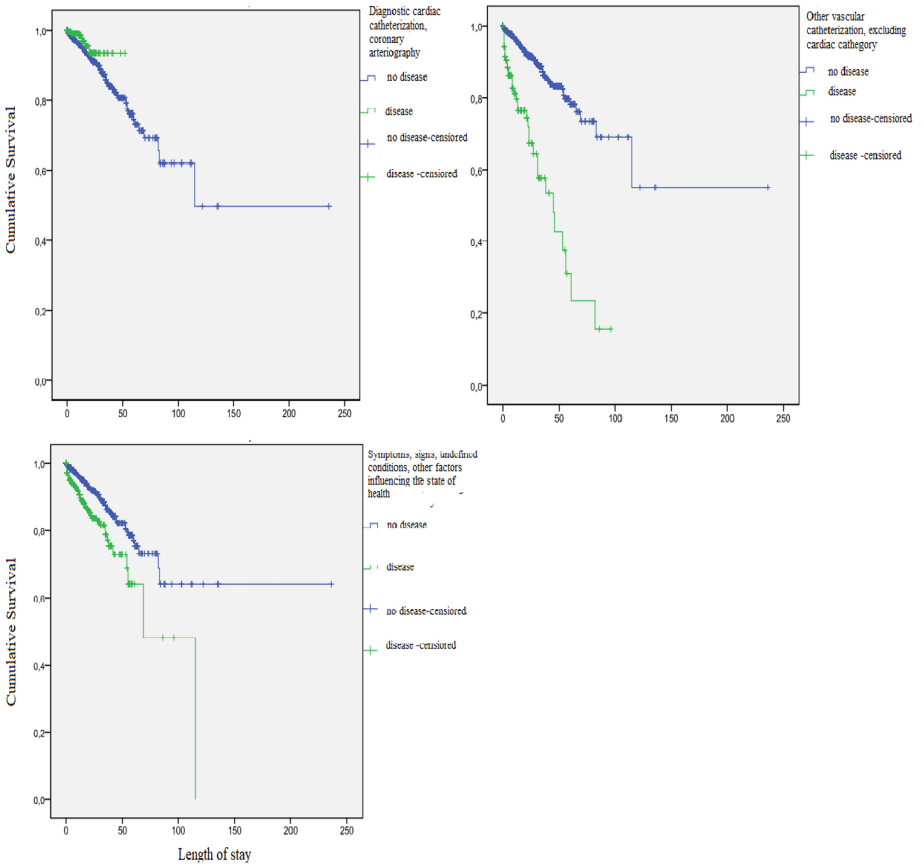

Fig. 2. Prognostic factors for the cardiovascular patients' hospital stays using Kaplan-Meier estimators.

3.2 Use of Logistic Regression Model in Sample I

Table 4. Prognostic factors for the analysis of cardiovascular patients' hospital stays, Logistic regression model.), demonstrates the results of a logistic regression model of the status variable (surviving/deceased) from the same variables considered in the previous Cox model to understand the role of censorship (more specifically, deaths) within this model [23].

If we compare the two methods, Table 3. Prognostic factors for the analysis of cardiovascular patients' hospital stays, Cox model.) makes evident that, concerning simple death events, the interventions of insertion, revision, replacement, or removal of a cardiac pacemaker or an automatic defibrillator, as well as diseases of the musculoskeletal system and connective tissue lose any statistical significance. Meanwhile, the other prognostic factors increasing the risk of death are much more relevant.

Conversely, "protective" factors have slightly lower effects in their odds ratio; however, they generally vary little from those of the Cox model. It can, then, be assumed that

Table 4. Prognostic factors for the analysis of cardiovascular patients' hospital stays, Logistic regression model.

