

## RESEARCH ARTICLE

# Estimation of patient flow in hospitals using up-to-date data. Application to bed demand prediction during pandemic waves

Daniel Garcia-Vicuña<sup>1</sup>, Ana López-Cheda<sup>2</sup>, María Amalia Jácome<sup>2</sup>, Fermin Mallor<sup>1\*</sup>

**1** Institute of Smart Cities, Public University of Navarrodpadre, Pamplona, Spain, **2** Departamento de Matemáticas, Research Group MODES, CITIC, Universidade da Coruña, A Coruña, Spain

\* [mallor@unavarra.es](mailto:mallor@unavarra.es)**OPEN ACCESS**

**Citation:** Garcia-Vicuña D, López-Cheda A, Jácome MA, Mallor F (2023) Estimation of patient flow in hospitals using up-to-date data. Application to bed demand prediction during pandemic waves. PLoS ONE 18(2): e0282331. <https://doi.org/10.1371/journal.pone.0282331>

**Editor:** Ayesha Maqbool, National University of Sciences and Technology NUST, PAKISTAN

**Received:** November 2, 2022

**Accepted:** February 13, 2023

**Published:** February 27, 2023

**Copyright:** © 2023 Garcia-Vicuña et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data underlying the results presented in the study are available from website <https://dm.udc.es/modes/es/node/256>, and they can be downloaded from [http://dm.udc.es/data/files/data\\_EMNP.zip](http://dm.udc.es/data/files/data_EMNP.zip).

**Funding:** DGV and FM acknowledge the support by grant PID2020-114031RB-I00 (AEI, FEDER EU) and by the Government of Navarre, 0011-3597-2020-000003 (COVID). ALC was sponsored by the BEATRIZ GALINDO JUNIOR Spanish Grant from MICINN (Ministerio de Ciencia e Innovación) with

## Abstract

Hospital bed demand forecast is a first-order concern for public health action to avoid health-care systems to be overwhelmed. Predictions are usually performed by estimating patients flow, that is, lengths of stay and branching probabilities. In most approaches in the literature, estimations rely on not updated published information or historical data. This may lead to unreliable estimates and biased forecasts during new or non-stationary situations. In this paper, we introduce a flexible adaptive procedure using only near-real-time information. Such method requires handling censored information from patients still in hospital. This approach allows the efficient estimation of the distributions of lengths of stay and probabilities used to represent the patient pathways. This is very relevant at the first stages of a pandemic, when there is much uncertainty and too few patients have completely observed pathways. Furthermore, the performance of the proposed method is assessed in an extensive simulation study in which the patient flow in a hospital during a pandemic wave is modelled. We further discuss the advantages and limitations of the method, as well as potential extensions.

## 1. Introduction

A key aspect in hospital management is planning strategies to avoid healthcare systems to be overwhelmed, which could involve an increment of the number of preventable deaths. During the COVID-19 pandemic, the explosive growth of the number of infected cases in a short period of time has caused massive strain on medical systems. Although a considerable number of restrictions has been adopted in most countries, hospitals worldwide have been overburdened. Most of the deaths were caused by the virulence of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2), but some may have been due to pandemic-associated overloads in hospital capacity [1–4].

The estimation of the hospital ward and Intensive Care Units (ICU) beds' demand is critical for making wise decisions about clinical operations and resource allocations. Menon et al. [5] expose the importance of estimating the critical care bed capacity, as well as developing an appropriate contingency planning. Specifically, they modelled the demand for critical care

code BGP18/00154. ALC and MAJ acknowledge partial support by the MICINN Grant PID2020-113578RB-I00 and partial support of Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2020-14). ALC and MJ wish to acknowledge the support received from the Centro de Investigación de Galicia "CITIC", funded by Xunta de Galicia and the European Union European Regional Development Fund (ERDF)-Galicia 2014-2020 Program, by grant ED431G 2019/01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

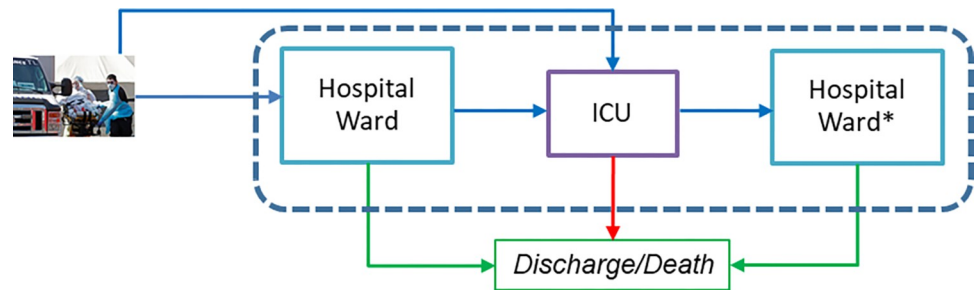
beds in England using a range of attack rates and pandemic durations. More recently, Gitto et al. [6] highlight the importance of having a straightforward and data-driven approach which provides accurate predictions of hospital bed demand. In the face of the COVID19 pandemic, a wide variety of recent studies is related to estimating the capacity of hospital and ICU beds around the world. Litton et al. [7] assessed the capacity of ICU beds in Australia, and they report that intensive care bed capacity could be near tripled in response to the expected increase in demand caused by COVID19. Besides, Barasa et al. [8] evaluate the capacity of the Kenyan health system in terms of general hospital and ICU beds. In Europe, Peña and Espinosa [9], Deschepper et al. [10], López-Cheda et al. [11], and Garcia-Vicuña et al. [12], among others, provide different tools for making predictions of the required number of beds in hospital wards and ICUs. This is essential to avoid important ethical dilemmas related to patient triage [13, 14].

Considering all this, it is vital to register not only the updated number of available beds, but also to forecast hospital bed demand. To develop these forecasts, Susceptible, Infected, Recovered (SIR) models [15, 16] or agent-based models (ABMs) [17, 18] have become common tools for estimating demand for hospital beds during the COVID-19 pandemic. The estimation of the number of hospitalized patients is the first step to forecast hospital bed demand in SIR and ABM models. Nonetheless, it is equally important to estimate how the trend of inpatients will be in the near future. This estimation is based on the distribution of the lengths of stay (LoS) of inpatients in hospital ward or ICU, as well as the probabilities of being transferred to the hospital ward or ICU. Discrete Event Simulation (DES) models are being used increasingly in health-care services for the dynamics of the inpatients [19–21]. They assume predetermined parametric models for the distribution of LoS in hospital ward and ICU. DES methods provide reliable and robust estimates, enabling to manage hospital resources in the most efficient way, only if the assumed models conform to the real trajectory of the inpatients in the hospital facilities. Consequently, it is crucial to obtain accurate estimations used in the simulation models to obtain solid forecasts which would support healthcare managers in optimal resource planning, especially in times of pandemics when healthcare resources are scarce.

In the literature, model parameters for the estimation of the lengths of stay are usually derived from published data [15] and using the health system's historical data [22]. This results in a non-dynamic static model. Nonetheless, the course of a pandemic is a non-stationary situation, in which hospitalization parameters may vary between different waves and places, and evolve over time. Integration of near-real-time hospital occupancy data into the model can have a large impact on improving forecast accuracy [23].

Hospital electronic health record systems provide patient-level information that allows knowing both the pathway of each released patient, and their current location (ward or ICU) if they have not been discharged. Each patient arriving at the hospital can be admitted to the hospital ward or directly to the ICU. Besides, those patients admitted to the wards may worsen their health status and require the transfer to the ICU. From both facilities, patients can die, so they abandon the system, or they can be discharged after improving their health status. In the last situation, patients in the ICU would be transferred to the hospital ward (we use the symbol \* to represent that those patients have been in the ICU before) until they get over the disease (see Fig 1).

In this paper, we consider the problem of estimating the distribution of variables associated with the pathway and LoS of patients in hospital ward and ICU dynamically. In contrast to standard analysis where data are analyzed after the end-of-study, in this application, the end-of-study is a moving target. We propose to estimate them along the time by using all available data collected during a moving time window, from the beginning of the pandemic the first infected patient was admitted to hospital to day  $t$  after the pandemic started.



**Fig 1. Representation of patient flow in the health system.** The symbol \* represents that those patients have been in the ICU before.

<https://doi.org/10.1371/journal.pone.0282331.g001>

The major challenge of the proposed methodology is how to handle the information from patients still hospitalized, since only the pathway of the discharged patients is fully known. This lack of complete information of the inpatients is due to not only censorship in the observed LoS but also to the fact that it is unknown which event will be observed in the future. The main contribution of this paper is the introduction of some new approaches which deal with this challenge.

The objectives in this work are twofold: (a) to propose two competitive methods to estimate the probability distributions included in the patient pathway, which take advantage of the incomplete information from patients still in hospital at the time of the estimation; (b) to compare these methods with alternatives that dismiss the valuable information of these inpatients. The performance of the proposed and alternative estimators is assessed in a simulation study. Furthermore, using the ICU bed prediction method in [12], the efficiency of these predictions on the accuracy of the statistical estimators is evaluated.

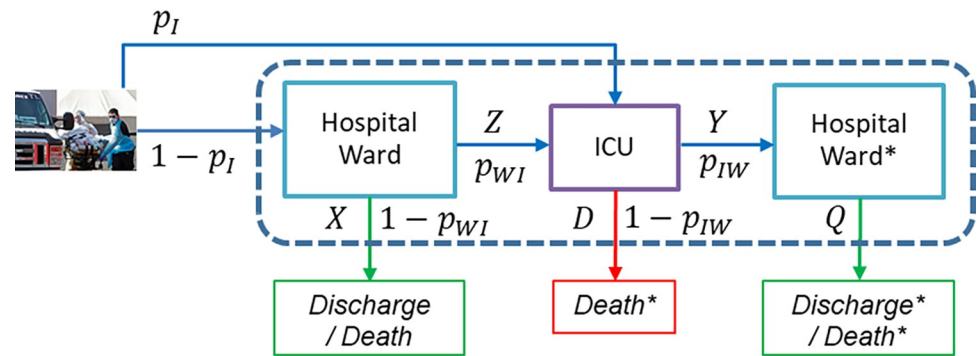
The rest of the paper is organized as follows. Section 2 introduces the new methodology and describes the notation. Section 3 presents the design of experiments for the simulations. The results obtained in the two simulation studies are included in Section 4. Finally, Section 5 ends the paper with the conclusions of this work.

## 2. Methodology

### 2.1. The estimation problem

We consider the problem of forecasting hospital bed demand by estimating the distribution of the LoS and probabilities associated with the pathway of patients hospitalized during a pandemic wave. Because LoS probability distributions and branching probabilities may vary between different waves and between different places, we propose a method to estimate them that uses all data collected from the time the first infected patient was admitted, until the present time. Patients who have already abandoned the hospital due to discharge or death provide complete information for the estimation of LoS probability distributions and branching probabilities, while patients who are still hospitalized provide censored information that may not even be known to which variables are referred to, as we explain below. At the beginning of a pandemic wave, there are few patients and most of them are still hospitalized, but their valuable information should not be disregarded by the statistical estimators. Fig 2 shows the same patient flow as Fig 1, including the LoS probability distributions and the branching probabilities that compose the patient pathway through the hospital.

For a patient who has been hospitalized in hospital ward, it is unknown whether he or she will be finally admitted to the ICU or not, so it is unknown whether the observed value of the



**Fig 2. Representation of patient flow in the health system showing the LoS variables and branching probabilities.**  $Z$ : time in hospital ward until admission to the ICU,  $X$ : time in hospital ward until discharge or death,  $Y$ : time in the ICU before being transferred to hospital ward,  $D$ : time in the ICU until death, and  $Q$ : time in the hospital ward after discharge from the ICU and the branching probabilities  $p_I$ : probability of direct admission to the ICU,  $p_{WI}$ : probability of admission to the ICU from the ward, and  $p_{IW}$ : probability of going to hospital ward from ICU. The symbol \* represents that those patients have been in the ICU before.

