

Article

Modeling Hospital Resource Management during the COVID-19 Pandemic: An Experimental Validation

J. M. Calabuig ¹, E. Jiménez-Fernández ^{2,*}, E. A. Sánchez-Pérez ¹ and S. Manzanares ³

¹ Instituto Universitario de Matemática Pura y Aplicada, IUMPA-UPV, Universitat Politècnica de València, Camí de Vera, s/n, 46022 Valencia, Spain; jmcabubu@mat.upv.es (J.M.C.); easancpe@mat.upv.es (E.A.S.-P.)

² Departamento de Teoría e Historia Económica, Universidad de Granada, Campus Cartuja, s/n, 18071 Granada, Spain

³ Dirección Médica, Hospital Virgen de las Nieves, Av. de las Fuerzas Armadas, 2, 18014 Granada, Spain; sebastian.manzanares.sspa@juntadeandalucia.es

* Correspondence: edjimfer@ugr.es

Abstract: One of the main challenges posed by the healthcare crisis generated by COVID-19 is to avoid hospital collapse. The occupation of hospital beds by patients diagnosed by COVID-19 implies the diversion or suspension of their use for other specialities. Therefore, it is useful to have information that allows efficient management of future hospital occupancy. This article presents a robust and simple model to show certain characteristics of the evolution of the dynamic process of bed occupancy by patients with COVID-19 in a hospital by means of an adaptation of Kaplan-Meier survival curves. To check this model, the evolution of the COVID-19 hospitalization process of two hospitals between 11 March and 15 June 2020 is analyzed. The information provided by the Kaplan-Meier curves allows forecasts of hospital occupancy in subsequent periods. The results shows an average deviation of 2.45 patients between predictions and actual occupancy in the period analyzed.

Keywords: health econometrics; COVID-19; Kaplan-Meier; decision; optimization



Citation: Calabuig, J. M., E. Jiménez-Fernández, E. A. Sánchez-Pérez, and S. Manzanares. 2021. Modeling Hospital Resource Management during the COVID-19 Pandemic: An Experimental Validation. *Econometrics* 9: 38. <https://doi.org/10.3390/econometrics9040038>

Academic Editor: Kajal Lahiri

Received: 27 August 2021

Accepted: 2 October 2021

Published: 14 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

One of the conclusions drawn from the analysis of previous health crises is that prevention strategies are necessary to cushion the impact of a pandemic of the characteristics of the COVID-19 disease (Lai et al. 2004; Zhao et al. 2020). Currently, prevention strategies are also key to control the occurrence of new outbreaks of this disease. Goscé et al. (2020) shows that immediate action in the early stages of the epidemic in the affected London boroughs would have slowed the spread. Unpredictability in the implementation of measures to contain an epidemic has the immediate consequence of overloading the healthcare system. The result is massive and unexpected hospital admissions, creating a national problem of unforeseeable dimensions.

Public health agencies report epidemic curves periodically, focusing on laboratory confirmed cases of COVID-19 using the various diagnostic tests. However, epidemic curves based on laboratory-confirmed cases show disease detection, but do not help to efficiently forecast hospital bed occupancy or intensive care unit (ICU) bed occupancy (García-Basteiro et al. 2020). The main challenge facing the governments of countries affected by the pandemic is to prevent the healthcare system from collapsing or becoming so stressed that it cannot adequately care for patients with COVID-19 or other pathologies.

Following the outbreak of the COVID-19 pandemic, a remarkable effort has been made to obtain reliable information on mortality and critical events in patients affected by the disease. For example, Vaid et al. (2020) show an application of machine learning techniques to this question, using Extreme Gradient Boosting together with comparable baseline methods to provide some prospective insight into relevant changes in patient status. These models are able to identify at-risk patients and underlying correlations

that were not explicitly observed before by experts. However, relevant to the present work, this provides valuable information on the status of the pandemic, but also useful data for the management of a hospital's resources: knowing how many intensive care beds could be available a week earlier could save lives. More information on machine learning techniques for management during the COVID-19 pandemic can be found in, for example, (Roimi et al. 2021) or (Domínguez-Olmedo et al. 2021) and the references therein. To improve prospective techniques and adapt current machine learning technology, other procedures that blend statistical analysis and artificial intelligence have also proved useful. The reader can find applications of the so-called state-space models for inference in the paper (Zhou and Ji 2020) and the references cited in it (see Imani and Ghoreishi 2021) for some applicable current developments of this technique. Other aspects of COVID-19 management that are relevant and affect public health have also been addressed using machine learning tools.

Reliability Analysis or Survival Analysis is the set of procedures used to analyze data in which the latent variable is the time that elapses from a well-defined initial instant to the occurrence of a certain event or final instant. This variable represents the time of life, for example relapse of a patient with respect to a certain pathology, death, or the occupation of a bed. From a mathematical point of view, this variable is represented by a non-negative random variable (Cox and Oakes 1984). One of the most popular models to describe how to estimate and graph survival curves is the Kaplan-Meier (KM) model (Kaplan and Meier 1958). We use this model to estimate hospital occupancy for patients who have been diagnosed with COVID-19.

The Kaplan-Meier curve of bed occupancy by COVID-19 patients (KM curve) represents the management capacity of hospital institutions, and can be considered as the "fingerprint" of the system's response to an increasing number of admissions. It acts as the transfer function in the control system that mathematically expresses the balance between inputs to the system (new patients) and outputs (discharged or deceased patients), which can be represented as a convolution of the inputs with the KM curve (Calabuig et al. 2020).¹ The KM curve we refer to is the one that represents the statistical distribution of the probability that an admitted patient remains in the hospital; in this sense, it can be understood as the survival function of the virus. Our treatment is independent of the disease process during the course of the patients' stay in the hospital, and of the factors that aggravate the disease, (Vincent and Taccone 2020), focusing attention on the statistical distribution of the time a patient stays in the hospital.

The aim of this paper is to present the survival analysis performed on the data of hospital admissions and discharges of COVID-19 patients during the second quarter of 2020 in two hospitals in Spain. The aim is to show how the procedure can be used systematically to provide simple and easy-to-analyze evidence of the state of the healthcare systems at each point in time and prospectively in a short interval of one-two weeks. The main methodological hypothesis that we assume for the application of the model is that, basically, hospital management will be similar in all the waves of contagion that are being suffered to the first one that was recorded before the summer of 2020 in Spain. However, the model can be improved by feeding back data in different periods. The implementation of a machine learning algorithm would provide more optimal information for future prediction of hospital occupancy. The unavailability of more recent data has prevented us from refreshing our model and, therefore, from applying these techniques with updated data. In any case, from the data available to us, it is possible to extract the relevant KM curves that allow us to predict the behaviour of hospitals in crisis situations, although the dynamics of admissions would not be the same (although similar). Moreover, although estimating the input function is the weak point of any model, the convolution of its extrapolation with the KM curve can help to dampen the error made in this approximation. This information provides us with an estimate of hospital occupancy, allowing us to adapt in the following weeks the future occupancy of other specialties to the estimated needs. Additionally, according to (O'Hagan and Stevens 2004), it is possible to estimate the average total cost

per patient and obtain an economic evaluation generated by the occupancy due to this disease. After the emergence of the COVID-19 crisis, numerous investigations on the estimation of the risk of severe disease in similar mathematical frameworks have been developed (see, for example, Liang et al. 2020; Zhao et al. 2020).

