

Article

Not peer-reviewed version

Impact and Effectiveness of COVID-19 Vaccines Using Machine Learning to Analyze a Time Series: A Population-Based Study

[Rafael Garcia-Carretero](#)*, [Maria Ordoñez-Garcia](#)*, Oscar Vazquez-Gomez, Belen Belen Rodriguez-Maya, [Ruth Gil-Prieto](#), [Angel Gil-de-Miguel](#)

Posted Date: 16 September 2024

doi: 10.20944/preprints202409.1247.v1

Keywords: COVID-19; vaccines; SARS-CoV-2; hospitalizations; mortality; machine learning



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Impact and effectiveness of COVID-19 vaccines using machine learning to analyze a time series: A population-based study

Rafael Garcia-Carretero ^{1,†,*} , Maria Ordoñez-García ^{2,†,*}, Oscar Vazquez-Gomez ¹ , Belen Rodriguez-Maya ¹, Ruth Gil-Prieto ³  and Angel Gil-de-Miguel ³ 

¹ Internal Medicine Department, Mostoles University Hospital, Rey Juan Carlos University, 29835 Mostoles (Spain); rgcarretero@salud.madrid.org

² Hematology Department, Mostoles University Hospital, Rey Juan Carlos University, 29835 Mostoles (Spain); maria.ordonez@salud.madrid.org

³ Department of Preventive Medicine and Public Health, Rey Juan Carlos University (Madrid, Spain)

* Correspondence: rgcarretero@salud.madrid.org; maria.ordonez@salud.madrid.org

† These authors contributed equally to this work.

Abstract: Background: Although confirmed cases of infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have been declining since late 2020 due to general vaccination, little research has been done regarding the impact of vaccines against SARS-CoV-2 in Spain in terms of hospitalizations and deaths. **Objective:** Our aim was to identify the reduction in severity and mortality of coronavirus disease 2019 (COVID-19) at a nationwide level due to vaccination. **Methods:** We designed a retrospective, population-based study to define waves of infection and to describe the characteristics of the hospitalized population. We also studied the rollout of vaccination and its relationship with the decline in hospitalizations and deaths. Finally, we developed two mathematical models to estimate non-vaccination scenarios using machine learning modeling (with the ElasticNet and RandomForest algorithms). The vaccination and non-vaccination scenarios were eventually compared to estimate the number of averted hospitalizations and deaths. **Results:** In total, 498,789 patients were included, with a global mortality of 14.3%. We identified six waves or epidemic outbreaks during the observed period. We established a strong relationship between the beginning of vaccination and the decline in both hospitalizations and deaths due to COVID-19 in all age groups. We also estimated that vaccination prevented 170,959 hospitalizations (CI 95% 77,844–264,075) and 24,546 deaths (CI 95% 2,548–46,543) in Spain between March 2021 and December 2021. We estimated a global reduction of 9.19% in total deaths during the first year of COVID-19 vaccination. **Conclusions:** Demographic and clinical profiles changed over the first months of the pandemic. Patients over 80 years old and other age groups obtained clinical benefit from early vaccination. The severity of COVID-19, in terms of hospitalizations and deaths, decreased due to vaccination.

Keywords: COVID-19; vaccines; SARS-CoV-2; hospitalizations; mortality; machine learning

1. Introduction

The coronavirus disease 2019 (COVID-19) pandemic has had a significant impact on the health of the population as well as significant implications in all sectors of society and the daily lives of citizens [1–4]. It is claimed that high levels of vaccination coverage, the characteristics of the omicron variant, and increased diagnostic testing likely contributed to the observed impact of the pandemic in the last months of 2021. In addition, there was a very high incidence of confirmed cases, but a majority of these had mild symptoms or were asymptomatic. This placed a significant strain on primary healthcare rather than in hospitals. Therefore, the percentage of hospital occupancy and intensive care unit (ICU) beds much lower than expected relative to what occurred over the remainder of the pandemic [5–8].

By February 2022, more than 92% of the Spanish population over the age of 12 was fully vaccinated [9]. Current evidence indicates that the various COVID-19 vaccines have had very high effectiveness in restricting moderate and severe forms of the disease and reducing lethality. Vaccines, while reducing the probability of infection, are less effective for completely preventing virus replication in

the upper respiratory mucosa of a vaccinated individual, which means that transmission is possible from vaccinated individuals who have been infected, even if the disease is mild or asymptomatic [10–15]. This makes it infeasible to aim for the virus's eradication at present. Therefore, researchers should focus their efforts toward reducing the severity of infections while maintaining a level of transmission that is manageable and does not generate an excessive burden on the healthcare system.

As noted, due to the increase in vaccination coverage and the immunity generated from natural infections, the majority of the population is protected against severe COVID-19 [16]. Data show that protection has been maintained even against a variant different enough antigenically from the previous ones to produce very high incidence rates in the population that previously had immunity.

Observational studies, such as case-control or cohort studies, are not always feasible, so several studies using alternative approaches have been conducted to demonstrate the effectiveness of vaccination [8,17]. Likewise, several meta-analyses have studied the effectiveness of vaccination from three aspects: efficacy against infection, efficacy against severe disease (and hence reduction of risk of hospital admission), and ability to reduce transmissibility of vaccinated individuals who become infected [13,15,16,18]. However, the impact of vaccination in terms of decreasing hospitalizations and deaths in a nationwide, population-based, epidemiological study has not yet been conducted in Spain.

1.1. Hypothesis and Objectives of Our Research

We designed a population-based study to assess vaccination as a major public health intervention. By this means, we investigated whether vaccines have been beneficial to the Spanish population. Our research objective was to determine whether vaccination reduced the number of hospitalizations and deaths. We conducted our study in three stages. First, we described the differences between two periods: the first months of the pandemic, in which no vaccination was present, and the last months of 2021, when a high proportion of the Spanish population was vaccinated. Second, we compared trends in hospitalizations and deaths with the vaccination rate. Finally, we assessed the effectiveness of vaccines against severe disease, in terms of the reduction of hospitalizations and mortality due to COVID-19. We estimated the number of averted hospitalizations and deaths. We also compared the evolution of the pandemic across two scenarios: vaccination (the observed scenario) and non-vaccination (an estimated scenario). The estimated scenario was fitted using time series and machine learning analyses.

2. Materials and Methods

2.1. Data Collection and Study Design

We designed a retrospective, population-based study using data collected from electronic health records. We collected data from the Spanish Minimum Basic Data Set at Hospitalization (MBDS-H), provided by the Spanish Ministry of Health [19]. We also collected data related to COVID-19 vaccination in the European Union/European Economic Area (EU/EEA) from the European Centre for Disease Prevention and Control [20]. Figure A1 shows a flowchart of the study.

MBDS-H is a mandatory administrative registry of hospital discharges that covers more than 95% of Spanish hospitals, including public centers in the National Spanish Health System and private hospitals. Nearly 97% of total hospital discharges are covered in the database. The MBDS-H is exclusively built from discharge reports. Microdata from patients include information on sex, age, dates of admission and discharge, type of discharge, primary and secondary diagnoses at discharge, length of stay, and surgical or obstetric procedures, among other data. Other administrative data were recorded by default, including the province where the hospitalization occurred, place of residence, and cost of hospitalization. By default, the Ministry of Health provides de-identified data to ensure patient privacy; thus, no names or personal information were recorded. The purpose of the MBDS-H is to facilitate the development of retrospective studies for calculating the burden of hospitalization and assessing risk factors from thousands of patients, i.e., enabling population-based studies. From

2016 onward, MBDS-H used the International Classification of Diseases, 10th edition, coding system. MBDS-H is considered a valuable system for the epidemiological analysis of any coded disease.

Vaccination data were downloaded from the European Centre for Disease Prevention and Control [20]. These data were collected through The European Surveillance System. All EU/EEA Member States are requested to report basic indicators on vaccination (vaccines categorized by manufacturer, number of doses administered, vaccinated population, etc.). Data are categorized by target and age group at a national level.

2.2. Inclusion and Exclusion Criteria

In this retrospective study, cases were collected from the MBDS-H from the Spanish Ministry of Health. We included all patients with the code for COVID-19 (U07.1) in any diagnostic position (either primary or secondary diagnosis) from 1 January 2020 to 31 December 2021.

All age groups were studied, with special emphasis on those older than 60 years of age. We analyzed the healthcare impacts in terms of mortality and ICU admission by dividing the population into age groups. Patients with incomplete data regarding ICU admission, mortality, length of stay, or COVID-19 disease were excluded.