<https://doi.org/10.1371/journal.pone.0282331.g002>

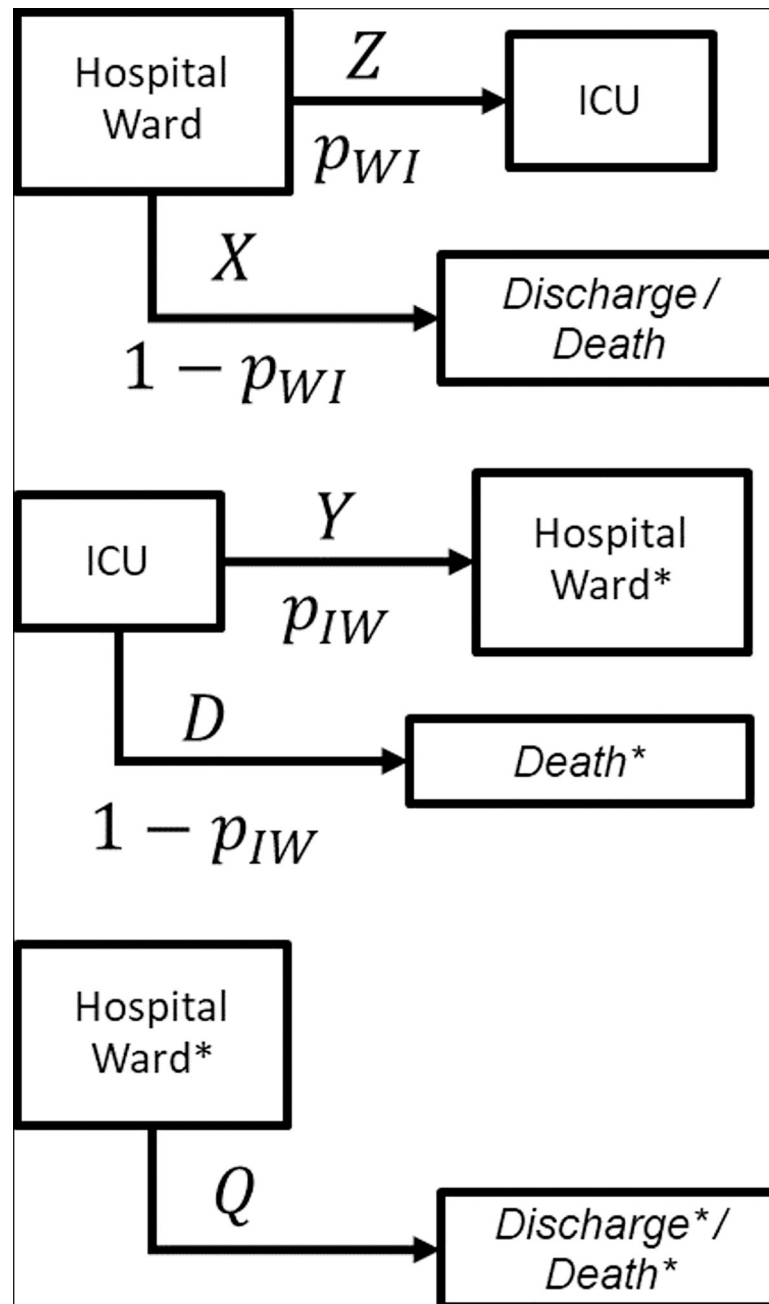
LoS in hospital ward is a censored observation for the variable  $Z$ , “time in hospital ward until admission to the ICU”, or for the variable  $X$ , “time in hospital ward until discharge or death without ICU admission”. In this section, we propose an estimation method for the probability distributions of these variables  $Z$  and  $X$ , as well as the probability of admission to the ICU from the ward,  $p_{WI}$ , that uses the information of all patients admitted to the hospital at the present time (Fig 3, top).

Observe that the same estimation methodology can be applied to the estimation of the probability distributions of  $Y$ , “time in the ICU before being transferred to hospital ward”, and  $D$ , “time in the ICU until death”, and  $p_{IW}$ , the probability of discharge to hospital ward (Fig 3, center). In this case, it is unknown whether a patient who is admitted to the ICU will evolve favourably until being transferred to the hospital ward or whether he or she will die in the ICU. Therefore, the observed LoS of these patients still in the ICU is censored, and it is unknown if it is a censored observation of  $Y$  or  $D$ . Finally, patients discharged from the ICU and still admitted to the hospital ward provide censored data for the variable  $Q$  “time in the hospital ward after being discharged from the ICU” (Fig 3, bottom). The estimation of the variable  $Q$  can be obtained with traditional methods that deal with censoring.

From this point onwards, the introduced notation and methods correspond to the estimation of the distribution of variables  $Z$  and  $X$  and probability  $p_{WI}$  (Fig 3, top) and refer to patients admitted to hospital ward, so patients admitted directly to ICU are not considered. Same methods with similar notation must be used with patients admitted to ICU for the estimation of the distribution of variables  $Y$  and  $D$  and probability  $p_{IW}$  (Fig 3, center). The estimation of the distribution of variables  $Q$  (Fig 3, bottom) can be performed with classical methods in survival analysis.

Time  $t = 0$  is set as the time the first patient is admitted to the hospital. At a fixed time  $t$ , each of the  $n(t)$  patients can be in one of the following sets: (a) ICU set: patients who have required ICU at some point, regardless if they are still in ICU, returned to hospital ward (Hospital Ward\* in Fig 2) or discharged (Discharge\*/Death\* in Fig 2); (b) HW set: patients without ICU admission who still are in the hospital ward (Hospital Ward in Fig 2); and (c) DIS set: patients without ICU admission already discharged (Discharge/Death in Fig 2).

For each patient  $i$ , with  $i = 1, \dots, n(t)$ , admitted to the hospital before time  $t$ , we define a vector  $\mathbf{u}_i(t) = [t_{HAi}, t_{HDi}, t_{IAi}, t_{IDi}]$  that contains four times:  $t_{HAi}$  the time of admission to hospital



**Fig 3. Flow diagram of patients in three situations.** Patients admitted to the hospital (top), patients admitted to the ICU (center), and patients discharged from the ICU (bottom).

<https://doi.org/10.1371/journal.pone.0282331.g003>

ward,  $t_{HD_i}$  the time of discharge from hospital ward,  $t_{IA_i}$  the time of admission to ICU, and  $t_{ID_i}$  the time of discharge from ICU. At a fixed time  $t$ , the patient  $i$  can be still in the hospital ward or in the ICU, so in these cases  $t_{HD_i} = t$  and  $t_{ID_i} = t$  respectively. Besides this, some of the times in  $\mathbf{u}_i(t)$  might remain unknown. For example, if the patient  $i$  was not admitted to ICU at time  $t$ , then  $t_{IA_i}$  and  $t_{ID_i}$  are unknown and they will be denoted as  $\emptyset$ .

The times in vector  $\mathbf{u}_i(t)$  enable the patients to be classified into the aforementioned sets:  
 (a) ICU set: patients with ICU admission divided into three subsets, those back to the hospital

ward from ICU with  $\mathbf{u}_i(t) = [t_{HA_i}, t, t_{IA_i}, t_{ID_i}]$ , those still in ICU with  $\mathbf{u}_i(t) = [t_{HA_i}, t, t_{IA_i}, t]$ , and with  $\mathbf{u}_i(t) = [t_{HA_i}, t_{HD_i}, t_{IA_i}, t_{ID_i}]$  those who have died in the ICU ( $t_{HD_i} = t_{ID_i}$ ) and those who have already been discharge from hospital ward ( $t_{HD_i} > t_{ID_i}$ ); (b) HW set: patients in hospital ward without ICU admission with admission dates  $\mathbf{u}_i(t) = [t_{HA_i}, t, \emptyset, \emptyset]$ ; and (c) DIS set: discharged patients who did not required ICU with  $\mathbf{u}_i(t) = [t_{HA_i}, t_{HD_i}, \emptyset, \emptyset]$ .

Let us define the indicator of the event ‘admission to ICU’, given by  $\delta_i(t) = 1$  if  $t_{IA_i}$  is known at time  $t$  (ICU set) and  $\delta_i(t) = 0$  otherwise (HW and DIS sets). Similarly, let us denote  $v_i(t)$  the indicator which reveals if the patient has been discharged directly from hospital ward or died at a time before  $t$ , so ICU admission will never be required. In other words,  $v_i(t) = 1$  if the patient belongs to DIS set ( $(t_{IA_i}, t_{ID_i}) = (\emptyset, \emptyset)$  and  $t_{HD_i}$  is known), and  $v_i(t) = 0$  otherwise (HW and ICU sets). We consider the trivariate variable  $O = (T, \delta, v)$ , where  $T$ , a variable related to the observed length of stay in hospital ward, may take the following values:

$$\begin{aligned} t_i &= t_{HD_i} - t_{HA_i} && \text{when } \delta_i(t) = 0 \text{ and } v_i(t) = 1 \text{ (DIS set)} \\ t_i &= t_{IA_i} - t_{HA_i} && \text{when } \delta_i(t) = 1 \text{ and } v_i(t) = 0 \text{ (ICU set)} \\ t_i &= t - t_{HA_i} && \text{when } \delta_i(t) = 0 \text{ and } v_i(t) = 0 \text{ (HW set)} \end{aligned} \tag{1}$$

Observe that, at a time  $t$ , value  $t_i$  of patients in DIS set provides an observation of variable  $X$ , value  $t_i$  of patients in ICU set provides an observation of variable  $Z$ , and value  $t_i$  of patients in HW set provides a censored observation for either variable  $X$  or variable  $Z$ .

For the rest of the paper, we consider the following notation:

- $\hat{p}_{WI}(t)$  the estimation of the probability  $p_{WI}$  at time  $t$ .
- $F_X(x), F_Z(z)$  the cumulative distribution function of variables  $X$  and  $Z$ , respectively.

### 2.2. Nonparametric methods using survival analysis

Survival analysis refers to the statistical methods used to analyze time-to-event data in the presence of censored observations. Note that the information related to patients in states ICU and DIS is complete for the estimation of the distributions of  $Z$  and  $X$ , respectively. However, for patients still in HW set at time  $t$ , we have right censored data since it is unknown whether they will require ICU or not, nor the final duration of the stay in hospital ward. It should be noted that, in the first weeks of the pandemic, HW set is expected to include most patients. All patients in hospital ward at time  $t$  provide valuable information for the estimation of  $p_{WI}$  and the distribution of  $X$  and  $Z$ . It is therefore essential to carry out a methodology which incorporates all the information contained in these censored observations.

**NP method.** Nonparametric (NP) methods for estimation have specific advantages such as flexibility and ease of computation, and are a popular choice for analysing survival data, such as the Kaplan-Meier estimator to estimate the survival function or the Nelson-Aalen estimator for the cumulative hazard function.

In classical survival analysis, it is assumed that all the individuals will experience the event of interest. That is the case when estimating the distribution of the variable  $Q$  (Fig 3, bottom) as the event of interest is ‘Discharge\*/Death\*’, and this model assumes that all patients in hospital ward coming from ICU will never require ICU again and leave hospital eventually. Therefore, classical nonparametric methods in survival analysis, such as Kaplan-Meier estimator, can be used to estimate the distribution of  $Q$ . However, that assumption does not hold for the estimation of the distribution of  $Z$ , ‘time in hospital ward until admission to the ICU’, since the event of interest is ‘ICU admission’ and not all the inpatients will require entering ICU. The same situation holds when estimating the distribution of  $X$ , ‘time in hospital ward until

discharge without ICU admission”, as not all the patients in hospital ward will be discharged without ICU admission. The individuals who are free of experiencing the event are called *long-term survivors*, or simply *cured subjects*. Note that here a *cured* individual refers to a subject who will not experience the event of interest, and this is not necessarily related to be *cured* in medical terms.

Mixture cure models (MCM) account for this situation since they consider that the population is a mixture of two groups of patients, the susceptible ones to the event of interest and the cured individuals (see [24–28] among others). The observed time of all cured individuals is censored, as the event will not occur and therefore it is never observed. Traditional MCM assume that cured individuals are unidentifiable, as censoring prevents from distinguishing which censored subjects are cured and which ones will experience the event in the end. Nonetheless, that is not the case in our context. MCM when the cured subjects are randomly identified addresses this situation, and it has received much attention in recent years (see [29–32]).

