Case Study: Hospital demand forecast during the COVID-19 pandemic (Catalonia - Spain)

Vázquez-Méndez, Asunción
sunivazquez@hotmail.com

González-Camblor, Manuel
manuelgonzcam@hotmail.com

Oliveras-Alsina, Pere
pereoliveras@uoc.edu

González-Corral, Jose Manuel
jmgcorral@gmail.com

May 19, 2020

Abstract

Background: The pandemic caused by COVID-19 has challenged hospital managers to deal with the planning of their resources to meet a never-before-seen demand for unscheduled hospitalizations and intensive care. In Spain, highly hit for the pandemic, Government set strict measures of social distancing.

Objective: Assist a professional in the management of hospital resources in order to define dimensions and make decisions given the extraordinary demand caused by COVID-19.

Design: SIR model parameters fitting based on local data.

Results: We provide the hospital with the parameters required to test different scenarios of social contact reduction.

Settings:

Region - 3 hospitals in Catalonia (Spain), covering 200.000 people.

Data period time - March 30th - April 14th.

Lock-down date - March 14th (tightened measures from March 30th to April 9th, nowadays facing measures lightening).

Limitations: Due to urgent need of insights, daily update of information and closeness to hospitalization census peak, we limited our analysis to the adjustment of the SIR model to data.

We didn't have access to ICU data in the moment when the analysis was performed so we couldn't fit severity parameters.

Conclusion: We assisted the hospital managers on providing paremeter values to launch SIR model simulations with the aim of making informed decisions on resources needs.

We found that lock-down impact on R_0 was about 50% reduction when starting to take effect.

1 Introduction

The pandemic caused by COVID-19 has placed the world's health care systems under a level of stress that is, in many cases, unprecedented. The urgent demand for tests, protective equipment, respirators and specialized medical facilities has forced governments and institutions to work in a coordinated manner to provide provisions to health care centers. It is the managers of the hospitals themselves, however, that must deal with the planning of their resources to meet a never-before-seen demand for unscheduled hospitalizations and intensive care.

These key actors in the crisis must be able to simulate scenarios for the pandemic's evolution in order to anticipate situations of extreme saturation through measures such as home hospitalization, referral to other centers or cancellation of interventions and non-urgent appointments.

We have collaborated with a network of hospitals in Catalonia (Spain) to provide them, even given the uncertainty regarding key aspects of the SARS-CoV-2 coronavirus, with a daily forecast of the hospital census and patients in intensive care, taking into account the approximate hospitalization date for the first person tested positive for COVID-19, the severity of the disease, and the isolation measures decreed. Our contribution is focused on the parameterization of the CHIME model (*COVID-19 Hospital Impact Model for Epidemics*) developed by Penn Medicine (University of Pennsylvania Health System)

2 State of the Art

Mathematical modeling has been a key factor in the development of Epidemiology. The first work on smallpox, formulated in the 18th century, paved the way, in the 20th century, for the birth of the deterministic modeling of epidemics. This has been fundamental in the fight against diseases such as malaria and measles.

Today, the mathematical models used in Epidemiology can be broadly classified into [2]:

- **Compartmental models**: divide the population into groups, depending on their status in the epidemic (Susceptible, Infected, Recovered, etc.) and model the transition of individuals between groups.
- **Models focused on individuals**: are based on numerical simulations that take into account each individual's behavior.
- **Hybrid models**: combine the two previous approaches.

2.1 Epidemiological calculators on the Internet

The rise of the Internet, together with the *opensource* movement, have democratized scientific knowledge and have helped to connect professionals from diverse disciplines in order to develop and evolve all kinds of models and applications.

For the subject at hand, we can find, at the click of a button, epidemiological calculators that integrate deterministic evolution models (SIR/SEIR) for the COVID-19 pandemic.

