Heliyon 9 (2023) e13545

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A multistate model and its standalone tool to predict hospital and ICU occupancy by patients with COVID-19

Miguel Lafuente [a,b], Francisco Javier López [a,b], Pedro Mariano Mateo [a,b,c], Ana Carmen Cebrián [a,b], Jesús Asín [a], José Antonio Moler [d], Ángel Borque-Fernando [e], Luis Mariano Esteban [b,f,*], Ana Pérez-Palomares [a,**], Gerardo Sanz [a,b]

[a] Department of Statistical Methods, Universidad de Zaragoza, C. Pedro Cerbuna 12, 50009 Zaragoza, Spain
[b] Institute for Biocomputation and Physics of Complex Systems-BIFI, Universidad de Zaragoza. C. de Mariano Esquillor Gómez, Edificio I+D, 50018 Zaragoza, Spain
[c] Centre Q-UPHS. Quantitative Methods for Uplifting the Performance of Health Services, Spain
[d] Department of Statistics and Operational Research, Universidad Pública de Navarra, Campus Arrosadía S/n, 31006 Pamplona, Spain
[e] Department of Urology, Miguel Servet University Hospital and IIS Aragón, Paseo Isabel La Católica 1-3, 50009 Zaragoza, Spain
[f] Department of Applied Mathematics, Escuela Universitaria Politécnica de La Almunia, University of Zaragoza, C/ Mayor 5, 50100 La Almunia de Doña Godina, Spain

ARTICLE INFO

Keywords:
COVID-19
Health resources
Multistate models
Hospital and ICU occupancy
Predictive tool

ABSTRACT

Objective: This study aims to build a multistate model and describe a predictive tool for estimating the daily number of intensive care unit (ICU) and hospital beds occupied by patients with coronavirus 2019 disease (COVID-19).
Material and methods: The estimation is based on the simulation of patient trajectories using a multistate model where the transition probabilities between states are estimated via competing risks and cure models. The input to the tool includes the dates of COVID-19 diagnosis, admission to hospital, admission to ICU, discharge from ICU and discharge from hospital or death of positive cases from a selected initial date to the current moment. Our tool is validated using 98,496 cases positive for severe acute respiratory coronavirus 2 extracted from the Aragón Healthcare Records Database from July 1, 2020 to February 28, 2021.
Results: The tool demonstrates good performance for the 7- and 14-days forecasts using the actual positive cases, and shows good accuracy among three scenarios corresponding to different stages of the pandemic: 1) up-scenario, 2) peak-scenario and 3) down-scenario. Long term predictions (two months) also show good accuracy, while those using Holt-Winters positive case estimates revealed acceptable accuracy to day 14 onwards, with relative errors of 8.8%.
Discussion: In the era of the COVID-19 pandemic, hospitals must evolve in a dynamic way. Our prediction tool is designed to predict hospital occupancy to improve healthcare resource management without information about clinical history of patients.

* Corresponding author. Escuela Universitaria Politécnica de La Almunia, Universidad de Zaragoza, C. Mayor 5, 50100 La Almunia de Doña Godina, Spain.
** Department of Statistical Methods, Universidad de Zaragoza, C. Pedro Cerbuna 12, 50009 Zaragoza, Spain
E-mail addresses: lmeste@unizar.es (L.M. Esteban), anapp@unizar.es (A. Pérez-Palomares).

https://doi.org/10.1016/j.heliyon.2023.e13545
Received 28 June 2022; Received in revised form 28 January 2023; Accepted 2 February 2023
Available online 5 February 2023

Conclusions: Our easy-to-use and freely accessible tool (https://github.com/peterman65) shows good performance and accuracy for forecasting the daily number of hospital and ICU beds required for patients with COVID-19.

1. Introduction

Hospital systems are grappling with the resource management challenges posed by outbreaks of coronavirus 2019 disease (COVID-19) [1]. The COVID-19 pandemic has subjected health systems around the world to unprecedented stress, requiring resources that have even exceeded their planned emergency capacity [2]. As a matter of fact, versatile models able to accurately predict the evolution of the pandemic at different scales (country, region or organization) are of great interest [3].

Substantial effort has been devoted to modeling the spread of the disease and different models to forecast the trends of the pandemic have been proposed. Gnanvi et al. [4] conducted an extensive review of different prediction techniques. Their results show that SIR (Susceptible, Infected and Recovered) models were very common (46.1%), 36.5% were statistical models (including Bayesian models), while artificial intelligence based models were only 6.7% at the time of the review. On the other hand, they claim that the use of larger databases, equivalent to longer study periods, provides more accurate predictions.

Compared to the large number of models for predicting COVID-19 cases, fewer resource management planning models have been developed for pandemics and, so, they constitute innovative research in this field [5].

Hospital management requires planning for different types of resources, such as medical staffing and necessary equipment, which are highly dependent on the number of beds occupied in the hospitalization wards and the ICU. Moreover, if widespread vaccination turns the COVID-19 pandemic into a seasonal circular disease, as has previously happened with other known coronaviruses [6,7], the management of hospital resources will always be conditioned by the occupation of beds by patients with COVID-19.

The prediction of periods of maximum and minimum hospital occupancy is of particular importance, not only for a better allocation of resources to times of greatest need, but also to recover and prioritize the care of non-COVID-19 pathologies in stages of low saturation [8]. Surgical scheduling could clearly benefit from these predictive models. Otherwise, attention to non-COVID-19 pathology is relegated to checking for lower COVID-19 saturation, which reduces the possibility of planning and responsiveness. This leaves the hospital ineffectively dealing with the constraints of a new COVID-19 incidence peak/wave. Similarly, human resource management would benefit from bed occupancy predictions, as vacations and other leave days can be scheduled during periods when COVID-19 occupancy is expected to be low.

Thus, predictions for COVID hospital occupancy should be used as an important input in hospital management not only to avoid the health system collapse but to recover the resource situation pre-COVID-19 [9]. While a large number of freely available tools have been designed for patient-level predictions [10–13], only a few predict hospital occupancy [9,14]. Therefore, tools focused on cohort-level predictions are needed for COVID-19 management.

The spread of COVID-19 has differed according to the waves and virus strains that have emerged; hence resource management must be dynamic and adaptable. Here, we propose an online dynamic statistical tool to forecast hospital and ICU bed occupancy by patients with COVID-19 in a health system, in the short or long term. Our predictive tool uses a multistate model where patients move between states. The transition probabilities between states are estimated using classical survival techniques, competing risks, and cure models.

Future hospital occupancy is related to patients who are currently in the system (which we call “COVID-19 Positive Census”) and those who may be introduced in the forecast period (“COVID-19 Incident Positive Cases”). Our model simulates the trajectories of both types of patients and these trajectories are then used to forecast the COVID-related hospital and ICU occupation from the current moment onwards. The follow-up of positive cases is defined by a trajectory that includes the dates of diagnosis, admission to hospital, admission to ICU, discharge from ICU and discharge from hospital or death.

Our tool has some advantages with respect to other models in the literature. First, it does not require assumptions about distributions or model parameters. Moreover, the required input data are ordinary information regarding admission and discharge dates that are usually available in health systems. The tool can be applied at any level, from hospital to regional or national level, regardless of its location, the type of hospitals or its internal structure, assuming that the population of potential patients is clearly defined. The flexibility of our tool makes it transferable to other health systems, which is an important aspect of COVID-19 management.

2. Related work

In the early months of the pandemic, efforts were made to predict hospital occupancy and capacity saturation. Several tools and papers were developed to assist hospital administrators in this process [15]. As stated before, most of the developed models were based on SIR type models or their extensions and parametric distributions were used to model virus spread and length of hospital stay (LoS) [10,16–21]. In these applications, input parameters, such as the percentage of patients admitted to hospital or the LoS (its mean/-median) in the hospital, are required to predict the hospital resources needed in different scenarios.

Predicting hospital and ICU occupancy requires accurate estimation of the LoS in hospital care. The LoS is dependent on the state of the patient, clinical care strategies, and resource availability. Works devoted to predicting the LoS have identified patient characteristics that produce high variability in LoS [16,17]. Most of them use a parametric approach, assuming a particular family of distributions for the LoS in their data.

Farcomeni et al. [22] developed another type of model based on regression and time-series methodology for short-term predictions

(1–3 days ahead) of ICU bed needs during the first epidemic wave in different regions of Italy. Goic et al. [23] provided a combined autoregressive machine-learning and epidemiological model for short-term forecasting (7 and 14 days) of ICU utilization at a regional level in Chile. These models have the advantage that they do not use the evolution of a patient, but their accuracy is limited for mid- and long-term forecasting.

The ability to predict hospital occupancy by means of SIR models and their extensions has been questioned due to the evolution of the pandemic. A limitation of SIR models is that they do not consider patient characteristics and, for example, the time from infection should be included in the models [24,25].

Stochastic models are an alternative to SIR. A multistate model is a stochastic model where the characteristics of a patient and the time spent in each state are incorporated into the model. Several authors have proposed a multistate model to predict hospital occupancy [9,14,26–28], that is, by simulating the trajectory of each patient.

In those works, in addition to needing some parameters, patients enter the system when they are admitted to hospital, while in our case the starting point is a positive diagnosis. In Ref. [27] a Poisson model was used to simulate the number of new hospital admissions during the 10 days of prediction and a non-parametric approach is used to estimate the transition probabilities between states. In Ref. [14] the transitions between the states are modeled by Cox regression models and competing-risk techniques were used to estimate the parameters of the model, considering age, sex, state at hospitalization and cumulative days in hospital as covariates. The arrival process of patients to be hospitalized is needed in the prediction period together with the clinical state of each patient. They provide an R software package to use the model.

In [26] a completely parametric model was considered, by using Population Growth (PG) models for the new admissions process and Weibull and Lognormal distributions for ICU and hospital LoS, respectively.

Caro et al. [9] provided a model for predicting hospital occupancy using a discretely integrated condition event simulation. This approach is based on parametric distributions for hospital length of stay where the input parameters or some information about the distributions is needed. On the other hand, it is a flexible model which can be used with Microsoft Excel.

Bekker et al. [28] proposed two independent models for ICU and non-ICU occupancy. For the arrival process (ICU and non-ICU arrivals) they developed a new technique using linear programming. Standard survival techniques were applied to estimate LoS for ICU and non-ICU. They used public data, so the time spent in hospital for a patient is not incorporated into the model and a residual length of stay is used instead. The model is not data driven since a tuning parameter is necessary and it is not estimated from the data. The accuracy of the model could be limited for mid- and long-term forecasting, but the advantage is that the model uses little data.