2.3. Definition of Waves

We categorized the pandemic following a previous epidemiological study [21]. Using only data from Spain, the authors split the entire pandemic period into outbreaks or epidemiological waves based on the 14-day cumulative incidence, which marked the turning point for each wave. Every turning point indicated the end of one wave and the beginning of the next one.

As mentioned in the introduction, the first and second stages of our study were descriptive. We analyzed the evolution of the pandemic and its outbreaks, comparing the first waves, during which time vaccination was absent, with the last waves of 2021, when vaccination was present. Herein, we describe the demographic and epidemiological differences between the two periods and their relationships with vaccination.

2.4. Vaccination Rollout

The Vaccination Strategy Against COVID-19 in Spain was developed by the Spanish Ministry of Health [22]. The workgroup prioritized certain age groups to receive the vaccine, based on the supply of doses and the availability of current evidence, taking into consideration the demographic characteristics of the Spanish population [23]. Ethical concerns and assessment of risk factors were also considered to prioritize certain age groups over others. The elderly and health-care workers were the first groups to receive the vaccine, and the rollout moved forward through the rest of the age groups over the course of 2021. We assessed the trends of vaccination over time using data from the European Centre for Disease Prevention and Control.

We split the population into age groups to assess the evolution of the pandemic in terms of hospitalizations and deaths. Then, we compared the vaccination rates to those trends by age group.

2.5. Estimated Scenarios of the Unvaccinated Population

Designing research to investigate the effectiveness of vaccination in terms of reducing hospitalizations is a valuable endeavor but can be challenging. As noted, our objective was to determine whether vaccination reduced the number of hospitalizations and deaths. There are two common options for the design of clinical research on vaccines. The first approach is a retrospective cohort study, in which patients are divided into two groups (e.g., vaccinated and unvaccinated). The other design is a matched case-control study, in which a subset of vaccinated patients are matched with an equal number of unvaccinated ones, based on relevant characteristics. In either design, the aim would be to compare the hospitalization and death rates and clinical characteristics between the two groups.

However, due to a lack of data, traditional research designs such as cohort and case-control studies would not be feasible in our context.

Thus, we designed a population-based, epidemiological, nationwide study to compare two scenarios: the observed one describing actual hospitalizations and deaths before and after vaccination, and an estimated one simulating the trends of the pandemic had there not been vaccination in 2021. The first scenario indicates that the outcomes in the first months of the pandemic were widely different from those of the last months of 2021 (with vaccination). Finally, we compared the two scenarios to extract the estimated effect of vaccination.

2.6. Mathematical modeling of hospitalizations and mortality

First, we transformed our data set into a time series. Once historical data on the number of deaths from COVID-19 were collected, we prepared the data for analysis by checking for missing values, outliers, and inconsistencies. Then, we aggregated data on daily hospitalizations and daily deaths into appropriate time intervals to have a time series. Exploratory analyses were used to understand the trends and patterns in the historical data. This involved creating visualizations and summary statistics to identify any seasonality or trends in the occurrence of COVID-19.

With a clean data set, the next stage was to select a forecasting method. Estimating population data, specifically the number of events due to COVID-19, is a common but challenging task in epidemiology and clinical research. To make such predictions, statistical methods and models to forecast future trends can be used, but choosing an appropriate forecasting method depends on the characteristics of the data. More specifically, the choice of method depends on the complexity of the data and the availability of relevant predictor variables. Then, after testing different methods of analysis (see Appendix B), we employed machine learning techniques to forecast hospitalizations and deaths in the absence of vaccination. We used two machine learning algorithms (described below) because these can capture non-obvious (both linear and nonlinear) patterns in data. We evaluated each model's performance using appropriate metrics such as mean absolute percentage error through cross-validation to ensure reliability for the training period (July 2020 to February 2021). Then we used each model to forecast future population hospitalizations and deaths for the desired time period (March to December 2021). We created point forecasts (single estimates) and prediction intervals (confidence intervals) to quantify the uncertainty in our predictions.

2.7. Machine Learning Algorithms

We employed two algorithms, ElasticNet (EN) and RandomForest (RF). The former assumes linearity and the latter makes no assumptions on linearity.

EN is a machine learning technique used for linear regression and feature selection. It combines two regularization methods: L1 and L2 regularization [24]. L1 encourages some feature coefficients to be exactly zero, effectively performing feature selection by eliminating less important features. L2, on the other hand, penalizes large coefficients and prevents overfitting. EN strikes a balance between these two regularization techniques by introducing a hyperparameter that controls the mix of L1 and L2 penalties. This hyperparameter, often denoted *alpha*, allows one to adjust the level of feature selection and regularization. A value of *alpha* equal to 0 corresponds to L2 regularization, while a value of 1 corresponds to L1 regularization. Any value in between blends the characteristics of both methods. EN is valuable when dealing with data sets containing many features, as it helps prevent overfitting and can automatically select the most relevant features. It is commonly used in predictive modeling, where the goal is to create accurate models that generalize well to new data while optimizing feature usage. Researchers and data scientists apply EN in various fields, including healthcare, finance, and natural language processing [25–27].

RF is a powerful machine learning technique used in various fields, including clinical research and engineering. It is essentially a collection of decision trees, where each tree is a simple predictive model [28,29]. It uses different subsets of the available data and features to create each decision

tree. What sets RF apart is its random nature. This randomness injects diversity into models. By combining predictions from multiple trees, RF models become robust and less prone to overfitting, which makes them excellent at making generalizations from data. For researchers, RF can be used to make predictions based on complex, multidimensional data. It is well-suited to handle both numerical and categorical data, which is key in fields such as healthcare, where patient information can include a mix of variables. Clinicians and engineers find RF useful for various applications [30,31], such as disease prediction, image analysis, and quality control in manufacturing. RF models are known for their versatility, reliability, and ability to produce accurate and interpretable results, which makes them a valuable tool for decision support and pattern recognition.

2.8. Fitting the Models

To fit the models, we first split our time series data set into a training and a testing set. The training set was the period between July 1, 2020, and February 29, 2021. We excluded the first wave (March to June 2020) because we considered it an outlier that could add noise to the final model. The testing set was not used to develop the models but for comparison purposes only. We made no assumptions on the likelihood, normality, or linearity of the training set. We fitted the models to the training set, tuning the hyperparameters of each algorithm to achieve the best accuracy. For EN we set *alpha*, and for RF, we set *mtry* and the numbers of trees. A key mathematical condition when tuning the models was that they should fit accurately with the observed data, i.e., the training data set.

We used the R package *randomforest* for the RF model and the R package *glmnet* for the EN model. We fitted the two models on time series of both hospitalizations and deaths. Thus, we computed four models. Once they were developed, we forecasted data for the next months: March 1, 2021, to December 31, 2021. Finally, by comparing the estimated hospitalizations and deaths (had vaccination never been implemented) with the observed data, we could explore the number of events that were averted in the last 10 months of 2021 due to vaccination.

2.9. Statistical analyses

All statistical and machine learning-based analyses were conducted using R language version 4.3.2 (Vienna, Austria)[32]. Statistical significance was set at 0.05.

3. Results

3.1. Global Overview of the Pandemic

We included data from 498,789 hospital admissions (Table 1), and excluded 113 patients due to inconsistent or incomplete data. We split the observation period into waves, as described above (Figure 1). The first waves included more than 115,000 hospitalizations, and this number dropped up to 50,000 in the fourth and the fifth waves. Men were admitted at higher rates than women (56.1%, $p=0.001$), with no changes in the distribution during the pandemic. The median age was 66, but this tended to decrease across the fourth and fifth waves (59 and 57, respectively). Length of stay, in both the standard hospitalization ward and in the ICU, was more heterogeneous, and no clear trend could be established. Although the global mortality ratio was 14.3%, we observed a decreasing trend from the first wave (18.2%) to the fourth and fifth waves (7.4% and 10.1%, respectively). Comorbidities such as type 2 diabetes, hypertension, coronary disease, dementia, kidney disease, malignancy (either solid tumor or hematological malignancy), and chronic respiratory disease showed a decreasing trend starting with the fourth wave. Other comorbidities, such as liver and cerebrovascular diseases, showed no changes. Obesity and heart failure showed a more heterogeneous trend.

