Journal of Statistics on Health Decision

Journal Pre-proof

Characterization of the epidemiological curve of the 1st wave of the pandemic - study of the determinants for death and hospitalization due to COVID-19 in Portugal

Francisca Freire, Pedro Sá-Couto, Vera Afreixo

DOI:	https://doi.org/10.34624/jshd.v3i2.24208
Reference:	24208
Received date:	31-03-2021
Revised date:	20-09-2021
Accepted date:	09-11-2021

Please cite this article as: Freire F, Sá-Couto P, Afreixo V. Characterization of the epidemiological curve of the 1st wave of the pandemic - study of the determinants for death and hospitalization due to COVID-19 in Portugal. *Journal of Statistics in Health Decision* (2021), https://doi.org/10.34624/jshd.v3i2.24208

TITLE

Characterization of the epidemiological curve of the 1st wave of the pandemic - study of the determinants for death and hospitalization due to COVID-19 in Portugal

RUNNING HEAD

Determinants for death and hospitalization due to COVID-19

AUTHORS

Francisca Freire - Department of Mathematics, University of Aveiro, Portugal
Pedro Sá-Couto - Department of Mathematics, University of Aveiro, Portugal
Center for Research & Development in Mathematics and Applications (CIDMA), University of Aveiro. Portugal.
Vera Afreixo - Department of Mathematics, University of Aveiro, Portugal
Center for Research & Development in Mathematics and Applications (CIDMA), University of Aveiro. Portugal.

CORRESPONDING AUTHOR:

Francisca Freire
franciscamf@ua.pt
962012040

KEYWORDS

COVID-19, Portugal, Mortality, Hospitalization, Logistic regression, Demographic factors, Comorbidities.

FUNDING:

This work is supported by The Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020

CONFLICTS OF INTEREST:

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS:

Conceptualization, F.F., V.A., P.S.C.; data curation, F.F.; formal analysis, F.F.; investigation, F.F., V.A., P.S.C.; methodology, F.F., V.A., P.S.C.; supervision, F.F., V.A., P.S.C.; validation, F.F., V.A., P.S.C.; visualization, F.F., V.A., P.S.C.; writing—original draft, F.F.; writing—review and editing, F.F., V.A., P.S.C. All authors have read and agree to the published version of the manuscript.

ABSTRACT

COVID-19 is an infectious disease caused by the SARS-CoV-2 virus that emerged in China and has since spread rapidly to all countries in the world. The aim of this study is to characterize the peak of the first wave of the pandemic in Portugal using some demographic and clinical determinants. The database studied contains epidemiological surveillance data of COVID-19 from Portugal until June 30, 2020. The peak of the pandemic was determined considering the number of diagnoses, deaths, and hospitalizations over time, and the estimated period for the first wave was between March 19 and April 24. Multivariable logistic regression models were performed to estimate the Odds Ratio (OR) and the respective 95% confidence intervals (95% CI) for this period and for each of the variables considered in relation to both study outcomes: mortality (5.8%) and hospitalization (18.8%). The main risk factors considered for the mortality outcome were: Age Group (≥ 80 years: OR = 11.98; 65-79 years: OR = 4.06; reference group: <64 years), Hospitalization (OR = 6.48) and Comorbidities (OR = 5.74). For the outcome hospitalization, the main risk factors were: Age Group (≥ 80 years: OR = 8.54; 65-79 years: OR = 3.90; reference group: <64 years), male gender (OR= 1.58) and Comorbidities (OR = 5.19). The proposed models presented high area under the curve (AUC) results (mortality: AUC=91.1%; and hospitalization: AUC=84.5%), with small amplitudes in the 95% CI. The present study demonstrated that, overall, the major risk factors associated with worse prognostics of COVID-19 were advanced age (over 65 years), and the existence of comorbidities. For the risk of hospitalizations, the male gender was also a significant risk factor.

Introduction

COVID-19 is an infectious disease caused by the SARS-CoV-2 virus that first appeared in late 2019 in Wuhan, China and has since spread rapidly to all countries around the world. In most cases, individuals who contract this virus show only mild to moderate symptoms, however, in individuals over the age of 60 or who have some comorbidity, there is a greater predisposition for the development of more serious conditions of the disease or even death (1,2).