When estimating the distribution of  $Z$ , “time in hospital ward until admission to the ICU”, all patients admitted to ICU ( $\delta_i(t) = 1$ ) are uncensored while those who have already been discharged from hospital at a time before  $t$  without ICU admission ( $v_i(t) = 1$ ) are cured from the event ‘ICU admission’ as they will never be admitted to ICU in the future. For a fixed time  $t$ , the *NP method* estimates the distribution function of  $Z$ ,  $F_Z(z) = P(Z \leq z)$ , nonparametrically using the estimator in Safari et al. [33] and the observations  $\{(t_i, \delta_i(t), v_i(t)), i = 1, \dots, n(t)\}$

$$\hat{F}_{Z,t}^{NP}(z) = 1 - \frac{\tilde{F}_{Z,t}(t_{n(t)}) - \tilde{F}_{Z,t}(z)}{\tilde{F}_{Z,t}(t_{n(t)})},$$

$$\text{where } \tilde{F}_{Z,t}(z) = 1 - \prod_{i=1}^{n(t)} \left\{ 1 - \frac{\delta_i(t)1(t_i \leq z)}{n(t) - i + 1 + \sum_{j=1}^{i-1} v_j(t)} \right\}$$

where  $t_1 \leq \dots \leq t_{n(t)}$  are the sorted observed times in Eq (1).

Similarly, when estimating the distribution of  $X$ , “time in hospital ward until discharge without ICU admission”, all patients discharged or dead without ICU admission ( $v_i(t) = 1$ ) are uncensored, and those admitted to ICU ( $\delta_i(t) = 1$ ) are cured from the event because they will never experience ‘discharge without ICU admission’. The distribution function of the time in hospital ward until discharge without ICU admission,  $F_X(x) = P(X \leq x)$  is estimated nonparametrically for a fixed time  $t$ , as follows;

$$\hat{F}_{X,t}^{NP}(x) = 1 - \frac{\tilde{F}_{X,t}(t_{n(t)}) - \tilde{F}_{X,t}(x)}{\tilde{F}_{X,t}(t_{n(t)})},$$

$$\text{where } \tilde{F}_{X,t}(x) = 1 - \prod_{i=1}^{n(t)} \left\{ 1 - \frac{v_i(t)1(t_i \leq x)}{n(t) - i + 1 + \sum_{j=1}^{i-1} \delta_j(t)} \right\}$$

Finally, the probability of requiring ICU from ward is estimated for a fixed time  $t$  using the nonparametric estimator in Safari [32]:

$$\hat{p}_{WI}^{NP}(t) = 1 - \prod_{i=1}^{n(t)} \left\{ 1 - \frac{\delta_i(t)}{n(t) - i + 1 + \sum_{j=1}^{i-1} v_j(t)} \right\}. \tag{2}$$

See [32, 33] for the consistency and order of convergence of estimators (1) and (2).

There are some nonparametric alternatives to estimate the probability  $p_{WI}$ , such as imputation methods [34] or a competing risks model [29]. Note that the first method is biased under the common assumption of independent censoring. Besides, the main disadvantage of the second approach is that, if the patient with the largest observed time is still in hospital ward and did not require ICU (HW set), then the estimator of  $p_{WI}$  is not unique, and only upper and lower bounds are provided [32].

Standard and routinely-implemented cure model methodologies, such as the mixture cure model based on the proportional hazards assumption [35–39] or the accelerated failure time model [40–44] are not discussed here since covariates are not considered in the model.

### 2.3. Parametric methods based on the EM algorithm

We denote as  $\mathbf{o}(t) = (o_1(t), \dots, o_i(t), \dots, o_{n(t)}(t))$  the realization of variable  $O$  in the  $n(t)$  patients admitted to the hospital since the beginning of the pandemic wave. We have developed an iterative procedure, based on the Expectation-Maximization (EM) algorithm, to estimate the distribution functions of the variables  $X$  and  $Z$  and the probability  $p_{WI}$ . First, an initial estimation of the parameters is carried out by using only the fully-known data, those observations with  $\delta_i(t) + v_i(t) = 1$ , that is, in DIS ( $v_i(t) = 1$ ) or ICU ( $\delta_i(t) = 1$ ) sets. In the main iteration, the estimated parameters are used to update the probability of being admitted to ICU for each patient in HW set. These updated probabilities allow the calculation of a new likelihood function for the parameters, which is maximized to obtain a new estimation of the probability distribution parameters. These two steps (updating ICU admission probabilities and getting and maximizing the new likelihood function) are repeated until the stopping criteria are satisfied.

We consider the following additional notation:

- $\theta_V$  the vector of parameters of the distribution function of a general variable  $V$ .
- $\hat{\theta}_V(t)$  the estimation of the vector of parameters  $\theta_V$  at time  $t$ .
- $F_{\theta_V}(v)$ ,  $f_{\theta_V}(v)$  the distribution and density function of a general variable  $V$  with parameters  $\theta_V$  respectively.
- $L_V(\theta_V|\mathbf{o}(t))$  the likelihood function of sample  $\mathbf{o}(t)$  used to estimate  $\theta_V$ .
- $\hat{\theta}_X^{(k)}(t)$  and  $\hat{\theta}_Z^{(k)}(t)$ : the estimation of vectors  $\theta_X$  and  $\theta_Z$  in the  $k$ -th iteration of the algorithm at time  $t$ .
- $\hat{p}_{WI}^{(k)}(t)$ : the estimation of the probability  $p_{WI}$  in the  $k$ -th iteration of the algorithm at time  $t$ .

The steps of the algorithm are detailed below in the *EM method*.

**EM method.** 1. *Initialization.* We set  $k = 0$  and estimate the parameters  $\theta_X$ ,  $\theta_Z$  and the probability  $p_{WI}$  by using the data in vector  $\mathbf{o}(t)$ :

$$\hat{p}_{WI}^{(0)}(t) = \frac{\sum_{i=1}^{n(t)} \delta_i(t)}{\sum_{i=1}^{n(t)} (\delta_i(t) + v_i(t))}, \tag{3}$$

$$\hat{\theta}_X^{(0)}(t) = \arg \max_{\theta_X} L_X^{(0)}(\theta_X|\mathbf{o}(t)), \text{ where } L_X^{(0)}(\theta_X|\mathbf{o}(t)) = \prod_{i=1}^{n(t)} f_{\theta_X}(t_i)^{v_i(t)}, \tag{4}$$



$$\hat{\theta}_Z^{(0)}(t) = \arg \max_{\theta_Z} L_Z^{(0)}(\theta_Z|\mathbf{o}(t)), \text{ where } L_Z^{(0)}(\theta_Z|\mathbf{o}(t)) = \prod_{i=1}^{n(t)} f_{\theta_Z}(t_i)^{\delta_i(t)}. \tag{5}$$

2. *Repeat until stop criteria are met.* Iteration  $k+1$ . From the  $k$ -th iteration,  $k \geq 0$ , the estimations  $\hat{\theta}_X^{(k)}(t)$ ,  $\hat{\theta}_Z^{(k)}(t)$  and  $\hat{p}_{WI}^{(k)}(t)$  are known. The iteration is divided in two steps: in the first one, the calculation of the expected value of the probability of admission to the ICU of each patient in HW set is carried out, which allows estimating the probability of admission to ICU,  $p_{WI}$ , and the expectation of the likelihood function. The second step computes the estimations of  $\theta_X$  and  $\theta_Z$  by maximizing the likelihood functions in the previous step.

**2.1. Expectation.** For each patient  $i$  in HW set, the probability  $\hat{p}_{WI,i}^{(k+1)}(t)$  of being admitted to ICU is updated as the posterior probability given the time  $t_i$  already spent at the hospital ward:

$$\begin{aligned} \hat{p}_{WI,i}^{(k+1)}(t) &\equiv P(\delta_i(s) = 1, s > t | \hat{p}_{WI}^{(k)}(t), \hat{\theta}_X^{(k)}(t), \hat{\theta}_Z^{(k)}(t)) \\ &= \frac{(1 - F_{\hat{\theta}_Z^{(k)}(t)}(t_i)) \hat{p}_{WI}^{(k)}(t)}{(1 - F_{\hat{\theta}_Z^{(k)}(t)}(t_i)) \hat{p}_{WI}^{(k)}(t) + (1 - F_{\hat{\theta}_X^{(k)}(t)}(t_i))(1 - \hat{p}_{WI}^{(k)}(t))}. \end{aligned}$$

Considering the updated probabilities of being admitted to ICU for each patient in HW set, we estimate the unconditional probability of admission to ICU:

$$\hat{p}_{WI}^{(k+1)}(t) = \frac{1}{n(t)} \sum_{i=1}^{n(t)} [\delta_i(t) + (1 - \delta_i(t))(1 - v_i(t)) \hat{p}_{WI,i}^{(k+1)}(t)]$$

and the likelihood functions of the sample as expected functions:

$$L_X^{(k+1)}(\theta_X|\mathbf{o}(t)) = E[L_X(\theta_X|\mathbf{o}(t))] = \prod_{i=1}^{n(t)} f_{\theta_X}(t_i)^{v_i(t)} \prod_{i=1}^{n(t)} [(1 - F_{\theta_X}(t_i))^{(1-\delta_i(t))(1-v_i(t))} (1 - \hat{p}_{WI,i}^{(k+1)}(t))],$$

$$L_Z^{(k+1)}(\theta_Z|\mathbf{o}(t)) = E[L_Z(\theta_Z|\mathbf{o}(t))] = \prod_{i=1}^{n(t)} f_{\theta_Z}(t_i)^{\delta_i(t)} \prod_{i=1}^{n(t)} [(1 - F_{\theta_Z}(t_i))^{(1-\delta_i(t))(1-v_i(t))} \hat{p}_{WI,i}^{(k+1)}(t)].$$

**2.2. Maximization.** The likelihood functions are maximized to find the parameter estimation:

$$\hat{\theta}_X^{(k+1)}(t) = \arg \max_{\theta_X} (L_X^{(k+1)}(\theta_X|\mathbf{o}(t))),$$

$$\hat{\theta}_Z^{(k+1)}(t) = \arg \max_{\theta_Z} (L_Z^{(k+1)}(\theta_Z|\mathbf{o}(t))).$$

3. *Stop criteria.* Let  $\epsilon_X$ ,  $\epsilon_Z$ , and  $\epsilon_{p_{WI}}$  be some fixed values that control the accuracy of the iterative calculations. Repeat Step 2 until the sequence of values of the estimated parameters

converges:

$$|\hat{\theta}_X^{(k+1)}(t) - \hat{\theta}_X^{(k)}(t)| \leq \epsilon_X,$$

$$|\hat{\theta}_Z^{(k+1)}(t) - \hat{\theta}_Z^{(k)}(t)| \leq \epsilon_Z,$$

$$|\hat{p}_{WI}^{(k+1)}(t) - \hat{p}_{WI}^{(k)}(t)| \leq \epsilon_{p_{WI}}.$$

The final estimates are  $\hat{p}_{WI}^{EM}(t) = \hat{p}_{WI}^{(k+1)}(t)$ ,  $\hat{\theta}_X^{EM}(t) = \hat{\theta}_X^{(k+1)}(t)$ , and  $\hat{\theta}_Z^{EM}(t) = \hat{\theta}_Z^{(k+1)}(t)$ , and then  $\hat{F}_{X,t}^{EM}(x) = F_{\hat{\theta}_X^{EM}(t)}(x)$  and  $\hat{F}_{Z,t}^{EM}(z) = F_{\hat{\theta}_Z^{EM}(t)}(z)$ .