In addition to the one we propose in this work, there are different methodologies that allow estimating the bed capacity of hospitals during the pandemic. Garcia-Vicuña et al. (2020) analyze on the one hand, the inflows of patients affected by SARS-CoV-2 where the Gompertz curve is proposed after studying models based on the growth curves of the twenty most affected countries until 15 June 2020. On the other hand, they analyze the implementation of the simulation model (DAS) applied in the first outbreak in the Autonomous Community of Navarra (Spain), based on different procedures according to the information available on hospitalized patients. Bhandari et al. (2020) analyzed the survival curves of 987 COVID-19 patients from SMS Medical College, Jaipur (Rajasthan, India) where censoring is also established, as in our case, by hospital discharge or death. Although the data reported in the cited article differ significantly from those analyzed in the present work, they do show some coincidences; it should be taken into account that the health systems of the two countries on which our analysis is based (Spain and India), are different.

Given the assumptions we make—which we have already explained—the following considerations must be taken into account for the application of our model.

1. It should be borne in mind that this is essentially an extrapolation, so the results should be interpreted with the greatest of reservations.
2. It should also be noted that, since data input was closed on that day, it is assumed that on August 31 the number of patients admitted for COVID-19 was zero.
3. All patients are considered to complete their hospital stay, i.e., there are no cases that are not followed up. This has relevance within the classic Kaplan-Meier model (empty set of censored population).
4. Some statistical aspects related to the reliability of the solution (error bars, expected deviations, etc.) are only partially presented, since we intend to convey the most relevant information of the solution of this complex problem. This information can be found in part in the tables presented at the end of the paper.

2. Materials and Methods

2.1. Method

The Kaplan-Meier survival model consists of estimating the probability that an individual who has a given property—which is lost over time—will continue to keep it over time t (Kaplan and Meier 1958). The time variable takes discrete values and is counted by days, starting at $t = 0$. If we write $P(t)$ for this probability, we can estimate it by

$$P(t) = \frac{n(t)}{N} = \frac{N - d(t)}{N}, \quad t \in \mathbb{N}, \quad (1)$$

where $N = n(0)$ is the total amount of individuals in the population, $n(t)$ is the number of individuals still having the property at the time t , and $d(t)$ is the number of individuals leaving the group—deceased, either not having the property or not being under control anymore—after the time t . The canonical example is given by the so-called survival curves: for example, the distribution of survival time of patients after heart transplantation.

In our context, we apply these curves for the case in which we consider a given population of COVID-19 patients admitted to a hospital, and $n(t)$ the number of these individuals still in the hospital after t days. This allows us to describe the hospital dynamics, i.e., the number of hospital beds that have been used by COVID-19 patients, when considering a flow of new cases. The function $P(t)$ is then an estimate of an ideal distribution function $\mathcal{P}(t)$ that would describe the global behavior of the flow of patients through the hospital. We can then apply the classical statistical framework to estimate some relevant parameters of the flow. Let us recall that, in general, if we have a continuous probability distribution

$p(t)$ and we want to determine the expectation of the value of any random variable $X(t)$ in an interval $[0, t]$, we can write it as the integral

$$\sum_{s=0}^t p(s) X(s) ds = E[X](t)$$

However, the Kaplan-Meier curve does not give such a distribution function but an estimate of the probability of the virus to survive in the control group of patients after t days, and so we have to rewrite this equation to get the desired estimated value of any random variable in an interval $[0, t]$. Let $[0, T]$ be a discrete interval measured in days. Let us consider a real function $J : [0, T] \rightarrow \mathbb{R}$ that gives any property that can be assumed to be proportional to the number of patients (for example, the number of beds to be used, hospital rooms, number of intensive-care ventilators, or the number of patients itself). Our model is based on the construction of an estimate of the total amount of this quantity by means of the estimate of the number of new cases entering the hospital, and its convolution with the “pressure function” J representing the “new needs” of the variable J at the time t . Let us compute the total amount H of the “need” J after some days $t = 0, 1, 2, \dots$. At the day $t = 1$, if there is a need equal in $t = 0$ equal to $J(0)$, it produces a real need given by the rate of the value of $J(0)$ for the estimate of the rate of the patients that are still at the hospital, $P(1)$. Therefore, we have that at the day $t = 1$ the estimate of H is given by

$$H(1) = J(0)P(1)$$

For $t = 2$, we have the part of $I(0)$ corresponding to the rate of patients that are still in the center, $P(2)$ —that is, $J(0)P(2)$ —, plus the part of $J(1)$ corresponding the rate of patients that have been $2 - 1 = 1$ days, that is $J(1)P(1)$, what gives a total amount of

$$H(2) = J(0)P(2) + J(1)P(1)$$

Following this rule, we can write the general formula of the accumulated need H at the day t as the convolution equation

$$H(t) = \sum_{s=0}^{t-1} J(s) \cdot P(t-s), \quad t = 1, 2, 3, \dots$$

Note that we are starting to compute J at $t = 0$ —to initialize the original need—, but the model starts to work at $t = 1$. Using this equation, if we consider $J(t)$ to be the function $I(t)$ of the number of new cases entering the hospital on the day t , we have that the estimate of the total number of infected patients $D(t)$ still in the hospital on the day t is given by the equation

$$D(t) = \sum_{s=0}^{t-1} I(s) \cdot P(t-s), \quad t = 1, 2, 3, \dots$$

Of course, we do not know the exact value of the function P , but we can use the estimate of the probability given by the formula we started from to obtain an estimate of D . Anyway, we can also give a parameterization of such a function P and calculate an error band, as will be seen later, although it is not necessary to use this method. This simple model allows us to calculate, from the information obtained during the first wave of the epidemic (before August 2020), the number of patients in the hospital at a given time, once a good estimate of the function of new cases I is known. Also, as has been explained above, the model can be used for doing a prospective of any particular service of the hospital, for example the number of occupied beds in rooms of intensive care. It is not our intention to explain here the construction and main properties of our adaptation of the Kaplan-Meier model to this special case; the interested reader can find all the explanations and main equations in (Calabuig et al. 2020) and the references therein. To estimate the occupancy,

we programmed an algorithm using R software (R Core Team 2020). In addition, we use several standard R packages for the computations and plots, as *survival* (Therneau and Grambsch 2000) or *ggplot2* (Wickham 2016).

2.2. Material

Based on the anonymized statistical data provided on admissions and discharges of COVID-19 patients from the Andalusian Health Service in the hospitals of Granada from March 2020 to September 2020, the Kaplan-Meier survival curve for the SARS-COV2 in these hospitals was calculated. The sample consists of 1.176 patients who were admitted between 11 March and 15 August, the date on which the de-escalation period ended in the Andalusia region.

Let us provide a basic statistical description of the database of admissions and discharges by group. The database consists of 625 men and 551 women. Figure 1a shows the box plot that allows us to visualize the differences and the distribution of observations by gender. Considering the age of the men and women hospitalized for COVID-19, the mean age of the 625 men is 65.4 years with a standard deviation of 15.7 and the mean age of the 551 women is 66.9 years and a standard deviation of 18.5. There were only two outliers in both men and women, very young patients. A computing unpaired two-samples Wilcoxon test provides significant differences between the ages at admission at 5% significance with a p-value of 0.03069. Figure 1b shows the box plot with respect to the number of days of hospital stay. An asymmetric distribution is observed with a large number of patients having prolonged stays. Patients who suffer a worsening of the disease and do not die, have prolonged stays in intensive care units and later in recovery units of more than 30 days. Table 1 shows the results for stays by gender in the period analyzed.