Mortality, hospitalizations, and ICU admissions are often used as clinical outcomes for assessing disease progression and severity. Age is referred in many studies as one of the most relevant risk factors for determining the clinical outcome in patients infected with the SARS-CoV-2 virus (2–4). The existence of comorbidities, in particular hypertension, diabetes and coronary heart disease is also associated with severe illnesses and worse prognosis in patients with COVID-19 (2,3,5). Although there is no consensus among the scientific community, some studies also consider that gender has an influence on the outcome of infected individuals (6).

In a meta-analysis by K. Dorjee et al (2020), seventy-seven studies comprising 38906 hospitalized patients were included, and was shown that individuals aged > 65 represented 85% (95% CI: 80–89) of deaths from COVID-19 (7). Furthermore, individuals aged > 60 also had a summary relative risk of 3.61 (95% CI: 2.96–4.39) for the occurrence of death and 1.57 (95% CI: 1.36–1.80) for the development of severe disease, being considered as a risk factor for both outcomes. In M. A. Barek et al. (2020), a meta-analysis was carried out with fifty-five studies from January 1, 2020 to May 23, 2020, including 10014 patients with SARS-CoV-2 infection (8). It was shown that individuals with comorbidities had an Odds Ratio of 3.13 (95% CI: 2.26–4.32), presenting a greater risk for the development of severe disease.

Although there are some previous studies regarding the epidemiological surveillance data of COVID-19 in Portugal, none of these studies specifically address the peak of the first wave of the pandemic, instead, they analyze the data as a whole (9). This study is particularly relevant in obtaining more specific information about the characteristics of the population that constitutes the epidemiological curve, a phase with special characteristics in which the growth in the number of cases is much higher (or even exponential). This information can also be useful in future waves of the pandemic, allowing for the comparison of different epidemiological peaks and inferring about the measures that could result in better clinical outcomes.

Therefore, the aim of this study is to characterize the peak of the first wave of the pandemic by showing the impact of available demographic and clinical determinants associated to the outcome of mortality and hospitalization.

Methods

Study Design

An observational retrospective study was conducted including all reported cases of COVID-19 until June 30, 2020 in Portugal.

Data Source

The database includes anonymized epidemiologic surveillance data provided by the General Health Directorate of Portugal (DGS) after attaining an authorization for the treatment of data by academics. This authorization was granted after request and submission of a form and an ethical committee authorization to perform research on the dataset (10). The data were collected via SINAVE (National System for Epidemiological Surveillance), a platform used for notification of epidemiological surveillance data corresponding to confirmed and validated cases of COVID-19 in Portugal until June 30, 2020.

General Characteristics and Outcomes

The outcomes considered were mortality and hospitalization, the first being the primary outcome. The mortality outcome was defined from the date of death provided in the database. Data provided by SINAVE include individual’s demographic (age, sex, region) and clinical characteristics (Hospitalizations, ICU admissions, Comorbidities).

Age was categorized into age groups: <64 years, 65-79 years and ≥ 80 years, the last two considered as possible risk factors. Similarly, geographic locations were categorized into five regions: "Lisbon and Tagus Valley (LTV)", "North", "Center", "South" and "Islands"; according to the information provided by DGS (11). The “Islands” region includes individuals from Madeira and Azores islands. The data on hospitalizations and comorbidities comprised individuals with an “Unknown” status that, for the purposes of this study, we considered as missing values. In addition, observations representing outliers relative to the dates of diagnosis and hospitalization were also excluded.

Considering that the first reported case of COVID-19 in Portugal occurred on March 2, 2020, the database was filtered to include only diagnosis dates as of this date and hospitalization dates after February 1, 2020. Moreover, despite the fact the database includes observations until June 30, 2020, the last recorded date of death was May 24 of the same year. Therefore, in order to normalize the peak curve estimation for the three parameters, the database was truncated only to comprise observations with diagnosis and hospitalization dates prior to May 25, 2020.

Statistical Analysis

For the epidemiological curves, we generated line plots with the absolute and absolute cumulative curves and relative frequencies for the dates of diagnosis, hospitalization, and death. The median value for the period of time associated to diagnosis was used to determine the peak of the 1st wave of the pandemic of COVID-19 in Portugal.

Descriptive statistics, such as absolute and relative frequencies, were calculated for all categorical variables. The chi-square test was also performed between the levels of the categorical variables (including Yates’s continuity correction).