Three of the most representative are:

- **Goh Epidemic Calculator**

  Developed by Gabriel Goh, researcher at OpenAI, this integrates a SEIR model (*Susceptible -> Exposed -> Infected -> Recovered*) and predicts the pandemic's evolution for a certain population.

  Its inputs are variables related to disease transmission and severity. It takes into account the control measures applied, which are aimed at reducing social contact.

- **COVID-19 Spread vs Healthcare Capacity Model** (Allison Hill)

  Also a SEIR model, but in this case different levels of infection severity are considered (textitI1: Mild, I2: Severe, I3: Critical).

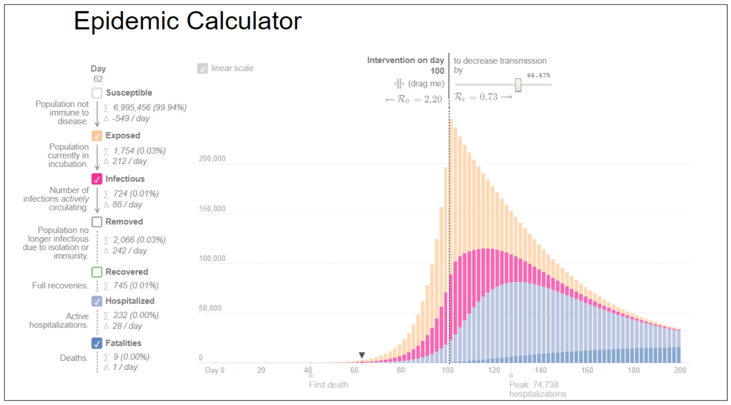

Figure 1: Goh Epidemic Calculator

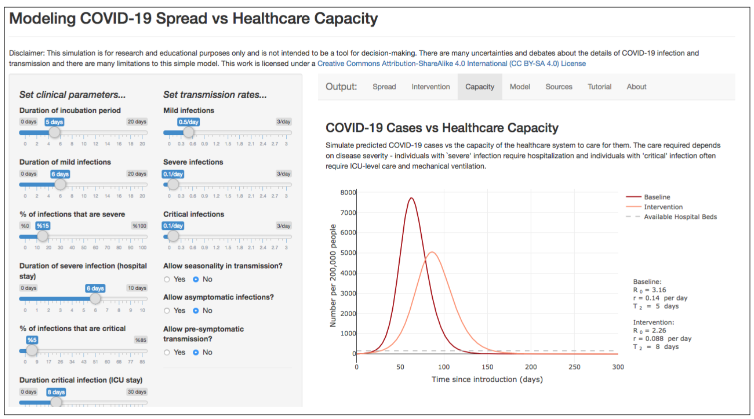

Figure 2: Hill Epidemiological Calculator

Consists of several tabs in which infection evolution can be observed without any control measure (*Spread*), as well as their influence according to the intensity with which they are applied (Intervention) and the saturation level of the health care centers (*Capacity*).

The author of the model is Allison Hill, a researcher at Harvard.

- **Penn University Locally Informed Simulation to Predict Hospital Capacity Needs**

  The Penn University epidemiological calculator is based on a SIR model and aims to predict the timing and magnitude of the peak of hospital admissions resulting from COVID-19. It also focuses on the volume of patients that will require intensive care and mechanical respiration.

Of course, the adoption of any of these tools to assist in crucial decision-making requires the informed use of a scientific procedure for adaptation to each specific case, and evaluation of how adequate the fit is.

These precautions are especially important at a time when research is moving forward at a frenetic pace due to the urgent need to anticipate what is going to happen: a large part of the new discoveries regarding COVID-19 are pending review.