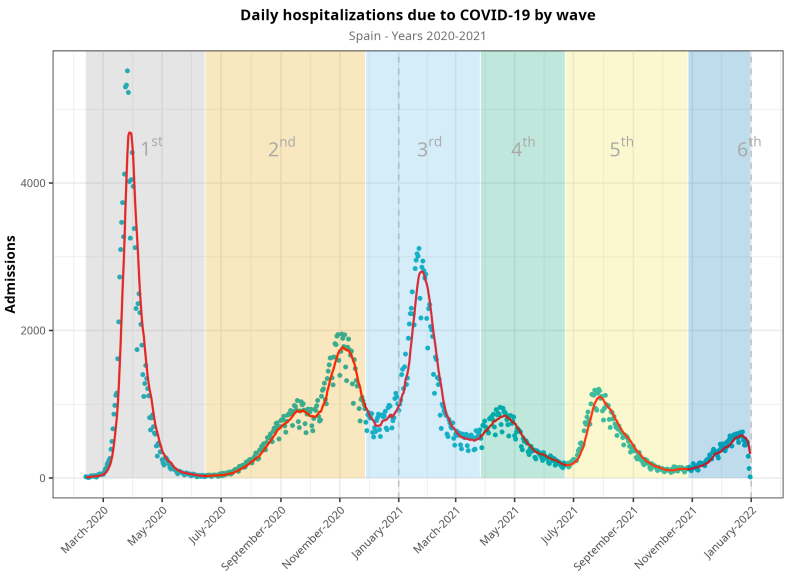


Figure 1. Evolution of hospitalizations during the COVID-19 pandemic in Spain from March 2020 to December 2021, split into waves.

Table 1. Epidemiological and demographic characteristics of patients hospitalized in Spain between 2020 and 2021.

	All waves	First	Second	Third	Fourth	Fifth	Sixth	P value
Hospital admissions	498,789	115,356	127,114	126,623	51,006	54,570	24,120	0.001
Sex (men)	56.1%	56.2%	55.5%	56.2%	57.5%	55.3%	56.2%	0.001
Age, median (IQR)	66 (28)	69 (25)	68 (28)	69 (25)	59 (24)	57 (38)	65 (28)	0.001
Hospital stay, median (IQR)	8 (9)	11.9 (9)	8 (9)	12.1 (9)	7 (7)	10.4 (8)	6 (7)	0.001
ICU								
Admissions	54,354	10,218	13,302	14,441	7,194	6,941	2,258	0.001
ICU (%)	10.9	8.9	10.5	11.4	14.1	12.7	9.4	0.001
ICU stay, median (IQR)	10 (21)	11 (22.3)	10 (21.4)	11 (21.5)	11 (21.4)	10 (18.7)	6 (9.1)	0.001
Mortality								
Deaths	71,437	21,037	18,229	20,400	3,793	5,487	2,491	0.001
Mortality rate (%)	14.3	18.2	14.3	16.1	7.4	10.1	10.3	0.001
Comorbidities								
Type 2 diabetes	21.7	20.6	23.2	24.3	18.3	18.3	21.4	0.001
Hypertension	33.8	35.2	34.7	37.2	31.1	25.0	31.6	0.001
AMI	7.1	7.3	7.4	7.9	5.3	5.7	7.5	0.001
CHF	7.2	6.5	7.8	8.4	4.2	6.8	7.9	0.001
Dementia	4.8	5.5	5.4	5.2	2.1	3.9	3.6	0.001
Kidney disease	10.5	10.3	11.2	11.7	6.5	9.7	11.1	0.001
Liver disease	0.4	0.4	0.4	0.5	0.3	0.4	0.5	0.001
Malignancy	5.6	5.4	5.8	6.0	4.2	5.3	7.4	0.001
Obesity	12.9	9.2	12.6	14.0	17.0	14.4	13.8	0.001
COPD	7.4	7.6	7.4	7.9	5.9	6.6	8.8	0.001
CEVD	0.7	0.7	0.8	0.9	0.5	0.6	0.8	0.001

ICU: intensive care unit. AMI: acute myocardial infarction. CHF: congestive heart failure. CEVD: cerebrovascular disease. COPD: chronic obstructive pulmonary disease. IQR: interquartile range. Age is expressed in years. Hospital/ICU stay is expressed in days.

Table 2 shows disaggregated data of hospital admissions, ICU admissions, and mortality by age group. These data are also represented in Figure A2.

Table 2. Outcomes in terms of total admissions, ICU admissions an mortality of hospitalized disaggregated by age groups.

	Total	First wave		Second wave		Third wave		Fourth wave		Fifth wave		Sixth wave		P-value
Admissions														
≤17	6,568	583	(8.9%)	16,22	(24.7%)	1,021	(15.5%)	657	(10%)	1,764	(26.9%)	921	(14%)	0.001
18–49	99,570	18,223	(18.3%)	23,525	(23.6%)	18,542	(18.6%)	13,669	(13.7%)	20,660	(20.7%)	4,951	(5%)	0.001
50–59	78,535	18,524	(23.6%)	19,534	(24.9%)	19,432	(24.7%)	11,560	(14.7%)	6,262	(8%)	3,223	(4.1%)	0.001
60–79	178,667	45,100	(25.2%)	44,052	(24.7%)	48,808	(27.3%)	18,198	(10.2%)	13,267	(7.4%)	9,242	(5.2%)	0.001
≥80	125,834	30,642	(24.4%)	35,998	(28.6%)	36,365	(28.9%)	5,806	(4.6%)	11,707	(9.3%)	5,316	(4.2%)	0.001
ICU														
≤17	490	95	(19.4%)	109	(22.2%)	74	(15.1%)	62	(12.7%)	111	(22.7%)	39	(8%)	0.001
18–49	10,227	1,493	(14.6%)	2,186	(21.4%)	1,957	(19.1%)	1,513	(14.8%)	2,633	(25.7%)	445	(4.4%)	0.001
50–59	11,187	2,117	(18.9%)	2,733	(24.4%)	2,994	(26.8%)	1,705	(15.2%)	1,240	(11.1%)	398	(3.6%)	0.001
60–79	28,548	5,862	(20.5%)	7,159	(25.1%)	8,300	(29.1%)	3,537	(12.4%)	2,526	(8.8%)	1,164	(4.1%)	0.001
≥80	2,372	342	(14.4%)	754	(31.8%)	687	(29%)	185	(7.8%)	248	(10.5%)	156	(6.6%)	0.001
Deaths														
≤17	27	4	(14.8%)	7	(25.9%)	6	(22.2%)	2	(7.4%)	6	(22.2%)	2	(7.4%)	0.001
18–49	1,282	345	(26.9%)	300	(23.4%)	253	(19.7%)	92	(7.2%)	222	(17.3%)	70	(5.5%)	0.001
59–59	3,274	905	(27.6%)	760	(23.2%)	861	(26.3%)	277	(8.5%)	343	(10.5%)	128	(3.9%)	0.001
60–79	25,427	7,937	(31.2%)	5,914	(23.3%)	7,074	(27.8%)	1,871	(7.4%)	1,733	(6.8%)	898	(3.5%)	0.001
≥80	40,855	11,707	(28.7%)	11,111	(27.2%)	12,041	(29.5%)	1,499	(3.7%)	3,127	(7.7%)	1,370	(3.4%)	0.001

Mortality ratios were calculated dividing the the events presented in a given age group in a wave by the total number of population in each group.

Among hospitalizations, the predominant age groups were >60 years old in the second and third waves. The group >80 years old dropped dramatically in fourth and fifth waves, and the group of 18- to 49-year-olds was predominant in the fifth wave. Regarding mortality, deaths in all ages dropped quite evenly, although the patients who were more affected were >60 years old (Figure 2). We display Figure 2 beginning with the second wave, as details of the following waves displayed are of interest to compare second and third waves on one side with fourth and fifth waves on the other.

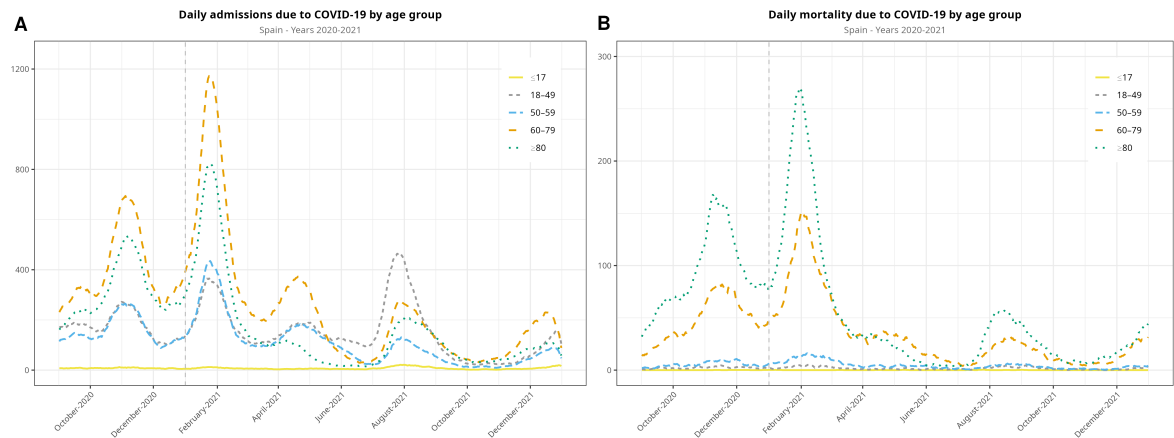


Figure 2. Evolution of the COVID-19 pandemic in terms of hospitalizations (a), and in-hospital deaths (b), disaggregated by age group, between September 2020 and December 2021 (the first wave is not included in the visualization).