The database was then divided into training (70%) and test (30%) subsets for analysis (12). After the division, the proportions of each outcome were verified in both datasets (training and test) to ensure that they are similar to those in the complete dataset.

Multivariable logistic regression models were performed using the training database and applying generalized linear models with a binomial distribution and logit link function to estimate the odds ratios (OR) and 95% confidence intervals (95% CI) for each outcome considered. Multicollinearity was assessed through calculation of the VIF (variance inflation factor). Using the test subset, we evaluated the performance of the model by determining the Area Under the Curve (AUC) and the respective 95% confidence interval.

The significance level was set at 5% and the statistical analysis performed using software R version 3.6.1 and RStudio version 1.2.5001 and the packages “pROC”, “ResourceSelection” and “car” and also “dplyr” and “ggplot2” for graphic support.

Results

Determination of the Epidemic Peak

To determine the peak of the pandemic, a total of 14476 individuals were included from March 1 to May 24, 2020. Figure 1A), along with the incidence and cumulative incidence curves, shows the median number of diagnoses during the period under analysis (median = 133). The peak of the curve is defined by the period in which the number of diagnoses was higher than the median value, and this corresponds to the period between March 19 and April 24. Using the same methodology, based on Figure 1B), the peak of hospitalizations was defined between March 11 and May 2, (corresponding to the time period in which the daily number of hospitalizations was greater than 17). The peak of deaths among infected with COVID-19 during the first wave of the pandemic (Figure 1C)) was likewise defined between March 25 and May 6 (corresponding to the time period in which the daily number of deaths was greater than 9). We can observe that there is a time lag between the peak of diagnosis (April 24), hospitalizations (May 2), and deaths (May 6).

Characterization of the Infected Population

The descriptive statistics includes all individuals diagnosed with COVID-19 between March 19 and April 24, corresponding to the first wave considering the number of diagnosis (n=10296). As shown in Table 1, 5902 (57.32%) women and 4394 (42.68%) men were included in this exploratory analysis. Moreover, 7632 (74.13%) individuals were under 64 years old, 1480 (14.37%) were between 65 and 79 years and 1184 (11.50%) were 80 years or over. Regarding geographic regions, the distribution of COVID-19 infections between regions was 24.57% in LTV, 57.80% in the north region, 13.89% in the center, 1.09% in the south and 2.65% in the islands. Considering the distribution of infection rates per 100,000 habitants were 68.70 in LTV (2530 individuals from a population of 3,682,860), 160.51 in the north of Portugal (5951 individuals from a population of 3,573,961), 87.92 in the center (1430 individuals from a population of 1,626,462), 12.36 in the south (112 individuals from a population of 906,060) and 52.93 in both Madeira and Azores Islands (273 individuals from a population of 496,921). The overall proportion of deaths in the data set was 5.6% (n=578), the proportion of hospitalizations were 18.8%, and 44.77% of the infected individuals presented at least one comorbidity.

For the mortality endpoint, it was possible to notice a reduced lethality in several groups, especially in individuals aged up to 64 years old (0.71%), with no comorbidities (0.40%) or hospitalizations (1.35%). On the other hand, individuals aged 80 and over are those with the highest lethality (30.60%), followed by individuals that were hospitalized with COVID-19 (23.97%) and those with any comorbidity (12.04%). The South region also presents a higher lethality (10.71%) when compared to the other Portuguese regions.

Considering the outcome hospitalization, it is possible to verify that the incidence is considerably higher in every analyzed group compared to the mortality endpoint, as expected. The relative frequency of hospitalizations reaches a maximum value of 59.46% for individuals aged over 80 and a minimum of 4.84% for individuals without comorbidities (see Table 1). It is also possible to see a very high incidence in the South region (39.29%), this being the third highest incidence, preceded by individuals aged between 65 and 79 (41.15%).