Prognostic factors	B	SE	p-values	EXP(β)
Diseases of the heart valves	–1,290	0,261	<0,001	0,275
Essential hypertension	–0,881	0,169	<0,001	0,414
Cardiac arrest and ventricular fibrillation	5,388	0,201	<0,001	21,867
Congestive heart failure, not due to hypertention	1,184	0,147	<0,001	3,266
Diagnostic cardiac catheterization, coronary arteriography	–0,981	0,266	<0,001	0,375
Other vascular catheterization, excluding cardiac cathegory	3,015	0,205	<0,001	20,388
Symptoms, signs, undefined conditions, other factors influencing the state of health	1,368	0,148	<0,001	3,926

the structure of the latter model is only partially influenced by trends in hospital deaths and that its results are sufficiently reliable to discern prognostic factors.

3.3 Experimental Procedure in Sample II

The cardio–circulatory disease variables initially included in the Cox model are as follows:

- Total number of diagnoses and procedures (procedures include both examinations and interventions);
- Variables related to subjects, such as sex and age classes;
- Variables concerning the type of hospital admission (in the case of surgical or medical hospitalization).

Once the likelihood tests were verified, it was found that the p-value was highly significant ($p = 0.0001$). In fact, the individual variables highlighted in Table 5 are all very significant (with the least significant variable being $p = 0.005$).

The variables that most increase the probability of death during hospitalization are as follows:

- Cardiac arrest (deaths more than 16 times greater than the lack of it);
- Non-cardiac vascular catheterization surgery (2.9 times higher probability);
- Acute cerebral vascular disease (2.38 times higher probability);
- Essential hypertension and hypertension with complications are relevant factors for lengthy hospital survival.

Table 5. Prognostic factors for the analysis of cardiovascular patients' hospital stays, Cox model.

Prognostic factors	B	SE	p-values	EXP(β)
Age (classes)	0,287	0,055	<0,001	1,332
Essential hypertension	–0,800	0,209	<0,001	0,450
Hypertension complications/ secondary hypertension	–0,588	0,178	0,001	0,555
Cardiac arrest and ventricular fibrillation	2,780	0,150	<0,001	16,112
Acute cerebrovascular disease	0,870	0,173	<0,001	2,387
Infectious and parasitic diseases	0,603	0,214	0,005	1,827
Diseases of the respiratory system	0,408	0,138	0,003	1,504
Other vascular catheterization, excluding cardiac cathegory	1,069	0,244	<0,001	2,912

Starting from the results described here, it is appropriate to observe how the effects of significant prognostic factors induced in hospitals remain in survival times. To do this, we will use Kaplan–Meier estimators [24], which are not constrained to fixed-time subintervals.

Therefore, they allow comparisons between modes of flexible prognostic factors using survival curves relative to the subjects who detect such rules. The first prognostic factor for which we cite the graph of the survival function (Fig. 3) is the age group [25].

The graph (Fig. 3) highlights that the hospital stays of patients suffering from cardiovascular diseases, distributed by age group, differ in two groups:

- Short- and medium-duration hospitalization (1–1 80 days) for the age groups 57–67, 68–77, 78–83, and > 84. Older patients are also likelier to die in the first 60 days.
- Long stay (180–365 days), which is part of the age groups < 44 and 45–56. We can add that people younger than 44 have more chance of survival than other classes from the first month of hospitalization; this probability increases over time compared to other patients. It was found that between 80 and 130 days of hospitalization, the likelihood of living increases for the 57–67 age group compared to the age group 45–56.

Concerning patients suffering from cardiovascular diseases who have cardiac arrest and ventricular fibrillation (Fig. 3) as their main diagnosis, we can reinforce the predictions of the Cox method (16 times greater probability of death). Namely, patients with such a diagnosis have a much higher probability of death than patients without such a diagnosis. We can add that the sample combined with cardiac arrest has a maximum hospital stay of 125 days (instead of months).

Regarding (Fig. 4) noncardiac vascular catheterization surgery (2.9 times higher probability), these patients have shorter hospital stays (maximum 140 days) and greater chances of death, up to 44 days of hospitalization.