3.2. Vaccination Rollout

Figure 3 plots the vaccination rollout in Spain, both globally and by age group. Vaccination began in December 2020 with the elderly and healthcare workers. By December 31, 2021, 80.3% of the whole Spanish population was fully vaccinated, i.e., having received the complete regimen, and 97.2% had received at least one dose. By April 2021, 75% of >80-year-olds and 48% of >60-year-olds were fully vaccinated. Figure A3 provides more detail on age groups regarding vaccination coverage over time.

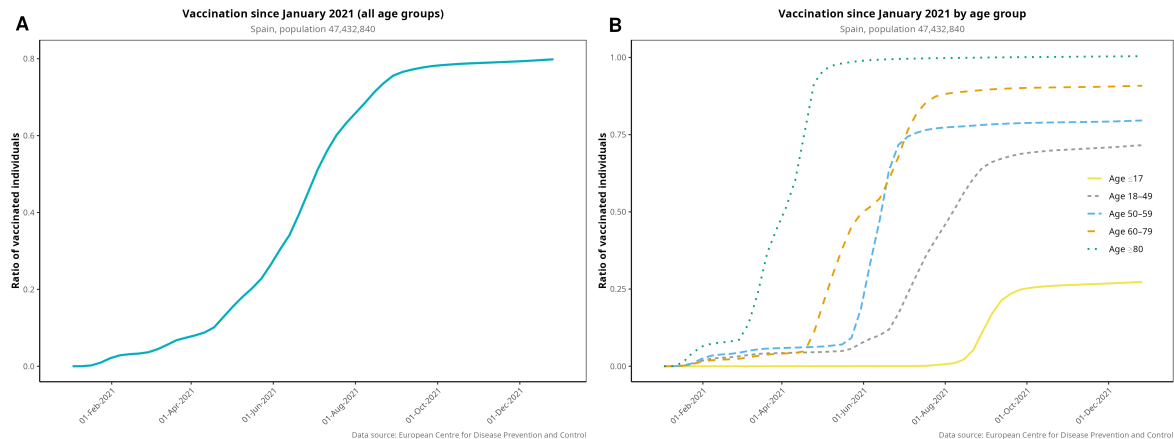


Figure 3. Vaccination rollout in Spain for the entire population (i.e., fully vaccinated individuals, (a)), and disaggregated by age group (b). Elderly patients were prioritized to receive the first dose of vaccine.

3.3. Hospitalizations and Deaths in an Estimated Scenario

As noted, our approach was to estimate both hospital admissions and mortality by parsing the time series to machine learning algorithms. We developed four models, one for hospitalizations and another for deaths using each algorithm. Figure 4 shows the observed and estimated scenarios. Figure A4

shows the models with confidence intervals. Table 3 shows the estimates of hospitalizations and deaths in the absence of vaccination. Using the RF model, we estimated that 251,830 hospitalizations and 37,673 deaths would have occurred in a non-vaccination scenario during the period between March and December 2021. Using the EN model, the estimated hospitalizations and deaths were 307,617 and 37,141, respectively. Compared to the observed data, we estimated that vaccination prevented 115,172 hospitalizations and 25,078 deaths with the RF model, and 170,959 hospitalizations and 24,546 deaths with the EN model. Finally, we plotted Figure 5, showing the cumulative hospitalizations and deaths with both the RF and EN models.

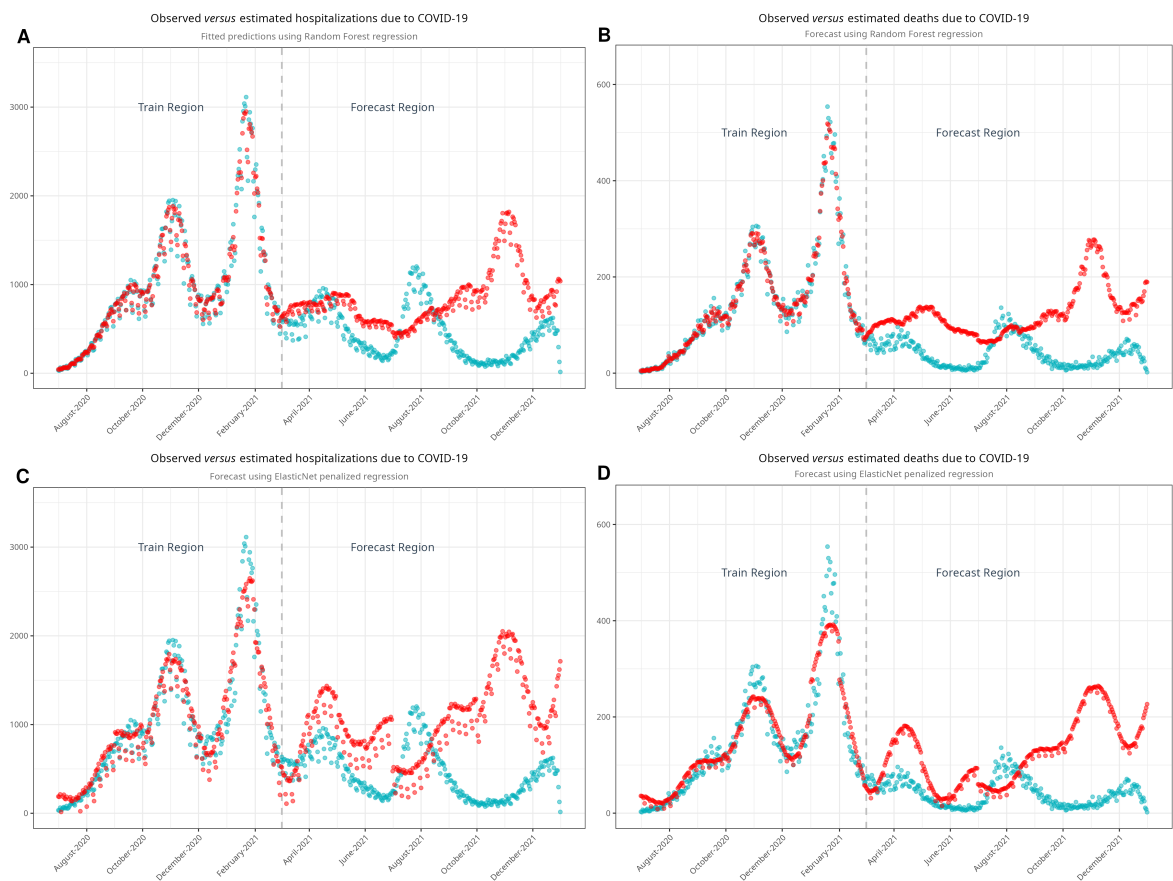


Figure 4. Models developed with RandomForest, (a) and (b), and with ElasticNet (c) and (d) to estimate non-vaccination scenarios. Blue dots represent the observed values while red dots represent estimated values. Note the good fitting of the model in the train region before it estimates the values in the forecast region.

Table 3. Observed, estimated, and averted events in the first year of vaccination.

	Hospitalizations		Deaths	
	Events	(95% CI)	Events	(95% CI)
RandomForest				
Observed	136,658	(NA)	12,595	(NA)
Estimated	251,830	(216,99–286,663)	37,673	(317,13–43,633)
Averted	115,172	(80,339–150,005)	25,078	(191,18–31,038)
ElasticNet				
Observed	136,658	(NA)	12,595	(NA)
Estimated	307,617	(214,502–400,733)	37,141	(15,143–59,138)
Averted	170,959	(77,844–264,075)	24,546	(2,548–46,543)

Estimation computed in the period between March, 2021 and December, 2021. We estimated data using two models: ElasticNet and RandomForest NA: non-applicable. CI: confidence interval.