Determination of Risk Factors for Mortality and Hospitalization

The multivariable logistic regression analysis for the outcomes mortality and hospitalization are presented in Table 2. For both outcomes, individuals aged 80 or more are the ones with the highest risk for the occurrence of death and hospitalization of approximately 12.0 and 8.5 times, respectively, when compared to individuals under 64 years. The existence of comorbidities also proved to be an important risk factor for the occurrence of both outcomes (Mortality: OR = 5.74, 95% CI: 3.33-10.64 and Hospitalization: OR = 5.19, 95% CI: 4.36-6.21). For the mortality outcome, individuals that have been hospitalized with COVID-19 also present a higher risk (OR=6.48, 95% CI: 4.87-8.71) for the occurrence of death when compared to those not hospitalized. Male gender is considered significant only for the outcome hospitalization, with an associated risk that is 1.6 times higher than that of female individuals. Likewise, the region was also considered to be significant only for the outcome hospitalization. The South region presents a risk of hospitalizations among those infected with COVID-19, 1.9 times higher than the LTV region. On the other hand, the North, Center, and Island regions have a lower risk of hospitalization compared to the reference region (OR = 0.60, 95% CI: 0.51-0.71; OR = 0.59, 95% CI: 0.46-0.74; and OR = 0.47, 95% CI: 0.28-0.76, respectively).

Performance Evaluation of the Regression Models

The roc curves presented in Figure 2 show a good performance for both models, with AUC = 0.911, 95% CI: 0.920-0.945 for the death predictor model and AUC = 0.845, 95% CI: 0.828-0.864 for the hospitalization predictor model.

Discussion

Although, currently, COVID-19 is an extremely addressed topic all over the world, in Portugal, there are few articles that analyze the epidemiological surveillance data made available by DGS, especially those of June 30. This study only addresses the peak of the first wave of the pandemic, instead of analyzing the whole available data, being unique in this sense. This can provide particularly relevant and specific information about the characteristics of the infected population during this phase of exponential growth of the pandemic. Also, it was especially relevant to define the main risk factors for both the occurrence of deaths and hospitalizations among those infected.

The first obstacle faced during the present study was the determination of the peak of the first wave of the pandemic, since there is no consensual date described in the literature. Therefore, to do so, we studied the epidemiological curves, analyzed the behavior of the cumulative curves, and used the median for delimiting the peak of the pandemic for each of the specified clinical outcomes (13,14). It was also possible to identify, a lag of approximately 12 days between the end of the peak of diagnoses and deaths from COVID-19. This is slightly above the average lag between daily COVID-19 cases and deaths reported by R. Jin et al. (2021), of approximately 8 days (SD= 4 days) for nineteen regions. However, these results are similar in some included countries such as Switzerland (15). The peak of the pandemic was defined based on the diagnosis wave so that we were able to have a control group for both outcomes.

In this study, the main risk factors established for both outcomes included old age (over 65 years) and the existence of comorbidities, which is concordant with other previously published articles addressing this matter (9). For the mortality outcome, hospitalization was also considered as a risk factor. In the study by PJ Nogueira et al. (2020), the risk factors obtained for mortality were: age (despite being more stratified, it is possible to verify a progressive increase in the OR with increasing age: 0–55 years: Reference.group 66–70 years: OR = 20.4; 71–75 years: OR = 34.0; 76–80 years OR = 50.9; 81–85 years: OR = 70.7; 86–90 years: OR = 83.2); the existence of comorbidities (with the OR calculated individually for each pathology); and also the male gender (OR = 1.47) that in our study was only considered as statistically significant for the outcome hospitalization (9).

Advanced age proved to be one of the main risk factors for both mortality and hospitalization with COVID-19 at the peak of the first wave of the pandemic.

Furthermore, this risk proved to be greater with increasing age. Individuals over the age of 65 are, in general, associated with worse prognosis (both hospitalization and mortality) being this risk even greater for individuals over 80 years of age (OR=11.98 for mortality; OR=8.54 for hospitalization). In Z. Zheng et al. (2020) age over 65 years old, was one of the main risk factors reported for disease progression in patients with COVID-19 with an OR = 6.06, 95% CI: 3.98-9.22 (16). This can be particularly serious in a country such as Portugal that has an extremely aged population with more than 2.2 million (22.0%) being over 65 years (data from 2019) (17).

The existence of comorbidities was another factor considered to have a major influence in determining the clinical outcome in patients infected with COVID-19. The existence of preconditions has been described as an important factor in determining the clinical outcome of patients with COVID-19 (16,18,19). In this study, even though the dataset provides information regarding some individual comorbidities, only the general variable was analyzed, since the individual data had several missing values that could give rise to biased estimates that do not reflect the real situation.

Although, in most studies, hospitalization is considered only as a clinical outcome (e.g. RC Menezes Soares, 2020), in this study, we considered it as a risk factor for the occurrence of death (20). A higher lethality was shown in infected patients who have been hospitalized. On the other hand, this outcome can also be considered as a confounding factor since hospitalized patients usually present some characteristics such as older age or comorbidities that can be the real risk factors for the occurrence of death.