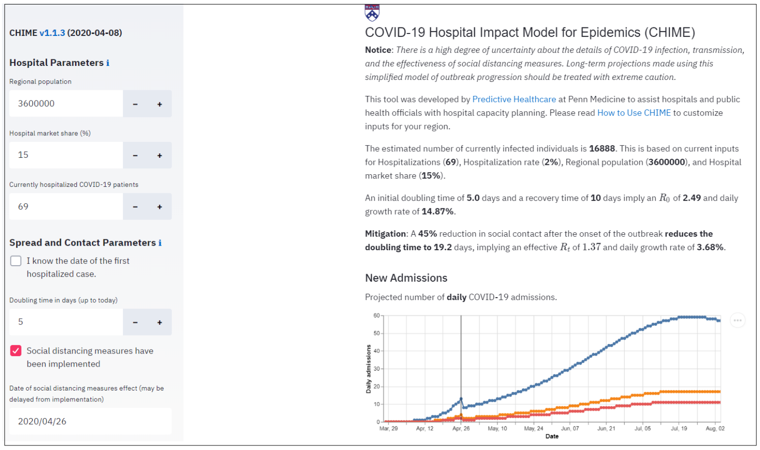

Figure 3: Penn Chime graphical interface

2.2 PENN CHIME

Developed by scientists at the University of Pennsylvania (Perelman School of Medicine and Penn Medicine Predictive Healthcare), Penn CHIME is a SIR model for estimating COVID-19's impact on the resources of the 3 hospitals in the area of influence.

The objective pursued was the prediction of the hospitalization peak for virus-positive patients and the number of ICU beds and ventilators needed to meet demand. The team worked against the clock: just two days after the team was commissioned to create projections of the pandemic's evolution, it built the model and integrated it into a scenario visualization tool.

This tool, available on the Internet, can be adapted to any hospital institution in the world. The code is open and available on GitHub. There is also a small community on Slack to share knowledge and consult with experts and other contributors in regards to the model's technical details and its application in other states and countries. For all these reasons, we chose to focus our efforts on parameterizing this model to the specific case of the hospitals with which we worked. In addition, the Goh model had been used for the hospital, and we think the possibility of comparing the results to be interesting.

3 SIR model: PENN CHIME

As we saw, the SIR epidemiological model is a *compartmental* model. Specifically, it divides the population into three groups:

- **S**usceptible individuals: those who lack immunity and therefore may become Infected if exposed to the infectious agent.
- **I**nfected individuals: those who suffer from the disease (whether or not they show symptoms) and are virus transmitters; that is, they can infect Susceptible individuals.
- **R**ecovered individuals: those who, having been infected, are no longer infected and considered themselves immune, either due to having overcome the disease or because they have died. (As of today, reinfection has not been ruled out for the SARS-CoV-2 coronavirus.)

With these premises, the growth and evolution of the disease, from an instant t to an instant $t{+}1$ is modeled according to the following equations:

$$S_{t+1} = S_t - \beta S_t I_t$$

$$I_{t+1} = I_t + \beta S_t I_t - \gamma I_t$$

$$R_{t+1} = R_t + \gamma I_t$$

β is the *effective contagion rate*: $\beta = \tau * c$, where τ is the *transmissibility* of the virus and c is the number of people exposed.

γ is the inverse of the *mean recovery time.*

Finally, the quotient $\frac{\beta}{\gamma}$ is what is known as R_0, the **basic reproductive number**. It represents the average number of people that will be infected by a single person.

As can be deduced, this number will be high if:

- the transmissibility of the pathogen is high,
- there are many people exposed, or
- the recovery time is long; that is, those infected are contagious for many days.

The **measures for restricting social contact** thus have a direct impact on R_0, as c (the number of people exposed) decreases.

The Penn CHIME model estimates the value R_0 and R_t (*contagion rate in moment* t) based on the **doubling time** at a given instant, T_d, and the **recovery time**, T_r:

$$R_0 = \frac{2^{\frac{1}{T_d}} - 1 + \frac{1}{T_r}}{\frac{1}{T_r}}$$

Based on the size of the population in the hospital's area of influence and on various indicators of disease severity, a predictive forecast is made for those admitted with COVID-19, as well as the number of people who will need intensive care.