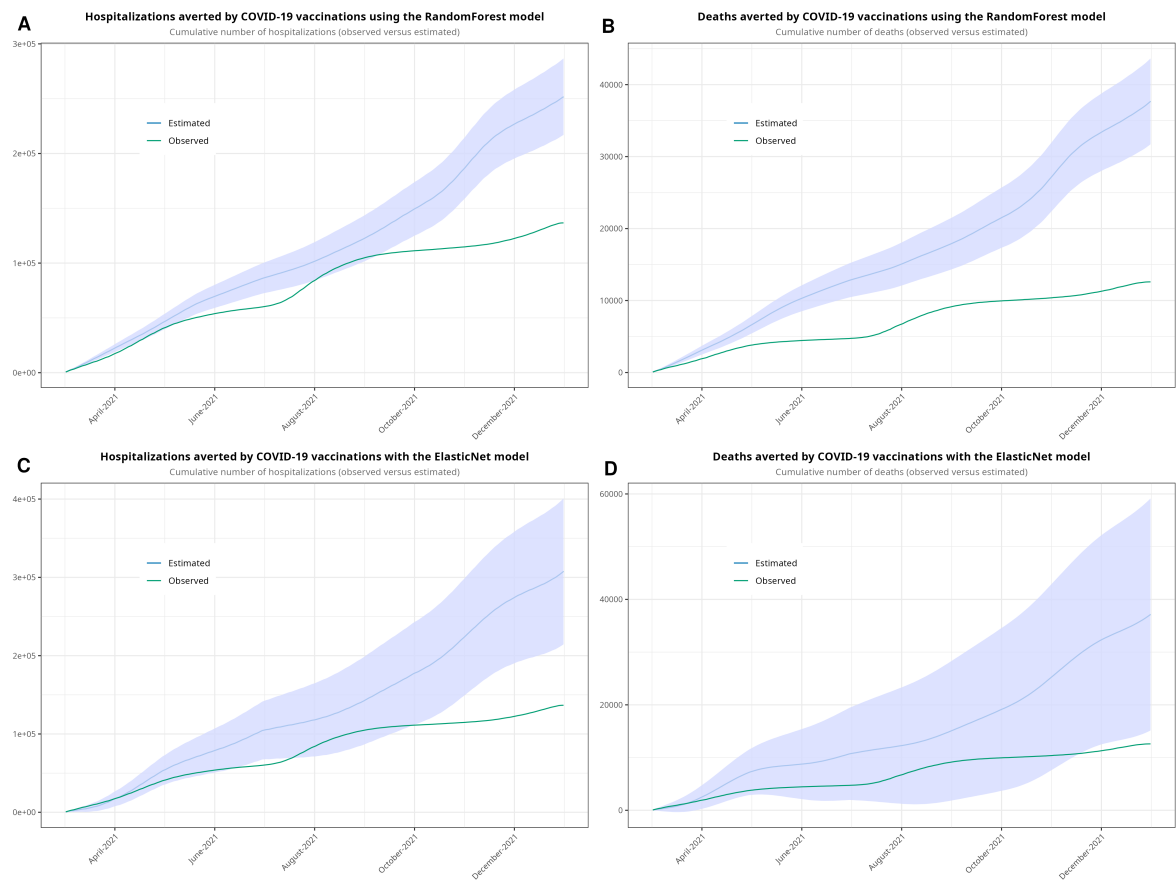


Figure 5. Cumulative sum with RandomForest, (a) and (b), showing the observed values and the estimates with 95% confidence intervals, and with ElasticNet, (c) and (d). We forecasted values from March 1 to December 31, 2021.

4. Discussion

4.1. Descriptive Analyses

We have described the high number of hospitalizations and deaths during the first waves of the pandemic in Spain. We have demonstrated how this trend began to decrease in March to April 2021 as a result of vaccination, which was the major public health intervention during the COVID-19 pandemic.

Overall, the first wave showed the highest number of hospitalizations, the highest mortality rate, the longest hospital and ICU lengths of stay, and the oldest patients. The fourth and the fifth waves showed a decreasing trend in terms of hospitalizations and mortality. In addition, these last waves showed an overall younger, healthier population.

While the sixth wave was included in our analyses, the results might not be reliable, as this wave ended mid-February 2022, and its results are not fully represented in tables and figures. However, the fourth and fifth waves showed that the demographic profile of hospitalized individuals changed with respect to the previous waves, showing a turning point in the evolution of the pandemic.

With respect to admissions by age group, the group of patients under 17 contributed only marginally during the observed period of the pandemic. However, patients over 60 years old were the largest group of those admitted to the hospital due to COVID-19. Patients between 18 and 59 years old were hospitalized at a lower rate. Additionally, most of the deaths occurred in patients >60 years old, particularly in patients over 80 years old, whereas mortality in the rest of the age groups was marginal, as seen in Figure 3.

4.2. Vaccination Rollout

Vaccination in Spain began in late December 2020, as soon as vaccines were proven to be safe and to offer significant protection against severe forms of COVID-19, as part of a European initiative [33]. Within only a few weeks from the beginning of vaccination (2.2% of total population by March and 10% by May 2021), we observed a rapid decline in both hospitalizations and deaths beginning in March and April 2021. We also observed a strong temporal correlation between decreasing hospitalizations and deaths on one side and the evolving vaccination rollout on the other (Figures 2 and 3). The decline in hospitalizations and deaths was first observed in patients over 80 years old, showing a relationship between vaccination and protection against both outcomes. This relationship can also be seen in the remaining age groups after the beginning of vaccination. This steady pace of vaccination consolidated the decline in the severity of the pandemic. Our data are also in line with other studies that have investigated the benefit of vaccination and its protective effects [34–36].

4.3. Modeling and Estimating Data in a Non-Vaccination Scenario

It can be challenging to quantify the impact of vaccination if an incomplete picture of the pandemic is obtained. Infections and confirmed cases are either often under-reported or underestimated [37,38]. For this reason, we relied on reported hospitalizations and deaths to determine this impact instead of trends of non-hospitalized, confirmed cases.

Our reference publication was Barandalla et al. [39], who developed simulated curves of hospitalization in the absence of vaccines and compared those curves with the observed incidence. That study investigated hospitalizations in Spain between February 2020 and June 2021. The authors estimated the expected hospitalizations during 2021 in the absence of vaccination, extrapolating data from the second wave. The scenario of an unvaccinated population was estimated to create a statistical model, as follows: the authors disaggregated the entire population curve across age groups and took the proportion of hospitalization of age groups of unvaccinated or less-vaccinated population as a reference. These proportions of hospitalizations were extrapolated to the remaining groups, and hence there were curves of real incidence of hospitalization and curves of expected hospitalization in the absence of vaccines for each age group. Finally, these curves were compared. Showing the decrease in incidence, they demonstrated the beneficial impact of vaccination on hospitalizations. Likewise, vaccine effectiveness against hospitalization in age groups ≥ 65 years old were estimated in a European study from October 2021 to March 2022 [40]. The reference group was the unvaccinated population. The authors performed a survival analysis using the Cox proportional hazards regression model to estimate the hazard ratios of hospitalization.

It is beyond the scope of this manuscript to discuss all studies that have used mathematical models to estimate mortality in the absence of vaccination, but it is worth mentioning some of

them. A mathematical model reported by Watson et al. [17] estimated that 14.4 million deaths were prevented in 185 countries in 2021. The authors used a framework based on a “susceptible-exposed-infections-recovered-susceptible” model to estimate a non-vaccination scenario. This model was fitted using MCMC, and the authors calculated the time-varying reproductive number to determine the estimated number of contagions. Havers et al. [8] conducted a cross-sectional study that included adults hospitalized with COVID-19, comparing vaccinated versus unvaccinated individuals. Both studies demonstrated the effectiveness of vaccination and its impact on the evolution of the COVID-19 pandemic using different mathematical approaches. Autoregressive time series modeling was assessed in other studies [41,42].

Machine learning has also been used to estimate the evolution of the COVID-19 pandemic in terms of confirmed cases: Kirbaş et al. [43] conducted a comparative study with different approaches, including ARIMA, neural networks, and long-short term memory (LSTM), to forecast the evolution of the pandemic. LSTM provided predictions with the best accuracy. Neural networks were used to study the dynamics of confirmed cases of COVID-19 by Nabi et al. [44]. Although deep learning (i.e., neural networks) seems to have better prediction accuracy than standard statistical methods (ARIMA) or machine learning, it entails high costs in computational resources and time [45].

Vaccination conferred sufficient protection to severe disease and altered the course of the COVID-19 pandemic. Given the conditions of the pandemic, measuring the impact of vaccination directly by comparing a vaccination scenario with a non-vaccination scenario was not possible at a nationwide level. This is why mathematical models are useful for estimating non-vaccination scenarios to achieve such comparisons. Thanks to our estimated scenarios, we could assess the impact of vaccination in Spain. Our approaches revealed estimations for the hospitalizations and deaths averted as a result of vaccination against SARS-CoV-2.