Table 1. Statistic summary COVID-19 occupation by gender.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
Man	1.00	6.00	9.00	12.75	15.00	107.00
Woman	1.00	4.00	8.00	10.64	13.00	73.00

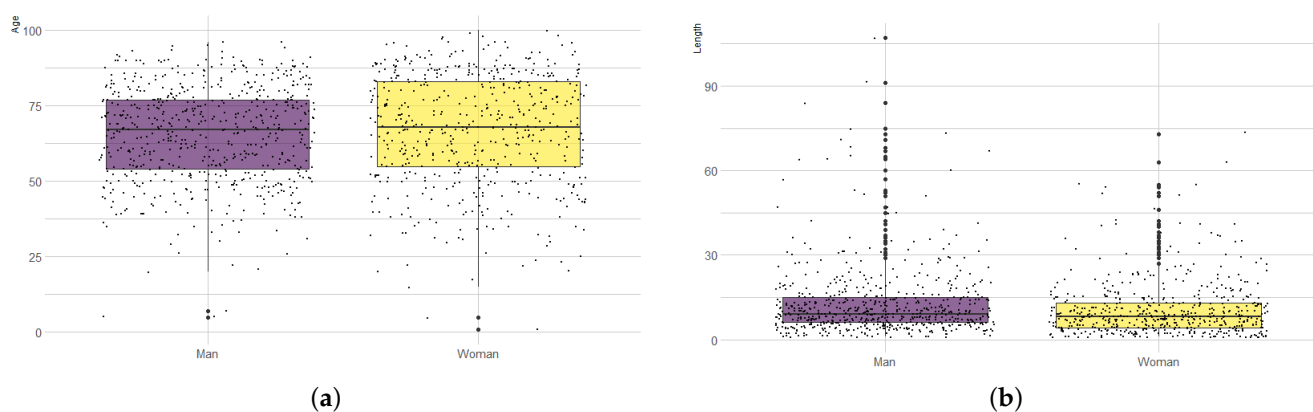


Figure 1. Gender Box plot of the age of patients admitted to hospitals in the Metropolitan Area of Granada by COVID-19. (a) Age, (b) Length.

3. Results

3.1. Kaplan-Meier Curves Obtained

The Kaplan-Meier curve for the period analyzed in the hospitals of Granada can be seen in Figure 2, and the results obtained are shown in Table 2. For example, the probability that a patient is occupying a bed for 8 days is between 0.47 and 0.54 assuming a significance level of 1%. For a set of observations of 1159 patients, the mean estimate for the survival

curve gives 11.764 days of average length of stay, with 9 days being the median estimate. The estimate of the deviation yields a value of 0.333 days.

Table 2. Results of the survival analysis for hospital occupancy by patients with COVID-19 in the hospitals of the Metropolitan Area of Granada between 11 March and 15 June 2020.

Time	N.Risk	N.Event	Surv	Std.Err	Lower	Upper
1	1159	50	0.96	0.01	0.94	0.97
2	1109	57	0.91	0.01	0.88	0.93
3	1052	57	0.86	0.01	0.83	0.88
4	995	83	0.79	0.02	0.75	0.82
5	912	74	0.72	0.02	0.69	0.76
6	838	79	0.65	0.02	0.62	0.69
7	759	89	0.58	0.03	0.54	0.61
8	670	82	0.51	0.03	0.47	0.54
9	588	63	0.45	0.03	0.41	0.49
10	525	71	0.39	0.04	0.35	0.43
11	454	55	0.34	0.04	0.31	0.38
12	399	40	0.31	0.04	0.28	0.34
13	359	41	0.27	0.05	0.24	0.31
14	318	34	0.25	0.05	0.21	0.28
15	284	37	0.21	0.06	0.18	0.24
16	247	21	0.19	0.06	0.17	0.23
17	226	18	0.18	0.06	0.15	0.21
18	208	21	0.16	0.07	0.13	0.19
19	187	14	0.15	0.07	0.12	0.18
20	173	20	0.13	0.08	0.11	0.16
21	153	14	0.12	0.08	0.10	0.15
22	139	11	0.11	0.08	0.09	0.14
23	128	15	0.10	0.09	0.08	0.12
24	113	10	0.09	0.09	0.07	0.11
25	103	5	0.08	0.10	0.07	0.11
26	98	6	0.08	0.10	0.06	0.10
27	92	9	0.07	0.11	0.05	0.09
28	83	4	0.07	0.11	0.05	0.09
29	79	7	0.06	0.11	0.05	0.08
30	72	3	0.06	0.12	0.04	0.08
31	69	9	0.05	0.13	0.04	0.07
32	60	5	0.05	0.13	0.03	0.07
33	55	1	0.05	0.13	0.03	0.06
34	54	3	0.04	0.14	0.03	0.06
35	51	6	0.04	0.15	0.03	0.06
36	45	3	0.04	0.15	0.02	0.05
37	42	3	0.03	0.16	0.02	0.05
38	39	2	0.03	0.16	0.02	0.05
39	37	1	0.03	0.16	0.02	0.05
40	36	1	0.03	0.17	0.02	0.05
41	35	5	0.03	0.18	0.02	0.04
42	30	2	0.02	0.19	0.01	0.04
45	28	2	0.02	0.19	0.01	0.04
46	26	1	0.02	0.20	0.01	0.03
47	25	2	0.02	0.21	0.01	0.03
51	23	2	0.02	0.22	0.01	0.03
52	21	2	0.02	0.23	0.01	0.03
53	19	1	0.02	0.23	0.01	0.03
54	18	1	0.01	0.24	0.01	0.03
55	17	2	0.01	0.26	0.01	0.02
57	15	1	0.01	0.27	0.01	0.02
60	14	1	0.01	0.28	0.01	0.02

Table 2. Cont.

Time	N.Risk	N.Event	Surv	Std.Err	Lower	Upper
63	13	1	0.01	0.29	0.00	0.02
64	12	2	0.01	0.31	0.00	0.02
65	10	1	0.01	0.33	0.00	0.02
67	9	1	0.01	0.35	0.00	0.02
68	8	1	0.01	0.38	0.00	0.01
71	7	1	0.01	0.41	0.00	0.01
73	6	2	0.00	0.50	0.00	0.01
75	4	1	0.00	0.58	0.00	0.01
84	3	1	0.00	0.71	0.00	0.01
91	2	1	0.00	1.00	0.00	0.01

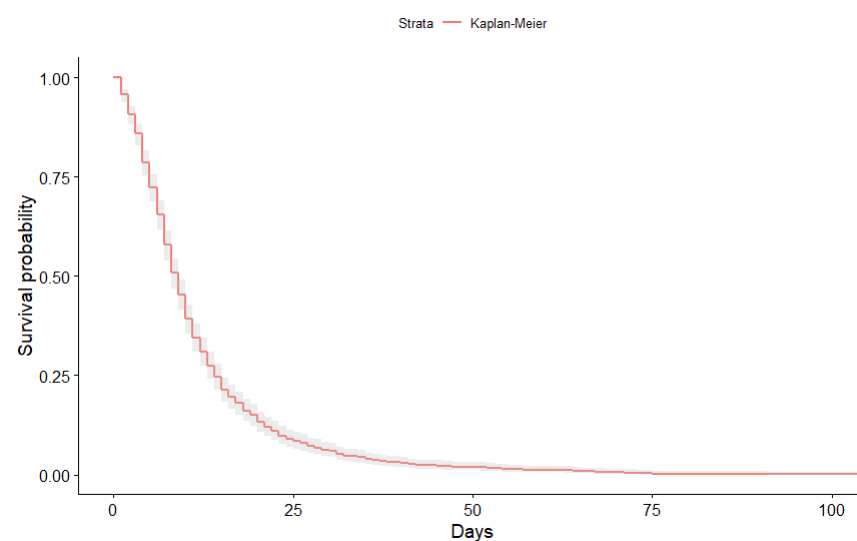


Figure 2. Kaplan-Meier curve for hospital occupancy by patients of the COVID-19 in the hospitals of the Metropolitan Area of Granada between 11 March–15 June 2020.