The gender variable showed significant results only regarding the outcome hospitalization (16). According to the present study, males have a 1.58 higher risk of hospitalization when compared to females. Similarly, V.Jain et al. reported being male as a non-significant risk factor for COVID-19 ICU hospitalization (OR=1.15, 95% CI: 0.89-1.48) (19).

Finally, taking into account the distinct regions, significant results were also detected regarding hospitalizations of infected individuals. In the North, Center, and Islands the risk was significantly lower when compared to the reference region, LTV (OR = 0.60, OR = 0.59, and OR = 0.47, respectively). In contrast, the South region had a significantly higher risk of hospitalization compared to LVT (OR = 1.91). To our knowledge, no other published studies present odds risk results about regions, representing a novelty and interesting information for public health decision makers. However, these results must be carefully interpreted since there are several factors, in addition to those included in the analysis, that may influence the decision to hospitalize a patient. For instance, the number of hospitalizations may be related to the specific guidelines defined for each hospital, the number of available beds for this purpose and the irregular distribution of the pandemic in Portugal, especially during this first wave. For example, the greater risk of hospitalizations in regions where there are a smaller number of infected individuals (e.g., South). In this situation the number of available beds may be higher and the decision to hospitalize may not be so restrictive. Therefore, the region should not be directly interpreted as a risk factor for the occurrence of hospitalization without further information regarding other potential confounding factors.

The main limitation of our study is related to the lack of consistency in data registration, which can lead to the existence of some abnormalities. When the data was made available, an alert was made by the Portuguese DGS to the fact that the total number of cases in the file does not coincide with those reported in the DGS daily bulletin since it is surveillance data from medical notifications. Nevertheless, there are strong criticisms in the literature regarding the quality of the data provided, namely, the discrepancy in the proportion of patients infected with comorbidities between the data made available in April and the data updated on the 30th of June, as well as the extreme increase in the proportion of missing values in this variable (21). In addition to the quality problems referred by C. Costa-Santos et al. (2020), there are many others reported by some researchers, such as the existence of pregnant men, a 134-year-old person and more than 4,000 missing data between the April and June database (22). Moreover, during this study it was possible to find some non-documented inconsistencies regarding the dates of diagnosis and hospitalization. In some cases, the presented dates were from before the beginning of the pandemic of COVID-19 in Portugal (March 2, 2020). Also, there was one month of missing data in the variable containing the dates of deaths (death dates were only reported until the 24th of May). In order to minimize the impact of these problems in the present study, the database was truncated so that observations with dates of diagnosis, hospitalization or death outside the acceptable limits were excluded from the analysis, and the variables of the specific comorbidities were not used in this study.

Conclusions

The present study allowed us, initially, to determine the peak of the first wave of the COVID-19 pandemic in Portugal and also to demonstrate that, in general, the major risk factors associated with worse prognostics of this disease, during this period, were advanced age (over 65 years), and the existence of comorbidities. For the risk of hospitalizations, the male gender was also considered a major factor.

Since this study only addresses a very specific fraction of the COVID-19 cases in Portugal during the first wave of the pandemic, it may be interesting, in future works, to perform similar analyzes for subsequent waves and to compare the results obtained from each analysis. This will allow a phase-specific characterization of the infected population and determination of the predominant risk factors in each of them.

Acknowledgements

Data used in this work was made available by the Portuguese Directorate-General of Health (DGS), from June the 30th.

References

1. World Health Organization. Coronavirus disease (COVID-19) [Internet]. 2020 [cited 2020 Dec 30]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19
2. Chams N, Chams S, Badran R, Shams A, Araji A, Raad M, et al. COVID-19: A Multidisciplinary Review. Front Public Heal. 2020;8:1–20.
3. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020;395(10229):1054–62.
4. Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol. 2021;18(3):141–54.
5. Awortwe C, Cascorbi I. Meta-analysis on outcome-worsening comorbidities of COVID-19 and related potential drug-drug interactions. Pharmacol Res. 2020;161(105250).
6. Hu T, Liu Y, Zhao M, Zhuang Q, Xu L, He Q. A comparison of COVID-19, SARS and MERS. PeerJ. 2020;8:1–30.
7. Dorjee K, Kim H, Bonomo E, Dolma R. Prevalence and predictors of death and severe disease in patients hospitalized due to COVID-19: A comprehensive systematic review and meta-analysis of 77 studies and 38,000 patients. PLoS One. 2020;15(12):1–27.
8. Barek MA, Aziz MA, Islam MS. Impact of age, sex, comorbidities and clinical symptoms on the severity of COVID-19 cases: A meta-analysis with 55 studies and 10014 cases. Heliyon. 2020;6(12):1–24.
9. Nogueira PJ, de Araújo Nobre M, Costa A, Ribeiro RM, Furtado C, Bacelar Nicolau L, et al. The Role of Health Preconditions on COVID-19 Deaths in Portugal: Evidence from Surveillance Data of the First 20293 Infection Cases. J Clin Med. 2020;9(8):15.