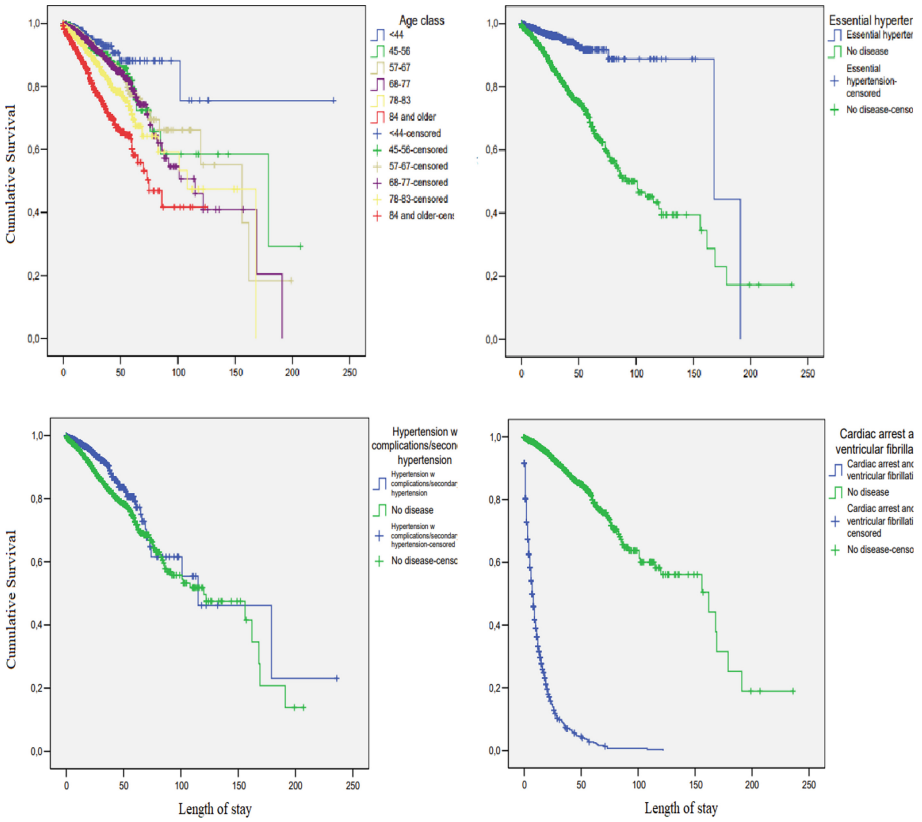

Fig. 3. Prognostic factors for the cardiovascular patients' hospital stays using Kaplan-Meier estimators.

Acute cerebral vascular disease patients (Fig. 4) are 2.4 times likelier to die. Patients with this disease die more frequently in the first 65 days. The situation of mortality greatly lessens as time progresses. The longest hospital stay for this type of diagnosis is 210 days.

However, patients with infectious and parasitic diseases are 1.8 times likelier to die in hospitals than patients without these issues. The most deaths occurred between the first and twentieth days of hospitalization. The duration of hospitalization in this circumstance reached 244 days.

Last, patients with respiratory diseases (Fig. 4) were likely to die 1.5 times more often than patients who do not have the disease. They die more between the first and the 54th days. Finally, such hospitalizations can last 247 days.