In Bicher [29] a harmonized model was developed based on three different models: SIR-X model, agent based SEIR model and an autoregressive model to predict the daily number of infections in Austria. With these projections, the authors provide forecasts for hospital occupation by estimating the hospitalization rate and the length of stay. The model is used for short term predictions.

Recently, machine learning and deep learning models have been applied to study the evolution of the pandemic [30–32]. The use of these techniques is also applied on health management and particularly on hospital occupancy, combining neural network model (NN) with a Susceptible-Exposed-Infected-Recovered model (SEIR) [33], or comparing Long Short-Term memory (LSTM) network, convolutional neural network (CNN) and their combination [34].

To complement the variety of proposed models in the literature, we propose a nonparametric approach since we want our tool to be broadly applicable, and because it is unlikely that a single parametric family of distributions will fit the observed data in a wide range of situations. In addition, our model is derived from very general information, such as a simple confirmed diagnosis of COVID-19, and is therefore independent of the initial categorization of the disease. We provide a user-friendly standalone tool that does not require any knowledge of programming. The tool is flexible since it allows the inclusion of covariates through groups. Moreover, users can introduce their own process of new positive cases or use the estimates provided by the tool, so it is a completely data driven model.

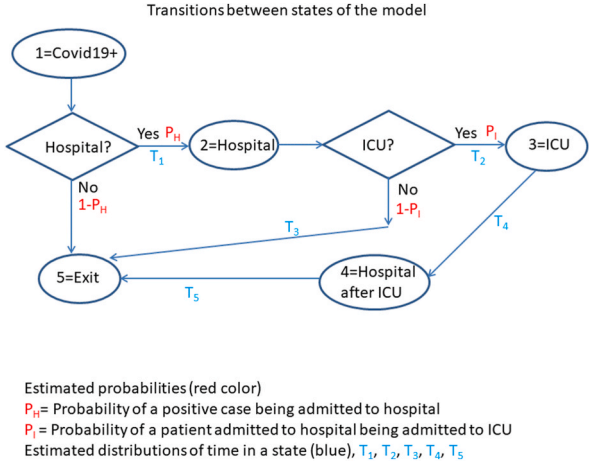

Fig. 1. Multistate structure.

3. Materials and methods

3.1. Outcomes

Study outcomes are two COVID-19 indicators: the daily number of hospital and ICU beds occupied by COVID-19 patients from the present day to a number *s* of days ahead. As hospital beds we consider ICU and non-ICU beds.

3.2. Multistate model

Multistate models describe the evolution of individuals through several states over the course of time. Each individual has her/his own path in the model and her/his own times in each state. In medicine, individuals are usually patients, which evolve through different stages of a disease, not necessarily in a one-directional fashion, or through different "states" in a hospital (e.g., hospital ward, surgery, ICU). While survival analysis has been present for more than 50 years in medical research, the use of multistate models is more recent, especially in hospital epidemiology [35,36].

Our model describes the evolution of individuals through different states from the time they test positive onward. To keep the model as simple and universal as possible we consider five states: 1 = Positive, 2 = Hospital, 3 = ICU, 4 = Hospital after ICU, 5 = Exit. The path for a newly positive case is shown in Fig. 1. Since our primary objective is hospital and ICU occupancy, we do not separate the patients by their final outcome (death or recovery).

In the model, the probability of being admitted to hospital (and thus occupying a bed) is denoted by P_H; the time between diagnosis and admittance to hospital is random and denoted by T_1. Once admitted to hospital (state 2), a patient may or may not be admitted to ICU. There is a probability P_I that a patient in hospital will be admitted to ICU at a random time (T_2) since admission to hospital. Patients not admitted to ICU spend some number of days (T_3) in state 2 and then leave the system (state 5). Patients admitted to ICU (state 3), stay for some time (T_4) before being discharged from the ICU (state 4). They spend some time (T_5) in the hospital after discharge from the ICU and then leave the system (state 5). Note that there is not a direct arrow from state 3 to state 5 for patients dying in the ICU; this transition is taken into account by letting T_5 be equal to 0 for those patients. Also, T_1 is 0 for patients admitted to hospital on their date of positive diagnosis and T_2 is 0 for patients admitted to ICU on their first day at hospital. Table 1 gives a summary of the states, probabilities and times in the model.

The distributions of the times in the different states of the model will be defined through the corresponding survival functions. The survival function of a random time until an event occurs is denoted by S(t), which represents the probability that the event has not occurred in the first t days. The survival function of times T_1, ….,T_5 will be denoted by S_1, …,S_5.

To keep the model as simple and easily exportable as possible, we do not consider readmissions. In fact, since data on readmissions are likely to be sparse, the estimations of recurrent transitions would not be very reliable.

Regarding its probabilistic structure, our model can be seen as a semi-Markov model. We assume that the future evolution of an individual depends on the number of days in the present state (positive, hospital, ICU, post-ICU) but not on the days spent in previous states. In particular, we do not assume that the distributions of the times have a constant failure rate, which is a harsh condition needed for the model to be Markovian [37]. Indeed, as we use a nonparametric approach, we do not impose any restriction on the distribution of the random times T_1, ….,T_5; also, no previous knowledge of these distributions is assumed, since the model will estimate the corresponding survival functions from the cohort data.

Table 1
States, probabilities and times in the model.

States of the model
1 = "Positive" = An individual with a positive test for SARS-CoV-2 but not admitted to hospital
2 = "Hospital" = A patient admitted to hospital but not admitted to ICU
3 = "ICU" = A patient admitted to ICU
4 = "Hospital after ICU" = A patient at hospital after being discharged from ICU
5 = "Exit" = A patient discharged from hospital or dead
Probabilities
P_H = transition probability from state 1 to 2, i.e., probability that a positive case is admitted to hospital
P_I = transition probability from state 2 to 3, i.e., probability that a patient admitted to hospital is admitted to ICU
Times
For patients admitted to hospital:
T_1: time between diagnosis and admission to hospital
For patients admitted to hospital who are not admitted to ICU:
T_3: time between admission to hospital and discharge
For patients admitted to ICU:
T_2: time between admission to hospital and admission to ICU
T_4: time between admission to ICU and discharge from ICU
T_5: time between discharge from ICU and discharge from hospital

3.3. Data structure for estimation

All parameters and probability distributions above are estimated by the model. The cohort used for estimation is formed by all the patients who tested positive between two fixed dates: t_I and t_F, where t_F is usually taken to equal t_0, the present day (see Fig. 2). We assume that we have all the information about the trajectory of the patients in the cohort since the day they tested positive up to the present day t_0. Thus, the follow-up period of the cohort is $[t_I, t_0]$. Note that, for patients still in the system on t_0, we have the time they spent on their previous states and a right-censored value for the time they spend in their current state.

We show in Table 2 an example of data for five individuals, where we are taking t_I = 2020-06-01, t_F = 2020-12-31 and t_0 = 2021-03-01. The first individual tested positive on September 30th, but was never admitted to hospital. Patient 2 tested positive on October 5th, was admitted to hospital on October 15th, has not been admitted to ICU but is still in hospital by March 1st. Patient 3 tested positive on November 10th, was admitted to hospital on November 25th, admitted to ICU on November 28th, and discharged from ICU and hospital on December 20th and January 12th, respectively. Patient 4 tested positive on November 25th, was admitted to hospital on November 30th and to ICU on December 15th, discharged from ICU on February 15th, but still in hospital on March 1st. Finally, patient 5 tested positive on January 12th, admitted to hospital on January 28th, and has not been admitted to ICU nor discharged by March 1st.

The cohort includes patients 1–4, since they tested positive between June 1st and December 31st; note that their dates after December 31st, such as patient 3 leaving the hospital on January 12th, are also used for estimation. Patient 5, who tested positive after December 31st is not included in the cohort.

Patient characteristics, such as age and comorbidities, may affect the length of the hospital stay [16]. Thus, the use of covariates can be useful to predict hospital and ICU occupancy. To ensure the generalizability of this model, we do not include specific covariates as predictor variables, as the corresponding data may not be available in all health systems. Instead, covariates can be incorporated into the model by defining groups of patients with similar characteristics, such as sex, age group, or level of risk according to comorbidities. Groups are taken as an input to the model.

Since we may have more information of patients already in the system (COVID-19 Positive Census), such as comorbidities, than for future positive cases (COVID-19 Incident Positive Cases) two different groupings for patients are considered. The first grouping (Grouping 1) will be used to simulate the evolution of patients in the COVID-19 Positive Census, those already in the system at t_0. The second grouping (Grouping 2) will be used to simulate the evolution of COVID-19 Incident Positive Cases, those who will test positive from day t_0+1 to t_0+s.

If groups are defined, the estimations of P_H, P_I, S_1, …,S_5 are carried out for each group in Groupings 1 and 2. That is, the cohort of patients will be split in groups as defined by Grouping 1, and separate estimations of P_H, P_I, S_1, …,S_5 will be obtained for each group. Likewise, the cohort will be split in groups as defined by Grouping 2, and separate estimations of P_H, P_I, S_1, …,S_5 will be obtained for each group.

For instance, we may consider the following groups given by age and sex in the COVID-19 Positive Census: G1.1: <61 years and male; G1.2: <61 years and female, G1.3: 61–80 years and male; G1.4: 61–80 years and female; G1.5:> 80 years and male; G1.6::> 80 years and female; and groups by sex for COVID-19 Incident Positive Cases: G2.1: male and G2.2: female.

We show in Table 3 the example of 5 individuals in Table 2 with two columns added indicating the groups where each individual belongs.

If a unique group is considered, the columns G1 and G2 must have a unique value for all individuals.

3.4. Model estimation

To estimate the parameters of the model, we consider nonparametric survival models, one for each state since different states have different characteristics.

For patients in state 1 we need to estimate the probability of being admitted to hospital, P_H, and the distribution of T_1, the time in state 1 for patients who will be admitted to hospital. We model this state as a cure model [38]. Cure models are a special type of survival analysis model wherein it is assumed that there is a proportion of subjects who will never experience the event of interest and thus the survival curve will eventually reach a plateau [39]. In our case, we consider that the cured patients are those who will not be admitted to hospital. Thus, the survival function of time in state 1 is:

$$P(Time\ in\ state\ 1 > t) = (1 - P_H) + P_H \cdot S_1(t)$$

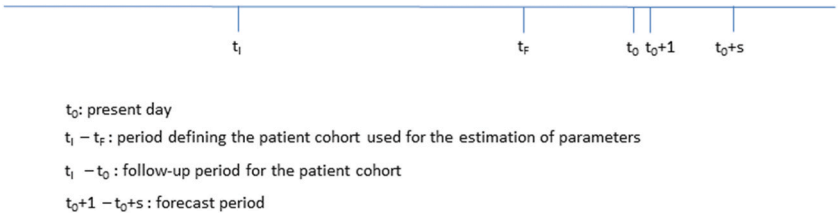

Fig. 2. Periods defined in the estimation and forecast procedures.