Another issue to take into account is described by the risk function², that is, the probability that a patient admitted for COVID-19 will leave the hospital on day t . While the survival function focuses primarily on the “non-occurrence” of the event (probability that the patient is occupying a bed on day t), the hazard function focuses on the “occurrence” of the event, i.e., that the patient leaves the hospital. This function provides useful information because it provides an occupancy rate. On the other hand, the cumulative risk function shown in Figure 3, provides the cumulative sum for the considered event (a patient leaves a free bed) on day t .

The results of the Kaplan-Meier curves and hazard ratios by gender and age groups are presented in the corresponding figures below. Figure 4 shows the Kaplan-Meier curve by gender for hospital occupancy by COVID-19 patients in hospitals in Granada between 11 March–15 June 2020. We used the Bonferroni method (see for example [Tripathi and Pandey 2017](#)) to verify that there are significant differences between both curves, being the probability of staying at least t days higher in men than in women. The computations show statistical differences between man and woman with a p-value equal to 0.00015. The risk functions are shown in Figure 5.

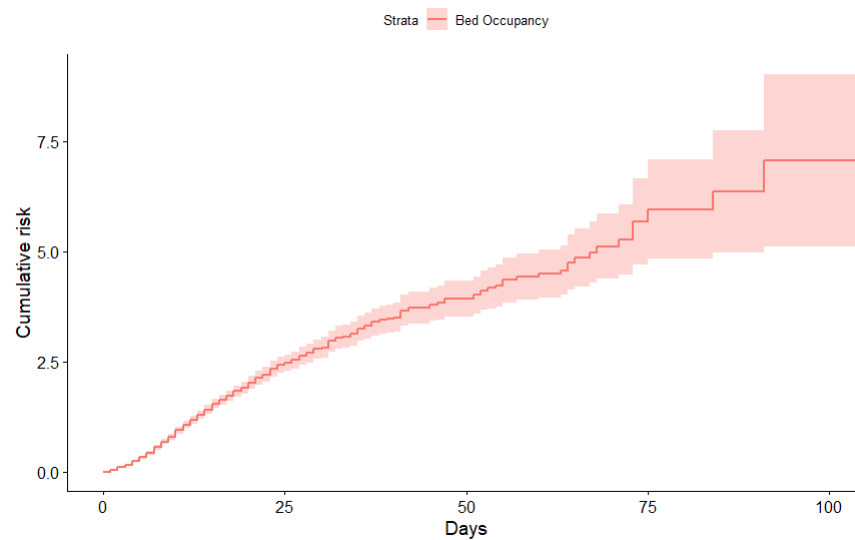


Figure 3. Cumulative Risk Curve for hospital occupancy by patients of the COVID-19 in Granada hospitals between 11 March–15 June 2020.

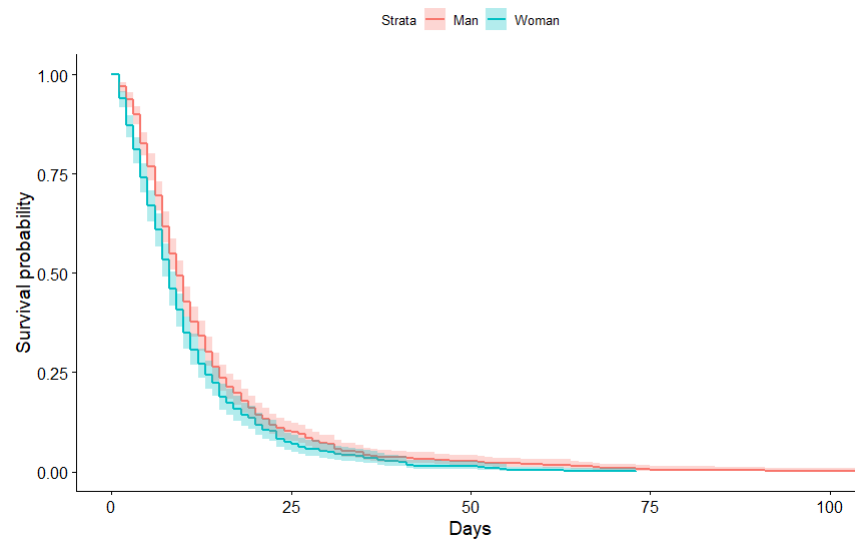


Figure 4. Kaplan-Meier curve by gender for hospital occupancy by COVID-19 patients in Granada hospitals between 11 March–15 June 2020.

Figure 6 shows the Kaplan-Meier curves by age ranges for hospital occupancy by COVID-19 patients in Granada hospitals between 11 March and 15 June 2020. Using the Bonferroni method, we studied whether there were significant differences between all of them. The only pair that does not present significant differences is between the curves of the age ranges 31–49 years and 50–65 years with a p-value of 0.3371. Therefore, these two age ranges can be analyzed together when estimating hospital occupancy. Additionally, the cumulative hazard function (Figure 7) shows us that the risk of leaving the hospital is high for patients aged 0–30 years after 25 days. Patients in the 31–49 age range have a constant risk of leaving the hospital after 55 days. The cumulative risk rises for patients over 65 years of age at 75 days and finally, at 90 days, for patients between 50 and 65 years of age.

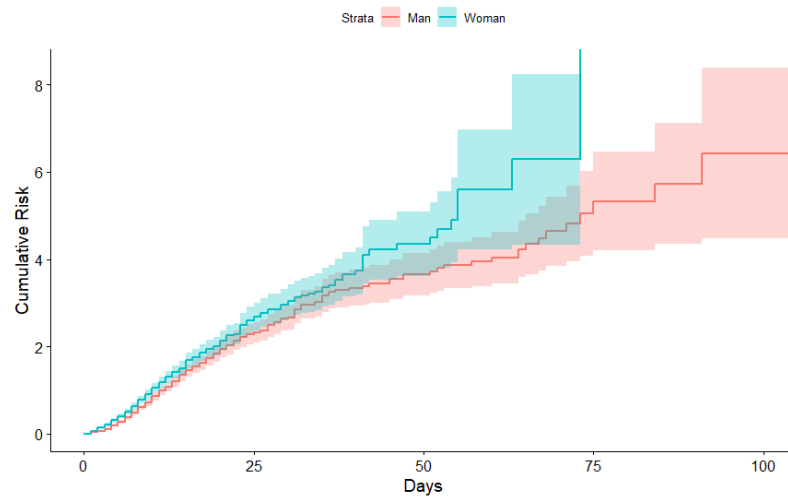


Figure 5. Risk curve by gender for hospital occupancy by COVID-19 patients in Granada hospitals between 11 March–15 June 2020.