10. Ministério da Saúde. Disponibilização de Dados [Internet]. Disponibilização de Dados - COVID-19. 2020 [cited 2020 Dec 28]. Available from: https://covid19.min-saude.pt/disponibilizacao-de-dados/
11. Ministério da Saúde. Ponto de Situação Atual em Portugal [Internet]. Ponto de Situação Atual em Portugal - COVID-19. 2021 [cited 2021 Feb 9]. Available from: https://covid19.min-saude.pt/ponto-de-situacao-atual-em-portugal/
12. Uçar MK, Nour M, Sindi H, Polat K. The Effect of Training and Testing Process on Machine Learning in Biomedical Datasets. Math Probl Eng. 2020;2020:17.
13. Spiers R. Getting ahead of the curve – what are epidemic curves and why are they important? [Internet]. Getting ahead of the curve – what are epidemic curves and why are they important? | HSC Public Health Agency. 2020 [cited 2021 Mar 18]. Available from: https://www.publichealth.hscni.net/node/5257
14. Torok M. Epidemic Curves Ahead. 2002.
15. Jin R. The lag between daily reported Covid-19 cases and deaths and its relationship to age. J Public health Res [Internet]. 2021 Mar 9; Available from: https://www.jphres.org/index.php/jphres/article/view/2049
16. Zheng Z, Peng F, Xu B. Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. J Infect. 2020;81(2):16–25.
17. População residente, média anual: total e por grupo etário [Internet]. PORDATA - População residente, média anual: total e por grupo etário. 2020 [cited 2021 Mar 17]. Available from: https://www.pordata.pt/Portugal/População+residente++média+anual+total+e+por+grupo+etário-10-1140
18. Gold MS, Sehayek D, Gabrielli S, Zhang X, McCusker C, Ben-Shoshan M. COVID-19 and comorbidities: a systematic review and meta-analysis. Postgrad Med. 2020;132(8):749–55.
19. Jain V, Yuan J-M. Predictive symptoms and comorbidities for severe COVID-19 and intensive care unit admission: a systematic review and meta-analysis. Int J

Public Heal. 2020;65(5):533–46.

20. Menezes Soares R de C, Mattos LR, Raposo LM. Risk Factors for Hospitalization and Mortality due to COVID-19 in Espírito Santo State, Brazil. Am J Trop Med Hyg. 2020;103(3):1184–90.

21. Costa-Santos C, Ribeiro-Vaz I, Monteiro-Soares M. The Hidden Factor—Low Quality of Data is a Major Peril in the Identification of Risk Factors for COVID-19 Deaths: A Comment on Nogueira, P.J., et al. "The Role of Health Preconditions on COVID-19 Deaths in Portugal: Evidence from Surveillance Data of the First 20293 Infection Cases". J Clini Med. 2020;9(11).

22. Novais V. Infetados antes de março e registos de mortos que desapareceram. Erros nas bases de dados da DGS. Infetados antes de março e registos de mortos que desapareceram Erros nas bases de dados da DGS – Observador [Internet]. 2020 Oct 9 [cited 2020 Dec 23]; Available from: https://observador.pt/2020/11/09/infetados-antes-marco-e-registos-de-mortos-que-desapareceram-erros-nas-bases-de-dados-da-dgs/

Tables

Table 1: Characterization of the population with COVID-19 during the peak of the 1st wave of the pandemic in Portugal for the outcomes mortality and hospitalization. All cases of COVID-19 were reported between March 2 and June 30, 2020.