Table 2
Example of data for 5 individuals. A. Hosp.-Date of admission to hospital, A. ICU- Date of admission to ICU; discharge from ICU, discharge from hospital.

ID	Positive	A. Hosp.	A. ICU	D. ICU	D. Hosp.
1	2020-09-30	NA	NA	NA	NA
2	2020-10-05	2020-10-15	NA	NA	NA
3	2020-11-10	2020-11-25	2020-11-28	2020-12-20	2021-01-12
4	2020-11-25	2020-11-30	2020-12-15	2021-02-15	NA
5	2021-01-12	2021-01-28	NA	NA	NA

Table 3
Example of data for 5 individuals. G1- Group number for grouping 1; G2- Group number for grouping 2; A. Hosp.-Date of admission to hospital, A. ICU- Date of admission to ICU; D. ICU- Date of discharge from ICU, D.Hosp - Date of discharge from hospital.

ID	G1	G2	Positive	A. Hosp.	A. ICU	D. ICU	D. Hosp.
1	3	1	2020-09-30	NA	NA	NA	NA
2	4	2	2020-10-05	2020-10-15	NA	NA	NA
3	1	1	2020-11-10	2020-11-25	2020-11-28	2020-12-20	2021-01-12
4	2	2	2021-11-25	2020-11-30	2020-12-15	2021-02-15	NA
5	5	1	2021-01-12	2021-01-28	NA	NA	NA

where S_1 is the survival function of time T_1. We use the mixture cure model approach of Taylor [40] to obtain estimates of P_H and S_1; the latter is estimated using a Kaplan-Meier type estimator.

For state 2, since there are more than one type of event, that is, being admitted to ICU or being discharged from hospital or death, a competing-risk model is used [41]. Competing risks occur when there are several outcomes which 'compete'. Traditional methods, such as Kaplan Meier method, are not designed to accommodate competing risks, and special techniques of survival analysis must be used to correctly estimate the marginal probability of an outcome in those cases. Competing-risk models have been used in the literature to analyze the evolution of COVID-19 patients [23,42]. In our case, the survival function of time in state 2 is:

$$P(Time\ in\ state\ 2 > t) = P_I \cdot S_2(t) + (1 - P_I) \cdot S_3(t)$$

where S_2 and S_3 are the survival functions of times T_2 (time until admission to the ICU) and T_3 (time until discharge). We use a standard semiparametric approach to obtain estimates of P_I, the probability of being admitted to the ICU and the survival functions S_2 and S_3.

For states 3 and 4, a standard survival model is used, since all the patients in these states will be eventually discharged from ICU and hospital, respectively. We use the Kaplan-Meier estimator [43] for the estimation of S_4 and S_5, the survival functions of T_4 and T_5. Table 4 shows a summary of the estimation methods.

If groupings are used, the estimations are carried out in each group separately.

3.5. Simulation and prediction

Once the probabilities and survival functions have been estimated, the model can be used to forecast hospital and ICU occupancy in the interval from day t_0+1 to day t_0+s. This is done via Monte Carlo simulation of the evolution of the patients in the system using two sources of information. First, the patients who are in the system on day t_0 (COVID-19 Positive Census), that is, positive cases who are in any of the states (1–4) of the multistate structure; and second, new positive cases from day t_0+1 to day t_0+s (COVID-19 Incident Positive Cases).

Each simulation run comprises the simulation of each patient's trajectory in the period t_0+1, …, t_0+s and the computation of the number of hospital and ICU beds occupied on day t by adding all patients in hospital or ICU on that day. We repeat this procedure nsim times (for instance with nsim equal to 2000, the simulation error is almost negligible) and use the nsim samples to estimate the expected number of ICU and hospital beds occupied on day t, as well as their standard deviation and 5% and 95% percentiles. The simulation of each patient is carried out as follows.

Table 4
Probabilities and time spent of the model, together with the statistical methods used for their estimation.

		Estimation methods
P_H	Probability of a positive case being admitted to hospital	Mixture cure model.
T_1	Time between diagnosis and admittance to hospital	Kaplan-Meier estimation of time distribution
P_I	Probability of a patient admitted to hospital being admitted to ICU	Competing-risk model with semiparametric approach
T_3	Time spent in hospital for a patient admitted to hospital but no to ICU	
T_2	Time spent in hospital (before ICU), for a patient admitted to ICU	
T_4	Time spent in ICU	Kaplan-Meier estimation
T_5	Time spent in hospital after ICU	

The simulation of the trajectory of the cases diagnosed between t_0+1 and t_0+s (COVID-19 Incident Positive Cases) is rather straightforward. Consider an individual who will test positive on day t_0+m. We flip a coin with probability of heads equal to P_H to know if the patient will be admitted to hospital. If the individual will not be admitted to hospital, no further action is needed. If the patient will be admitted to hospital, then a random value of T_1, simulated using the estimation of the survival function S_1, sets the date of admission to hospital equal to t_0+m + T_1. Then, we flip a coin with probability of heads equal to P_I to determine if the patient will be admitted to the ICU. If the patient is not admitted to ICU, a random value of T_3 is simulated, defining the discharge date from hospital as t_0 +m + T_1+T_3; if the patient is admitted to ICU, values for T_2, T_4 and T_5 are drawn setting the date of admission to ICU equal to t_0+m + T_1+T_2 and the dates of discharge from ICU and from hospital equal to t_0+m + T_1+T_2 +T_4 and t_0+m + T_1+T_2+T_4+T_5, respectively.

The model needs a prediction of the daily number of new COVID-19 Incident Positive Cases in the period [t_0+1, t_0+s]. This can be taken as an input provided by the users or estimated by the model using the daily number of positive cases until t_0 and applying the Holt-Winters (H–W) methodology. The Holt-Winters method is an extension of the exponential smoothing method to allow forecasting of data with a trend and to capture seasonality. Exponential smoothing methods provide forecasts that are weighted averages of past observations, with the weights decaying exponentially as the observations get older; see Refs. [44,45] for more details. Prediction for data with no seasonality is also available. In the case that groups are used, the Holt-Winters method is applied separately to each group, in Grouping 2, to obtain the prediction of the number of infected individuals in each group. If the prediction of new positive cases is taken as an input to the model, a separate prediction must be entered for each group.

The simulation of the trajectory of patients already in the system on day t_0 (COVID-19 Positive Census) is similar, but conditional probabilities instead of raw probabilities must be used for their first step in the system. This is because we are not assuming loss of memory, so the number of days in their present state is needed for simulating the remaining time in the state and the following transition. The simulation depends on the state where each patient is on day t_0. For a patient in state 1 (who tested positive m days before t_0), we flip a coin with probability of heads equal to

$$P_{H|m} = \frac{P(T_1 > m)P_H}{P(T_1 > m)P_H + 1 - P_H}$$

If the patient is admitted to hospital, the time to admission will be obtained from the conditional distribution of T_1 given that $T_1 >$ m; that is,

$$P(T_1 = m + v \mid T_1 > m) = \frac{P(T_1 = m + v)}{P(T_1 > m)}$$

and the admission to hospital date will be t_0+v. Once in hospital, the rest of the patient's trajectory is simulated using unconditional probabilities. For a patient in state 2, admitted to hospital m days before t_0 but not to the ICU, the probability of being admitted to ICU is

$$P_{I|m} = \frac{P(T_2 > m)P_I}{P(T_2 > m)P_I + P(T_3 > m)(1 - P_I)}$$

If the patient is not to be admitted to ICU, he/she will be discharged on day t_0+v, where v is drawn with probability

$$\frac{P(T_3 = m + v)}{P(T_3 > m)}$$

Otherwise, the patient will be admitted to ICU on day t_0+v, where v is drawn with probability

$$\frac{P(T_2 = m + v)}{P(T_2 > m)}$$

and the rest of her/his trajectory will be simulated with unconditional probabilities. For a patient in state 3 (admitted to ICU on day t_0-m), the day of discharge from ICU will be t_0+v, where v is drawn with probability

$$\frac{P(T_4 = m + v)}{P(T_4 > m)}$$

and the discharge date from hospital will be simulated from the unconditional distribution of T_5. Last, a patient in state 4 (discharged from ICU on day t_0-m) will be discharged from hospital on day t_0+v where v is drawn with probability

$$\frac{P(T_5 = m + v)}{P(T_5 > m)}$$

3.6. Validation

The model validation can be performed by "predicting" the hospital and ICU occupancy of a past [t_0+1,t_0+s] period where the actual values are known.

It is noteworthy that to run the model we need to know, together with the number of positive cases at t0 ("COVID-19 Positive Census"), the number of new positive cases on each day of the prediction period [t0+1,t0+s] ("COVID-19 Incident Positive Cases"),

which will be unknown in real applications. As mentioned above, if no prediction of that number is available, our tool obtains a prediction using Holt-Winters exponential smoothing method. The validation procedure is carried out in two different ways. First, in order to evaluate the accuracy of the model, we compute the error, with respect to the actual figures of the hospital and ICU occupancy, of the model predictions when the actual number of positive cases in the period [t0+1,t0+s] is used to simulate the trajectories of "COVID-19 Incident Positive Cases". This is the model intrinsic error and it represents the minimum error that can be achieved. Second, since in real applications the number of "COVID-19 Incident Positive Cases" will be unknown, we also compute the error when the model is run using the Holt-Winters prediction of the number "COVID-19 Incident Positive Cases". The two main measures used to assess the performance of the model are the mean absolute error (MAE) defined as

$$MAE = \frac{1}{s}\sum_{i=t_0+1}^{t_0+s} |A_i - P_i|$$

and the mean absolute percentage error (MAPE), defined as

$$MAPE\ (\%) = \frac{1}{s}\sum_{i=t_0+1}^{t_0+s} \frac{|A_i - P_i|}{A_i} 100\%$$

where A_i is the actual daily number of hospital (ICU) occupied beds on day i and P_i is the corresponding predicted number.