The table 1 presents the complete list of parameters [3].

To these, we must add the reference date, *current date*, which is related to the volume of patients admitted. Furthermore, in one of the model's evolutions, the *date on which the social distancing measures took effect* was added as an input; we will examine this in detail in the corresponding section.

4 Our Contribution

Our work with the Penn CHIME model has focused on the proper selection of parameters to generate forecasts. Hospital management has provided us with information regarding the reference population, as well as objective data on admissions and tests.

We have processed this data in order to construct a time series for the hospital census of COVID-19 patients, based on which we have assessed the adjustment for the model (square root of the mean square error).

HOSPITAL	
Regional Population	Population size in the hospital area (initial value for S - Susceptible).
Hospital Market Share (%)	Quota of patients for the hospital in the region (can be estimated as the number of beds available with respect to the total number of beds in the region).
Currently Hospitalized COVID-19 patients	Number of COVID-19 patients admitted, at a given time.
EXPANSION AND CONTAGION	
Date of the first hospitalized case	*(Optional) Date of first hospitalization for COVID-19.*
Doubling time before current date	*(Does not apply if the first date of hospitalization for COVID-19 is known)* Number of days for the number infected to double. Related to R_0 and pushes the rate of new cases in the first phase of expansion.
Social distancing (% reduction in social contact)	The estimate of how much social contact is reduced in the region compared to no social distancing at all.
SEVERITY OF THE DISEASE	
Hospitalization % (total infection)	Percentage of infected cases requiring hospitalization.
ICU % (total infection)	Percentage of infected cases requiring intensive care.
Ventilated % (total infections)	Percentage of infected cases requiring mechanical ventilation.
Infectious days	Number of days that an infected person can infect another.
Average Hospital Length of stay (days)	Average number of days of admission for COVID-19 patients.
Average Days in ICU	Average number of ICU admission days for COVID-19 patients.
Average Days on Ventilator	Average number of days for which COVID-19 patients require mechanical ventilation.

Table 1: CHIME model parameters

We have also relied on the information managed by the hospital, published by the Generalitat de Catalunya and other sources, to define a range of possible values for the *doubling time* and *days of infection*.

We have been in continuous contact with the hospital, and have shared the optimal configuration, calibrated with the passage of time, with the hospital staff.

Coincidentally, the study began when the peak of new admissions was very near, meaning that the true application of the simulation was not prediction, but rather verification and anticipation of the resource decongestion rate, as well as analysis of the impact of modifications in social distancing policies.

5 Data

5.1 Description and transformation

The data used to determine the true census of hospitalized COVID-19 patients was as follows:

- **Admissions**: history of patient hospitalizations and discharges, since 2019, updated every few days.The fields most relevant for our study were:

  StartDate | *EndDate* | *FileID* | *Diagnosis* | *BedId* | *Age* | *CausDischarge*

  If a patient did not show an *EndDate*, it means that she remained hospitalized.

- **PCR test**: list of tests by date and file number.

  Up to three tests were performed per patient, and the results could be *Positive*, *Undetectable* or *No result*.

Construction of the COVID-19 patient time series required:

1. determination of the patients admitted on a specific date.
2. flagging of the positives for COVID-19.

Thus, on a specific date f computation of those admitted for COVID-19 was determined by the number of different *FileID* that they fulfilled:

- $StartDate \leq f$
- $EndDate > f$ or $EndDate = null$
- *Positive* result in any of the tests carried out.

That is, only those with a positive test were considered COVID-19 patients, and not all those who had, at the outset, a diagnosis compatible with the disease: pneumonia, respiratory failure, etc.

We created an interactive dashboard with the aim of allow the hospital to validate how we were interpreting data.