**EMNP method.** Different estimators considered for the initialization step ( $k = 0$ ) result in a different method for the final estimators of the parameters  $\theta_X$  and  $\theta_Z$  and the probability of admission to ICU from ward,  $p_{WI}$ . The EMNP method combines both the EM algorithm and the nonparametric approach. This integrated approach is intended to consider the flexibility of the NP method and the efficiency of the EM algorithm. Specifically, the probability  $p_{WI}$  is initially estimated using the NP estimator in Eq (2), that is,  $\hat{p}_{WI}^{(0)}(t) = \hat{p}_{WI}^{NP}(t)$ . The initial values for the parameters  $\theta_X$  and  $\theta_Z$  are those in Eqs (4)–(5). The other steps in the EM method remain unchanged. Finally, the EMNP estimators are  $\hat{p}_{WI}^{EMNP}(t)$ ,  $\hat{\theta}_X^{EMNP}(t)$ , and  $\hat{\theta}_Z^{EMNP}(t)$ , given by the EM algorithm, and then  $\hat{F}_{X,t}^{EMNP}(x) = F_{\hat{\theta}_X^{EMNP}(t)}(x)$  and  $\hat{F}_{Z,t}^{EMNP}(z) = F_{\hat{\theta}_Z^{EMNP}(t)}(z)$ .

### 2.4. Naïve alternative methods

We present three naïve alternative methods for estimating the distribution parameters  $\theta_X$  and  $\theta_Z$  and the probability  $p_{WI}$ . They do not consider censored observations, that is, patients in HW set by time  $t$ . Although it results, at the beginning of the pandemic wave, in possibly biased estimates, this negative effect tends to fade away at advanced stages of the pandemic, as the number of censored observations decreases. The first method only uses complete information (CI method), that is, it only includes those patients admitted to the hospital whose values of the vector  $u_i(t)$  are completely known. In an attempt to increase the sample size, we define two estimation procedures that somehow include the censored observations given by patients still in hospital ward who have not required ICU yet (HW set). On the one hand, by assuming that all these patients in HW set will not require ICU in the future (I method). This method is expected to be biased as long as the assumption is not true. The last estimation procedure (IP method) reduces estimation bias by considering all the patients with complete information and some of the patients currently admitted in the hospital ward with unknown entrance to the ICU.

**CI method.** Only patients who entered ICU or have been discharged are included in the estimations, so patients in HW set by time  $t$  are dismissed. This results in  $\hat{p}_{WI}^{CI}(t) = \hat{p}_{WI}^{(0)}(t)$ ,  $\hat{\theta}_X^{CI}(t) = \hat{\theta}_X^{(0)}(t)$ , and  $\hat{\theta}_Z^{CI}(t) = \hat{\theta}_Z^{(0)}(t)$ , the initial estimations for the EM method in Eqs (3)–(5). So  $\hat{F}_{X,t}^{CI}(x) = F_{\hat{\theta}_X^{CI}(t)}(x)$  and  $\hat{F}_{Z,t}^{CI}(z) = F_{\hat{\theta}_Z^{CI}(t)}(z)$ .

This approach of omitting the observations in HW set raises several issues. First, it leads to the loss of valuable information. Second, the result of ignoring these censored observations is an underestimation of the distributions of  $Z$  and  $X$ , since only the patients who have been quickly discharged or transferred to ICU will be considered in the procedure. This underestimation, of considerable magnitude at early stages of the pandemic given the large number of censored observations in the data, will ease over time as the proportion of censored observations decreases.

**I method.** The estimation of parameter  $\theta_Z$  for the distribution of  $Z$ , the length of stay in hospital ward until ICU admission, is the same as in the CI method,  $\hat{\theta}_Z^I(t) = \hat{\theta}_Z^{(0)}(t)$  in Eq (5) ( $\hat{F}_{Z,t}^I(z) = F_{\hat{\theta}_Z^I(t)}(z)$ ). As for the estimation of  $\theta_X$  and the probability  $p_{WI}$ , this method seeks to include the censored information of the patients in HW set. The final event of these inpatients remains unknown by time  $t$ , but most patients still in HW set are expected not to require ICU in the future. This method oversimplifies the model by assuming that none of these patients in HW set will be admitted to the ICU. Therefore, the probability  $p_{WI}$  is estimated empirically at time  $t$  as follows:

$$\hat{p}_{WI}^I(t) = \frac{\sum_{i=1}^{n(t)} \delta_i(t)}{n(t)} \tag{6}$$

Regarding the estimation of parameter  $\theta_X$  for the distribution of the length of stay in hospital ward until discharge,  $X$ , all the observed LoS of the patients in HW set ( $(\delta_i(t), v_i(t)) = (0,0)$ ) by time  $t$ ,  $t_i = t - t_{HAi}$ , are considered as censored observations of variable  $X$ :

$$\begin{aligned} \hat{\theta}_X^I(t) &= \arg \max_{\theta_X} L_X^I(\theta_X | \mathbf{o}(t)), \text{ where } L_X^I(\theta_X | \mathbf{o}(t)) \\ &= \prod_{i=1}^{n(t)} f_{\theta_X}(t_i)^{v_i(t)} \prod_{i=1}^{n(t)} (1 - F_{\theta_X}(t_i))^{(1-\delta_i(t))(1-v_i(t))} \end{aligned} \tag{7}$$

Therefore,  $\hat{F}_{X,t}^I(x) = F_{\hat{\theta}_X^I(t)}(x)$ . Note that these I estimators are biased. In fact, both the probability of being transferred to the ICU from ward ( $p_{WI}$ ) and the time until transfer to the ICU ( $Z$ ) are underestimated. Observe that some patients in HW set will require admission to the ICU so their observed LoS  $t_i$ , used in Eq (7) as censored observations of  $X$ , are actually censored values for variable  $Z$ . This yields biased estimates of the parameters  $\theta_X$  and  $\theta_Z$ . In turn,  $\hat{p}_{WI}^I(t)$  underestimates the probability of admission to ICU from ward as only patients in ICU set are included in the numerator of Eq (6), while some patients in the HW set will be admitted to ICU as well. Nonetheless, the estimations will improve as pandemic advances, and ICU and DIS sets grow in size with respect to HW set.

**IP method.** In order to reduce the bias in the *I method* resulting from dismissing the patients in HW set, a subset of the hospitalized inpatients is included in the estimation procedure, those who are more likely to have complete information in their pathways in the short term. This approach does not consider the patients admitted to HW, ICU and DIS sets in the last  $d$  days, where  $d$  is calculated as the percentile  $P$  of the probability distribution of  $Z$ , estimated at time  $t$  considering all patients in ICU set:

$$d(t) = F_{\hat{\theta}_Z^{CI}(t)}^{-1}(P). \tag{8}$$

The estimation procedure resembles the *I method* where the datasets HW, ICU and DIS are now replaced with  $HWd = HW \{i | t - t_{HAi} < d(t)\}$ ,  $ICUd = ICU \{i | t - t_{HAi} < d(t)\}$ , and  $DISd = DIS \{i | t - t_{HAi} < d(t)\}$ . The bias is reduced, because the  $HWd$  set now includes patients with a small probability of being transferred to ICU, at the expense of estimating with fewer observations in  $ICUd$  and  $DISd$  sets.

### 3. Simulation studies

Two simulation studies have been carried out to compare the estimation methods presented in Section 2, and to determine their impact on the predictions of bed occupancy obtained using those estimated distributions and probabilities as simulation inputs. In particular, the goal is to

test the performance of the proposed estimation methods in Subsection 2.2 (*NP method*) and Subsection 2.3 (*EM method* and *EMNP method*), and their comparison with the methods in Subsection 2.4 (*CI method*, *I method* and *IP method*) that dismiss incomplete information. In this section, we describe the simulation model and the experimental design that have been carried out to assess the accuracy of both the estimation of  $p_{WI}$  and  $F_Z(z)$  and the prediction of hospital resources needed to care for all patients, specifically the number of ICU beds required. The results related to the estimation accuracy are shown in Subsection 4.1 and the impact on the precision of the predictions, in Subsection 4.2. All methods and simulations have been programmed using Python 3.7.

This section is organised as follows. We first present the mathematical modelling of hospital dynamics using a DES model in Subsection 3.1. Subsection 3.2 describes how the DES model simulates the patient arrival process in order to generate different pandemic waves. In Subsection 3.3, we explain how to simulate the pathway and LoS for each patient at the hospital. Moreover, in Subsection 3.4, we present how to generate the remaining pathway and LoS of patients that are admitted in the hospital at specific time  $t$  and for those who will arrive in the future. The latter allows different scenarios to be projected into the future based on the hospital's situation at a specific point in time during the pandemic.

### 3.1. The discrete event simulation model

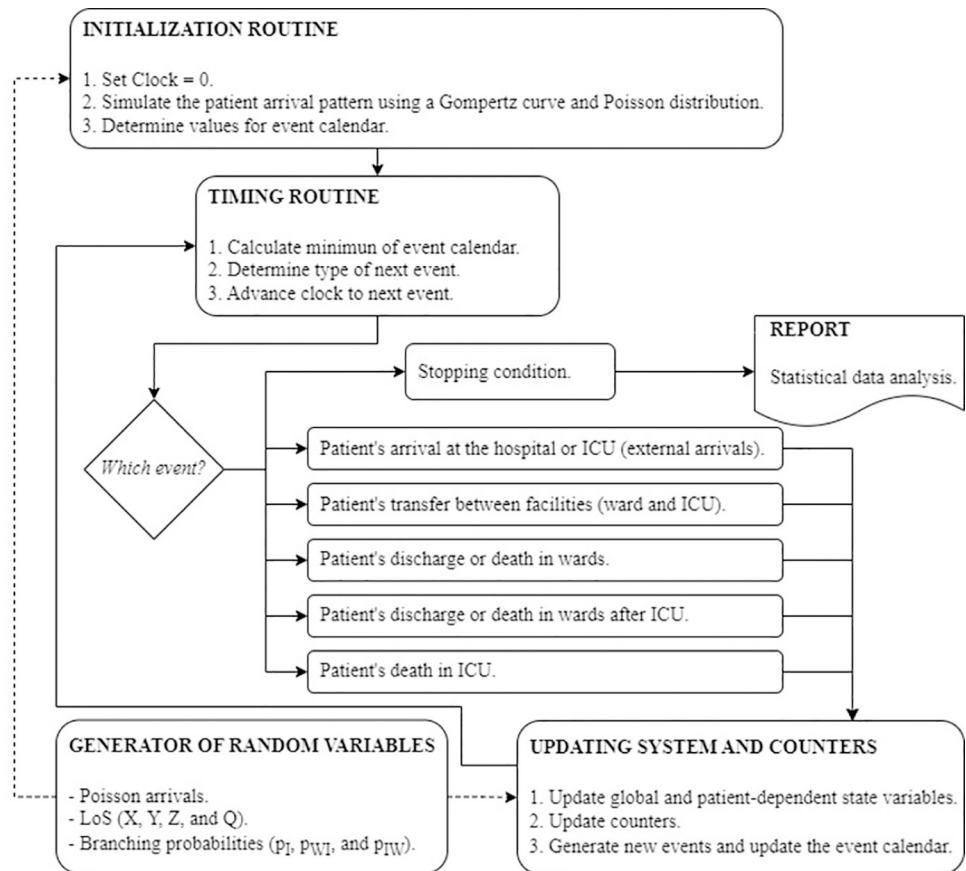
A DES model is developed to assess the accuracy of the estimators. DES models create entities that are transformed by several processes until they exit the modelling system. In our simulation model, the entities are the COVID-19 patients and the processes are the health care received in the hospital ward and/or ICU. In this way, the DES model is able to reproduce the hospital admission of patients during a pandemic wave and the trajectory in the hospital for each patient. The simulation model represents patient flow through the different hospitalization routes; that is, the area enclosed by dashed lines in Fig 2.

The system is described by a set of state variables, which provide at any time a complete representation of the simulated system, and the set of events, which modify the value of the state variables. We consider two global state variables, number of beds occupied by COVID-19 patients in hospital wards and the ICU, and two patient-dependent state variables, the admission place at time  $t$  (ward without a previous stay in ICU, ICU and ward after transferral from ICU) and the time at which patient enters the current admission place.

The events that modify the state variables are the following five: a new patient admission to the hospital, a patient transfer from ward to ICU, a patient discharge in ward, a patient discharge in ward after ICU and a patient discharge in ICU. Fig 4 outlines the DES model of the health system. A complete description of the DES model, and how each state variable is updated as each type of event occurs is presented in [12].