3.7. The tool

We have implemented our model in a standalone application that runs on any platform supporting the Java runtime environments version 1.8.1 or greater. All the required libraries, together with a user manual and a sample dataset, are included in the distributed executable version of the tool (https://github.com/peterman65).

The tool includes all the steps of the model, from estimation to forecasting (Fig. 3). The inputs to the tool are the dates t_I, t_F, t_0 and a file with information concerning patients' trajectories from t_I to t_0 and the groups of patients. In addition, a file containing a prediction of the number of new positive cases in the forecast period, if available, is an input to the tool; if such a prediction is not available, then the tool will compute its own prediction as explained above.

The tool also includes a validation option using the user's previous data (see the user manual for details). In the validation option, the MAE and MAPE of predictions are shown, together with their standard deviation. Moreover, since the tool provides day by day predictions, other performance measures, e.g. giving different weights to overestimation or underestimation errors, can be computed.

The tool uses R language programming, version 4.0.2 (The R Foundation for Statistical Computing, Vienna, Austria), for the estimation procedures of the model, but this is transparent to the user, so no knowledge on R is needed for running the tool. Furthermore, no R installation is required.

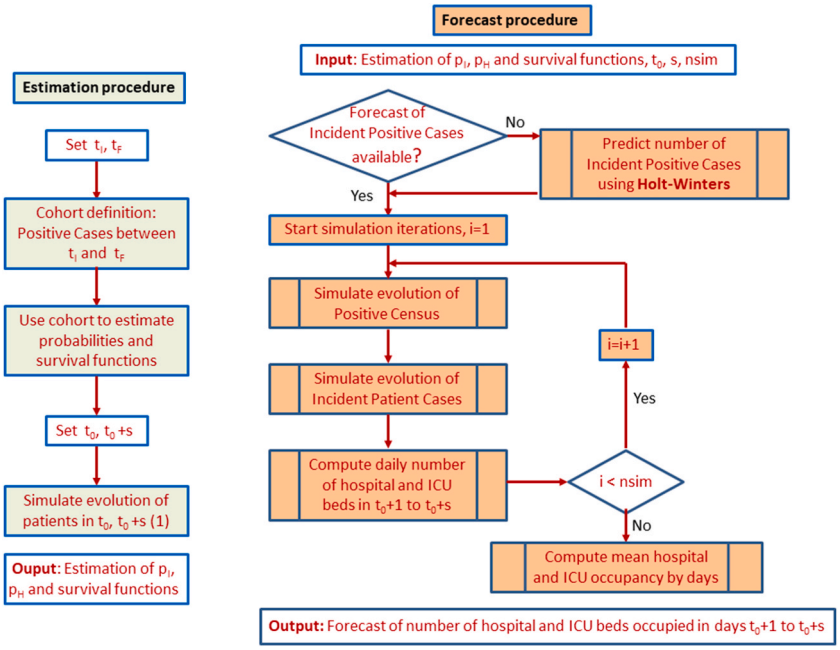

Fig. 3. Methodology flow chart to predict occupancy.

4. Results and discussion

4.1. Patient data

For validating the model, we use data collected from July 1, 2020 to February 28, 2021 in the region of Aragón (Spain), in Northeastern Spain, which has 1,328,753 inhabitants (January 1, 2020). Aragón, as the rest of Spain and Europe, experienced the first pandemic wave between February and May 2020. After a period of full lockdown, community transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) decreased markedly [46]. After several local outbreaks in June and July, SARS-CoV-2 transmission became widespread, and Aragón had the highest midsummer incidence in the European region [47]. Since then, several pandemic waves have occurred in the rest of Europe.

The database of Aragón health system (SALUD) covers the entire population and is composed of 21 hospitals (13 public and 8 private).

The criteria for admission and hospital management of patients with SARS-CoV-2 infection were based on the recommendations published by the Spanish Ministry of Health, which included COVID-19 emergency management and COVID-19 clinical management [48]. Hospitalization was considered chronologically related to a SARS-CoV-2 infection, when it occurred within the first 30 days after the first positive SARS-CoV-2 test. Our primary data source was the Aragón Healthcare Records Database. We collected data relative to the dates of diagnosis, admission to hospital, admission to ICU, discharge from ICU and discharge from hospital. To consider patient groups we also collected patient demographic information, including sex and age.

Our entire cohort are all patients with SARS-CoV-2 infection confirmed from July 1, 2020 to February 28, 2021. The descriptive characteristics are displayed in Table 5. The daily series of new positive cases, hospitalizations and ICU admissions are shown in Fig. 4.

4.2. Model validation

We evaluate the performance of the model in three scenarios with forecasts for 7- and 14-days ahead, and a long-term forecast for 52 days, see Table 6. The choice of 7- and 14-days is based on their closeness to the mean stay times in non-ICU and ICU beds, respectively (see Table 5) and, also, because they are useful time periods for scheduling hospital resources. The three scenarios corresponded to different stages of the pandemic shown in Fig. 5: 1) up-scenario, from January 8 to January 21, 2021; 2) peak-scenario from January 22 to February 4, 2021; and 3) down-scenario from February 15 to February 28, 2021. For each scenario, we set t_I as July 1, 2020 and t_F as the day before the starting date of the forecast period.

Table 7 below summarizes the mean actual number of beds in the prediction period (Actual Occupancy Mean) and the MAE, described in Section 3.6, for the two types of predictions, the model intrinsic error (Actual Positives) and the error using H–W predictions (H–W positives), in forecasts for 7 (7 d) and 14 days (14 d) in the three scenarios defined in Table 6 and the average of the three scenarios.

In Supplementary material, Table S1 shows the analogous results of Table 7, considering the model with groups of patients according to age (<61 years, 61–80 years and >80 years) and sex.

Fig. 6 shows the performance of our simulation tool for estimating the 14-day hospital bed occupancy for the up, peak and down scenarios, considering estimations as a single group. Figure S1 in the supplementary material depicts the results for six groups according to sex and age and Figure S2 shows the results for ICU occupancy by groups.

As it is shown in Table 7, the model intrinsic error showed good accuracy for the three scenarios, with relative errors varying from 2.2% (10.6 beds in hospital of 472.8, in the up-scenario) to 7.4% (6.1 beds of 80.8 in the peak-scenario). Remarkably, the magnitude of

Table 5
Cohort distribution. For Total, Sex and Sex-Age group, the number of positive cases (Positive), patients admitted to hospital (In hospital) and patients admitted to ICU (ICU). In the last two columns, the percentage of patients in hospital with respect to the number of positive cases and the percentages of ICU patients with respect to positive and in hospital cases are shown in brackets. For Age and LoS median value and, in brackets, Q1 and Q3.

	Positive	In hospital	ICU
Total	98,496	9940 (10.09%)	703 (0.71%, 7.07%)
Sex			
Female	51,654	4675 (9.05%)	228 (0.44%, 4.88%)
Male	46,766	5256 (11.24%)	475 (1.01%, 9.04%)
Sex-Age group			
Female (0-60 y)	37,609	1313 (3.49%)	77 (0.20%, 5.86%)
Male (0-60 y)	35,129	1731 (4.93%)	152 (0.43%, 8.78%)
Female (61 y- 80 y)	7934	1464 (18.45%)	142 (1.79%, 9.70%)
Male (61 y- 80 y)	7950	2037 (25.62%)	305 (3.84%, 14.97%)
Female (>80 y)	5586	1858 (33.26%)	6 (0.11%, 0.32%)
Male (>80 y)	3021	1422 (47.07%)	13 (0.43%, 0.91%)
Age (years)	44 (25–61)	72 (57–84)	65 (57–72)
Female	45 (26–62)	75 (58–86)	66 (57–72)
Male	44 (24–60)	69 (56–82)	65 (58–72)
Length of stay (days)		8 (5–14)	17 (9–31)
Female	–	8 (5–14)	17 (9–28)
Male	–	9 (5–15)	18 (9–31)

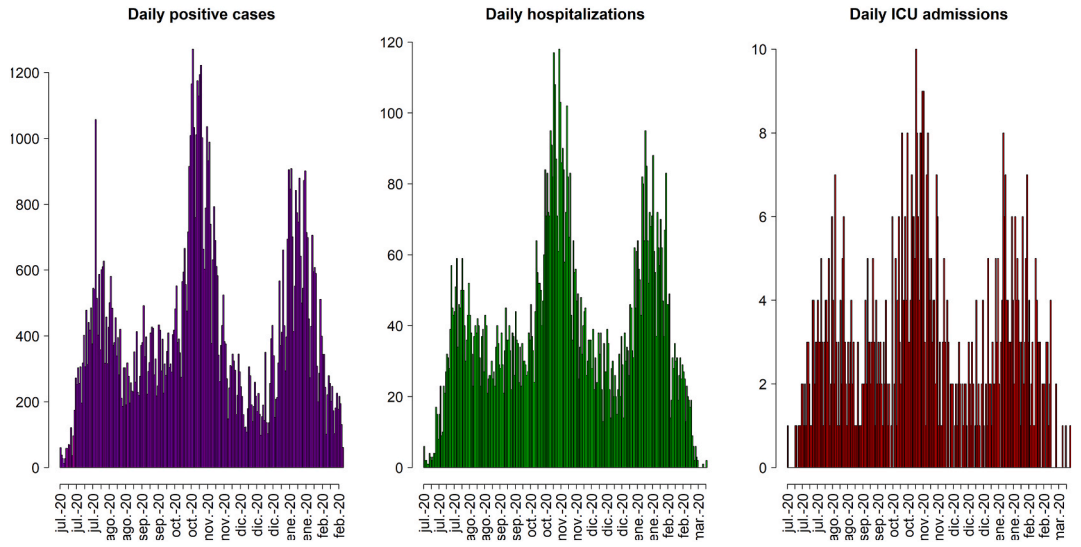

Fig. 4. Actual positive cases, hospital and ICU bed occupancy between July 1, 2020 to February 28, 2021.

Table 6
Cohort and forecast period in the scenarios considered in the validation design.