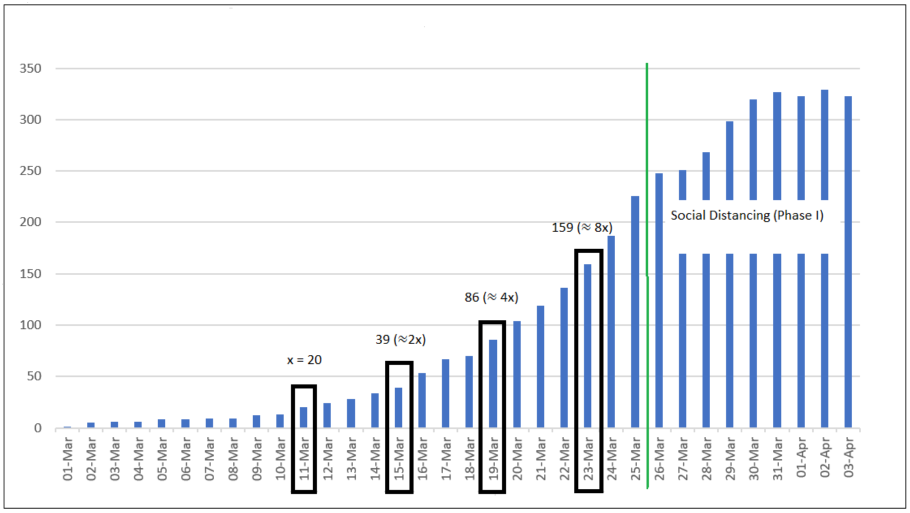

Figure 4: Doubling time for admitted patients

5.2 Problems associated with the nature of the data

As mentioned, flagging of the positives was directly associated with the test results. Despite the adoption of objective determination criteria, it was sometimes necessary to repeat the test several times, and the result could take several days. This caused the information to change in the subsequent data updates that we received.

In general, the number of positives increased with each new version. The final days of data were particularly affected, and could be subject to an increase of 15% to 20%.

The data for the final days of each update were thus not used to monitor the model fit.

6 Parameterization

- The hospital's reference population is 200,000 inhabitants (100% Market Share).
- Hospital management estimated that the percentage of infected patients requiring hospitalization was around 2.5%; of these, 10% may require intensive care (and mechanical ventilation).
- The average length of stay was about 10 days (14 for severe, 9 for mild).

6.1 Date of first admission vs. Doubling Time

The **first PCR tests** began to be administered on **March 17**, however, many of the patients who tested positive on that date and on subsequent dates had already been hospitalized for some time. Determination of the date of the first case may thus be inaccurate, so we opted for the doubling time criteria: that is, the time it takes for the number of people infected to double. In this case, our reference is the number of people hospitalized (which represents a percentage of the total number of people infected).

In the dates prior to confinement, we can observe that **the number of people hospitalized doubled every 4 days** (Figure 4).

R0 (~ 2.3)	Td=2	Td=3	Td=4	Td=5	Td=6	Td=7	Td=8
Id=5	3.071	2.300	1.946	1.743	1.612	1.520	1.453
Id=6	3.485	2.560	2.135	1.892	1.735	1.625	1.543
Id=7	3.899	2.819	2.324	2.041	1.857	1.729	1.634
Id=8	4.314	3.079	2.514	2.190	1.980	1.833	1.724
Id=9	4.728	3.339	2.703	2.338	2.102	1.937	1.815
Id=10	5.142	3.599	2.892	2.487	2.225	2.041	1.905
Id=11	5.556	3.859	3.081	2.636	2.347	2.145	1.996
Id=12	5.971	4.119	3.270	2.784	2.470	2.249	2.086
Id=13	6.385	4.379	3.460	2.933	2.592	2.353	2.177
Id=14	6.799	4.639	3.649	3.082	2.714	2.457	2.267

Figure 5: Relationship between R0, doubling time and recovery days

Often, the doubling time observed in those hospitalized tends to be significantly less than the doubling time for positives [3]; for the Penn Chime model, this parameter, together with the recovery time, provides the value R_0.