### 3.2. Patient arrival process

Let  $T_{End}$  be the simulation horizon time for the pandemic wave, and  $G(t)$  the cumulated number of hospitalized patients at time  $t$  for  $t = 1, \dots, T_{End}$ . In this study,  $G(t)$  is simulated using Population Growth (PG) models. This methodology provides methods for modelling the number of cumulative positive cases, hospitalizations, and other pandemic variables. Some examples of PG models that have been found in the literature are the Gompertz [45], the Richards [46], the Stannard [47], and the logistic model [48]. Gompertz model shows a better fit to data of daily COVID-19 new cases as well as better predictive capacity than other PG models [12]. Therefore, the arrival of patients at the hospital are generated using the Gompertz model, via the equation proposed by Zwietering et al. [49] who rewrote the original one [45] to ease the



**Fig 4. Discrete event simulation model.** Flow diagram of the main components of the DES model highlighting the five events that modify the two types of state variables.

<https://doi.org/10.1371/journal.pone.0282331.g004>

biological interpretation of its parameter. The arrival curve,  $G(t)$ , is generated with the following Gompertz model:

$$G(t) = 5000 \exp(-\exp(2.0743 - 0.0678t)) \tag{9}$$

The selected curve,  $G(t)$ , in Eq (9) models a cumulative number of 5,000 patients and a duration of 60 days, where duration is defined as the time elapsed from the admission of 5% to 95% of the total number of patients (see Fig 5).

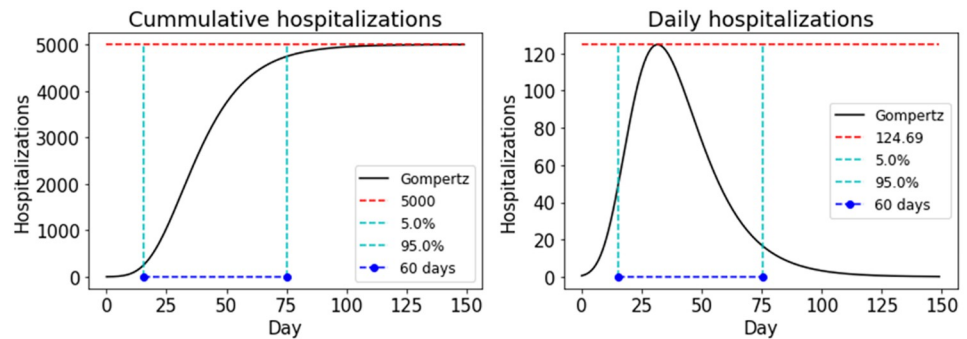
From the Gompertz-type hospitalization curve the expected number of daily hospitalizations is calculated as  $\lambda(t) = G(t) - G(t-1)$ . The number of daily hospitalizations at day  $t$ ,  $H(t)$ , for  $t = 1, \dots, T_{End}$  is simulated from a Poisson distribution with mean  $\lambda(t)$ :

$$P(H(t) = k) = \frac{e^{-\lambda(t)} \lambda(t)^k}{k!}, t = 1, \dots, T_{End} \tag{10}$$

Therefore, in each of the simulated scenarios, patient arrival pattern is different.

### 3.3. Flow of patients in the hospital

For each patient arriving at the hospital, a pathway is simulated reproducing the patient pathway outlined in Fig 1. Each patient can be admitted to the hospital ward or directly to the ICU. The probability of direct admission to ICU upon arrival is  $p_I = 0.028$ . Besides, those patients



**Fig 5. Gompertz curve generated to model a pandemic scenario.** This scenario has 5,000 cumulative hospitalizations and 60 days of duration. The left-hand side shows cumulative hospitalizations for the selected scenario while the right-hand side shows daily ones, that is, the derivative curve.

<https://doi.org/10.1371/journal.pone.0282331.g005>

admitted to the wards may worsen their health status and require the transfer to the ICU. The probability of a patient initially admitted to a ward requiring transfer to ICU was set  $p_{WI} = 0.088$ . From both hospital ward and ICU, patients can die, so they abandon the system, or they can be discharged after improving their health status. In the last situation, patients in the ICU would be transferred to the hospital ward until they get over the disease. The probability of a patient being transferred from ICU to hospital ward is  $p_{IW} = 0.816$ .

In the simulation experiments, probability distributions for the LoS are assumed to be Weibull  $W(\alpha, \beta)$ , where  $\alpha$  is the scale parameter and  $\beta$  is the shape parameter, and time is measured in days: LoS in the hospital ward of a patient not needing ICU, variable  $X$ , is distributed as  $W(10.2, 1.25)$ , the time spent by a patient in the hospital ward before transfer to the ICU, variable  $Z$ , is distributed as  $W(4.1, 1.15)$ . In addition, the LoS of a patient in the ICU, both variables  $Y$  and  $D$ , are distributed as  $W(17.3, 1.1)$ . Finally, the LoS of a patient in hospital ward after being discharged from ICU, variable  $Q$ , is distributed as  $W(11.85, 1.4)$ .

All these selected values are estimations based on real patients during a COVID-19 pandemic wave [12].

### 3.4. Simulating future hospital patient-flow

At a specific day of a pandemic wave, prediction of the resources needed for patients care, such as ICU beds, might be of interest. In this study, the simulated pandemic wave is referred to as Reference Scenario (RS), and the specific day is called Simulation Starting Point (SSP).

For prediction of bed occupancy at time  $t$ , the future pathways for the inpatients must be simulated by estimating all the distributions and probabilities involved in the patient's pathway (see Fig 2) with the information available up to that specific day  $t$ .

Accurate prediction of ICU bed occupancy relies on the efficient estimates of all the probabilities and distributions in patients flow (see Fig 2). The goal of this study is limited to assess the influence of the estimation of  $p_{WI}$  and the distribution of variable  $Z$  in predicting ICU bed occupancy. For this reason, and in order to avoid extra variability into the simulation study so the differences in the estimations of bed occupancy and prediction capability are only assigned to the differences in the estimation of  $p_{WI}$  and the distribution of variable  $Z$ , patients pathways are simulated using the estimated values of  $p_{WI}$  and the distribution of variable  $Z$  using the methods of Section 2, the other probabilities and distributions involved in patients pathways (see Fig 2) are considered as known, and given by the models in Subsection 3.3.

Future pathways must be simulated for the three possible types of patients: patients in hospital ward at SSP day, patients currently admitted to the ICU at SSP day, and future patients admitted in the coming days.

A hospital pathway is simulated for each patient  $i$  currently in hospital ward for  $t_i$  days as follows. The patient is admitted to the ICU with probability

$$\hat{p}_{WI,i}(t) = \frac{(1 - \hat{F}_{Z,t}(t_i))\hat{p}_{WI}(t)}{(1 - \hat{F}_{Z,t}(t_i))\hat{p}_{WI}(t) + (1 - F_X(t_i))(1 - \hat{p}_{WI}(t))},$$

where  $\hat{p}_{WI}(t)$  and the function  $\hat{F}_{Z,t}(z)$  are the estimations computed with the methods in Section 2. If the patient requires ICU admission, then the simulated time in hospital ward left to ICU admission is  $z_i - t_i$ , where  $z_i$  is generated from the conditional distribution  $Z|Z > t_i$ , that is,  $(\hat{F}_{Z,t}(z_i) - \hat{F}_{Z,t}(t_i))/(1 - \hat{F}_{Z,t}(t_i))$ . If the patient  $i$  does not require ICU care, the hospital discharge will occur after a time  $x_i - t_i$ , where  $x_i$  is sampled from the conditional distribution  $X|X > t_i$ , that is,  $(F_X(x_i) - F_X(t_i))/(1 - F_X(t_i))$ .

The pathway of patients in ICU at SSP day is generated as in Subsection 3.3. For the simulation of future inpatients, the arrival curve  $G(t)$  must be previously estimated. In this study the patient arrival process is the same as the one used for the simulated pandemic wave RS and given by Eqs (9) and (10), to avoid introducing more variability into the simulation study. Once the future patient arrives, the pathway is simulated as in Subsection 3.3 using the models therein, except  $p_{WI}$  and the probability distribution of variable  $Z$  which are estimated with the methods in Section 2.

This simulation can be performed using different days as SSP. Subsection 4.2 shows the results obtained with four different days. It can be observed how the predictions change as more data is available for the estimations.

## 4. Results

This section presents the results obtained in the two simulation studies. First, in Subsection 4.1, we show the accuracy of the estimators as the pandemic progresses. Second, in Subsection 4.2, we include the impact of the estimates of  $p_{WI}$  and the distribution of variable  $Z$  on the simulation output. Specifically, we study the accuracy obtained in predicting the number of occupied ICU beds during a generated pandemic scenario.

In the use of the *IP method*, a value for the P percentile is needed in the computation of  $d$  in Eq (8). The two following percentiles have been chosen for the estimations: 50<sup>th</sup> percentile (*IQ2 method*) and 75<sup>th</sup> percentile (*IQ3 method*). In the use of the *EM* based procedures in Subsection 2.3 (*EM method* and *EMNP method*), we set  $\varepsilon_X = \varepsilon_Z = \varepsilon_{p_{WI}} = 0.01$  as stop criteria of the EM algorithm.

### 4.1. Estimation accuracy

To assess the accuracy of the estimations we generated 100 different pandemic waves, with  $T_{End} = 150$ , according to Subsection 3.2 and Subsection 3.3. In each scenario and for each day  $t = 1, \dots, 80$ , we estimated the probability of admission to ICU from hospital ward  $p_{WI}$ , and the distribution of variable  $Z$ , time to transfer to ICU from wards, with the information provided by the corresponding  $n(t)$  patients during the first 80 days. In order to compare the methods to estimate  $F_Z(z)$ , we computed  $\hat{\mu}_Z(t)$  the mean of the estimated distributions  $\hat{F}_{Z,t}(z)$  at time  $t$  and compared it to the real mean,  $\mu_Z = E(Z) = 3.9$  days. Besides, we approximated the integrated squared error (ISE) between the estimated curve  $\hat{F}_{Z,t}(z)$  and the true distribution function  $F_Z(z)$ .

Fig 6 shows the evolution over time of the estimation of  $p_{WB}$ , the estimated mean value of variable  $Z$  and the ISE of the estimators of  $F_Z(z)$ . The results show the median value of the 100 scenarios for each day. All methods provide results that converge to the true value of the estimated parameter (red color). Note that the sample size  $n(t)$  increases with  $t$ , so this convergence shows the consistency of the methods. The top graph in Fig 7 shows the convergence of the ISE as the sample size increases. The bottom graph in Fig 7 displays the computational effort to calculate the EMNP estimator, which moderately increases with the sample size. The computer used in the experiment was an Intel (R) Xeon (R) CPU E5-1630 v4 3.70GHz with 64.0 GB RAM. Even for large sample sizes, such as 5000, it takes about three minutes to get the estimation, which is affordable because it only needs to be done once a day, in order to predict the bed occupancy level.

Both the *EMNP method* and the *EM method* have a fast convergence in all simulated cases, which turns out to be relevant when the simulation model is used as a prediction tool for the resources needed in the future, as we expose in the next subsection. The *NP method* provides the third best results, improving alternative naïve methods. Nonetheless, the bias of these latter estimators, due to information from patients in hospital ward is dismissed, tends to fade away at advanced stages of the pandemic, as the proportion of these censored observations decreases. It should also be highlighted that the estimator *IQ3* outperforms the estimator *IQ2*.