	Cohort period		Forecast period			
	t_I	t_F	7-day ahead		14-day ahead	
			t_0+1	t_0+7	t_0+1	t_o+14
Up-scenario	July 1, 2020	January 7, 2021	January 8, 2021	January 14, 2021	January 8, 2021	January 21, 2021
Peak-scenario	July 1, 2020	January 21, 2021	January 22, 2021	January 28, 2021	January 22, 2021	February 4, 2021
Down-scenario	July 1, 2020	February 14, 2021	February 15, 2021	February 21, 2021	February 15, 2021	February 28, 2021
			52-day ahead			
	t_I	t_F	t_0+1	t_0+52		
Long-run period	July 1, 2020	January 7, 2021	January 8, 2021	February 28, 2021		

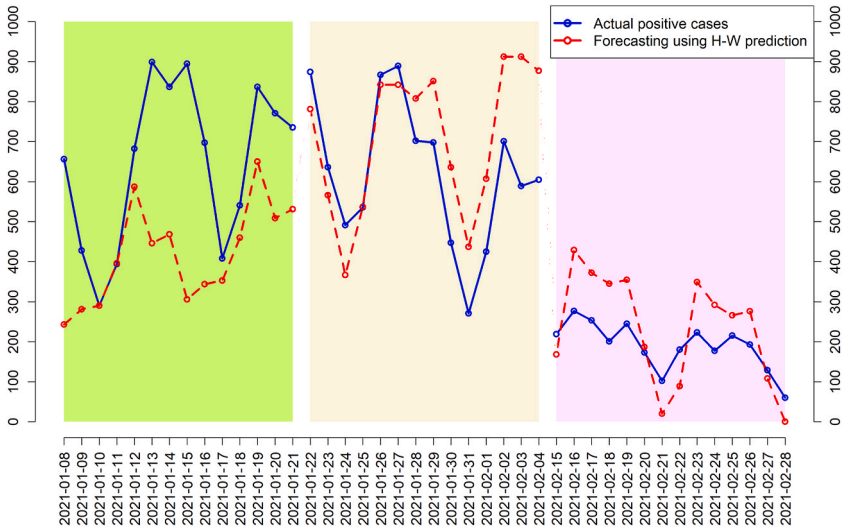

Fig. 5. Actual positive cases and Holt-Winters positive cases estimation in 1: Up-scenario, cohort period: July 1, 2020 to January 7, 2021. Forecasting period: January 8, 2021 to January 21, 2021.2: Peak-scenario: cohort period: July 1, 2020 to January 21, 2021. Forecasting period: January 22, 2021 to February 4, 2021.3: Down-scenario: cohort period: July 1, 2020 to February 14, 2021. Forecasting period: February 15, 2021 to February 28, 2021. H–W: Holt-Winters.

Table 7
Mean actual occupancy and MAE (standard deviation in brackets) of hospital and ICU occupancy forecasts for 7 days (7 d) and 14 days (14 d) in the three scenarios defined in Table 6 and the average of the three scenarios.

	HOSPITAL						ICU					
	Actual Occupancy. Mean		Actual positives. MAE		H–W positives. MAE		Actual Occupancy. Mean		Actual positives. MAE		H–W positives. MAE	
	7 d	14 d	7 d	14 d	7 d	14 d	7 d	14 d	7 d	14 d	7 d	14 d
Up Scenario	472.8 (26.4)	531.4 (70.2)	10.6 (9.1)	12.2 (9.8)	41.2 (19.9)	85.9 (53.0)	47.3 (2.2)	53.6 (8.5)	2.2 (1.0)	2.9 (1.8)	1.6 (0.8)	4.8 (5.5)
Peak Scenario	713.0 (24.4)	717.7 (23.7)	25.5 (16.7)	29.7 (19.0)	24.2 (14.8)	39.5 (32.8)	79.7 (2.0)	80.8 (2.3)	3.1 (1.5)	6.1 (3.8)	2.7 (1.2)	6.0 (4.3)
Down Scenario	532.3 (43.9)	485.8 (58.6)	25.5 (13.3)	18.7 (13.6)	36.4 (20.2)	28.6 (18.2)	76.1 (2.2)	73.3 (4.0)	3.3 (1.9)	2.7 (1.7)	3.2 (1.7)	2.2 (1.6)
Average all scenarios	572.7 (107.2)	578.3 (114,2)	20.5 (15.1)	20.2 (16.3)	33.9 (19.8)	51.3 (44.9)	67.7 (14.7)	69.2 (14.7)	2.9 (1.6)	2.4 (1.5)	3.9 (3.0)	4.1 (4.4)

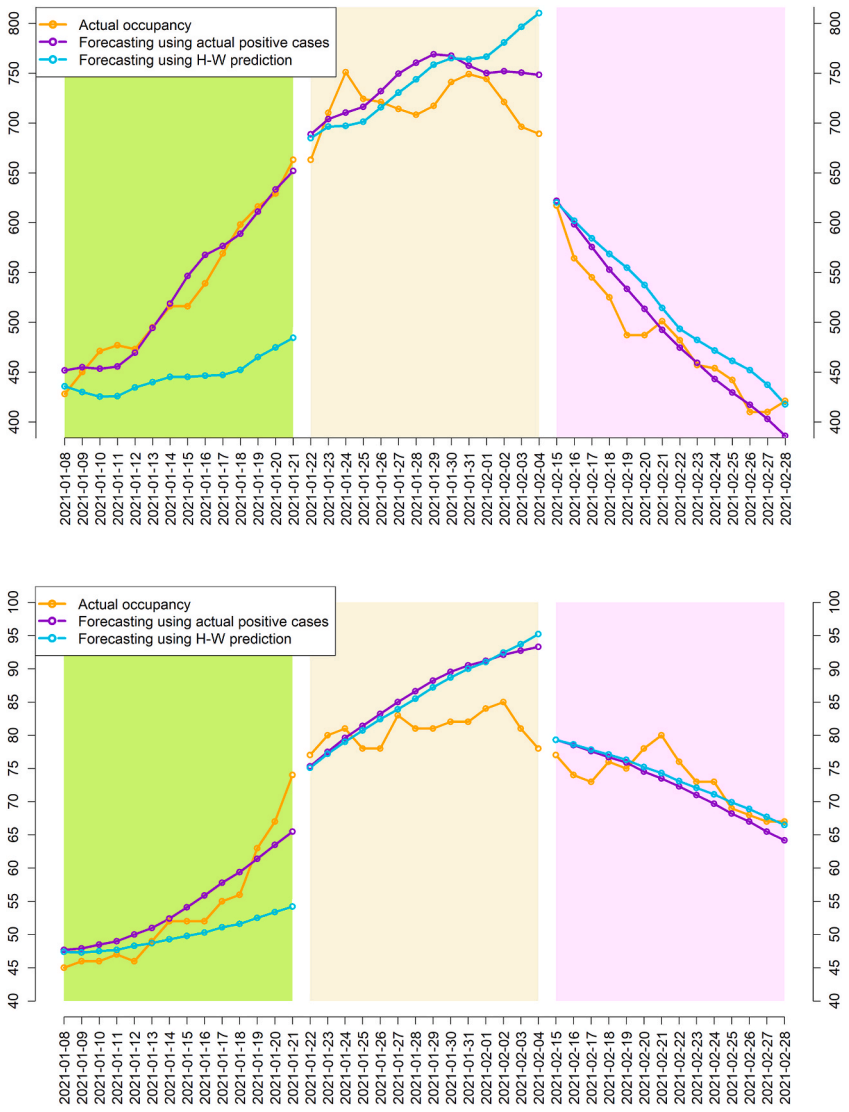

Fig. 6. Actual mean occupancy and predicted occupancy in hospital (top panel) and in ICU (bottom panel) in 1: Up-scenario, cohort period: July 1, 2020 to January 7, 2021. Forecasting period: January 8, 2021 to January 21, 2021.2: Peak-scenario: cohort period: July 1, 2020 to January 21, 2021. Forecasting period: January 22, 2021 to February 4, 2021.3: Down-scenario: cohort period: July 1, 2020 to February 14, 2021. Forecasting period: February 15, 2021 to February 28, 2021. H–W: Holt-Winters.

the errors in 7 and 14 days is similar except for ICU occupancy in the peak-scenario. The occupation of hospital and ICU beds during the COVID-19 pandemic has shown great variability within a few weeks or even days. Therefore, it has been very difficult to forecast occupancy by "standard methods", which usually assume that the system is in some sort of equilibrium. These accurate predictions allow the management team to foresee the scope of the actions required (i.e., the search of available beds).

When the input of the model are Holt-Winters predictions of the "COVID-19 Incident Positive Cases", predictions of Hospital and ICU occupancy are very close to the actual values. In fact, the average relative errors for hospital occupancy prediction at 7 and 14 days are satisfactory (5.9%, 8.8%), with daily average errors over the three scenarios of 33.9 and 51.3 beds. Analogously the average errors for ICU occupancy are 3.9 and 4.1 at 7 and 14 days over the three scenarios.

For the forecast in the long-run period we use actual new positive cases and compute MAEs and MAPEs with and without groups of patients. The results are shown in Table 8. For hospital occupancy, the relative prediction error was less than 4% which means a good performance for long-term predictions; the relative prediction error was 5.9% for ICU occupancy.

4.2.1. Ablation and sensitivity analysis

Table 9 summarizes the results of an additional analysis to evaluate the different sources of error in the model. Since the model consists of several stages, the aim is to evaluate how the error of the model changes when one or more stages are omitted. The model setup "Hospitalized" starts in state 2 in Fig. 1, so that the input of the model is the number of patients admitted to hospital, and the model setup "Without ICU" removes states 3 and 4 in Fig. 1 because ICU and non-ICU patients are not separated. The full model, "Positives", is also included in the table as a reference.

For the "Hospitalized" model, in the first part (Actual values), the actual number of hospital admissions is used throughout the forecast period whereas in HW values the predictions of hospital admissions with the HW method are used. Fig. 7 shows both series in the prediction periods. Whereas in "Positives" and "Without ICU" actual values and HW values refer to positive patients. As in Table 6, the MAE for the number of beds in hospital and in ICU for the three different scenarios, two prediction periods (7 and 14 days) and the two types of errors (model intrinsic error and the error using HW predictions) are summarized.

When actual values are used, for each forecast period and scenario, the MAE is quite similar regardless of the model setup. This entails two consequences: first, that including state 1 in the model adds little error and allows the user to anticipate bed occupancy since the COVID-19 diagnosis; second, the estimation procedure of transitions between states with the initial cohort is quite stable and faithfully reflects patient evolution even if data about ICU were unavailable. On the other hand, when HW values are used, for each model setup, MAE depends on the scenario. As a matter of fact, while the "Hospitalized" model setup reduces the MAE in the Up scenario, it dramatically increases it in the Peak scenario. This can be explained for the change of trend in the curve of positive cases in the days previous to the forecast period in the latter scenario. This change had not yet affected hospital admissions, so Holt-Winters prediction of new admissions do not take it into account, resulting in large errors in its predictions and, as a consequence, in the forecasts of bed occupancy. However, the Up period did not experience a change of trend, so starting the model in state 2 eliminates the error of estimation in state 1.