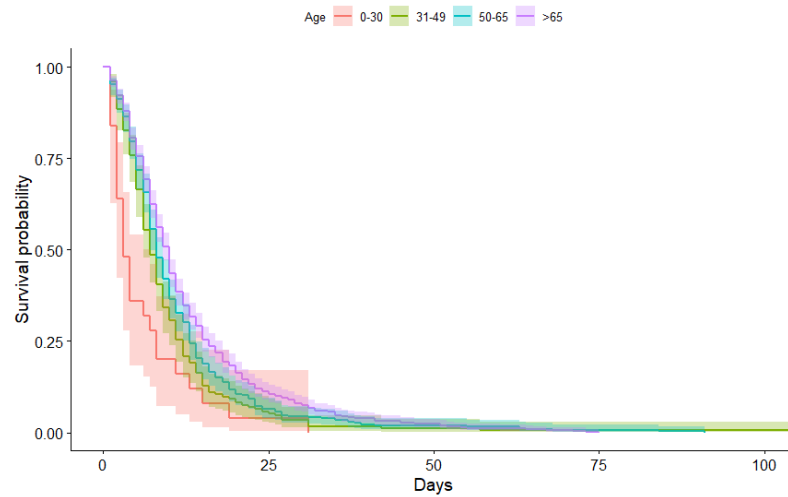


Figure 6. Kaplan-Meier curve for hospital occupancy by patients of the COVID-19 in Granada hospitals between 11 March–15 June 2020.

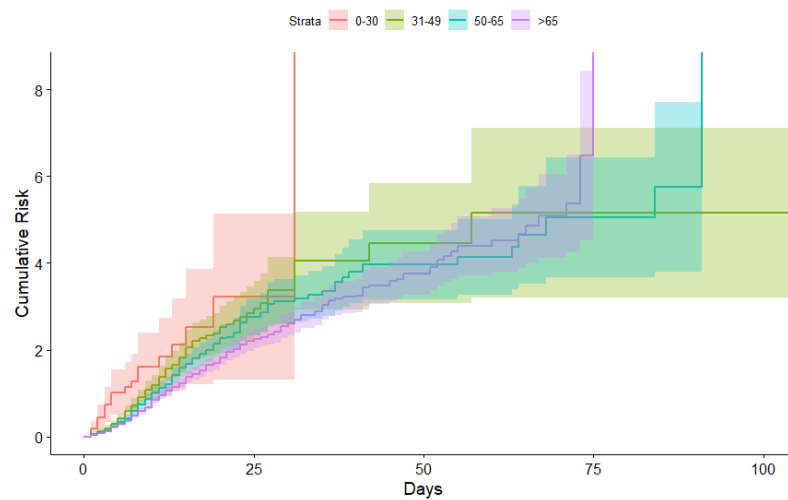


Figure 7. Risk curve for hospital occupancy by patients of the COVID-19 in Granada hospitals between 11 March–15 June 2020.

3.2. Occupancy in Intensive Care Units

Patient occupancy in intensive care units (ICU) represents the most sensitive problem in the management of this disease for several reasons. Firstly, because the number of beds is more limited. Secondly, because the occupancy rate has remained around 75% between 2013 and 2018, a level close to saturation (Eurostat 2020). The third and most worrying reason lies in the fact that patients in the Intensive Care Units of the COVID-19 have a much longer average length of stay. Consequently, the analysis of KM curves in these units becomes more necessary than in the rest.

We analyzed the KM curve of virus survival in the Intensive Care Units (ICU). We test our model based on the data provided by the Andalusian Health Service for the ICU of the Virgen de las Nieves Hospital during the months of March to June. The sample consisted of 75 patients who were admitted between 20 March and 15 August. Figure 8 shows the KM curve for the analyzed unit. On this case a manifest prolongation of stays is observed, for example, the probability of being at least 19 days in the ICU is 50% and between 9 and 11 days is 75%. The estimation of the cumulative hazard function estimated by the maximum likelihood method is shown in Figure 9 and Table 3 shows the events, risk and confidence interval for the probability that a patient is at least t days occupying an ICU bed. The mean length of stay in this unit is 20.40 days, with a standard deviation of 2.32 and a median of 14 days.

Table 3. Results of the survival analysis for ICU occupancy by COVID-19 patients at the Virgen de las Nieves Hospital.

Time	N.Risk	N.Event	Surv	Std.Err	Lower	Upper
1	75.00	6.00	0.92	0.03	0.79	0.97
2	69.00	3.00	0.88	0.04	0.74	0.95
3	66.00	2.00	0.85	0.05	0.71	0.93
4	64.00	4.00	0.80	0.06	0.65	0.89
5	60.00	4.00	0.75	0.07	0.59	0.85
6	56.00	1.00	0.73	0.07	0.58	0.84
7	55.00	2.00	0.71	0.07	0.55	0.82
8	53.00	3.00	0.67	0.08	0.51	0.79
9	50.00	1.00	0.65	0.08	0.49	0.77
11	49.00	4.00	0.60	0.09	0.44	0.73
12	45.00	1.00	0.59	0.10	0.43	0.72
13	44.00	3.00	0.55	0.11	0.39	0.68
14	41.00	5.00	0.48	0.12	0.33	0.62
16	36.00	2.00	0.45	0.13	0.30	0.59
18	34.00	4.00	0.40	0.14	0.26	0.54
19	30.00	3.00	0.36	0.15	0.22	0.50
21	27.00	2.00	0.33	0.16	0.20	0.47
22	25.00	3.00	0.29	0.18	0.17	0.43
23	22.00	1.00	0.28	0.19	0.16	0.42
25	21.00	1.00	0.27	0.19	0.15	0.40
26	20.00	1.00	0.25	0.20	0.14	0.39
27	19.00	1.00	0.24	0.21	0.13	0.37
28	18.00	1.00	0.23	0.21	0.12	0.36
30	17.00	1.00	0.21	0.22	0.11	0.34
32	16.00	1.00	0.20	0.23	0.10	0.33
33	15.00	2.00	0.17	0.25	0.08	0.30
34	13.00	1.00	0.16	0.26	0.07	0.28
35	12.00	1.00	0.15	0.28	0.06	0.27
37	11.00	1.00	0.13	0.29	0.05	0.25
38	10.00	1.00	0.12	0.31	0.05	0.23

Table 3. *Cont.*

Time	N.Risk	N.Event	Surv	Std.Err	Lower	Upper
39	9.00	1.00	0.11	0.33	0.04	0.22
43	8.00	1.00	0.09	0.36	0.03	0.20
44	7.00	1.00	0.08	0.39	0.02	0.18
47	6.00	1.00	0.07	0.43	0.02	0.17
49	5.00	1.00	0.05	0.49	0.01	0.15
78	4.00	1.00	0.04	0.57	0.01	0.13
83	3.00	1.00	0.03	0.70	0.00	0.11
84	2.00	1.00	0.01	0.99	0.00	0.09

Analogous to what happened in the occupancy of beds by patients in the COVID-19, Figures 10 and 11 show a significant difference in occupancy by gender and age interval. As in the previous case, men have longer stays and, on the other hand, patients over 50 years of age show the same behaviour.