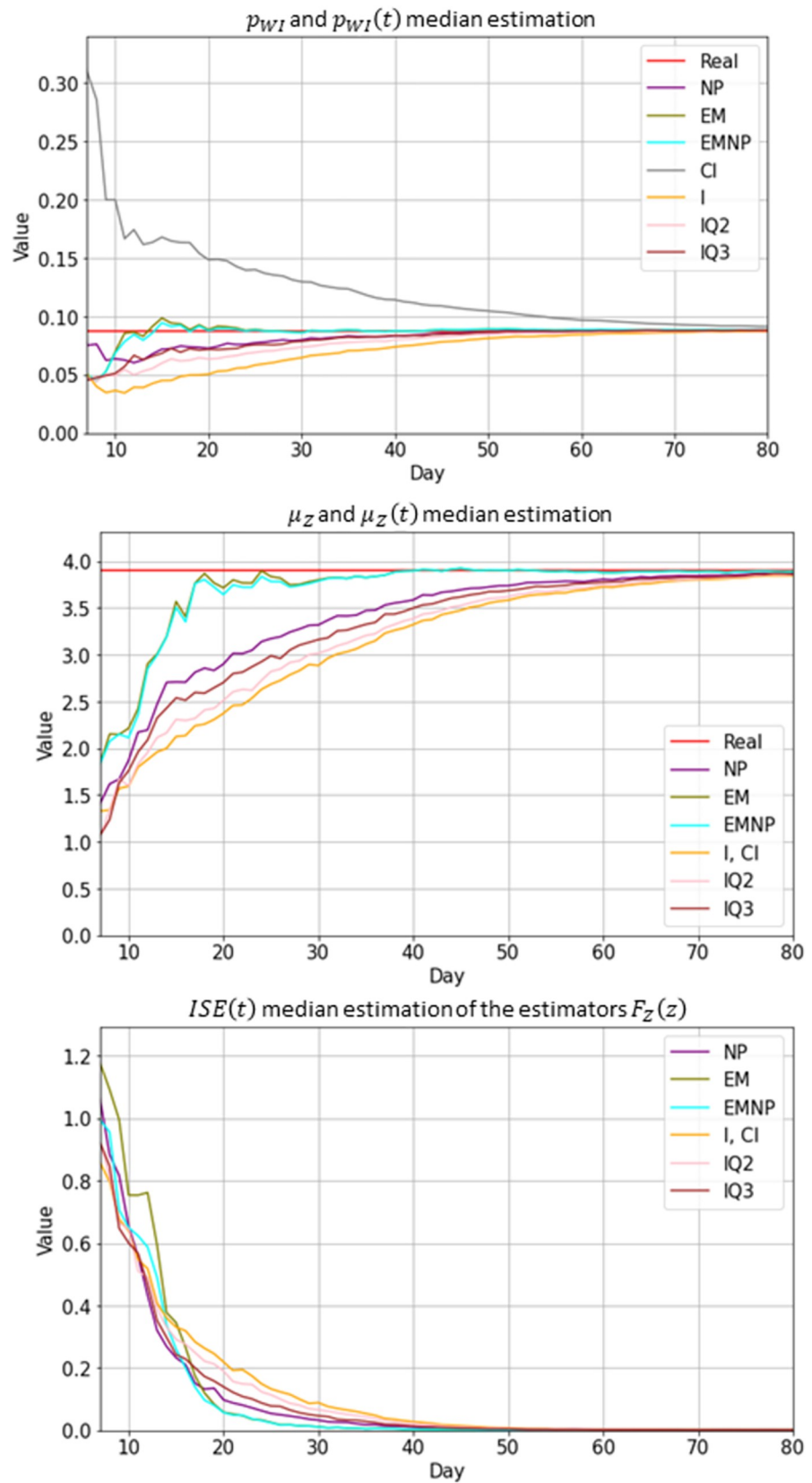
#### 4.2. Impact on the simulation output. Bed occupancy prediction accuracy

Simulation is used to predict the future bed occupancy level during the course of a fixed generated pandemic wave (RS), when the pandemic is on the 15<sup>th</sup>, 20<sup>th</sup>, 25<sup>th</sup>, and 30<sup>th</sup> days (SSP). For each SSP, we generated 500 future courses of the pandemic by simulating future hospital patient-flows, as described in Subsection 3.4. The predictions of ICU beds demand are obtained by the statistical analysis of the output of these 500 runs. The corresponding bed occupancy forecasts for each method are compared with those obtained from 500 future developments of the pandemic generated by simulating patient pathways based on the true values of the parameters and probabilities in Subsection 3.3.

Fig 8 shows twenty-eight predictions of ICU bed occupancy considering all methods from 4 different days (15<sup>th</sup>, 20<sup>th</sup>, 25<sup>th</sup>, and 30<sup>th</sup>). Note that these days are quite far away from the peak occupancy (45<sup>th</sup> day), with expected ICU bed occupancy of 176 beds and 90% centred prediction interval (154,198). The green line in each graph represents the evolution of the simulated pandemic up to the SSP (black dot). For each method, the 5<sup>th</sup> percentile (P5) and the 95<sup>th</sup> percentile (P95) of the predicted ICU occupancy levels are plotted using orange lines, whereas the blue lines represent the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the predictions when pandemic is simulated using the real parameter values (denoted by the letter R). As the pandemic progresses, the predictions of ICU bed occupancy of all methods approach the real occupancy rate. However, the *EMNP method* is the closest one for all the four different estimating days. Based on the results, we can also conclude that the *EM method* performs almost as well as the *EMNP method*. Besides, the *NP method* again improves the naïve methods that do not use all available information, and the *IQ3 method* has a better behavior than the *IQ2 method*. Finally, the *IC method* clearly overestimates ICU bed occupancy while *I method* underestimates ICU bed occupancy for all the SSP times considered.

In addition, we have also studied the errors in the predictions of the maximum number of occupied ICU beds and the day on which the maximum occurs. Fig 9 shows boxplots representing the estimation errors between the predictions when simulating with each estimation method and the predictions when simulating using the actual values of the parameters, calculated for each of the 500 simulations at different times (15<sup>th</sup>, 20<sup>th</sup>, 25<sup>th</sup>, and 30<sup>th</sup> day). Positive

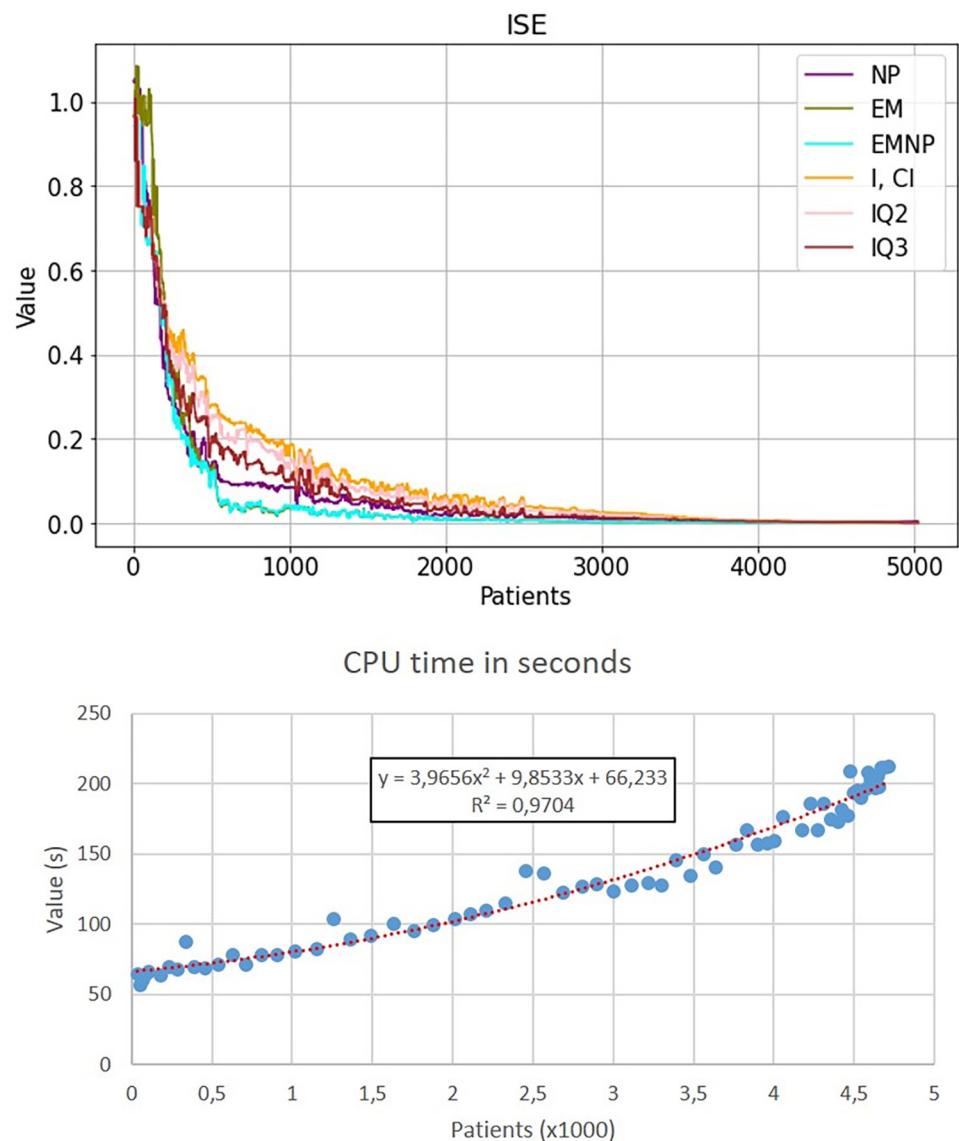




**Fig 6. The evolution over time of the estimations using different methods.** Median of the estimations of the probability  $p_{WI}$  (top), median of the estimated mean times in hospital ward until admission to ICU,  $\mu_Z$  (center), and ISE of the estimators of the distribution function of variable  $Z$  (bottom) over time  $t$  with all the methods, computed with 100 simulated pandemics. The real values of  $p_{WI}$  and  $\mu_Z$  are included for reference (red line). The horizontal axes represent the time (days) during the pandemic.

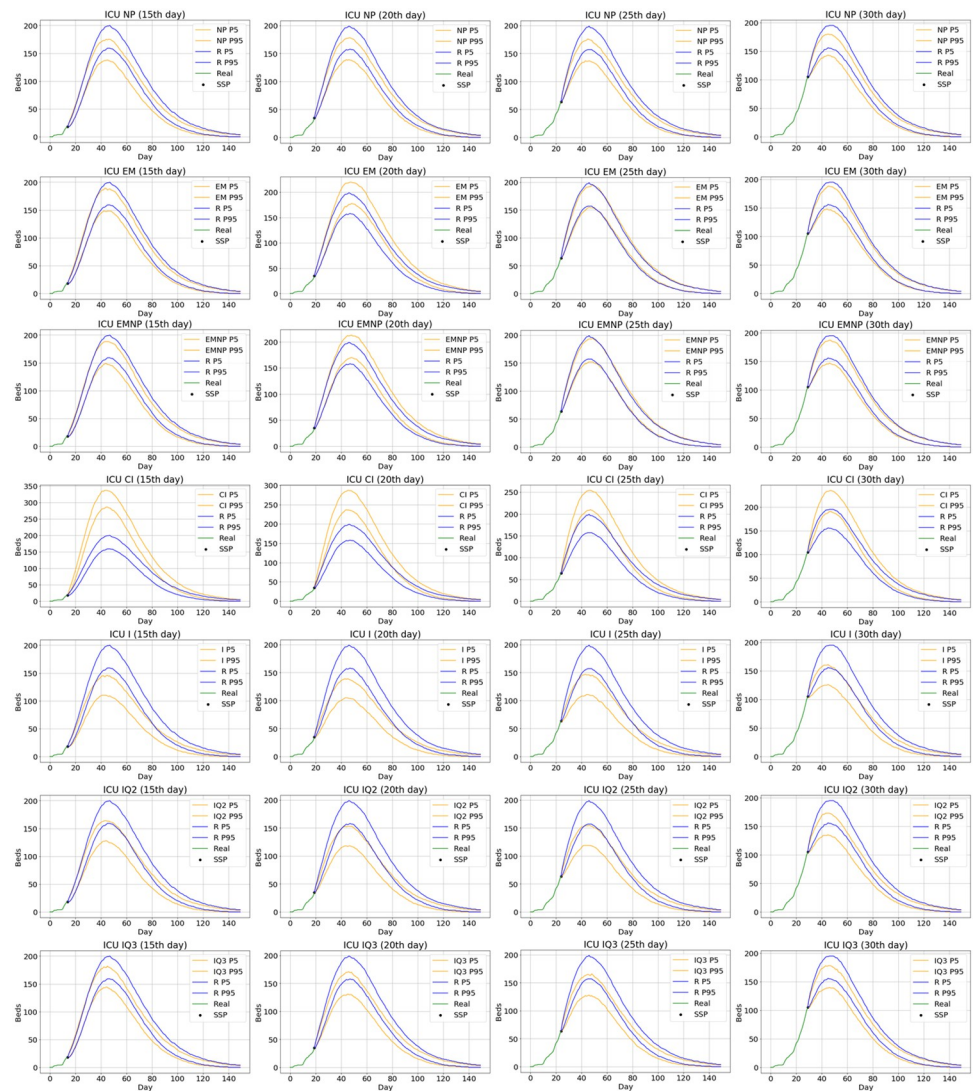
<https://doi.org/10.1371/journal.pone.0282331.g006>

differences indicate an overestimation while negative differences indicate an underestimation. When predicting the maximum number of occupied ICU beds, we can observe that the predictions improve as the prediction day advances, and the *EM* method and the *EMNP* method outperform all other approaches, with average errors of 10.184 and 11.574 beds respectively when



**Fig 7. Illustration of convergence in experiments.** Relationship between the error and the sample size with which the parameter estimation is done (top), and CPU time (in seconds) needed to obtain parameter estimations using the *EMNP* method as a function of the sample size (bottom).