As a sensitivity analysis of the model, the effect of nsim (number of replications) is studied considering the application of the model with nsim values between 200 and 3000, in each of the 3 scenarios, using as input both the actual positive cases and those predicted by HW. It has been found that in all scenario-input combinations similar MAE are obtained for all nsim values. The conclusion is that the prediction results do not vary with nsim. Details can be seen in Fig. 8, which plots the MAE value versus nsim for each scenario.

In addition, to analyze the influence of the estimation period on the accuracy of the predictions, we have considered different periods. The full period used in the construction of our model (6 months), together with periods of 4, 3 and 2 months. Table 10 shows the results for the different periods considered for the 14-day forecasts. We found similar results in the full period, 4 and 3 months predictions while the results in the 2-month case are clearly different. Note that an estimation period of less than 3 months may be too short to know the evolution of the trajectories, taking into account, for instance, that the length of stay in ICU has a Q3 of 31 days.

Table 8
Mean Actual Occupancy, MAEs and MAPEs (standard deviation in brackets) for hospital and ICU occupancy in the long-run period.

	HOSPITAL			ICU		
	Actual Occupancy. Mean	Actual positives. MAE	Actual positives. MAPE	Actual Occupancy mean	Actual positives. MAE	Actual positives. MAPE
No groups	596.5 (109.7)	21.8 (15.8)	3.8 (2.7)	71.6 (12.5)	4.2 (3.2)	5.9 (4.0)
Groups						
Female (<61 y)	58.5 (10.8)	8.1 (4.6)	13.8 (7.1)	8.3 (2.2)	1.5 (1.4)	16.8 (11.7)
Male (<61 y)	76.5 (17.3)	11.1 (6.7)	15.4 (10.8)	13.8 (3.9)	1.8 (1.3)	15.4 (13.5)
Female (61- 80 y)	109.5 (21.9)	12.0 (7.4)	10.7 (5.7)	13.8 (2.8)	3.2 (2.1)	23.6 (15.6)
Male (61- 80 y)	158.1 (24.8)	14.5 (9.2)	9.1 (5.3)	35.2 (6.3)	4.9 (3.4)	13.4 (9.4)
Female (>80 y)	106.4 (27.7)	8.0 (5.1)	7.3 (3.7)	0.3 (0.4)	0.4 (0.2)	[a]
Male (>80 y)	87.5 (15.7)	7 (4.8)	8.6 (6.9)	0.1 (0.4)	0.5 (0.2)	[a]

[a] >80 y patients were not eligible for ICU.

Table 9
MAE for hospital and ICU occupancy in the study of sources of error.

Model setup	Positives				Hospitalized				Without ICU	
	7 days		14 days		7 days		14 days		7 days	14days
	H	ICU	H	ICU	H	ICU	H	ICU	H	H
Actual values										
UP	10.6	2.2	12.2	2.9	10.6	2.4	11.1	3.0	11.7	13.8
PEAK	25.6	3.1	29.7	6.2	23.8	2.8	24.6	5.5	24.3	25.0
DOWN	25.5	3.3	18.7	4.2	24.8	3.3	19.5	2.4	26.2	18.9
Mean	20.6	2.9	20.2	4.4	19.7	2.8	18.4	3.6	20.7	19.2
HW values										
UP	41.2	1.6	85.9	4.8	11.5	2.3	20.3	3.1	37.1	79.3
PEAK	24.3	2.7	39.5	6.0	71.6	4.8	184.3	13.9	24.2	34.5
DOWN	36.4	3.2	28.6	2.2	59.1	3.4	51.1	2.7	38.6	30.8
Mean	34.0	2.5	51.3	4.3	47.4	3.5	85.2	6.6	33.3	48.2

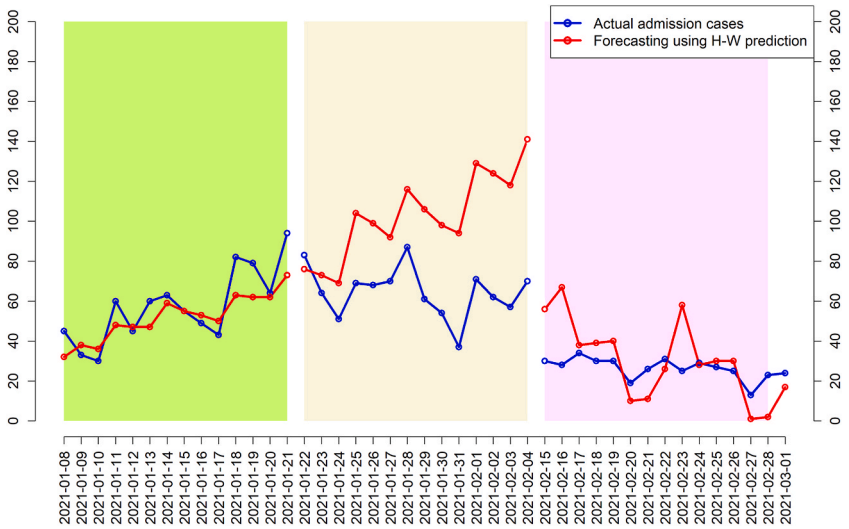

Fig. 7. Actual hospital admission cases and Holt-Winters cases estimation in 1: Up-scenario, cohort period: July 1, 2020 to January 7, 2021. Forecasting period: January 8, 2021 to January 21, 2021.2: Peak-scenario: cohort period: July 1, 2020 to January 21, 2021. Forecasting period: January 22, 2021 to February 4, 2021.3: Down-scenario: cohort period: July 1, 2020 to February 14, 2021. Forecasting period: February 15, 2021 to February 28, 2021. H–W: Holt-Winters.

4.3. Comparison study

In this section we compare the performance of our model with others which have appeared in the literature. A realistic comparison between models requires their application on the same population and on the same period, since the behavior of the models depend strongly on the epidemiological situation. Since data used in other works are not available, a direct comparison with the reported errors in similar works is not possible. Taking this into consideration, we compare the results of our model using our data with those reported in Refs. [14,23,27,28]. In order to make the comparisons as fair as possible, we have computed predictions with the same length of forecast periods as those considered in each work, starting on the first day of each scenario and averaging the errors over the three scenarios.

Table 11 shows the mean errors (MAPE) given in Goic et al. [23] for ICU occupancy for one- and two-week horizons together with our average errors over the three scenarios for one and two weeks, considering the Holt-Winters predictions for positive cases. We observe similar behavior in 7 days predictions andt a better performance of our model in 14 days.

Deschepper et al. [27] considered a model with 4 states: Non-ICU, ICU Midcare, ICU Standard and ICU ventilated. For comparison purposes, we have taken the number of patients in ICU as the sum of patients in the different ICU wards, and the number of patients in hospital as the sum of ICU and non-ICU patients. They provide predictions for the number of patients in each state, 10 days ahead from April 20, 2020 (period 1) and from April 27, 2020 (period 2). Since they report actual information of hospital occupancy, we have computed the relative errors and, then, the mean error for both periods has been obtained. The comparison with our tool using Holt-Winters predictions for positive cases is in Table 12, where we can see lower errors with our method.

In Table 13 we compare our results with those given in Bekker et al. [28]. They compute predictions for ICU and non-ICU occupation, in 3 and 7 days ahead and they report errors by using the actual hospital (ICU and non-ICU) arrival process and the predictions

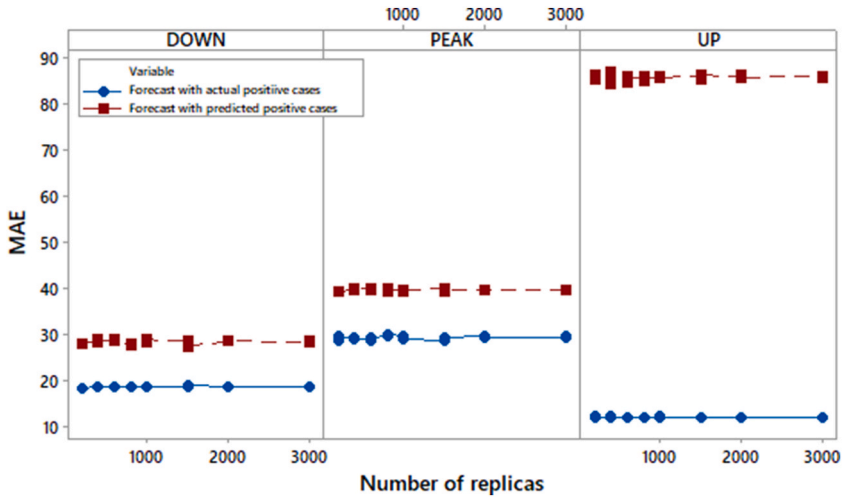

Fig. 8. Sensitivity analysis of MAE depending on number of replications (nsim) in the simulation process.

Table 10
MAE for hospital and ICU occupancy forecasts for 14 days, in the sensitivity study.

	UP				PEAK				DOWN			
	Hospital		ICU		Hospital		ICU		Hospital		ICU	
	actual	HW	actual	HW	actual	HW	actual	HW	actual	HW	actual	HW
Full Period	*12.1*	*85.8*	*3.0*	*4.8*	*29.4*	*39.8*	*6.1*	*6.0*	*18.7*	*28.2*	*2.9*	*2.2*
4-month	*11.7*	*86.1*	*2.3*	*5.8*	*29.0*	*37.8*	*6.1*	*5.8*	*19.6*	*27.0*	*2.9*	*2.2*
3-month	*11.8*	*87.6*	*2.4*	*6.1*	*30.2*	*38.3*	*6.6*	*6.4*	*20.4*	*25.3*	*3.0*	*2.2*
2-month	*20.7*	*76.5*	*6.3*	*9.8*	*47.9*	*55.4*	*2.7*	*2.6*	*18.6*	*34.8*	*3.7*	*2.7*

Table 11
Comparison study with ICU occupancy reported in Goic et al. [23].

	Goic et al. [23]		Our model	
	7 days	14 days	7 days	14 days
MAPE	4.11%	9.03%	3.69%	6.09%

Table 12
Comparison study in terms of MAPE with Hospital and ICU occupancy in [27].