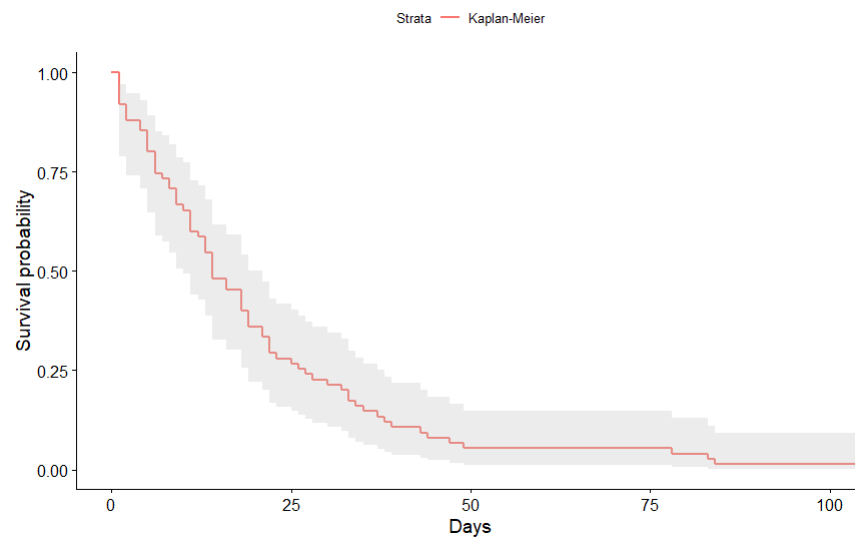


Figure 8. KM curve for ICU occupancy by COVID-19 patients in the Virgen de las Nieves hospital.

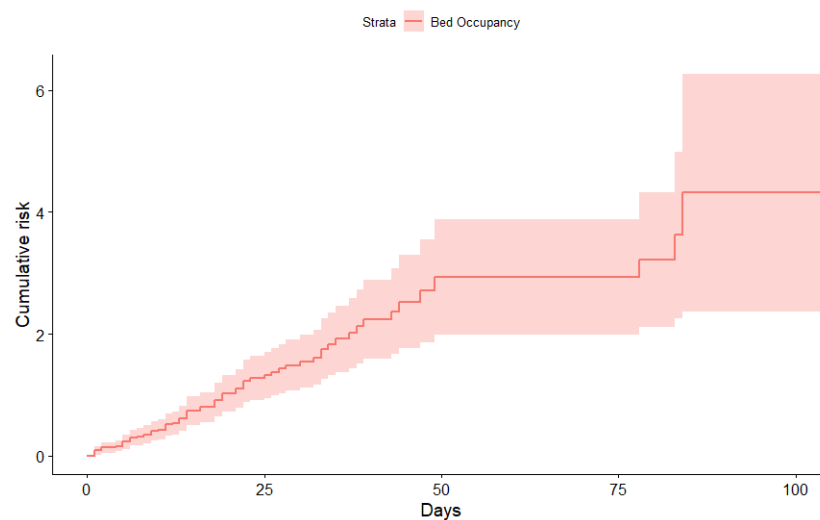


Figure 9. Risk curve for ICU occupancy by COVID-19 patients in the Virgen de las Nieves hospital.

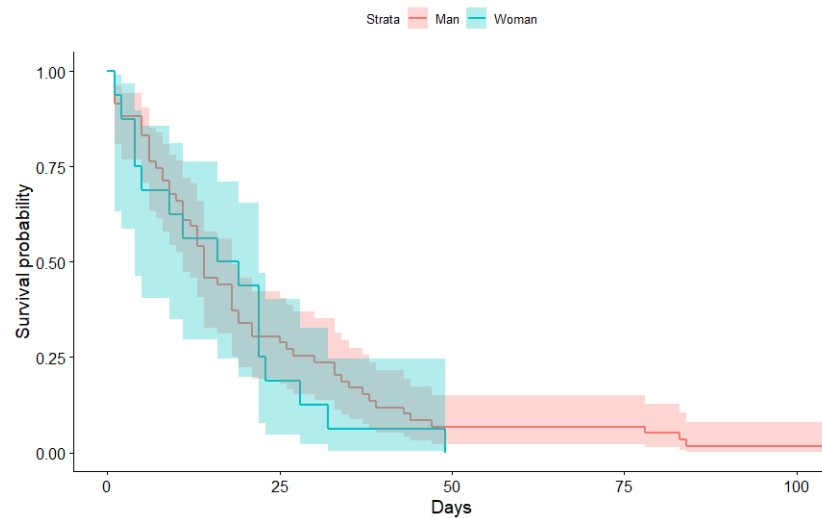


Figure 10. KM curve for ICU occupancy differentiated by gender in COVID-19 patients in the Virgen de las Nieves hospital.

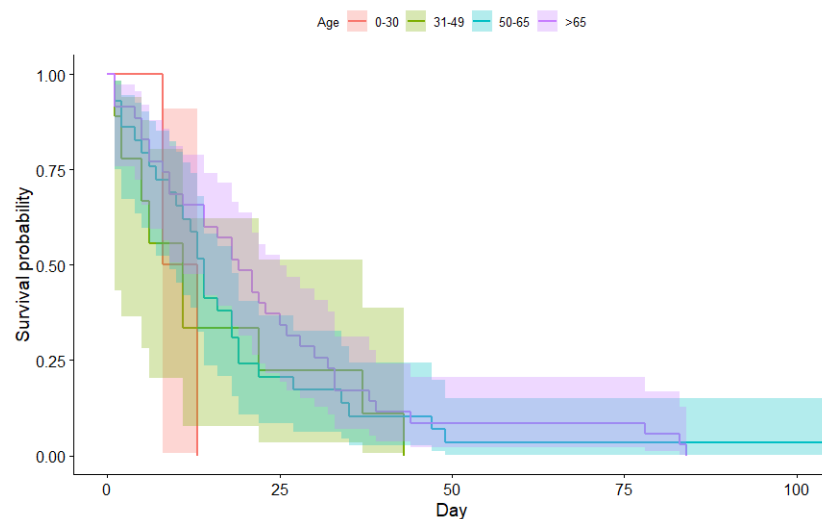


Figure 11. Risk curve for ICU occupancy differentiated by age intervals for the COVID-19 patients in the Virgen de las Nieves hospital.

3.3. Estimated Hospital Occupancy

Based on the data provided on admissions and discharges of patients with COVID-19 by the Andalusian Health Service during the months of March to June (first wave of COVID-19 infections), Kaplan-Meier curves were calculated to analyze the incidence on occupancy in hospitals in Granada. According to (Calabuig et al. 2020), it has been explained that an estimate of hospital occupancy can be made by convolving the KM curve with the estimated admissions function.

We will show how we can estimate the prediction error using validation or test sets. To do this, we use patient admission and discharge data from the COVID-19 between 16 June and 9 September to test the goodness of fit. The survival curve calculated in the period from 11 March to 15 June provides a COVID-19 patient occupancy model. With this nonparametric estimation we can approximate the balance of patient occupancy between 16 June and 9 September from the convolution of a (linear) extrapolation of the admissions during the previous period with the KM curve. This estimate can be compared with the actual COVID-19 patient occupancy data for the same period. Figure 12 shows the calculated simulated balance curve (blue line). On the other hand, the red curve represents the real balance for the analyzed period. It can be seen that, although linear extrapolation

is a rather poor model for estimating the number of new patients, the result is quite good, revealing that convolution with the KM curve gives a robust model that provides reasonable results even with naive estimates of the admission functions. To quantify the error of the prediction, we use the root mean square error

$$RMSE = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n} \quad (2)$$

where y_i represents the real balance of occupancy and \hat{y}_i denote the simulated occupancy for each day $i \in \{1, \dots, n\}$. We obtain a value for the RMSE of 2.4245 patients. The simulated curve is higher than the real one because in the analyzed period the incidence in occupancy was much lower. The knowledge of treatments for the fight against the disease has led a decrease in patient stays, reducing them to 7.88 days. Even so, the deviation obtained by the RMSE is sufficiently small and does not significantly affect the occupancy of a hospital floor.