4.4. Limitations

We estimated how the pandemic would have evolved if no vaccines had been available by estimating a new scenario, but we did not include non-pharmaceutical interventions, viral variants, or limitations on mobility that could have altered the viral evolution in the absence of vaccination. It is of interest to mention that the last waves of 2021 in Spain, which were primarily caused by the omicron variant and its descendants (B.1.1.529), presented different characteristics from the previous waves [37,38], but its impact was not included in our model. In addition, forecasting using the time series signature can be very accurate, particularly when time-based patterns are present in the underlying data. As with most uses of machine learning, the prediction is only as good as the patterns in the data. Forecasting using this approach may not be suitable where patterns are not present or when the future is highly uncertain (i.e., past results are not a suitable predictor of future performance). We could not use ARIMA or MCMC to create the estimated scenario, so we did not compare different approaches. Although it has been found that mortality due to COVID-19 may have been under-reported [46], in Spain almost all deaths occurred in-hospital, so our data can be considered reliable. This is key when modeling and fitting a machine learning algorithm because the final model can only be as good as the provided data.

5. Conclusions

We fitted mathematical models to estimate both hospitalizations and deaths due to COVID-19 in a non-vaccination scenario. We determined the impact of vaccination by estimating the hospitalizations and deaths that otherwise could have occurred if vaccines had not been administered. Vaccination altered the evolution of the COVID-19 pandemic and prevented up to 24,546 deaths in Spain in 2021. They reduced not only mortality but also the number of hospitalizations and the burden of the pandemic. Their protective effect was observable shortly after the beginning of vaccination for each age group. Machine learning approaches can be useful in uncertain contexts because a time series signature can provide accurate forecasts.

Author Contributions: Conceptualization, R.G-C. and M.O-G.; methodology, R.G-C. and O.V-G; software, R.G-C.; validation, R.G-C. and M.O-G.; formal analysis, R.G-C. and B.R-M.; investigation, R.G-C. and M.O-G.; resources, X.X.; data curation, R.G-C. and B.R-M.; writing—original draft preparation, R.G-C. and O.V-G; writing—review and editing, R.G-P. and A.G-M.; visualization, R.G-C.; supervision, R.G-P. and A.G-M.; project administration,R.G-P. and A.G-M.; funding acquisition, R.G-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was approved by the Ethical Board of Universidad Rey Juan Carlos (ID number 2610202334423). No identifying information was included in the manuscript. Because the authors used historical data, informed consent was not necessary. All procedures involving human participants were conducted in accordance with the ethical standards of the responsible institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed Consent Statement: Not applicable.

Data Availability Statement: A contract signed with the Spanish Health Ministry, which provided the dataset, prohibits the authors from providing their data to any other researcher. Furthermore, the authors must destroy the database upon the conclusion of their investigation. The database cannot be uploaded to any public repository.

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

Abbreviations

The following abbreviations are used in this manuscript:

ARIMA	Auto-Regressive Integrated Moving Average
COVID-2019	CoronaVirus Disease 2019
EN	ElasticNet
EU/EEA	European Union/European Economic Area
LSTM	Long-Short Term Memory
MBDS-H	Minimum Basic Data Set at Hospitalization
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
MCMC	Markov Chain with Monte Carlo
RF	RandomForest

Appendix A. Descriptive Analyses

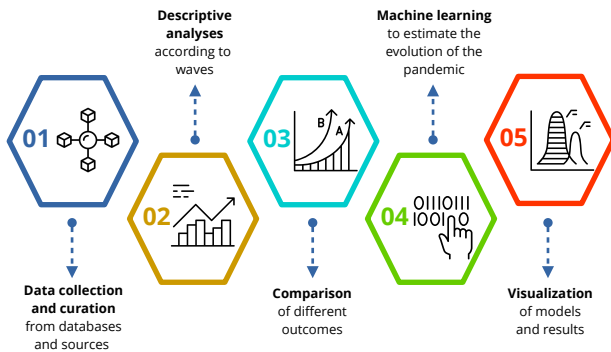


Figure A1. Flowchart and study design.

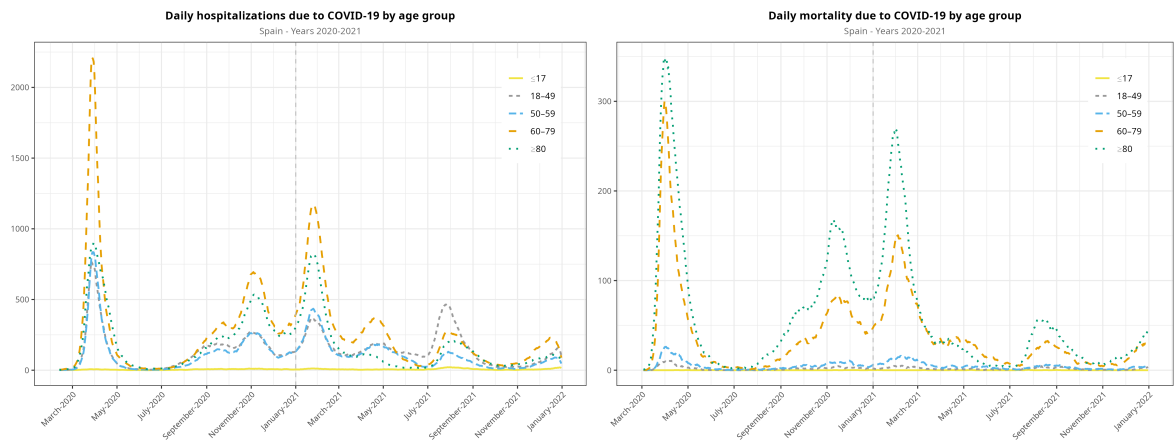


Figure A2. Evolution of the COVID-19 pandemic in terms of hospitalizations and deaths. All waves from the observation period are included.

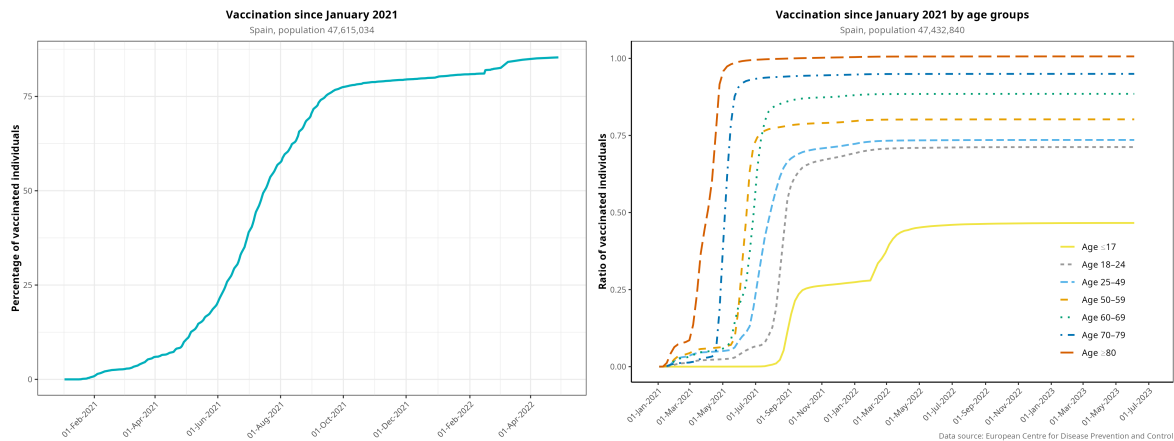


Figure A3. Vaccination rollout in Spain for the entire population (fully vaccinated individuals), and disaggregated by age group. Elderly patients were prioritized for the first dose of the vaccine.