The public health management body in Catalonia (CatSalud) estimated [1] that R_0 **had an approximate value of 2.3**. We generated the table of R_0's for a wide range of doubling and recovery time values (table 5); the observed value of 4 days of doubling time corresponds to **about 7 days of recovery time**.

As we can see, there are other parameter combinations that give an R_0 result close to 2.3. The prediction will be similar for any of these, but our main objective was to feed the model with a proper R_0 and not going further in infectious day or doubling time estimation.

6.2 Initial phase and reduction of social contact

The model requires the establishment of a reference date that is prior to the date that measures to reduce social contact were put into effect; that is, at the time when the pandemic was spreading naturally. For example, on **March 11th**, in full exponential growth, there were **20 people** admitted with COVID-19.

CatSalud [1] estimated that from March 23rd onwards, the effects of confinement began to be seen. However, in the case of our reference hospital, the decrease in the growth rate appears to be around March 26th/27th, as we can see in Figure 6.

6.2.1 Reduction of social contact

The percentage of reduction in social contact, obligated by the government-established confinement, modifies the value of R_0, with a direct impact on the rate of contagion.

This occurred in two phases of distinct intensity and duration: the initial phase, affecting mobility, schools, shops, etc. and another, more restrictive, phase, which limited the rest of non-essential activities (industries deemed non-essential) and lasted two weeks, before reverting back to the measures of the previous phase.

The ability to generate predictions with different restriction levels is one of the great strengths of the CHIME calculator, since it allows the epidemic's growth to be compared according to different strategies that could be adopted by governments.

In fact, organizations such as the Imperial College [4], have carried out studies to assess the opportunity and effects of adopting more or less restrictive measures to limit economic activity, school

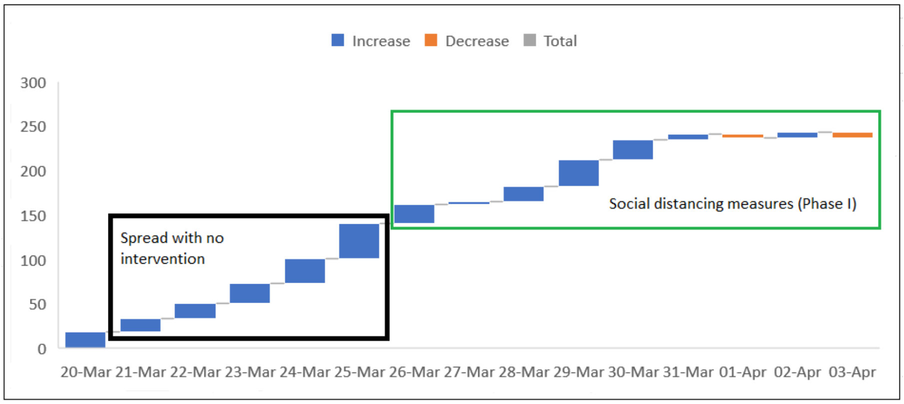

Figure 6: Evolution of admitted patients

closings, and the confinement of the elderly, in combination with hygiene measures, information provided to citizens, etc.

To aid in calculating the effective impact of the reduction in contact during the pandemic, Google created a mobility report for each country, which was broken down by region.

The mobility report results for Catalonia indicate the following:

- an 81% reduction in non-food and leisure shops.
- a 73% reduction in transportation stations.
- a 65% reduction in workplaces.
- a 63% reduction in parks.
- a 24% reduction in pharmacies and food stores.

In contrast, mobility in residential areas increased by 30%.

There are obvious limitations in establishing a reduction percentage that affects everyone equally, during a pandemic in which highly localized foci have been detected; we opted to carry out **different simulations with values of decreased social contact between 40% and 70%**.

7 Results

To execute the different simulations, we used the Python version of the original model, with some improvements from its contributors.