Variables	n(%)	Mortality		Hospitalization	
		Cases(%)	χ^2 Test	Cases(%)	χ^2 Test
Age Group					
<64 years	7632(74.13)	54(0.71)	$X^2(2) = 1825.7$ $p < 0.0001$	627(8.22)	$X^2(2) = 1825.7$ $p < 0.0001$
65-79 years	1480(14.37)	161(10.88)		609(41.15)	
≥ 80 years	1184(11.50)	363(30.66)		704(59.46)	
Gender					
Female	5902(57.32)	287(4.86)	$X^2(1) = 14.393$ $p < 0.0001$	944(15.99)	$X^2(1) = 14.393$ $p = 0.0001$
Male	4394(42.68)	291(6.62)		996(22.67)	
Region					
LTV	2530(24.57)	127(5.02)	$X^2(4) = 14.189$ $p = 0.007$	537(21.23)	$X^2(4) = 14.189$ $p = 0.007$
North	5951(57.80)	339(5.70)		1021(17.16)	
Center	1430(13.89)	93(6.50)		308(21.54)	
South	112(1.09)	12(10.71)		44(39.29)	
Islands	273(2.65)	7(2.56)		30(10.99)	
Hospitalization					
No	8356(81.16)	113(1.35)	$X^2(1) = 1515.7$ $p < 0.0001$	-	-
Yes	1940(18.84)	465(23.97)		-	-
Comorbidities					
No	5686(55.23)	23(0.40)	$X^2(1) = 648.19$ $p < 0.0001$	275(4.84)	$X^2(1) = 648.19$ $p < 0.0001$
Yes	4610(44.77)	555(12.04)		1665(36.12)	

LTV: Lisbon and Tagus Valley; OR: Odds Ratio; 95%CI: 95% Confidence Interval;

Table 2: Analysis of multivariable logistic regression for mortality and hospitalization outcomes in individuals infected with COVID-19 at the peak of the 1st wave of the pandemic in Portugal. All cases of COVID-19 were reported between March 2 and June 30, 2020.

Variables	**Mortality(n=7207)**		**Hospitalization(n=7207)**	
	OR	**95%CI**	**OR**	**95%CI**
Age Group				
<64 years (Ref.G)	1.000		1.000	
65-79 years	4.064 ***	2.762; 6.085	3.899 ***	3.287; 4.628
≥ 80 years	11.978 ***	8.291; 17.689	8.537 ***	7.065; 10.335
Gender				
Female (Ref.G)	1.000			
Male	1.245 ·	0.976; 1.590	1.575 ***	1.368; 1.813
Region				
LTV (Ref.G)	1.000			
North	1.243	0.925; 1.683	0.602 ***	0.510; 0.710
Center	1.000	0.675; 1.476	0.586 ***	0.464; 0.738
South	1.301	0.493; 3.196	1.912 *	1.027; 3.506
Islands	0.685	0.196; 1.833	0.466 **	0.275; 0.763
Hospitalization				
No (Ref.G)	1.000			
Yes	6.479 ***	4.874; 8.711	-	-
Comorbidities				
No (Ref.G)	1.000		1.000	
Yes	5.735***	3.331; 10.642	5.193 ***	4.356; 6.212

LTV: Lisbon and Tagus Valley; OR: Odds Ratio; 95%CI: 95% Confidence Interval; Ref. G.: Reference group;