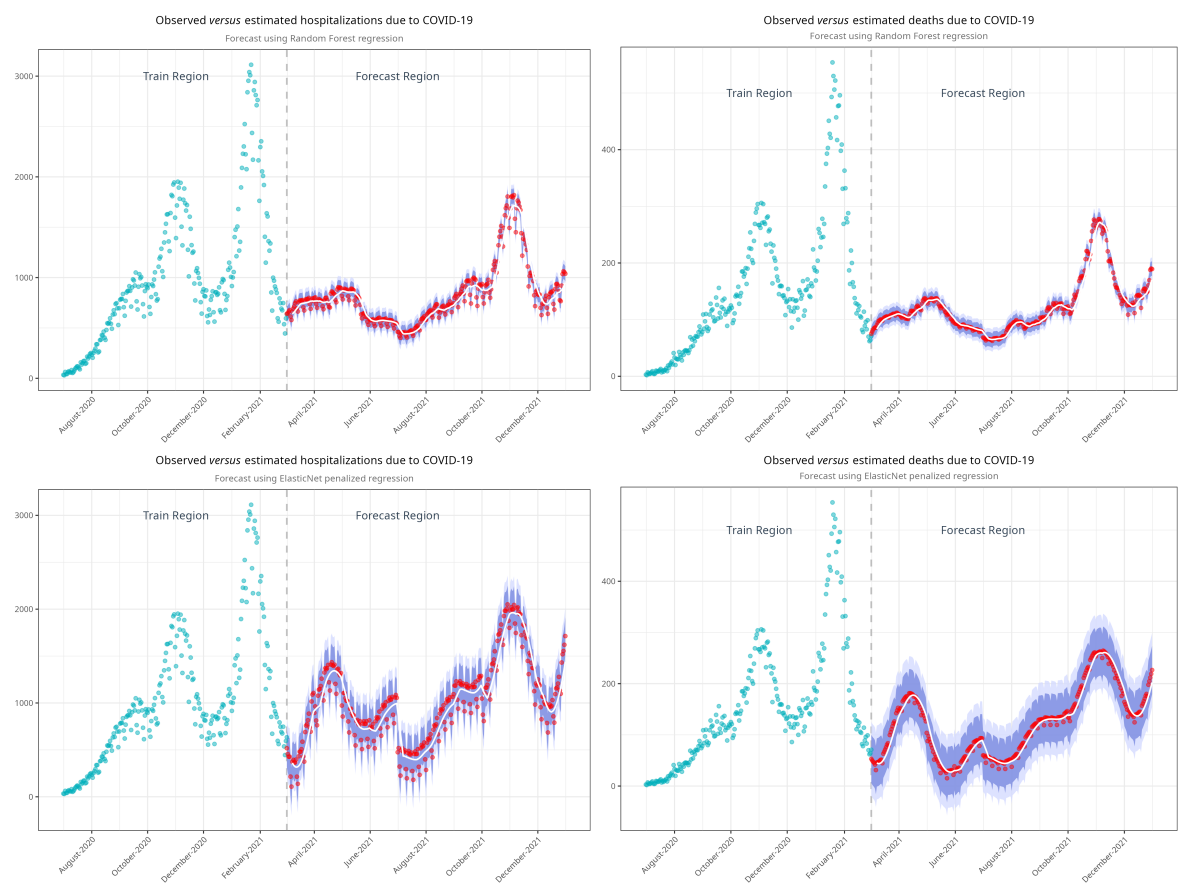


Figure A4. Models developed with RandomForest (a) and (b) and with ElasticNet (c) and (d) to estimate non-vaccination scenarios. Blue dots represent the observed values, while the smooth curves represent the estimated values and prediction intervals to account for variance between the model predictions and the observed data.

Appendix B. Mathematical Modeling of Hospitalizations and Mortality

First, we transformed our data set into a time series. Among the methods used to analyze time series, traditional statistical models such as auto-regressive (AR) models can be specified as linear regression on the lags of the time series. For example, an AR model only looks at the relationship between lags of a series and its future values. Seasonality and trend are key components of a given time series. While the trend represents a gradual change in the data, depicting long-term growth or decline, seasonality describes the short-term patterns that occur within a single unit of time and repeats indefinitely. Another useful technique is Markov Chain with Monte Carlo (MCMC) simulation, which describes dynamic changes based on the state of a given value and the chance of its transition. The MCMC algorithm explores the parameter space to find values that maximize the likelihood of the observed data.

Once historical data on the number of deaths from COVID-19 were collected, we prepared the data for analysis by checking for missing values, outliers, and inconsistencies. Then, we aggregated data on daily hospitalizations and daily deaths into appropriate time intervals to have a time series. Exploratory analyses were used to understand the trends and patterns in the historical data. This involved creating visualizations and summary statistics to identify any seasonality or trends in the occurrence of COVID-19.

With a clean data set, the next stage was to select a forecasting method. Estimating population data, specifically the number of events due to COVID-19, is a common but challenging task in epidemiology and clinical research. To make such predictions, statistical methods and models to forecast future trends can be used, but choosing an appropriate forecasting method depends on the characteristics

of the data. More specifically, the choice of method depends on the complexity of the data and the availability of relevant predictor variables. As noted, common methods include time series analysis, regression analysis, and machine learning techniques. We discarded time series analysis methods such as Auto-Regressive Integrated Moving Average (ARIMA) because we were unable to capture and project time-dependent patterns in the data, given their nature. Specifically, ARIMA estimations did not converge in our data set, likely because the developed model was not a good fit for the data. Likewise, MCMC required some assumptions that could not be fulfilled. The characteristics of our data and the underlying dynamics of COVID-19 hospitalizations did not justify the choice of the normal distribution and the assumption of independence in time steps. In addition, adjustments and the fine-tuning of some parameters based on the likelihood of our data and the prior distributions were either too complex or unavailable.

We used two machine learning algorithms because these can capture non-obvious (both linear and nonlinear) patterns in data. We evaluated each model's performance using appropriate metrics such as mean absolute percentage error through cross-validation to ensure reliability for the training period (July 2020 to February 2021). Then we used each model to forecast future population hospitalizations and deaths for the desired time period (March to December 2021). We created point forecasts (single estimates) and prediction intervals (confidence intervals) to quantify the uncertainty in our predictions.

References

1. Hosseinzadeh, P.; Zareipour, M.; Baljani, E.; Moradali, M. Social Consequences of the COVID-19 Pandemic. A Systematic Review. *Invest Educ Enferm* **2022**, *40*, e10.
2. Mofijur, M.; Fattah, I.; Alam, M.; Islam, A.; Ong, H.; Rahman, S.; others. Impact of COVID-19 on the social, economic, environmental and energy domains: Lessons learnt from a global pandemic. *Sustainable Production and Consumption* **2021**, *26*, 343–359.
3. Osofsky, J.; Osofsky, H.; Mamon, L. Psychological and social impact of COVID-19. *Psychological Trauma: Theory, Research, Practice, and Policy* **2020**, *12*, 468–469.
4. Saladino, V.; Algeri, D.; Auriemma, V. The Psychological and Social Impact of Covid-19: New Perspectives of Well-Being. *Frontiers in Psychology* **2020**, *11*. [cited 2023 Oct 15].
5. Patel, B.; Murphy, R.; Karanth, S.; Shiffaraw, S.; Peters, R.; Hohmann, S.; others. Surge in Incidence and Coronavirus Disease 2019 Hospital Risk of Death, United States, September 2020 to March 2021. *Open Forum Infect Dis* **2022**, *9*. [cited 2023 Oct 15].
6. Delahoy, M.; Ujamaa, D.; Whitaker, M.; O'Halloran, A.; Anglin, O.; Burns, E.; others. Hospitalizations Associated with COVID-19 Among Children and Adolescents—COVID-NET, 14 States, March 1, 2020–August 14, 2021. *MMWR Morb Mortal Wkly Rep* **2021**, *70*, 1255–1260.
7. Taylor, C. COVID-19—Associated Hospitalizations Among Adults During SARS-CoV-2 Delta and Omicron Variant Predominance, by Race/Ethnicity and Vaccination Status—COVID-NET, 14 States, July 2021–January 2022. *MMWR Morb Mortal Wkly Rep* **2022**, *71*. [cited 2023 Oct 15].
8. Havers, F.; Pham, H.; Taylor, C.; Whitaker, M.; Patel, K.; Anglin, O.; others. COVID-19-Associated Hospitalizations Among Vaccinated and Unvaccinated Adults 18 Years or Older in 13 US States, January 2021 to April 2022. *JAMA Internal Medicine* **2022**, *182*, 1071–1081.
9. de Sanidad, M. Situación actual Coronavirus. [cited 2023 Jan 24].
10. Stefanelli, P.; Trentini, F.; Petrone, D.; Mammone, A.; Ambrosio, L.; Manica, M.; others. Tracking the progressive spread of the SARS-CoV-2 Omicron variant in Italy, December 2021 to January 2022. *Euro Surveill* **2022**, *27*, 2200125.
11. Assessment of the further spread and potential impact of the SARS-CoV-2 Omicron variant of concern in the EU/EEA, 19th update, 2022. [cited 2023 Oct 15].
12. Markov, P.; Ghafari, M.; Beer, M.; Lythgoe, K.; Simmonds, P.; Stilianakis, N.; others. The evolution of SARS-CoV-2. *Nat Rev Microbiol* **2023**, *21*, 361–379.
13. Soheili, M.; Khateri, S.; Moradpour, F.; Mohammadzadeh, P.; Zareie, M.; Mortazavi, S.; others. The efficacy and effectiveness of COVID-19 vaccines around the world: a mini-review and meta-analysis. *Annals of Clinical Microbiology and Antimicrobials* **2023**, *22*, 42.