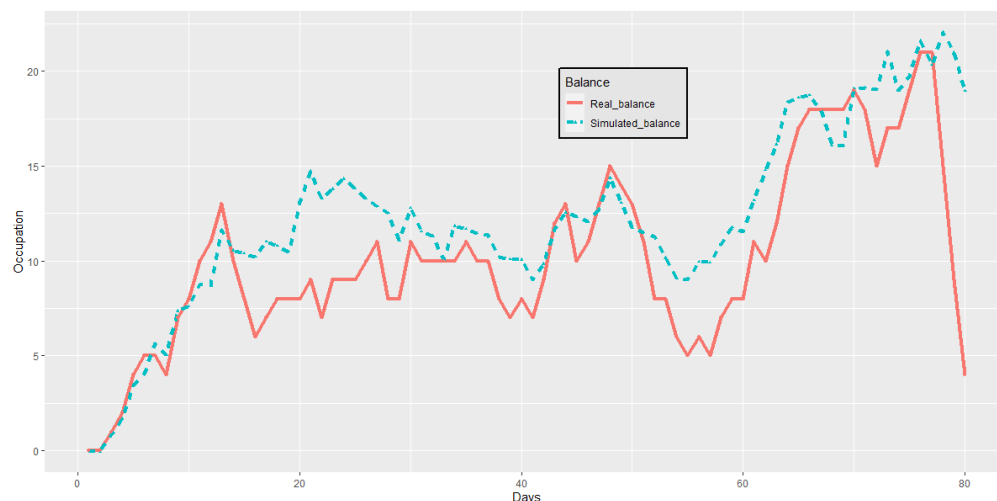


Figure 12. Actual and simulated COVID-19 patient occupancy balance between 16 June and 9 September in the hospitals of Granada.

Some researchers have endeavored to compute good estimates of this function using, for example, Gompertz-type curves, as done by (García-Basteiro et al. 2020), or SIR-type models as in (Cooper et al. 2020) or (Varotsos and Krapivin 2020). However, this is, in fact, the weak point of the models, since, as it has been shown during the whole crisis, it is really difficult to model the spread of the virus and, therefore, the number of new infections. Therefore, our model, which provides relatively good estimates with poor estimates of the admission function, could be useful at least for short-term forecasting.

On the basis of admissions to the ICU units of the Virgen de las Nieves hospital, from 16 August to 15 October 2020, we tested the goodness of fit of the model. In order to do it, we first estimated the discharges from this unit through convolution with the Kaplan-Meier curves, that were obtained in the first period using ICU data. Using this estimate, the balance of admissions and discharges or exitus of this unit can be calculated. Thus, a simulated bed occupancy curve for the ICU units of the hospital can be obtained. On the other hand, we have the real data, which correspond to the discharges or exitus in the same period. Figure 13 shows both balances. We can calculate the root mean square error (RMSE) between the actual values y_i of ICU occupancy and the estimated values \hat{y}_i . We obtain a value for RMSE of 2.1533 patients. As with non-ICU bed occupancy, we obtain an overestimation of the model. This variation could be partly due to a substantial improvement in survival among SARS-COV2 patients admitted to critical care, (Dennis et al. 2021).

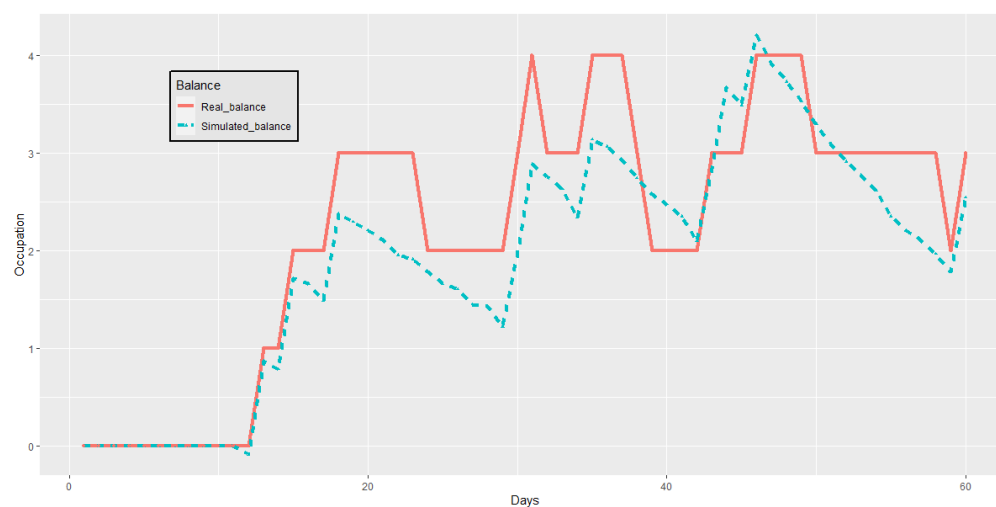


Figure 13. Real and simulated balances of occupancy in the ICUs by COVID-19 between 16 August and 15 October at the Virgen de las Nieves hospital.

4. Conclusions

This paper presents a methodology to provide a short-term estimate of hospital bed occupancy. When contagion in a given region is communal, hospitals suffer high stress, generating serious occupancy problems. Predicting patient occupancy allows healthcare managers to update hospital activity in the short term in order to optimize available resources.

The Kaplan-Meier curve of the bed occupancy by COVID-19 patients, together with convolution-type mathematical operators, provide a prediction of the balance of available beds in the following weeks. Additionally, the hazard function estimates the probability that a patient will leave a bed free on day t . Based on the data provided by the Andalusian Health Service on admissions and discharges of patients with COVID-19 in the hospitals of Granada from March to August, Kaplan-Meier curves and virus survival risk curves are calculated. The sample consists of 1176 patients who were admitted between March 11 and August 15. Assuming that no pharmacological solutions and treatments have been found to significantly reduce the severity of the disease, it can be expected that the temporal distribution of bed occupancy by COVID-19 patients will not change significantly in the coming months. In order to test the goodness of fit of the model, we estimate the balances based on the discharges of new patients with COVID-19 between August 16 and October 15 at the Hospital Virgen de las Nieves in the city of Granada.

The comparison between the estimated occupancy balances and the actual occupancy balances gives satisfactory results. The experiment was repeated with the occupancy of patients in the intensive care units, obtaining also satisfactory comparative results. The comparative study between COVID-19 patients in the ICUs and the rest of the COVID-19 patients is necessary, since the descriptive statistics show different results. The most seriously ill patients with this disease require longer hospitalizations and more exhaustive care, and therefore this group and the management of the corresponding service should be studied independently. It should be noted that the estimation of ICUs occupancy is crucial to analyze the activity of a hospital and its elasticity to provide a suitable response in a crisis. In summary, we can affirm that the models presented here and based on survival curves can be useful tools for the management of occupancy in the different units of a hospital, as well as for the management of all human resources and general services.