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figures

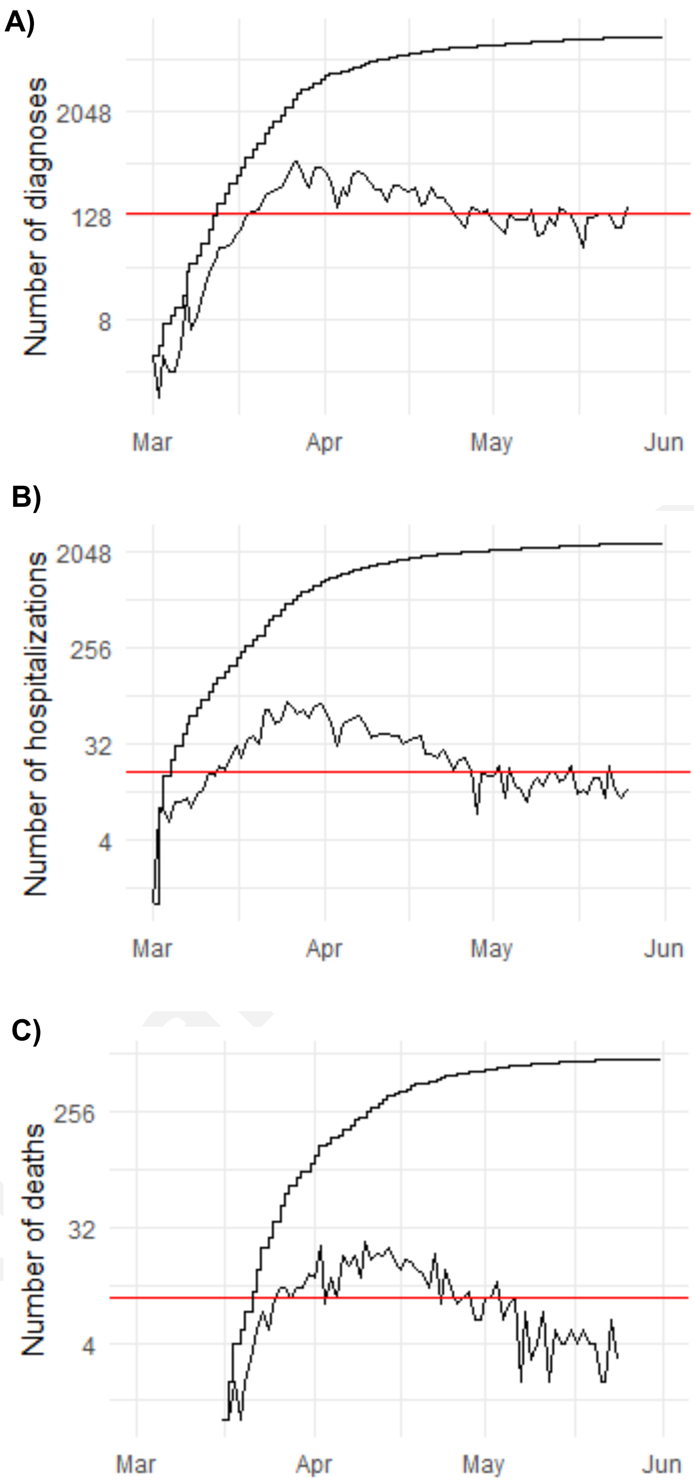

Figure 1: Evolution of the COVID-19 pandemic in Portugal between March 1 and May 24 including cumulative curves considering: A) number of diagnosis (median=133); B) number of hospitalizations (median=17) and C) number of deaths (median=9).

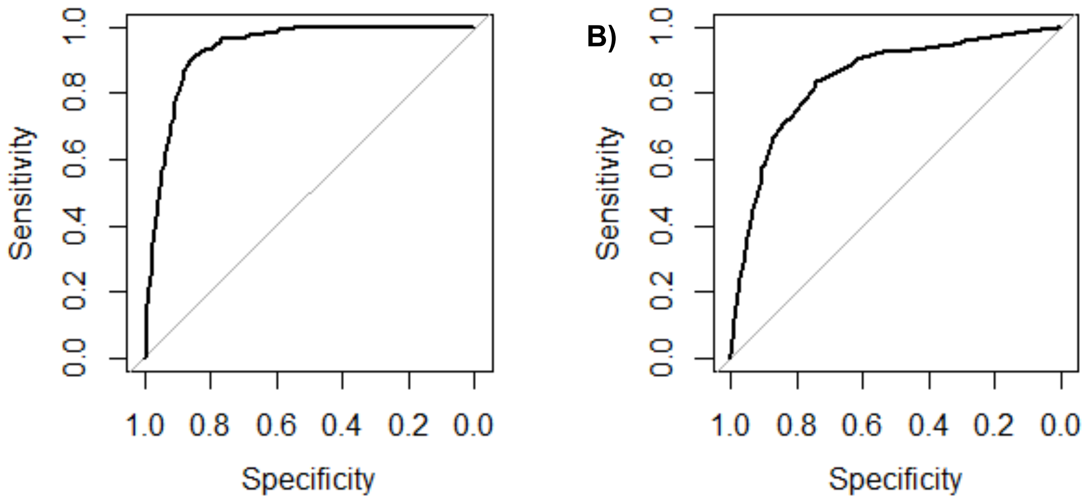

Figure 2: ROC curves with representation of sensitivity and specificity for regression models A) to predict mortality and B) to predict hospitalizations.