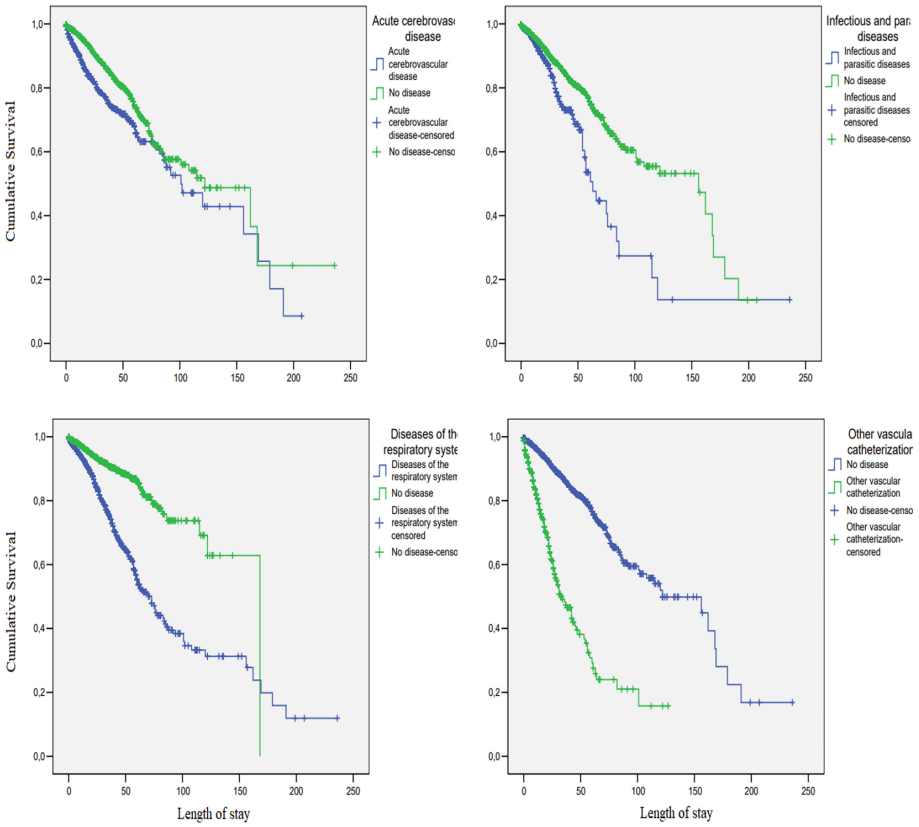

Fig. 4. Prognostic factors for the cardiovascular patients' hospital stays using Kaplan-Meier estimators.

4 Conclusion

The aim of this study was to statistically determine whether clinical and non-clinical factors impacted the duration of hospital admissions for cardiovascular diseases in Puglia between 2009 and 2014 (last year of free-use data).

The significant results described here are all related to diagnostic categories or interventions, particularly for the important nosological category of prognostic factors. It is worth noting that literature suggests a strong link between longer hospitalizations and less successful clinical interventions.

The analysis of both samples confirmed the clinical relevance and statistical significance of three prognostic factors: essential hypertension, cardiac arrest and ventricular fibrillation, and other vascular catheterization (excluding cardiac category). This demonstrates the robustness of the results.

In the analysis, we also included the 'age class of patients' factor ina second representative sample to investigate whether hospital admissions for this diagnostic category allow for more choices that are less dependent on clinical motivations. We conducted

further analysis of data parallel to that presented in Sample I to verify the duration of hospitalizations following this diagnostic category.

This work provides tools and information to recognise symptoms that exacerbate the mentioned diagnoses. It aims to prevent the worst outcomes and to prepare for serious emergencies and acute interventions for certain age groups and disease combinations. Based on these epidemiological results, it is appropriate to expect Apulian hospitals to activate intervention identification protocols for the six diagnoses mentioned in this work and for the age groups at greatest risk (68–77, 78–83, and > 84).