<https://doi.org/10.1371/journal.pone.0282331.g007>



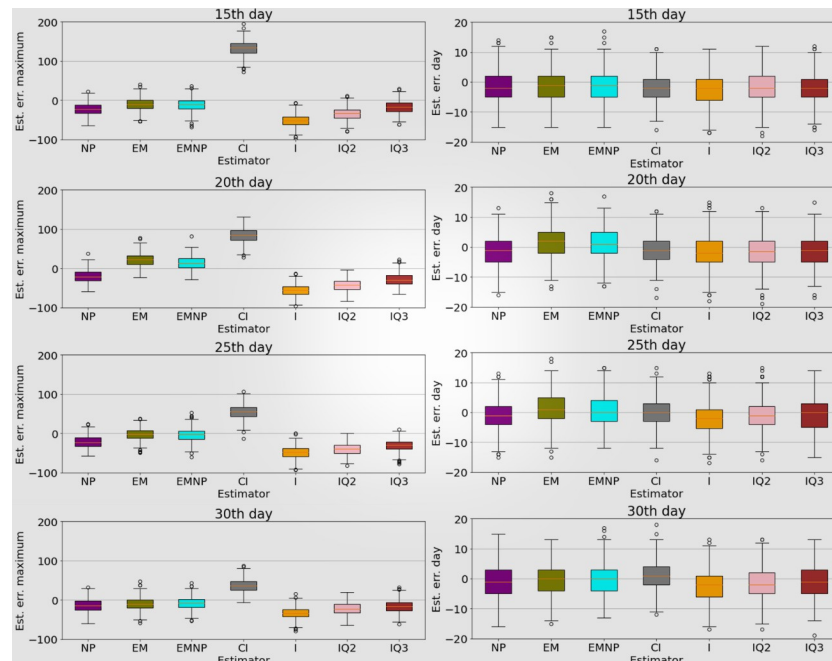
**Fig 8. Twenty-eight predictions of ICU bed occupancy considering all methods from 4 different days.** Prediction (5<sup>th</sup> and 95<sup>th</sup> percentiles) of ICU bed occupancy on the 15<sup>th</sup>, 20<sup>th</sup>, 25<sup>th</sup>, and 30<sup>th</sup> days of the pandemic with all the methods compared to prediction (5<sup>th</sup> and 95<sup>th</sup> percentiles) with actual parameters, denoted with letter R.

<https://doi.org/10.1371/journal.pone.0282331.g008>

estimating the 15<sup>th</sup> day, and 9.204 and 8.628 beds when estimating the 30<sup>th</sup> day. However, the results for the estimation of the day of maximum occupancy are very similar for all methods and all the days considered.

## 5. Discussion

In this work, we consider a DES model to forecast ICU bed demand via simulation of inpatient's future pathways. The simulation of patients' flow is carried out when all the distributions and branching probabilities in the patients' pathways are estimated. We introduced different methods to estimate efficiently these probabilities and lengths of stay, and showed how to apply them to estimate the probability that an inpatient in hospital ward will be admitted to ICU, and the distribution of the time in hospital ward until admission to ICU. The proposed methods can also be applied to estimate all other distributions and probabilities that



**Fig 9. Estimation errors between the predictions when simulating with each estimation method and with actual values.** Analysis of the maximum bed occupancy in the ICU, and the day on which the maximum occurs. For days 15<sup>th</sup>, 20<sup>th</sup>, 25<sup>th</sup>, and 30<sup>th</sup>, the estimation errors are shown for each of the 500 simulations between the values obtained with each method and the predictions with the real parameters.

<https://doi.org/10.1371/journal.pone.0282331.g009>

define the pathway of a patient, such as the probability of dying in ICU or the time in ICU until discharge. The great advantage of the proposed methods is that the estimation does not rely on published data across heterogeneous populations or health system's historical data, but on the real patients admitted to the hospital during the period when prediction is of interest. If information is updated frequently, then hospitalization and bed demand forecasts will be more accurate. The second advantage is that partial information provided by inpatients still in hospital at the time of estimation is included in the estimation procedures, which increases efficiency and reliability of the results. Note that the main challenge of using up-to-date patient-level information is that data provided by patients still in hospital ward is censored, as the future path of these patients remains unknown. Methods that take advantage of the partial information associated to these patients using mixture cure models (MCM) have been shown to be more efficient than naïve methods that do not use survival analysis techniques.

The EM and EMNP estimators can be applied in other contexts, for example, to estimate the parameters of stochastic compartmental models used to represent the spread of a pandemic [16, 18, 50]. Some compartmental models extend the original SIR model introducing more compartments such as Exposed, Quarantined, Hospitalized, etc. [50, 51]. The patient pathway through these compartments can be similar to those represented in Fig 3, and then susceptible to applying the estimators presented in this research. For example, an Infected patient can transit to the Recovery compartment or to the Hospital compartment [18]. Our estimators can estimate the probability distributions of both time until recovery and time to hospital admission, and transition probabilities, by using up-to-date data, which allows for model calibration at the first stages of the pandemic when the uncertainty is the greatest. Other context of possible application of both estimators is reliability, where data coming from both laboratory tests and field observation are usually censored. Maximum-likelihood-based

estimators have been proposed in the literature to deal with these data, see for example [52, 53], which assume a type-II censored scheme, which differs from ours.

The simulation results for the different scenarios show that the *NP method*, that does not assume any parametric form for the distribution of the times, not only provides more flexibility to the model but it converges faster than the parametric approaches. Note that fast convergence is relevant to have reliable estimates at early stages of the pandemic, so the simulation model could be used as a prediction tool for the hospital bed occupancy in the near future. The other methodology that gives very good results is the proposed *EMNP method*, which combines the nonparametric approaches based on the MCM of the aforementioned *NP method*, and assumes a parametric form for the distribution of the lengths of stay, with parameters estimated using the EM algorithm. The behavior of this new *EMNP method* has been assessed in a simulation study. As expected, the *EMNP method* outperforms the other approaches, as the distribution of the simulated data fulfils the parametric assumption in the *EMNP method*.

It is important to note that a variety of limitations exists for the proposed methods. One limitation concerns the simulated pathway for the patients, specifically, the number of admissions to ICU. In the simulated model, a patient is assumed to require ICU once at most. Although this is the most common case for inpatients in a hospital, it might not be realistic for all the subjects. The model can be extended to include more than one admission to ICU. However, increasing the possible number of times in ICU would make the model to become considerably complex.

A second limitation is that the proposed methods do not consider level-patient characteristics like infection severity, age, comorbidity status and diagnostic testing results. All the methods can be extended to incorporate these characteristics in the estimations as covariates. If only one covariate is to be included, the *NP method* and the *EMNP method* can be extended following Safari et al. [30, 33]. When there are many covariates, the sparseness of data gives rise to the well known “curse of dimensionality”, which implies that massive amounts of data will be required for accurate estimate as the number of covariates increases. Different approaches are available in the literature, which enable handling multiple covariates when estimating nonparametrically under censoring [54–56]. Alternative approaches to extend the *NP method* and the *EMNP method* to multiple covariates are the proportional hazards model [35–39] or the accelerated failure time model [40–44]. However, these extensions are beyond the scope of this study and considering these approaches is left for future work.

The efficiency of the proposed methods depends strongly on the quality of the patient-level information provided by the hospital electronic health record systems. The information should be uploaded frequently into the system, and the model updated accordingly. Notwithstanding, the proposed methods are more dynamic and adaptive than any other approach based on historical data. They are flexible and can be updated when new data are available.

In conclusion, the proposed *NP method* and *EMNP method* provide a useful, efficient, adaptive and easily applicable methodology for estimating the distribution times and branching probabilities in inpatients’ pathways. These methods achieve good performance without relying on comparable historical data that may not be available or may not be realistic. In addition, they are flexible and can be further extended to accommodate multiple patient-level characteristics. The provided estimates can be used subsequently in DES for modelling the demand for critical care beds. As a result, the proposed methods are useful tools to forecast bed occupancy, and we hope they are helpful to improve making decisions in hospital management.

## Author Contributions

**Conceptualization:** María Amalia Jácome, Fermin Mallor.

**Data curation:** Daniel Garcia-Vicuña, Ana López-Cheda.

**Formal analysis:** Daniel Garcia-Vicuña, Ana López-Cheda.

**Funding acquisition:** María Amalia Jácome, Fermin Mallor.

**Investigation:** Daniel Garcia-Vicuña, Ana López-Cheda.

**Methodology:** Daniel Garcia-Vicuña, Ana López-Cheda, María Amalia Jácome, Fermin Mallor.

**Project administration:** Fermin Mallor.

**Software:** Daniel Garcia-Vicuña.

**Supervision:** María Amalia Jácome, Fermin Mallor.

**Validation:** Ana López-Cheda, María Amalia Jácome.

**Writing – original draft:** Daniel Garcia-Vicuña, Fermin Mallor.

**Writing – review & editing:** Daniel Garcia-Vicuña, Ana López-Cheda, María Amalia Jácome, Fermin Mallor.

## References

1. Bravata DM, Perkins AJ, Myers LJ, Arling G, Zhang Y, Zillich AJ, et al. Association of Intensive Care Unit Patient Load and Demand With Mortality Rates in US Department of Veterans Affairs Hospitals During the COVID-19 Pandemic. *JAMA Netw open* 2021; 4(1):e2034266. <https://doi.org/10.1001/jamanetworkopen.2020.34266> PMID: 33464319
2. Janke AT, Mei H, Rothenberg C, Becher RD, Lin Z, Venkatesh AK. Analysis of Hospital Resource Availability and COVID-19 Mortality Across the United States. *J Hosp Med* 2021; 16(4):211–214. <https://doi.org/10.12788/jhm.3539> PMID: 33496664
3. Castagna F, Xue X, Saeed O, Kataria R, Puius YA, Patel SR, et al. Hospital bed occupancy rate is an independent risk factor for COVID-19 inpatient mortality: a pandemic epicentre cohort study. *BMJ Open* 2022; 12:e058171. <https://doi.org/10.1136/bmjopen-2021-058171> PMID: 35168984
4. Soria A, Galimberti S, Lapadula G, Visco F, Ardini A, Valsecchi MG, et al. The high volume of patients admitted during the SARS-CoV-2 pandemic has an independent harmful impact on in-hospital mortality from COVID-19. *PLoS One* 2021; 16(1):e0246170. <https://doi.org/10.1371/journal.pone.0246170> PMID: 33507954
5. Menon DK, Taylor BL, Ridley SA. Modelling the impact of an influenza pandemic on critical care services in England. *Anaesthesia* 2005; 60(10):952–954. <https://doi.org/10.1111/j.1365-2044.2005.04372.x> PMID: 16179037
6. Gitto S, Di Mauro C, Ancarani A, Mancuso P. Forecasting national and regional level intensive care unit bed demand during COVID-19: The case of Italy. *PLoS One* 2021; 16(2):e0247726. <https://doi.org/10.1371/journal.pone.0247726> PMID: 33630972
7. Litton E, Bucci T, Chavan S, Ho YY, Holley A, Howard G, et al. Surge capacity of intensive care units in case of acute increase in demand caused by COVID-19 in Australia. *Med J Aust* 2020; 212(10):463–467. <https://doi.org/10.5694/mja2.50596> PMID: 32306408
8. Barasa EW, Ouma PO, Okiro EA. Assessing the hospital surge capacity of the Kenyan health system in the face of the COVID-19 pandemic. *PLoS One* 2020; 15(7):e0236308. <https://doi.org/10.1371/journal.pone.0236308> PMID: 32687538
9. Peña VH, Espinosa A. Predictive modeling to estimate the demand for intensive care hospital beds nationwide in the context of the COVID-19 pandemic. *Medwave* 2020; 20(9):e8039. <https://doi.org/10.5867/medwave.2020.09.8039> PMID: 33031358
10. Deschepper M, Eeckloo K, Malfait S, Benoit D, Callens S, Vansteelandt S. Prediction of hospital bed capacity during the COVID-19 pandemic. *BMC Health Serv Res* 2021; 21:468. <https://doi.org/10.1186/s12913-021-06492-3> PMID: 34006279
11. López-Cheda A, Jácome MA, Cao R, De Salazar PM. Estimating lengths-of-stay of hospitalized COVID-19 patients using a non-parametric model: a case study in Galicia (Spain). *Epidemiol Infect* 2021; 149, e102:1–8. <https://doi.org/10.1017/S0950268821000959> PMID: 33902779