14. Harder, T.; Koch, J.; Vygen-Bonnet, S.; Külper-Schiek, W.; Pilic, A.; Reda, S.; others. Efficacy and effectiveness of COVID-19 vaccines against SARS-CoV-2 infection: interim results of a living systematic review, 1 January to 14 May 2021. *Eurosurveillance* **2021**, *26*. [cited 2023 Oct 15].
15. Graña, C.; Ghosn, L.; Evrenoglou, T.; Jarde, A.; Minozzi, S.; Bergman, H.; others. Efficacy and safety of COVID-19 vaccines. *Cochrane Database Syst Rev* **2022**, *12*, CD015477.
16. Yang, Z.; Jiang, Y.; Li, F.; Liu, D.; Lin, T.; Zhao, Z.; others. Efficacy of SARS-CoV-2 vaccines and the dose-response relationship with three major antibodies: a systematic review and meta-analysis of randomised controlled trials. *The Lancet Microbe* **2023**, *4*, e236–e246.
17. Watson, O.; Barnsley, G.; Toor, J.; Hogan, A.; Winskill, P.; Ghani, A. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. *Lancet Infect Dis* **2022**, *22*, 1293–1302.
18. Zeng, B.; Gao, L.; Zhou, Q.; Yu, K.; Sun, F. Effectiveness of COVID-19 vaccines against SARS-CoV-2 variants of concern: a systematic review and meta-analysis. *BMC Medicine* **2022**, *20*, 200.
19. de Sanidad Consumo y Bienestar Social, M. Portal Estadístico. Area de Inteligencia de Gestion. [cited 2019 Jul 6].
20. Data on COVID-19 vaccination in the EU/EEA, 2023. [cited 2023 Oct 15].
21. Informe no 128 Situación de COVID-19 en España a 10 de mayo de 2022, 2022. [cited 2022 Jun 9].
22. Gobierno de España. Estrategia de vacunación COVID-19.
23. Rodriguez-Maroto, G.; Atienza-Diez, I.; Ares, S.; Manrubia, S. Vaccination strategies in structured populations under partial immunity and reinfection. *medRxiv* **2021**, p. 2021.11.23.21266766. [cited 2022 Jul 1].
24. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B (Methodological)* **1996**, *58*, 267–288.
25. Garcia-Carretero, R.; Vigil-Medina, L.; Barquero-Perez, O.; Ramos-Lopez, J. Relevant Features in Nonalcoholic Steatohepatitis Determined Using Machine Learning for Feature Selection. *Metabolic Syndrome and Related Disorders* **2019**, *17*, 444–451.
26. Garcia-Carretero, R.; Barquero-Perez, O.; Mora-Jimenez, I.; Soguero-Ruiz, C.; Goya-Esteban, R.; Ramos-Lopez, J. Identification of clinically relevant features in hypertensive patients using penalized regression: a case study of cardiovascular events. *Medical & Biological Engineering & Computing* **2019**, *57*, 2011–2026.
27. Garcia-Carretero, R.; Vigil-Medina, L.; Barquero-Perez, O.; Mora-Jimenez, I.; Soguero-Ruiz, C.; Goya-Esteban, R.; others. Logistic LASSO and Elastic Net to Characterize Vitamin D Deficiency in a Hypertensive Obese Population. *Metabolic Syndrome and Related Disorders* **2020**, *18*, 79–85.
28. Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5–32.
29. Cutler, D.; Edwards Jr, T.; Beard, K.; Cutler, A.; Hess, K.; Gibson, J.; others. Random forests for classification in ecology. *Ecology* **2007**, *88*, 2783–2792.
30. Garcia-Carretero, R.; Vazquez-Gomez, O.; Lopez-Lomba, M.; Gil-Prieto, R.; Gil-de Miguel, A. Insulin Resistance and Metabolic Syndrome as Risk Factors for Hospitalization in Patients with COVID-19: Pilot Study on the Use of Machine Learning. *Metab Syndr Relat Disord* **2023**, *21*, 443–452.
31. Garcia-Carretero, R.; Vigil-Medina, L.; Mora-Jimenez, I.; Soguero-Ruiz, C.; Barquero-Perez, O.; Ramos-Lopez, J. Use of a K-nearest neighbors model to predict the development of type 2 diabetes within 2 years in an obese, hypertensive population. *Medical & Biological Engineering & Computing* **2020**, pp. 1–12.
32. Team, R.C. R: A Language and Environment for Statistical Computing, 2020. R Foundation for Statistical Computing.
33. Farsalinos, K.; Poulas, K.; Kouretas, D.; Vantarakis, A.; Leotsinidis, M.; Kouvelas, D.; others. Improved strategies to counter the COVID-19 pandemic: Lockdowns vs. primary and community healthcare. *Toxicology Reports* **2021**, *8*, 1–9.
34. Connors, M.; Graham, B.; Lane, H.; Fauci, A. SARS-CoV-2 Vaccines: Much Accomplished, Much to Learn. *Ann Intern Med* **2021**, *174*, 687–690.
35. Kustin, T.; Harel, N.; Finkel, U.; Perchik, S.; Harari, S.; Tahor, M.; others. Evidence for increased breakthrough rates of SARS-CoV-2 variants of concern in BNT162b2-mRNA-vaccinated individuals. *Nat Med* **2021**, *27*, 1379–1384.
36. Wang, Z.; Muecksch, F.; Schaefer-Babajew, D.; Finkin, S.; Viant, C.; Gaebler, C.; others. Naturally enhanced neutralizing breadth against SARS-CoV-2 one year after infection. *Nature* **2021**, *595*, 426–431.

37. Garcia-Carretero, R.; Vazquez-Gomez, O.; Ordoñez-Garcia, M.; Garrido-Peño, N.; Gil-Prieto, R.; Gil-de Miguel, A. Differences in Trends in Admissions and Outcomes among Patients from a Secondary Hospital in Madrid during the COVID-19 Pandemic: A Hospital-Based Epidemiological Analysis (2020-2022). *Viruses* **2023**, *15*, 1616.
38. Garcia-Carretero, R.; Vazquez-Gomez, O.; Gil-Prieto, R.; Gil-de Miguel, A. Hospitalization burden and epidemiology of the COVID-19 pandemic in Spain (2020-2021). *BMC Infect Dis* **2023**, *23*, 476.
39. Barandalla, I.; Alvarez, C.; Barreiro, P.; de Mendoza, C.; González-Crespo, R.; Soriano, V. Impact of scaling up SARS-CoV-2 vaccination on COVID-19 hospitalizations in Spain. *International Journal of Infectious Diseases* **2021**, *112*, 81–88.
40. Sentís, A.; Kislalya, I.; Nicolay, N.; Meijerink, H.; Starrfelt, J.; Martínez-Baz, I.; others. Estimation of COVID-19 vaccine effectiveness against hospitalisation in individuals aged > 65 years using electronic health registries; a pilot study in four EU/EEA countries, October 2021 to March 2022. *Eurosurveillance* **2022**, *27*, 2200551.
41. Chyon, F.; Suman, M.; Fahim, M.; Ahmmed, M. Time series analysis and predicting COVID-19 affected patients by ARIMA model using machine learning. *J Virol Methods* **2022**, *301*, 114433.
42. Maleki, M.; Mahmoudi, M.; Wraith, D.; Pho, K. Time series modelling to forecast the confirmed and recovered cases of COVID-19. *Travel Med Infect Dis* **2020**, *37*, 101742.
43. Kırbaş, İ.; Sözen, A.; Tuncer, A.; Kazancıoğlu, F. Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. *Chaos Solitons Fractals* **2020**, *138*, 110015.
44. Nabi, K.; Tahmid, M.; Rafi, A.; Kader, M.; Haider, M. Forecasting COVID-19 cases: A comparative analysis between recurrent and convolutional neural networks. *Results Phys* **2021**, *24*, 104137.
45. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V.; Semenoglou, A.; Mulder, G.; Nikolopoulos, K. Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward. *Journal of the Operational Research Society* **2023**, *74*, 840–859.
46. Whittaker, C.; Walker, P.G.; Alhaffar, M.; Hamlet, A.; Djaafara, B.A.; Ghani, A.; Ferguson, N.; Dahab, M.; Checchi, F.; Watson, O.J. Under-reporting of deaths limits our understanding of true burden of covid-19. *bmj* **2021**, 375.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.