	Deschepper et al. [27]			Our model
	Period 1	Period 2	Mean value	10 days
Hospital	21.84%	27.10%	24.47%	6.87%
ICU	10.20%	62.10%	36.15%	4.11%

Table 13
Comparison study in terms of MAE/mean with ICU and Non_ICU occupancy reported in Ref. [24]. Actual and predicted, in Ref. [28], stands for the use of the actual arrival process to hospital or the predicted values as input in the model, respectively; whereas in our model refers to actual or predicted positive cases.

	Bekker et al. [28]				Our model			
	3 days		7 days		3 days		7 days	
	Actual	Pred.	Actual	Pred.	Actual	Pred.	Actual	Pred.
Non-ICU	6%	8%	7%	13%	3.78%	5.05%	3.85%	7.14%
ICU	2%	3%	2%	9%	4.22%	3.91%	4.33%	3.69%

of the process. The measure used in this work is the ratio of MAE and the actual mean occupancy in the prediction period. The comparison results (Table 13) show that from actual to predicted data, our model has a smaller variation regardless of the prediction days. Besides this, Non-ICU forecasts are more accurate in our model, regardless of the input data, and, also, ICU forecasts are better in 7 days.

Table 14 corresponds to the comparison with Roimi et al. [14]. In that paper, the authors consider 3 clinical states: moderate, severe and critical in terms of the Israel Ministry of Health criteria. We have identified critical patients with our ICU patients. The predictions reported in the work are computed, on the one hand, considering no more future incoming patients, which is called Snapshot, simulating an emptying process of the hospital and, on the other hand, for hospital arrivals occupancy, that is, the evolution of the hospital arrivals in the prediction period. This corresponds with the model setup "Hospitalized" studied in Table 9 where state 1 in Fig. 1 is eliminated.

The forecast periods in Roimi et al. [14] have length 16 and 32 days for Snapshot, and 64 days for hospital arrivals. We have computed predictions with the same number of days for Snapshot, from February 11, 2021 to February 28, 2021 (16 days, Snapshot 1) and from January 28, 2021 to February 28, 2021 (32 days, Snapshot 2) and, finally, for hospital arrivals from December 27, 2020 to February 28, 2021 (64 days).

The accuracy of the results in Ref. [14] is measured in terms of absolute errors (MAE) and since the actual occupancy is not reported, the relative errors (MAPE) can not be computed. Instead, we have shown in Table 14 absolute errors (MAE) and a relative error in terms of the maximum occupation. Note that although our MAE are bigger than those reported by Roimi et al. [14], hospital and ICU occupancy in our case is much higher than in that work; when the MAE are normalized by the maximum occupancy, the result shows that our model gets better predictions.

4.4. Limitations

Our study has some limitations. We did not consider recurrent transitions into the states; in particular, we did not consider readmissions. Although readmissions due to COVID-19 can be recorded, it is not clear if a readmission is due to a worsening of the patient's primary infection by COVID-19, long run effects of the virus or complications of other pathologies.

Another limitation of our study is that the accuracy of the predictions are highly dependent on a good prediction of new positive cases. We used the HW predictions to illustrate our analysis and to provide estimates to validate the model, but we found some scenarios where long run prediction is not too good.

Finally, our tool was validated only with Aragón data. We hope that the interest and flexibility of our tool will be an incentive to validate it in other contexts. The main challenge for its use is having an up-to-date database with the information of positive cases.

5. Conclusions

We developed a free online statistical tool (https://github.com/peterman65) to forecast the number of occupied COVID-19 hospital and ICU beds in a health system. The input data for the tool is the evolution of a cohort of diagnosed positive cases, including at least dates of diagnosis, admission to hospital, admission to ICU, discharge from ICU, and discharge from hospital (possibly censored). In addition, the tool can consider groups of patients with similar characteristics (sex, age or comorbidities) and give predictions for each group.

The tool was based on a multistate model wherein the transition probabilities between states were estimated using statistical techniques for survival analyses, including cure model and competing risks. The tool processes new positive cases during the forecasting period; which can be provided either by the user or estimated by the tool using time-series methodology. We validate the performance of the prediction tool using patients with confirmed SARS-CoV-2 infections, from July 1, 2020 to February 28, 2021, who were extracted from the Aragón Healthcare Records Database, which includes information on 21 hospitals. We obtained good performance for predictions of hospital and ICU occupancy at 7, 14 and 52 days. Yet, as expected, we obtained a less accurate prediction using the positive case estimates provided by the Holt-Winters technique, especially from the 7th day prediction.

As a strength, our model was based on a nonparametric approach. This allows it to adapt to possible changes in a dynamic way, since it does not require assumptions about distributions or model parameters. COVID-19 has spread in several waves, each with its own specificities. The model incorporates these specificities into its estimates when a cohort of patients from the new wave feeds the model. Thus, the model is flexible enough to fit the probability transition between the states to changes in the evolution of COVID-19.

The inclusion of groups is another strength as health measures and vaccination policies are likely to change the structure of future

Table 14
Comparison study with ICU and Hospital occupancy reported in Roimi et al. [14].

	Maximum occupancy				MAE				MAE/Maximum occupancy			
	Hospital		ICU		Hospital		ICU		Hospital		ICU	
	[14]	Our	[14]	Our	[14]	Our	[14]	Our	[14]	Our	[14]	Our
Arrivals	100	716	25	80	4.72	17.21	1.68	3.98	4.72%	2.40%	6.72%	4.97%
Snapshot 1	90	622	15	90	3.15	20.19	1.47	4.44	3.50%	3.25%	9.80%	4.93%
Snapshot 2	85	631	20	82	3.13	12.38	1.98	2.83	3.68%	1.96%	9.90%	3.45%

positive cases, so groups can be used to account for these changes. In addition, grouping can be used for a tighter prediction of hospital occupancy in a scenario-based decision-making process, defined through the evolution of positive cases in groups.

The tool can be applied to obtain short and mid-term predictions with a good accuracy in different situations, as shown in Section 4.3. It also can be used for long-term forecasts in different hypothetical scenarios by taking the corresponding set of new positive cases; for instance, it can be used to predict the consequences (in terms of hospital and ICU occupancy) of a new wave of infections.

In addition, the input of the tool can be the positive cases or the hospital arrivals, see the paragraph "Ablation and sensitivity analysis". This is a generalization with respect to the majority of the models developed in the literature, where the new hospitalizations are used as the starting point of the model.

Finally, as in any statistical model, the cohort used for estimation should reflect the behavior of the forecast period. It is therefore important to validate the tool to check whether there have been any structural changes in recent weeks that may make the prediction less accurate. As discussed above, changes in the age structure of positive cases can be dealt with by choosing appropriate groups in the set of new positive cases. The validation option of the tool allows the user to detect such situations.

The information provided by the model, the prediction of the daily number of ICU and non-ICU beds occupied by COVID-19 patients will help hospital managers to make decisions about the required beds to face future admissions.

Author contribution statement

Miguel Lafuente; Francisco Javier López; Pedro Mariano Mateo; Ana Carmen Cebrián: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Jesús Asín; José Antonio Moler; Ángel Borque-Fernando; Luis Mariano Esteban; Ana Pérez-Palomares; Gerardo Sanz: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This work was supported by Gobierno de Aragón [E46-20R] and Ministerio de Ciencia e Innovación [PID2020-116873GB-I00].

Data availability statement

The authors do not have permission to share data.

Declaration of interest's statement

The authors declare no conflict of interest.

Informed consent statement

Due to the retrospective, observational nature of this study, the data could be fully anonymized and informed consent was waived.

Acknowledgments

The authors wish to thank the health system staff for the excellent medical care provided to their patients at considerable personal risk. The authors also thank the anonymous reviewers for their careful reading and valuable comments and suggestions, which have improved the presentation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e13545.

References

[1] C.R. Butler, S.P. Wong, A.G. Wightman, A.M. O'Hare, US clinicians' experiences and perspectives on resource limitation and patient care during the COVID-19 pandemic, JAMA Netw. Open 3 (11) (2020), e2027315, https://doi.org/10.1001/jamanetworkopen.2020.27315.

[2] D. Bertsimas, L. Boussioux, R. Cory-Wright, A. Delarue, V. Digalakis, et al., From predictions to prescriptions: a data-driven response to COVID-19, Health Care Manag. Sci. 24 (2021) 253–272, https://doi.org/10.1007/s10729-020-09542-0.

[3] S. Larabi-Marie-Sainte, S. Alhalawani, S. Shaheen, K. Almustafa, T. Saba, et al., Forecasting COVID19 parameters using time-series: KSA, USA, Spain and Brazil comparative case study, Heliyon 8 (2022), e09578, https://doi.org/10.1016/j.heliyon.2022.e09578.

[4] J.E. Gnanvi, K.V. Salako, G.B. Kotanmi, R.G. Kakaï, On the reliability of predictions on COVID-19 dynamics: a systematic and critical review of modelling techniques, Infect Dis Model 6 (2021) 258–272, https://doi.org/10.1016/j.idm.2020.12.008.

[5] W. Zhang, S. Liu, N. Osgood, H. Zhu, Y. Qian Y, P. Jia, Using simulation modelling and systems science to help contain COVID-19: a systematic review, Syst. Res. Behav. Sci. (2022) 1–28, https://doi.org/10.1002/sres.2897.

[6] S. Nickbakhsh, A. Ho, D.F.P. Marques, J. McMenamin, R. Gunson, et al., Epidemiology of seasonal coronaviruses: establishing the context for the emergence of coronavirus disease 2019, J. Infect. Dis. 222 (1) (2020) 17–25, https://doi.org/10.1093/infdis/jiaa185.

[7] Y. Li, X. Wang, H. Nair, Global seasonality of human seasonal coronaviruses: a clue for postpandemic circulating season of severe acute respiratory syndrome coronavirus 2? J. Infect. Dis. 222 (7) (2020) 1090–1097, https://doi.org/10.1093/infdis/jiaa436.

[8] S. Kashyap S. Gombar, S. Yadlowsky, A. Callahan, J. Fries, et al., Measure what matters: counts of hospitalized patients are a better metric for health system capacity planning for a reopening, J. Am. Med. Inf. Assoc. 27 (7) (2020) 1026–1031, https://doi.org/10.1093/jamia/ocaa076.

[9] J. Caro, J. Möller, V. Santhirapala, H. Gill, J. Johnston, et al., Predicting hospital resource use during COVID-19 surges: a simple but flexible discretely integrated condition event simulation of individual patient-hospital trajectories, Value Health 24 (11) (2021) 1570–1577, https://doi.org/10.1016/j.jval.2021.05.023.