Author Contributions: Conceptualization, J.M.C., E.J.-F., E.A.S.-P. and S.M.; methodology, J.M.C. and E.A.S.-P.; software, E.J.-F. and E.A.S.-P.; validation, E.J.-F. and E.A.S.-P.; formal analysis, J.M.C., E.J.-F. and E.A.S.-P.; investigation, J.M.C., E.J.-F., E.A.S.-P. and S.M.; resources, S.M.; data curation, E.J.-F.; writing—original draft preparation, E.J.-F. and E.A.S.-P.; writing—review and editing, E.J.-F. and E.A.S.-P.; visualization, E.J.-F.; supervision, E.J.-F. and E.A.S.-P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request due to restrictions eg privacy. The data presented in this study are available on request from the corresponding author.

Acknowledgments: Servicio Andaluz de Salud (SAS), Hospital Virgen de las Nieves, Hospital San Cecilio, M. Angeles García, (gerente *Hospital Virgen de las Nieves*), Manuel Reyes (gerente *Hospital San Cecilio*), *Hospital Virgen de las Nieves*), Cristina Carvajal (directora médica *Hospital Virgen de las Nieves*), Equipo TIC team Hospitales de Granada: Rosario Moreno, Teresa Jiménez, María del Mar Ceres.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ By convolution we mean the mathematical operator that transforms the input function F_c and the Kaplan-Meier (KM) function into a third function, which represents quantitatively how the inputs accumulate in the system and progressively leave it after a certain period of time.
- ² In the case we are dealing with we assume that there is only one risk, the risk that a patient leaves a bed free for any reason. Thus, the risk is interpreted as $(1 - P)$ being P the probability corresponding to the Kaplan-Meier distribution.

References

- Bhandari, Sudhir, Tak Amit, Singhal Sanjay, Shukla Jyotsna, Shaktawat Ajit Singh, Gupta Jitendra, Patel Bhoopendra, Kakkar Shivankan, Dube Amitabh, Dia Sunita, and et al. 2020. Patient Flow Dynamics in Hospital Systems During Times of COVID-19: Cox Proportional Hazard Regression Analysis. *Frontiers in Public Health* 8: 820. [CrossRef] [PubMed]
- Calabuig, Jose M., Luis M. García-Raffi, Albert García-Valiente, and Enrique A. Sánchez-Pérez. 2020. Kaplan-Meier type survival curves for COVID-19: A health data based decision-making tool. *Frontiers in Public Health*. [CrossRef]
- Cooper, Ian, Argha Mondal, and Chris G. Antonopoulos. 2020. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals* 139: 110057.
- Cox, David Roxbee, and David Oakes. 1984. *Analysis of Survival Data*. London: Chapman & Hall.
- Dennis, John M., Andrew P. McGovern, Sebastian J. Vollmer, and Bilal A. Mateen. 2021. Improving Survival of Critical Care Patients with Coronavirus Disease 2019 in England: A National Cohort Study, March to June 2020. *Critical Care Medicine* 49: 209–14. [CrossRef] [PubMed]
- Domínguez-Olmedo, Juan L., Alvaro Gragera-Martínez, Jacinto Mata, and Victoria Pachón Alvarez. 2021. Machine Learning Applied to Clinical Laboratory Data in Spain for COVID-19 Outcome Prediction: Model Development and Validation. *Journal of Medical Internet Research* 23: e26211. [CrossRef] [PubMed]
- Eurostat. 2020. Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php/Healthcare_resource_statistics_-_beds (accessed on 1 June 2021).
- García-Basteiro, Alberto L., Carlos Chaccour, Caterina Guinovart, Anna Lluçà, Joe Brew, Antoni Trilla, and Antoni Plasencia. 2020. Monitoring the COVID-19 epidemic in the context of widespread local transmission. *The Lancet Respiratory Medicine* 8: 440–42. [CrossRef]
- García-Vicuña, Daniel Laida Esparza, and Fermin Mallor. 2020. Hospital preparedness in epidemics by using simulation. The case of COVID-19. *medRxiv*. [CrossRef]
- Goscé, Lara, Andrew Phillips, Paula Spinola, Kishi Gupta, and Ibrahim Abubakar. 2020. Modelling SARS-COV2 Spread in London: Approaches to Lift the Lockdown. *Journal of Infection* 81: 260–65. [CrossRef]
- Imani, Mahdi, and Seyede Fatameh Ghoreishi. 2021. Two-Stage Bayesian Optimization for Scalable Inference in State-Space Models. *IEEE Transactions on Neural Networks and Learning Systems*. [CrossRef] [PubMed]
- Kaplan, Edward L., and Paul Meier. 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53: 457–81. [CrossRef]
- Lai, P. C., C. M. Wong, A. J. Hedley, S. V. Lo, P. Y. Leung, J. Kong, and G. M. Leung. 2004. Understanding the Spatial Clustering of Severe Acute Respiratory Syndrome (SARS) in Hong Kong. *Environmental Health Perspectives* 112: 1550–56. [CrossRef] [PubMed]
- Liang, Wenhua, Hengrui Liang, Limin Ou, Binfeng Chen, Ailan Chen, Caichen Li, Yimin Li, Weijie Guan, Ling Sang, Jiatao Lu, and et al. 2020. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Internal Medicine* 180: 1081–89. [CrossRef]
- O’Hagan, Anthony and John W. Stevens. 2004. On estimators of medical costs with censored data. *Journal of Health Economics* 23: 615–25. [CrossRef] [PubMed]
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online: <https://www.R-project.org/> (accessed on 1 June 2021).

- Roimi, Michael, Rom Gutman, Jonathan Somer, Asaf Ben Arie, Ido Calman, Yaron Bar-Lavie, Udi Gelbshtein, Sigal Liverant-Taub, Arnona Ziv, Danny Eytan, and et al. 2021. Development and validation of a machine learning model predicting illness trajectory and hospital utilization of COVID-19 patients: A nationwide study. *Journal of the American Medical Informatics Association* 28: 1188–96. [[CrossRef](#)] [[PubMed](#)]
- Therneau, Terry M., and Patricia M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. New York: Springer. ISBN 0-387-98784-3.
- Tripathi, Arunabh, and Anant Pandey. 2017. Post-hoc comparison in survival analysis: An easy approach. *Journal of Biosciences and Medicines* 5: 112–19. [[CrossRef](#)]
- Vaid, Akhil, Sulaiman Somani, Adam J. Russak, Jessica K. De Freitas, Fayzan F. Chaudhry, Ishan Paranjpe, Kipp W. Johnson, Samuel J. Lee, Riccardo Miotto, Felix Richter, and et al. 2020. Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation. *Journal of Medical Internet Research* 6: e24018. [[CrossRef](#)] [[PubMed](#)]
- Varotsos, Costas, and Vladimir F. Krapivin. 2020. A new model for the spread of COVID-19 and the improvement of safety. *Safety Science* 132: 104962. [[CrossRef](#)] [[PubMed](#)]
- Vincent, Jean-Louis, and Fabio S. Taccone. 2020. Understanding pathways to death in patients with COVID-19. *The Lancet Respiratory Medicine* 8: 430–32. [[CrossRef](#)]
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Zhao, Chenkai, Yueqin Xu, Xu Zhang, Yaping Zhong, Li Long, Wenzhi Zhan, Tingting Xu, Chen Zhan, Yuehan Chen, Jinghai Zhu, and et al. 2020. Public health initiatives from hospitalized patients with COVID-19, China. *Journal of Infection and Public Health* 13: 1229–36. [[CrossRef](#)] [[PubMed](#)]
- Zhou, Tianjan, and Yuan Ji. 2020. Semiparametric Bayesian inference for the transmission dynamics of COVID-19 with a state-space model. *Contemporary Clinical Trials* 97: 106146. [[CrossRef](#)] [[PubMed](#)]