12. Garcia-Vicuña D, Esparza L, Mallor F. Hospital preparedness during epidemics using simulation: the case of COVID-19. *Cent Eur J Oper Res* 2021; 30(1):213–249. <https://doi.org/10.1007/s10100-021-00779-w> PMID: 34602855
13. Azcarate C, Esparza L, Mallor F. The problem of the last bed: Contextualization and a new simulation framework for analyzing physician decisions. *Omega* 2020; 96:102120. <https://doi.org/10.1016/j.omega.2019.102120>
14. Garcia-Vicuña D, Esparza L, Mallor F. Safely learning intensive care unit management by using a management flight simulator. *Oper Res Heal Care* 2020; 27:100274. <https://doi.org/10.1016/j.orhc.2020.100274>
15. Weissman GE, Crane-Droesch A, Chivers C, Luong TB, Hanish A, Levy MZ, et al. Locally informed simulation to predict hospital capacity needs during the covid-19 pandemic. *Ann Intern Med* 2020; 173(1):21–28. <https://doi.org/10.7326/M20-1260> PMID: 32259197
16. Watson GL, Xiong D, Zhang L, Zoller JA, Shamsioian J, Sundin P, et al. Pandemic velocity: Forecasting COVID-19 in the US with a machine learning & Bayesian time series compartmental model. *PLoS Comput Biol* 2021; 17(3):e1008837. <https://doi.org/10.1371/journal.pcbi.1008837> PMID: 33780443
17. Hadley E, Rhea S, Jones K, Li L, Stoner M, Bobashev G. Enhancing the prediction of hospitalization from a COVID-19 agent-based model: A Bayesian method for model parameter estimation. *PLoS One* 2022; 17(3):e0264704. <https://doi.org/10.1371/journal.pone.0264704> PMID: 35231066
18. Valles TE, Shoenhard H, Zinski J, Trick S, Porter MA, Lindstrom MR. Networks of necessity: Simulating COVID-19 mitigation strategies for disabled people and their caregivers. *PLoS Comput Biol* 2022; 18(5):e1010042. <https://doi.org/10.1371/journal.pcbi.1010042> PMID: 35584133
19. Brailsford SC, Harper PR, Patel B, Pitt M. An analysis of the academic literature on simulation and modelling in health care. *J Simul* 2009; 3(3):130–140. <https://doi.org/10.1057/jos.2009.10>
20. Katsaliaki K, Mustafee N. Applications of simulation within the healthcare context. *J Oper Res Soc* 2011; 62(8):1431–1451. <https://doi.org/10.1057/jors.2010.20> PMID: 32226177
21. Currie CSM, Fowler JW, Kotiadis K, Monks T, Onggo BS, Robertson DA, et al. How simulation modelling can help reduce the impact of COVID-19. *J Simul* 2020; 14(2):83–97. <https://doi.org/10.1080/17477778.2020.1751570>
22. Rees EM, Nightingale ES, Jafari Y, Waterlow NR, Clifford S, Carl CA, et al. COVID-19 length of hospital stay: a systematic review and data synthesis. *BMC Med* 2020; 18:270. <https://doi.org/10.1186/s12916-020-01726-3> PMID: 32878619
23. Preiss A, Hadley E, Jones K, Stoner MCD, Kery C, Baumgartner P, et al. Incorporation of near-real-time hospital occupancy data to improve hospitalization forecast accuracy during the COVID-19 pandemic. *Infect Dis Model* 2022; 7(1):277–285. <https://doi.org/10.1016/j.idm.2022.01.003> PMID: 35136849
24. Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J R Stat Soc Ser B* 1949; 11(1):15–53. <https://doi.org/10.1111/J.2517-6161.1949.TB00020.X>
25. López-Cheda A, Cao R, Jácome MA, Van Keilegom I. Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Comput Stat Data Anal* 2017; 105:144–165. <https://doi.org/10.1016/j.csda.2016.08.002>
26. López-Cheda A, Jácome MA, Cao R. Nonparametric latency estimation for mixture cure models. *TEST* 2017; 26(2):353–376. <https://doi.org/10.1007/s11749-016-0515-1>
27. Amico M, Van Keilegom I. Cure Models in Survival Analysis. *Annu Rev Stat Its Appl* 2018; 5:311–342. <https://doi.org/10.1146/annurev-statistics-031017-100101>
28. Peng Y, Yu B. *Cure Models*. 1st Editio. New York: CRC Press, 2021.
29. Betensky RA, Schoenfeld DA. Nonparametric estimation in a cure model with random cure times. *Biometrics* 2001; 57(1):282–286. <https://doi.org/10.1111/j.0006-341x.2001.00282.x> PMID: 11252611
30. Safari WC, López-de-Ullibarri I, Jácome MA. A product-limit estimator of the conditional survival function when cure status is partially known. *Biometrical J* 2021; 63(5):984–1005. <https://doi.org/10.1002/bimj.202000173> PMID: 33646606
31. Bernhardt PW. A flexible cure rate model with dependent censoring and a known cure threshold. *Stat Med* 2016; 35(25):4607–4623. <https://doi.org/10.1002/sim.7014> PMID: 27311808
32. Safari WC, López-de-Ullibarri I, Jácome MA. Nonparametric kernel estimation of the probability of cure in a mixture cure model when the cure status is partially observed. *Stat Methods Med Res* 2022; 31(11):2164–2188. <https://doi.org/10.1177/09622802221115880> PMID: 35912505
33. Safari WC, López-de-Ullibarri I, Jácome MA. Nonparametric estimation of mixture cure models when the cure status is partially known. 2023. *Accepted in Lifetime Data Analysis*.
34. Aerts M, Claeskens G, Hens N, Molenberghs G. Local multiple imputation. *Biometrika* 2002; 89(2):375–388. <https://doi.org/10.1093/biomet/89.2.375>

35. Bin Fang H, Gang LI, Sun J. Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. *Scand J Stat* 2005; 32(1):59–75. <https://doi.org/10.1111/j.1467-9469.2005.00415.x>
36. Kuk AYC, Chen C-H. A mixture model combining logistic regression with proportional hazards regression. *Biometrika* 1992; 79(3):531–541. <https://doi.org/10.1093/biomet/79.3.531>
37. Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. *Biometrics* 2000; 56(1):237–243. <https://doi.org/10.1111/j.0006-341x.2000.00237.x> PMID: 10783801
38. Peng Y. Fitting semiparametric cure models. *Comput Stat Data Anal* 2003; 41(3–4):481–490. [https://doi.org/10.1016/S0167-9473\(02\)00184-6](https://doi.org/10.1016/S0167-9473(02)00184-6)
39. Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics* 2000; 56(1):227–236. <https://doi.org/10.1111/j.0006-341x.2000.00227.x> PMID: 10783800
40. Li CS, Taylor JMG. A semi-parametric accelerated failure time cure model. *Stat Med* 2002; 21(21):3235–3247. <https://doi.org/10.1002/sim.1260> PMID: 12375301
41. Li CS, Taylor JMG. Smoothing covariate effects in cure models. *Commun Stat* 2002; 31(3):477–493. <https://doi.org/10.1081/STA-120002860>
42. Lu W. Efficient estimation for an accelerated failure time model with a cure fraction. *Stat Sin* 2010; 20:661–674. PMID: 20414460
43. Zhang J, Peng Y. An alternative estimation method for the accelerated failure time frailty model. *Comput Stat Data Anal* 2007; 51(9):4413–4423. <https://doi.org/10.1016/j.csda.2006.06.017>
44. Zhang J, Peng Y. A new estimation method for the semiparametric accelerated failure time mixture cure model. *Stat Med* 2007; 26(16):3157–3171. <https://doi.org/10.1002/sim.2748> PMID: 17094075
45. Gompertz B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philos Trans R Soc London B Biol Sci* 1825; 182:513–585. <https://doi.org/10.1098/rstl.1825.0026>
46. Richards FJ. A flexible growth function for empirical use. *J Exp Bot* 1959; 10(2):290–301. <https://doi.org/10.1093/jxb/10.2.290>
47. Stannard CJ, Williams AP, Gibbs PA. Temperature/growth relationships for psychrotrophic food-spoilage bacteria. *Food Microbiol* 1985; 2(2):115–122. [https://doi.org/10.1016/S0740-0020\(85\)80004-6](https://doi.org/10.1016/S0740-0020(85)80004-6)
48. Ricker WE. Growth rates and models. *Fish Physiol* 1979; 8:677–743. [https://doi.org/10.1016/S1546-5098\(08\)60034-5](https://doi.org/10.1016/S1546-5098(08)60034-5)
49. Zwietering MH, Jongenburger I, Rombouts FM, van 't Riet K. Modeling of the bacterial growth curve. *Appl Environ Microbiol* 1990; 56(6):1875–1881. <https://doi.org/10.1128/aem.56.6.1875-1881.1990> PMID: 16348228
50. Anderson SC, Edwards AM, Yerlanov M, Mulberry N, Stockdale JE, Iyaniwura SA, et al. Quantifying the impact of COVID-19 control measures using a Bayesian model of physical distancing. *PLoS Comput Biol* 2020; 16(12):e1008274. <https://doi.org/10.1371/journal.pcbi.1008274> PMID: 33270633
51. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med* 2020; 26:855–860. <https://doi.org/10.1038/s41591-020-0883-7> PMID: 32322102
52. Tashkandy YA, Almetwally EM, Ragab R, Gemeay AM, Abd El-Raouf MM, Khosa SK, et al. Statistical inferences for the extended inverse Weibull distribution under progressive type-II censored sample with applications. *Alexandria Eng J* 2023; 65:493–502. <https://doi.org/10.1016/j.aej.2022.09.023>
53. Alrumayh A, Weera W, Khogeer HA, Almetwally EM. Optimal analysis of adaptive type-II progressive censored for new unit-lindley model. *J King Saud Univ—Sci* 2023; 35(2):102462. <https://doi.org/10.1016/j.jksus.2022.102462>
54. Liang HY, de Uña-Álvarez J, Iglesias-Pérez M del C. Asymptotic properties of conditional distribution estimator with truncated, censored and dependent data. *Test* 2012; 21(4):790–810. <https://doi.org/10.1007/s11749-012-0281-7>
55. Li Q, Racine JS. Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *J Bus Econ Stat* 2008; 26(4):423–434. <https://doi.org/10.1198/073500107000000250>
56. Amico M, Van Keilegom I, Legrand C. The single-index/Cox mixture cure model. *Biometrics* 2019; 75(2):452–462. <https://doi.org/10.1111/biom.12999> PMID: 30430553