[10] S.R. Knight, A. Ho, R. Pius, I. Buchan, G. Carson, et al., Risk stratification of patients admitted to hospital with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score, BMJ 370 (2020) m3339, https://doi.org/10.1136/bmj.m3339.

[11] M.W. Kattan, J. Xinge, A. Milinovich, J. Adegboye, A. Duggal, et al., An algorithm for classifying patients most likely to develop severe coronavirus disease 2019 illness, Crit Care Explor 2 (12) (2020), e0300, https://doi.org/10.1097/CCE.0000000000000300.

[12] R. Aznar-Gimeno, L.M. Esteban, G. Labata-Lezaun, R. del Hoyo, D. Abadia, et al., A clinical decision web to predict ICU admission or death for patients hospitalised with COVID-19 using machine learning algorithms, Int. J. Environ. Res. Publ. Health 18 (16) (2021) 8677, https://doi.org/10.3390/ijerph18168677.

[13] M. El Halabi, J. Feghall, J. Bahk, P. Tallón de Lara, B. Naraslmhan, et al., A novel evidence-based predictor tool for hospitalization and length of stay: insights from COVID-19 patients in New York City, Intern. Emerg. Med. 17 (2022) 1879–1889, https://doi.org/10.1007/s11739-022-03014-9.

[14] M. Roimi, R. Gutman, J. Somer, A.B. Arie, I. Calman, et al., Development and validation of a machine learning model predicting illness trajectory and hospital utilization of COVID-19 patients: a nationwide study, J. Am. Med. Inf. Assoc. 28 (6) (2021) 1188–1196, https://doi.org/10.1093/jamia/ocab005.

[15] M.G. Klein, C.J. Cheng, E. Lii, K. Mao, H. Mesbahi, et al., COVID-19 models for hospital surge capacity planning: a systematic review, Disaster Med. Public Health Prep. 16 (1) (2022) 390–397, https://doi.org/10.1017/dmp.2020.332.

[16] E.M. Rees, E.S. Nightindale, Y. Jafari, N.R. Waterlow, S. Clifford, et al., COVID-19 length of hospital stay: a systematic review and data synthesis, BMC Med. 18 (2020) 270, https://doi.org/10.1186/s12916-020-01726-3.

[17] Y. Alimohamadi, E. Mansouri, M. Sepandi, M. Sharafod, M. Arshade, et al., Hospital length of stay for COVID-19 patients: a systematic review and meta-analysis, Multidiscip. Respir. Med. 17 (1) (2022) 856, https://doi.org/10.4081/mrm.2022.856.

[18] G. Rainisch, E.A. Undurraga, G. Chowell, A dynamic modeling tool forestimating healthcare demand from the COVID19 epidemic and evaluating population-wide interventions, Int. J. Infect. Dis. 96 (2020) 376–386, https://doi.org/10.1016/j.ijid.2020.05.043.

[19] K.J. Locey, T.A. Webb, J. Khan, A.K. Antony, B. Hota, An interactive tool to forecast US hospital needs in the coronavirus 2019 pandemic, JAMIA Open 3 (4) (2020) 506–512, https://doi.org/10.1093/jamiaopen/ooaa045.

[20] P.Y. Boëlle, T. Delory, X. Maynadier, C. Janssen, R. Piarroux, et al., Trajectories of hospitalization in COVID-19 patients: an observational study in France, J. Clin. Med. 9 (10) (2020) 3148, https://doi.org/10.3390/jcm9103148.

[21] H. Diaz, G. España, N. Castañeda, L. Rodríguez, F. de la Hoz, Dynamical characteristics of the COVID-19 epidemic: estimation from cases in Colombia, Int. J. Infect. Dis. 105 (2021) 26–31, https://doi.org/10.1016/j.ijid.2021.01.053.

[22] A. Farcomeni, A. Maruotti, F. Divino, G. Jona-Lasinio, G. Lovison, An ensemble approach to short-term forecast of COVID-19 intensive care occupancy in Italian regions, Biom. J. 63 (3) (2021) 503–513, https://doi.org/10.1002/bimj.202000189.

[23] M. Goic M, M.S. Bozanic-Leal, M. Badal, L.J. Basso, COVID-19: short-term forecast of ICU beds in times of crisis, PLoS One 16 (1) (2021), e0245272, https://doi.org/10.1371/journal.pone.0245272.

[24] A. Comunian, R. Gaburro, M. Giudici, Inversion of a SIR-based model: a critical analysis about the application to COVID-19 epidemic, Physica D 413 (2020), 132674, https://doi.org/10.1016/j.physd.2020.132674.

[25] R.H. Stern, Locally informed simulation to predict hospital capacity needs during the COVID-19 pandemic, Ann. Intern. Med. 173 (8) (2020) 679–680, https://doi.org/10.7326/L20-1061.

[26] Garcia-Vicuña, L. Esparza, F. Mallor, Hospital preparedness during epidemics using simulation: the case of COVID-19, Cent. Eur. J. Oper. Res. 30 (2022) 213–249, https://doi.org/10.1007/s10100-021-00779-w.

[27] M. Deschepper, K. Eeckloo, S. Malfait, D. Benoit, S. Callens, et al., Prediction of hospital bed capacity during the COVID- 19 pandemic, BMC Health Serv. Res. 21 (1) (2021) 468, https://doi.org/10.1186/s12913-021-06492-3.

[28] R. Bekker, M. Broek, G. Koole, Modeling COVID-19 hospital admissions and occupancy in The Netherlands, Eur. J. Oper. Res. 304 (1) (2023) 207–218, https://doi.org/10.1016/j.ejor.2021.12.044D.

[29] M. Bicher, M. Zuba, L. Rainer, F. Bachner, C. Rippinger, et al., Supporting COVID-19 policy-making with a predictive epidemiological multi-model warning system, Commun. Med. 2 (2022) 157, https://doi.org/10.1038/s43856-022-00219-z.

[30] S. Shafiekhani, P. Namdar, S. Rafiei, A COVID-19 forecasting system for hospital needs using ANFIS and LSTM models: a graphical user interface unit, Digit. Health 8 (2022), 20552076221085057, https://doi.org/10.1177/20552076221085057.

[31] W. Jin, S. Dong, C. Yu, Q. Luo, A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning, Comput. Biol. Med. 146 (2022), 105560, https://doi.org/10.1016/j.compbiomed.2022.105560.

[32] M.B. Palermo, L. Micol, C. André, R. da Rosa, Tracking machine learning models for pandemic scenarios: a systematic review of machine learning models that predict local and global evolution of pandemics, Netw. Model. Anal. Health 11 (2022) 40, https://doi.org/10.1007/s13721-022-00384-0.

[33] R. Delli Compagni, Z. Cheng, S. Russo, T.P. Van Boecke, A hybrid Neural Network-SEIR model for forecasting intensive care occupancy in Switzerland during COVID-19 epidemics, PLoS One 17 (3) (2022), e0263789, https://doi.org/10.1371/journal.pone.0263789.

[34] L. Xu, R. Magar R, A.B. Farimani, Forecasting COVID-19 new cases using deep learning methods, Comput. Biol. Med. 144 (2022), 105342, https://doi.org/10.1016/j.compbiomed.2022.105342.

[35] M.J. Crowther, P.C. Lambert, Parametric multistate survival models: flexible modelling allowing transition-specific distributions with application to estimating clinically useful measures of effect differences, Stat. Med. 36 (29) (2017) 4719–4742, https://doi.org/10.1002/sim.7448.

[36] J. Beyersmann, M. Wolkewitz, A. Allignol, N. Grambauer M. Schumacher, Application of multistate models in hospital epidemiology: advances and challenges, Biom. J. 53 (2) (2017) 332–350, https://doi.org/10.1002/bimj.201000146.

[37] L. Yang, Y. Chen, X. Jiang, H. Tatano, Multistate models for the recovery process in the COVID-19 context: an empirical study of Chinese enterprises, Int. J. Disaster Risk Sci 13 (2022) 401–414, https://doi.org/10.1007/s13753-022-00414-5.

[38] Y. Peng, B. Yu, Cure Models. Methods, Applications and Implementation, CRC Press. Chapman & Hall, 2021.

[39] M. Pedrosa-Laza, A. López-Cheda, R. Cao, Cure models to estimate time until hospitalization due to COVID-19. A case study in Galicia (NW Spain), Appl. Intell. 52 (1) (2022) 794–807, https://doi.org/10.1007/s10489-021-02311-8.

[40] J.M.G. Taylor, Semi-parametric estimation in failure time mixture models, Biometrics 51 (3) (1995) 899–907, https://doi.org/10.2307/2532991.

[41] J. Beyersman, T.H. Scheike, Classical Regression Models for Competing Risks, Champan & Hall/CRC, 2013.

[42] B. Vekaria, C. Overton, A. Wisniowski, S. Ahmad, A. Aparicio-Castro, et al., Hospital length of stay for COVID-19 patients: data-driven methods for forward planning, BMC Infect. Dis. 21 (2021) 700, https://doi.org/10.1186/s12879-021-06371-6.

[43] D.W. Hosmer Jr., S. Lemeshow, S. May, Applied Survival Analysis: Regression Modeling of Time- To-Event Data, John Wiley, New York, 1999.

[44] C.C. Holt, Forecasting trends and seasonal by exponentially weighted averages, Int. J. Forecast. 20 (1) (1957) 5–10, https://doi.org/10.1016/j.ijforecast.2003.09.015.

[45] P.R. Winters, Forecasting sales by exponentially weighted moving averages, Manag. Sci. 6 (3) (1960) 324–342, https://doi.org/10.1287/mnsc.6.3.324.

[46] A. Tobías, Evaluation of the lockdowns for the SARS-CoV-2 epidemic in Italy and Spain after one month follow up, Sci. Total Environ. 725 (2020), 138539, https://doi.org/10.1016/j.scitotenv.2020.138539.
[47] The national COVID-19 outbreak monitoring group, COVID-19 outbreaks in a transmission control scenario: challenges posed by social and leisure activities, and for workers in vulnerable conditions, Spain, early summer Eurosurveillance 25 (35) (2020), 2001545, https://doi.org/10.2807/1560-7917.ES.2020.25.35.2001545.
[48] Ministerio de Sanidad de España. Documentos técnicos para profesionales. https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos.htm (Accessed 18 December 2022).