References

1. Wilkins, E., et al.: European Cardiovascular Disease Statistics 2017. European Heart Network, Brussels (2017)
2. Roth, G.A., et al.: Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet **392**, 1736–1788 (2018)
3. Centers for Disease Control and Prevention). The Burden of Chronic Diseases and Their Risk Factors—National and State Perspectives; CDC: Washington, DC, USA (2004)
4. Moriyama, H., Kohno, T., Kohsaka, S., Shiraishi, Y., Fukuoka, R., Nagatomo, Y., et al.: Length of hospital stay and its impact on subsequent early readmission in patients with acute heart failure: a report from the WET-HF Registry. Heart Vessels **34**(11), 1777–1788 (2019). https://doi.org/10.1007/s00380-019-01432-y
5. Timmis, A., et al.: European society of cardiology, on behalf of the Atlas Writing group, European society of cardiology: cardiovascular disease statistics 2021. Eur. Heart J. **43**(8), 716–799 (2022). https://doi.org/10.1093/eurheartj/ehab892
6. New York Heart Association, New York Heart Association. Criteria Committee. Nomenclature and criteria for diagnosis of diseases of the heart and great vessels. Little, Brown Medical Division (1979)
7. Wright, S.P., Verouhis, D., Gamble, G., Swedberg, K., Sharpe, N., Doughty, R.N.: Factors influencing the length of hospital stay of patients with heart failure. Eur. J. Heart Fail. **5**(2), 201–209 (2003). https://doi.org/10.1016/S1388-9842(02)00201-5
8. Formiga, F., Chivite, D., Manito, N., Mestre, A.R., Llopis, F., Pujol, R.: Admission characteristics predicting longer length of stay among elderly patients hospitalized for decompensated heart failure. Eur. J. Intern. Med. **19**(3), 198–202 (2008). https://doi.org/10.1016/j.ejim.2007.09.007
9. Mitani, H., Funakubo, M., Sato, N., Murayama, H., Rached, R., Matsui, N., et al.: In-hospital resource utilization, worsening heart failure, and factors associated with length of hospital stay in patients with hospitalized heart failure: a Japanese database cohort study. J. Cardiol. (2020). https://doi.org/10.1016/j.jjcc.2020.05.010
10. McPherson, K.: Length of stay and health outcome. Br. Med. J. (Clin. Res. Ed.) **288**(6434), 1854–1855 (1984). https://doi.org/10.1136/bmj.288.6434.1854
11. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc. **53**, 457–481 (1958). Tripepi, G., Catalano, F.: (2004)
12. Cox, D.R.: Regression models and life-tables (with discussion). J. Royal Stat. Soc. B **34**, 187–220 (1972)
13. Cox, D.R.: Partial likelihood. Biometrika **62**, 269–276 (1975)
14. Delvecchio, F.: Analisi dei tempi di sopravvivenza. Dipartimento di Scienze Statistiche dell'Università di Bari, Appunti a uso degli studenti (2001)

15. Cox, D.R., Snell, E.J.: Applied Statistics: Principles and Examples. Chapman & Hall, Londra (1981)
16. Kaplan-Meier analysis. Giornale Italiano di Nefrologia **21**(6), 540–546. Van Dijk, P.C., Jager, K.J., Zwinderman, A.H., Zoccali, C., Dekker, F.W.: The analysis of survival data in nephrology: basic concepts and methods of Cox regression. Kidney Int. **74**(6), 705–709 (2008)
17. Berkson J.: Application of the logistic function to bio-assay. J. Am. Stat. Assoc. **39**(227), 357–365 (1944). https://doi.org/10.2307/2280041
18. Lenz, S.T.: Alan Agresti (2013): Categorical data analysis. Stat Papers **57**, 849–850 (2016). https://doi.org/10.1007/s00362-015-0733-8
19. Provenzano, F., Tripepi, G., Zoccali, C.: How to measure effects in clinical research (corrected). G. Ital. Nefrol. **27**(3), 296–300 (2010)
20. Almashrafi, A., Elmontsri, M., Aylin, P.: Systematic review of factors influencing length of stay in ICU after adult cardiac surgery. BMC Health Serv. Res. **16**, 318 (2016). https://doi.org/10.1186/s12913-016-1591-3
21. Barili, F., et al.: An original model to predict Intensive Care Unit length-of stay after cardiac surgery in a competing risk framework. Int. J. Cardiol. **168**, 219–225 (2013). https://doi.org/10.1016/j.ijcard.2012.09.091
22. Cox, D.R., Snell, E.J.: The Analysis of Binary Data, 2nd edn. Chapman and Hall, London (1989)
23. Hosmer, D.W., Lemeshow, S.: Applied Logistic Regression, 2nd edn. Wiley, New York (2000)
24. Goel, M.K., Khanna, P., Kishore, J.: Understanding survival analysis: Kaplan-Meier estimate. Int. J. Ayurveda Res. **1**(4), 274–278 (2010). https://doi.org/10.4103/0974-7788.76794
25. Marschall, A., et al.: Pacemaker therapy in the elderly and very elderly: comorbidity-burden vs age, as prognostic factors for excess of length of in-hospital stay, complications and mortality. Euro. Heart J. **42**(Supplement_1), October 2021. ehab724.2820.https://doi.org/10.1093/eurheartj/ehab724.2820