Our development consisted of transforming the model into a Sklearn Regressor; this allowed us to use the available methods for error calculation (RSME), etc.

The model fit, in the pandemic free growth phase, is very high; this allows us confidence in the selection of the parameters made, especially in the R_0 value provided by CatSalud.

In regard to the different reduction scenarios, the results of the CHIME model, when **compared to the actual data, reflected an effective reduction of close to 50%** (Figure 7). In this scenario, the *curve* "flattened", avoiding a very alarming growth in the number of cases, only with slightly less restrictive measures (40% - peak of 464 infected on April 21st).

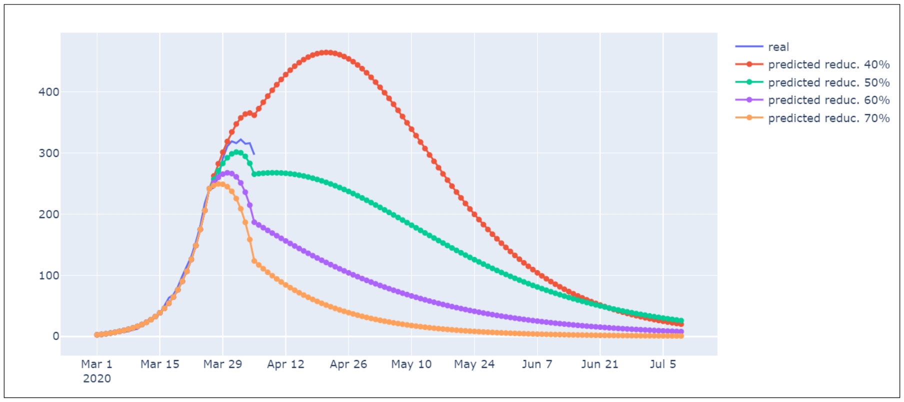

Figure 7: Social contact reduction scenarios

The **real peak** of those infected occurred in the hospital on **around April 2nd, with 322 people hospitalized**; table 2 shows when this maximum would have been reached and with what values for the different scenarios, as well as the estimated mean error.

% Reduction	Peak Date	Peak Hospitalizations	RSME
40	2020-04-21	464.40	19.00
50	2020-04-01	301.42	11.71
60	2020-03-30	267.52	34.48
70	2020-03-28	249.28	54.31

Table 2: Prediction error according to distancing

Most likely, the impact of the measures expected by the experts, was not drastic enough. A tightening of these measures was thus established, halting all non-essential economic activity for 2 weeks. These measures took effect on March 29th, and had an effect not so much on the maximum number of people hospitalized as on the rate of decline in new cases.

With these new measures, it was no longer possible to continue adjusting the model as it was configured.

8 Conclusions

The results of the PENN CHIME predictive model were compared with other predictive models created internally by the Department of Epidemiology and hospital management support, as well as those created by regional entities such as CatSalut.

Using these predictive models, the possible pandemic's behaviour could be predicted in the short term, and different scenarios could be simulated depending on the variables introduced into the model. Special interest was given to the social distancing variable and to how it affected the prediction of cases and the curve's evolution. We found that the model adjusted pretty well setting a 50% social contact reduction.

Based on the different predictive scenarios, the number of operational beds needed was calculated, as well as the time required for them to be available (space adjustments, acquisition of new units). According to the curve's evolution, these openings were likewise adapted, as well as the needs of the healthcare personnel.

References

[1] Generalitat de Catalunya: CatSalud. Coronavirus SARS-CoV-2 situació actual a Catalunya: 24 de març de 2020.

[2] M. González-Arroyo. *Modelización y simulación en epidemiología*. Trabajo de Fin de Grado, Universidad Complutense de Madrid, 2017.

[3] Penn Medicine. The CHIME App: Model parameters.

[4] Gemma Nedjati-Gilani et al. Neil M Ferguson, Daniel Laydon. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand.