Taylor & Francis
Taylor & Francis Group

Current Medical Research and Opinion

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/icmo20

The importance of association of comorbidities on COVID-19 outcomes: a machine learning approach

José Carlos Arévalo-Lorido, Juana Carretero-Gómez, Jose Manuel Casas-Rojo, Juan Miguel Antón-Santos, José Antonio Melero-Bermejo, Maria Dolores López-Carmona, Lidia Cobos Palacios, Jaime Sanz-Cánovas, Paula Maria Pesqueira-Fontán, Andrés Alberto de la Peña-Fernández, Navas-Maria de la Sierra Alcántara, Gema Maria García-García, José David Torres Peña, Jeffrey Oskar Magallanes-Gamboa, Rosa Fernández-Madera-Martinez, Javier Fernández-Fernández, Manuel Rubio-Rivas, Guillermo Maestro-de la Calle, Eva Cervilla-Muñoz, Antonio Ramos-Martínez, Manuel Méndez-Bailón, José Manuel Ramos-Rincón, Ricardo Gómez-Huelgas & On Behalf of SEMI-COVID-19 Network

To cite this article: José Carlos Arévalo-Lorido, Juana Carretero-Gómez, Jose Manuel Casas-Rojo, Juan Miguel Antón-Santos, José Antonio Melero-Bermejo, Maria Dolores López-Carmona, Lidia Cobos Palacios, Jaime Sanz-Cánovas, Paula Maria Pesqueira-Fontán, Andrés Alberto de la Peña-Fernández, Navas-Maria de la Sierra Alcántara, Gema Maria García-García, José David Torres Peña, Jeffrey Oskar Magallanes-Gamboa, Rosa Fernández-Madera-Martinez, Javier Fernández-Fernández, Manuel Rubio-Rivas, Guillermo Maestro-de la Calle, Eva Cervilla-Muñoz, Antonio Ramos-Martínez, Manuel Méndez-Bailón, José Manuel Ramos-Rincón, Ricardo Gómez-Huelgas & On Behalf of SEMI-COVID-19 Network (2022): The importance of association of comorbidities on COVID-19 outcomes: a machine learning approach, Current Medical Research and Opinion, DOI: 10.1080/03007995.2022.2029382

To link to this article: https://doi.org/10.1080/03007995.2022.2029382

The importance of association of comorbidities on COVID-19 outcomes: a machine learning approach

José Carlos Arévalo-Lorido[1] , Juana Carretero-Gómez[1] , Jose Manuel Casas-Rojo[2] , Juan Miguel Antón-Santos[2] , José Antonio Melero-Bermejo[2], Maria Dolores López-Carmona[3], Lidia Cobos Palacios[3], Jaime Sanz-Cánovas[3], Paula Maria Pesqueira-Fontán[4], Andrés Alberto de la Peña-Fernández[5], Navas- Maria de la Sierra Alcántara[6], Gema Maria García-García[1], José David Torres Peña[7,17,18], Jeffrey Oskar Magallanes-Gamboa[8], Rosa Fernández-Madera-Martinez[9], Javier Fernández-Fernández[10], Manuel Rubio-Rivas[11,19], Guillermo Maestro-de la Calle[12], Eva Cervilla-Muñoz[13], Antonio Ramos-Martínez[14], Manuel Méndez-Bailón[15], José Manuel Ramos-Rincón[16], Ricardo Gómez-Huelgas[3,20]; On Behalf of SEMI-COVID-19 Network

1. Complejo Hospitalario Universitario de Badajoz, Spain

2. Hospital Universitario Infanta Cristina de Parla. Madrid, Spain

3. Hospital Regional Universitario de Málaga, Spain

4. Hospital Clínico Universitario de Santiago de Compostela, A Coruña, Spain

5. Hospital Universitario Son Llàtzer. Palma de Mallorca. Illes Balears, Spain

6. Hospital Infanta Margarita de Cabra. Córdoba, Spain

7. Hospital Universitario Reina Sofía. Córdoba, Spain

8. Hospital Nuestra Señora del Prado. Talavera de la Reina. Toledo, Spain

9. Hospital de Cabueñes. Asturias, Spain

10. Hospital de Mataró. Barcelona, Spain

11. Hospital Universitario de Bellvitge. Barcelona, Spain

12. Hospital Universitario 12 de Octubre. Madrid, Spain

13. Hospital Universitario Gregorio Marañón. Madrid, Spain

14. Hospital Universitario Puerta de Hierro. Madrid, Spain

15. Hospital Clínico Universitario San Carlos. Madrid, Spain

16. Universidad Miguel Hernandez. Elche. Alicante, Spain

17. Lipids and Atherosclerosis Unit, Department of Internal Medicine, Maimonides Biomedical Research Institute of Cordoba (IMIBIC), Spain

18. CIBER Fisiopatología de la Obesidad y Nutrición (CIBEROBN), Instituto de Salud Carlos III, Madrid, Spain

19. Bellvitge Biomedical Research Institute-IDIBELL, Barcelona, Spain

20. Biomedical Research Institute of Málaga (IBIMA), Málaga, Spain

Correspondence:

José Carlos Arévalo-Lorido

Servicio de Medicina Interna. Complejo Hospitalario Universitario de Badajoz. Avda de Elvas s/n. 06080 Badajoz, Spain

Tel.: 0034 626 853 325

Email: joscarlor@gmail.com

Transparency

Declaration of funding

There was no sponsorship/funding for this manuscript.

Declaration of financial/other relationships

The authors of the present manuscript declare they have not any conflict of interest. Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

Author contributions

JC Arevalo-Lorido, Juana Carretero-Gomez, José Manuel Casas-Rojo and Ricardo Gómez-Huelgas were involved in the conception, design and interpretation of the data. José Manuel Casas-Rojo was additionally involved in data curation; José Carlos Arévalo-Lorido was additionally involved in the analysis of the data. All the authors were involved in the drafting of the paper or in its critical review for intellectual content; and the final approval of the version to be published. All authors agree to be accountable for all aspects of the work.

Acknowledgments

Authors gratefully acknowledge all the investigators who participate in the SEMI-COVID-19 registry.

Abstract

Background: The individual influence of a variety of comorbidities on COVID-19 patient outcomes has already been analyzed in previous works in an isolated way. We aim to determine if different associations of diseases influence the outcomes of inpatients with COVID-19.

Methods: Retrospective cohort multicenter study based on clinical practice. Data were taken from the SEMI-COVID-19 Registry, which includes most consecutive patients with confirmed COVID-19 hospitalized and discharged in Spain. Two machine learning algorithms were applied in order to classify comorbidities and patients (Random Forest -RF algorithm, and Gaussian mixed model by clustering -GMM-). The primary endpoint was a composite of either, all-cause death or intensive care unit admission during the period of hospitalization. The sample was randomly divided into training and test sets to determine the most important comorbidities related to the primary endpoint, grow several clusters with these comorbidities based on a discriminant analysis and GMM, and compare these clusters.

Results: A total of 16,455 inpatients (57·4% women and 42·6% men) were analyzed. According to the RF algorithm, the most important comorbidities were heart failure/atrial fibrillation (HF/AF), vascular diseases, and neurodegenerative diseases. There were six clusters: three included patients who met the primary endpoint (clusters 4, 5, and 6) and three included patients who did not (clusters 1, 2, and 3). Patients with HF/AF, vascular diseases, and neurodegenerative diseases were distributed among clusters 3, 4 and 5. Patients in cluster 5 also had kidney, liver, and acid peptic diseases as well as chronic obstructive pulmonary disease; it was the cluster with the worst prognosis.

Conclusion: The interplay of several comorbidities may affect the outcome and complications of inpatients with COVID-19.

Keywords

COVID-19; SARS-CoV-2; Comorbidity; Cluster analysis; Machine learning;

Short title: Role of comorbidities in COVID-19

Introduction

Many patients who had critical COVID-19 had underlying illnesses such as cardiovascular disease, or neoplasms[1]. Previous studies in patients with influenza A virus subtype H7N9 infection have shown that the existence of any comorbidity has been related to a three- to four-fold increase in risk of acute respiratory distress syndrome[2]. In the case of COVID-19, evidence from the pandemic has clearly demonstrated that patients with certain comorbidities are at a much greater risk of dying[3,4]. Given that many of these diseases are strongly associated with each other, it is common for patients to have multiple comorbidities.

To date, these diseases have been analyzed independently of one another, despite the fact that multiple comorbidities may merge different mechanisms that affect the clinical course, and prognosis of COVID-19. In light of this lack of evidence on combinations of comorbidities, this study aims to observe if different associations of diseases influence COVID-19 outcomes. We hypothesize that certain combinations of diseases may be more harmful than others and could determine poor outcomes for patients who have them.

Material and methods

This is an ongoing retrospective cohort study based on clinical practice that includes most consecutive patients with confirmed COVID-19 who were hospitalized and discharged in Spain from March 1, 2020. The first patient was included on March 24, 2020. Our analyses included patients hospitalized to November 23, 2020.

Patient selection and data collection

Patients are recruited through a registry (SEMI-COVID-19) sponsored by the Spanish Society of Internal Medicine which has been approved by the Provincial Research Ethics Committee of Málaga, Spain. Personal data are processed in strict compliance with Spanish Law 14/2007, of July 3, on Biomedical Research; and Spanish Organic Law 3/2018, of December 5, on the Protection of Personal Data and the Guarantee of Digital Rights.

All consecutive patients with confirmed SARS-CoV-2 infection who have been discharged or died after hospital admission are eligible for inclusion. COVID-19 was confirmed either by a positive result on real-time polymerase chain reaction testing of a nasopharyngeal or sputum sample or by a positive result on serological testing and consistent clinical presentation. Inclusion criteria for the registry are: a) age $\geq$ 18 years, b) confirmed COVID-19 diagnosis, c) first admission in a Spanish hospital participating in the study, and d) hospital discharge or in-hospital death. Exclusion criteria are subsequent admissions of the same patient and denial or withdrawal of informed consent. More information on the registry is available in a previously published study[5].

For the purpose of simplifying the analysis, we categorized the diseases related to patients' medical history into 13 groups of comorbidities (Table 1) as follow: neurodegenerative diseases, chronic kidney diseases (CKD), autoimmune rheumatic diseases, mental health conditions, obstructive lung diseases, human immunodeficiency virus (HIV) infection and acquired immunodeficiency syndrome (AIDS), hematologic malignancies, solid malignant neoplasms, cardiovascular risk factors, heart failure and atrial fibrillation, vascular diseases, digestive disorders and obesity and obstructive sleep apnea (OSA).

Study outcomes

The primary endpoint was a composite of either, all-cause death or intensive care unit (ICU) admission during the period of hospitalization. Secondary endpoints included all-cause death or ICU admission separately as well as complications during hospitalization, including acute lung injury (ALI) (Berlin criteria[6]), heart failure (according to ESC guidelines[7]), thromboembolic disease, acute kidney injury (AKI) (creatinine $\geq 1{\cdot}5$ times baseline values or an increase greater than $0{\cdot}3$ mg/dL within any 48-hour period or urine volume $< 0{\cdot}5$ mL/kg for six to twelve hours), and disseminated intravascular coagulation (ISTH criteria[8]).

According to the primary endpoint, two different methods were used to first determine which comorbidities would be primarily chosen in a decision tree using a random forest algorithm and then to determine which groups of comorbidities were more likely to precipitate one of the primary endpoint events using a discriminant analysis with Gaussian mixture model (GMM) clustering. For both algorithms, the sample was randomly divided into a training and test sets. The training sets included two-thirds of the sample in the case of clustering and 70% of the sample in the case of random forest intending to avoid coincidences among them. The test sets included the remaining portion of the samples. We grew the two algorithms with the training sets and proved the results in the test sets.

Random forest

Random forests (RF) are a classification and regression method based on the aggregation of a large number of decision trees[9]. RF is a very powerful ensemble machine learning algorithm which works by creating multiple decision trees and then combining the output generated by each of them (each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction). Split selection was carried out based on the decrease of Gini impurity (DGI), which is the procedure followed in the most commonly used type of RF. This version of RF is used in the 'randomForest' package[10]. In our analysis, the number of trees was set to ntree=1000 and the number of candidate predictors considered at each split was set to the default value of mtry=2 in the main analysis.

For this RF, two types of variable importance measures (VIMs) were considered: the mean DGI and the unscaled permutation-based importance measure. Both are implemented in the 'importance' function of the 'randomForest' package[11]

Cluster generation

The GMM is a probabilistic model for representing subpopulations that are normally distributed which form part of an overall population. The most frequently used distribution in modeling real-world unimodal data is Gaussian distribution. Modeling multimodal data using a combination of multiple unimodal Gaussian distributions is intuitively reasonable. The mclust package in r was used for this purpose[12]. We performed a mixture discriminant analysis (MDA, a technique that is used by the researcher to analyze the research data when the criterion or the dependent variable is categorical, in our case presence/absence of primary end-point) in which the data determined the best suited parametrization of the covariance and the number of mixture components[13].

To generate the model, the previously defined comorbidity variables were first converted into numerical values (Table 1) and to avoid overrepresentation of variables with more categories and their ordinality, then they were normalized (rescaling by the minimum and range of the vector, to make all the elements lie between 0 and 1 thus bringing all the values of numeric columns in the variable to a common scale).

To select the best Gaussian parameters of covariance with the lowest error, we performed a cross-validation process with the training and test sets. In our case, the best model was the EEV (which belongs to the general family with Equal volume, Equal size, and Variable orientation), with three clusters for patients who experienced an endpoint event (clusters 4 to 6) and three clusters for those who did not (clusters 1 to 3), by the mean of discriminant analysis.

The classification error into the clusters among training and test sets was 0·26.

Statistical analysis

Once the clusters were created, comparisons among the comorbidities by cluster was carried out using the chi-square test. We calculated the residuals in order to observe the contribution of the variable to the magnitude of the value obtained using the chi-square test. Other qualitative variables beyond those included in the clusters were also compared using the chi-square test and were shown as absolute values and percentages. Quantitative variables were first analyzed for normality using the Kolmogorov–Smirnov test and were then analyzed for equality of variances using Levene's test. As a result, all variables had nonparametric distribution and therefore we compared them using ANOVA and Welch's t-test for those with inequality of variances and with the Kruskal–Wallis one-way analysis of variance test for the rest. The values were shown as

medians and interquartile ranges. Lastly, a Kaplan-Meier curve was created for all-cause death by clusters and differences were using the log-rank test.

All analyses were conducted using R version 3.3.2 (R Core Team 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/).

Reporting of the study conforms to broad EQUATOR guidelines[14]

Results

A total of 16455 inpatients (57·4% men and 42·6% women) with a median age of 66·4 (27·7) years (supplementary figure 1) were analyzed. Of the total study population, 4241 died or were admitted to the ICU; the remainder were discharged alive.

Random forest

The error rate of the prediction was 25·5%. The model performed well on the train and test set, yielding an accuracy of 0·74.

The figure 2 shows the importance of the comorbidities selected. Depending on the method, HF and AF, CVRF, vascular diseases, and neurodegenerative diseases are the most frequently used variables in decision-making when determining which patients will meet the primary endpoint.

Clusters

A total of six clusters were defined. The most salient characteristics of the clusters are shown in tables 2 and 3 and figure 2.

Cluster 1 included 8765 patients (44·1% females). The most significant comorbidities were asthma, obesity, OSA, and PD. The cluster had a low rate of CVRF (mainly hypertension) and the patients' biochemical data showed less inflammation than patients in other clusters. Cluster 2 included 798 patients (44·9% females). The main comorbidities were HIV infection, hematologic malignancies, solid malignant neoplasms, and rheumatic disorders. Very few patients in this cluster had cardiovascular disease or CVRF. Cluster 3 included 2651 patients (46·4% females). The main comorbidities were HF and AF, several CVRF, vascular diseases, and neurodegenerative diseases. Patients in this cluster also had a considerable rate of rheumatic diseases. Cluster 4 included 3557 patients (36·4% females). Like cluster 3, the main comorbidities were CVRF, vascular diseases, HF/AF, and neurodegenerative diseases. However, unlike cluster 3, there was also a high rate of obesity, OSA, COPD, and depression in cluster 4. Cluster 5 included 457 patients (38·1% females). This cluster also had varied vascular diseases, HF/AF, and COPD. In contrast to the previous clusters, digestive disorders and CKD were predominant. Lastly, cluster 6 included 227 patients (37·9%

females). The main comorbidities were hematologic malignancies and rheumatic diseases. Some patients also had HIV infection and HF/AF.

Overall, patients in clusters 4, 5, and 6 were older and had more comorbidity (Charlson index). They presented with a higher inflammatory burden on their laboratory findings as well as more disturbances in data related to coagulation (table 3).

Outcomes

Figure 3 shows the rate of several complications by cluster. HF was more prevalent in clusters 5 and 6 and ALI in clusters 4 and 5. Cluster 5 had the highest rate of AKI and clusters 5 and 6 had the highest rate of DIC. Lastly, thrombotic disorders were more prevalent in cluster 2, 4, and 6, with more endpoint events due to pulmonary embolism (PE) than deep vein thrombosis (DVT). In regard to mortality, cluster 5 had the worst outcome (p = ·00). Since clusters 4, 5 and 6 included older patients than the remainder, we have stratified them according to the age (Supplementary table 1). The outcomes kept homogeneous to the total sample with similar significance including the age.

Discussion

This study shows the relative importance of several diseases regarding the prognosis in patients hospitalized for COVID-19. In our RF algorythm, the most common comorbidities related to a poor prognosis were those related to CVRF as well as obesity, obstructive respiratory diseases, HF/AF, and neurodegenerative diseases. Although this algorithm is subject to the limitations derived from including several diseases in the same category without taking into account their different severity, its results provide information on the different importance that these categories have with respect to a poorer prognosis, and complement the information derived from the clusters.

The presence of these comorbidities was similar to those reported in China[1] and New York City[15] except for neurodegenerative diseases, which were slightly more prevalent in our series[16].

Additionally, our results also showed the influence of age on the prognosis of COVID19. The clusters which experienced the primary end-point had higher age than the remainder. Convincing support from around the world suggests that age itself is a very significant risk factor for severe COVID-19 disease. However, this relationship with clinical severity is complicated to elucidate because many of the subjacent medical conditions that grow risk for severe illness from COVID-19 are more widespread with increasing age, especially in a sample of hospitalized people where finding healthy aging is quite strange. In fact, it was also reported that biological aging was an hepful predictor of disease severity after implementing biological age evaluations comprised of chronological age and nine biomarkers[17]. We have actually stratified our results by age and obtained similar findings being unable to do away with its influence on them.

As like the age, it should be noted the significant differences regarding the sex. Our study showed how males were predominant, mainly in the clusters with poor prognosis. Fewer women, both young and old, are dying than age-matched males. Beside hormone differences, which, however, do not appear to be the only factor, there are different potential mechanisms that may explain why women are less prone to severe COVID-19 infections such as ACE/ACE2 ratio and the transmembrane protease serine 2 up regulation in men[18].

Interestingly, our study reveals that the interdependence of different comorbidities may affect outcomes in a different manner than a single comorbidity on its own would. In our cohort of patients, the main diseases that determined outcomes according to the decision tree were HF and AF, vascular diseases, and neurodegenerative diseases. In this regard, it should be noted that CVRF, and even cardiovascular events—which have been previously identified as contributors to a worse outcome[19]—only seem to lead to worse outcomes in clusters 4 and 5. This is in contrast to what was observed in cluster 3, in which the endpoint was not reached. It may be due to the absence of other diseases such as lung disease or obesity in cluster 3 (comorbidities did represent in clusters 4 and 5 which have been related to higher mortality in COVID-19 patients[17]). Clusters 4 and 5 had higher rates of AKI and ALI, probably related to the elevated pro-inflammatory status that was observed on these patients' laboratory findings and that is related to severe COVID-19 which is characterized by a systemic cytokine release syndrome (CRS), increased levels of LDH and CRP, hypoalbuminemia, deepening decrease in lymphocyte counts and immune exhaustion of T cells[20].

HF and AF have also been shown to contribute to poor prognosis[21]. In our cohort, the highest rates of these diseases were in the cluster 5. Nevertheless, a higher rate was also present in the cluster 3, which had a better prognosis. The absence of obesity, OSA, pulmonary obstructive diseases, and malignancies in cluster 3 could be linked to their lower inflammatory response and lower rates of lung injury.

With respect to neurodegenerative diseases, data from previous studies suggest that patients with underlying neurologic impairment are vulnerable to more severe COVID-19[16]. These disorders were present in clusters 3 to 6. However, as mentioned above, only patients in clusters 4 to 6 had a worst prognosis and again, depending on the association among the diseases the patients had, the prognosis changed.

Our study found that the prevalence of certain complications varied depending on the cluster patients were classified into (figure 2). Overall, the rates of complications were higher in clusters 4 to 6 except for the case of thrombotic events, whose highest prevalence was in cluster 2. Cluster 2 (and 6) comprised patients with hematologic malignancies, solid malignant neoplasms, rheumatic diseases, and HIV infection. It is well known that these illnesses are linked to thrombotic events and thus, special care must be taken with treatment strategies regarding these patients. On the contrary, patients

in cluster 4 did not have these disorders. The presence of individuals with HF, obesity, and COPD—which have also been linked to thrombotic complications—may explain the higher rate of these complications. Additionally, these patients had the highest platelet-to-lymphocyte ratios and neutrophil-to-lymphocyte ratios, which have been described as risk factors for venous thrombosis[22] since activation of endothelium, platelets, and leukocytes leads to increased local and systemic generation of thrombin, which in turn leads to settling of fibrin, microangiopathy, and subsequent organ damage[20].

Finally, the role of liver and gastrointestinal diseases in outcomes merits mention. The majority of patients with these diseases were in cluster 5. Previous works have also reported a worse prognosis among patients with liver disease[23]. This poor outcome can be attributed to many factors, including low levels of albumin[24] (indeed, this cluster had the lowest level of albumin), but also due to their altered immune function and that they are more vulnerable to decompensation or development of acute-on-chronic liver failure with bacterial, fungal or viral infection[17].

Despite the potential sex and age confounders, our study is the first to attempt to understand the influence of various clinical entities in the same patient on the prognosis of Sars-Cov-2 infection. Our results show how the interaction of several comorbidities in the same patient can modify the inflammatory response and the profile of complications as well as the mortality, and that it is possible that there are comorbidities more involved than others in the development of such complications and mortality.

This study has several limitations. First, it is an observational retrospective cohort study conducted during a global pandemic, so there may be additional or unmeasured confounding factors. Additionally, as the registry lacked a control group, we were not able to know the prognosis and complications of the clusters generated in patients without COVID19. Second, misclassification errors in comorbidities could have occurred, as classification depended on the researchers' judgment, though findings that linked the different diseases to laboratory results were congruent. In this regard, the information about the clinical severity of each comorbidity was lacking which may influence on the different prognosis of the patients. Similarly, the classification of the diseases into categories may limit the study's interpretation. Finally, the time from hospital admission to ICU admission was not available, so it was not possible to create a complete Kaplan-Meier curve and regression analysis for the primary endpoint.

Conclusions

Our study shows the greater importance of some diseases in establishing a worse prognosis in patients hospitalized for COVID-19, such as those related to cardiovascular disease, obesity or neurodegenerative diseases. But it also shows how the interaction between several different comorbidities and not the presence of just one of them, can affect the results and complications of

hospitalized patients with COVID-19. In light of this finding, it is essential to accurately assess all underlying comorbidities in these patients, but not to correlate them with poor outcome as single entities, but rather to assess the possible prognostic value of groups of various comorbidities.

References

1. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et.al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020;395:507-513. doi: 10.1016/S0140-6736(20)30211-7.
2. Gao HN, Lu HZ, Cao B, Du B, Shang H, Gan JH, et al. Clinical Findings in 111 Cases of Influenza A (H7N9) Virus Infection. N Engl J Med 2013; 368:2277-2285. doi: 10.1056/NEJMoa1305584.
3. Emami A, Javanmardi F, Pirbonyeh N, Akbari A. Prevalence of Underlying Diseases in Hospitalized Patients with COVID-19: a Systematic Review and Meta-Analysis. Arch Acad Emerg Med. 2020;8:e35.
4. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. Int J Infect Dis. 2020;94:91-95. doi: 10.1016/j.ijid.2020.03.017.
5. Casas-Rojo JM, Antón-Santos JM, Millán-Núñez-Cortés J, Lumbreras-Bermejo C, Ramos-Rincón JM, Roy-Vallejo E, et al; en nombre del Grupo SEMI-COVID-19 Network. Características clínicas de los pacientes hospitalizados con COVID-19 en España: resultados del Registro SEMI-COVID-19. Rev Clin Española 2020;8:480-94. doi: 10.1016/j.rce.2020.07.003
6. ARDS Definition Task Force, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, Fan E, et al. Acute respiratory distress syndrome: the Berlin Definition. JAMA 2012;307:2526-33. doi: 10.1001/jama.2012.5669.
7. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al; Authors/Task Force Members; Document Reviewers. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail. 2016;18:891-975. doi: 10.1002/ejhf.592.
8. Taylor FB Jr, Toh CH, Hoots WK, Wada H, Levi M. Towards definition, clinical and laboratory criteria, and a scoring system for disseminated intravascular coagulation - On behalf of the Scientific Subcommittee on Disseminated Intravascular Coagulation (DIC) of

the International Society on Thrombosis and Haemostasis (ISTH). Thromb Haemost 2001;86:1327-1330.

9. Breiman L. Random forests. Machine Learning 2001;45:5–32. https://doi.org/10.1023/A:1010933404324
10. Liaw A, Wiener M . Classification and Regression by randomForest. R News 2002;2:18–22. http://CRAN.R-project.org/doc/Rnews/
11. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources, and a solution. BMC Bioinformatics 2007;8:25. doi: 10.1186/1471-2105-8-25.
12. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. R J. 2016;8:289-317.
13. Fraley C, Raftery A. E. Model-based clustering, discriminant analysis and density estimation. J Am Stat Assoc 2002;97:611-631. doi: 10.1198/016214502760047131
14. Simera I, Moher D, Hoey J, Schulz KF, Altman DG.. A catalogue of reporting guidelines for health research. Eur J Clin Invest. 2010;40:35-53.
15. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW; the Northwell COVID-19 Research Consortium. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA 2020;323:2052-2059. doi: 10.1001/jama.2020.6775.
16. Herman C, Mayer K, Sarwal A. Scoping review of prevalence of neurologic comorbidities in patients hospitalized for COVID-19. Neurology 2020;95:77-84. doi: 10.1212/WNL.0000000000009673.
17. Gao YD, Ding M, Dong X, Zhang JJ, Kursat Azkur A, Azkur D, et al. Risk factors for severe and critically ill COVID-19 patients: A review. Allergy. 2021;76:428-455.
18. Penna C, Mercurio V, Tocchetti CG, Pagliaro P. Sex-related differences in COVID-19 lethality. Br J Pharmacol. 2020;177:4375-4385.
19. 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;395:1054-1062. doi: 10.1016/S0140-6736(20)30566-3.
20. Sokolowska M, Lukasik ZM, Agache I, Akdis CA, Akdis D, Akdis M, et al. Immunology of COVID-19: Mechanisms, clinical outcome, diagnostics, and perspectives-A report of the European Academy of Allergy and Clinical Immunology (EAACI). Allergy. 2020;75:2445-2476.
21. Inciardi RM, Adamo M, Lupi L, Cani DS, Di Pasquale M, Tomasoni D, et al. Characteristics and outcomes of patients hospitalized for COVID-19 and cardiac disease in Northern Italy. Eur Heart J. 2020 May 14;41:1821-1829. doi: 10.1093/eurheartj/ehaa388.

22. Artoni A, Abbattista M, Bucciarelli P, Gianniello F, Scalambrino E, Pappalardo E, et al. Platelet to Lymphocyte Ratio and Neutrophil to Lymphocyte Ratio as Risk Factors for Venous Thrombosis. Clin Appl Thromb Hemost. 2018;24:808-814. doi: 10.1177/1076029617733039.
23. Cabibbo G, Rizzo GEM, Stornello C, Craxì A. SARS-CoV-2 infection in patients with a normal or abnormal liver. J Viral Hepat 2021;28:4-11. doi: 10.1111/jvh.13440.
24. Kumar-M P, Mishra S, Jha DK, Shukla J, Choudhury A, Mohindra R, et al. Coronavirus disease (COVID-19) and the liver: a comprehensive systematic review and meta-analysis. Hepatol Int. 2020;14:711-722. doi: 10.1007/s12072-020-10071-9.

Table 1: Categorization of the diseasaes related to patients' medical history into groups of comorbidities and their numerical equivalences to perform the clusters.

COMORBIDITY	DEFINITION
Neurodegenerative diseases	A range of conditions which primarily affect the neurons in the human brain, including dementia, Parkinson's disease, etc.
• 0	Absent
• 1	Present
Chronic kidney diseases (CKD)	Prolonged functional or structural kidney abnormalities that have been ongoing for more than three weeks
• 0	Absent
• 1	CKD without renal replacement therapy
• 2	CKD with renal replacement therapy
Autoinmune rheumatic diseases	Conditions such as systemic lupus erythematosus in which the connective tissues are most frequently targeted
• 0	Absent
• 1	Present
Mental health conditions	Only including depression, generalized anxiety disorder (GAD) and panic disorder (PD)
• 0	Absent
• 1	Depression
• 2	Either GAD or PD
Obstructive lung diseases	Respiratory diseases characterized by airway obstruction such as chronic bronchitis, diagnosed according to clinical criteria; chronic obstructive pulmonary disease (COPD), diagnosed according to clinical and spirometry criteria; and asthma, diagnosed according to reversible airflow obstruction
• 0	Absent
• 1	Chronic bronchitis/COPD
• 2	Asthma
Human immunodeficiency virus (HIV) infection and acquired immunodeficiency syndrome	-

(AIDS)	
• 0	Absent
• 1	AIDS
• 2	HIV infection without AIDS criteria
Hematologic malignancies	Only including leukemia and lymphomas diagnosed according to clinical and histological criteria
• 0	Absent
• 1	Leukemia (any kind)
• 2	Lymphoma (any kind)
Solid malignant neoplasms	Excluding lymphoma and diagnosed according to histological criteria, with or without metastatic disease
• 0	Absent
• 1	Solid neoplasm without metastases
• 2	Neoplasms with metastatic disease
Cardiovascular risk factors	Essential hypertension, dyslipidemia, and diabetes mellitus (DM) diagnosed according to clinical practice guidelines or whether the patient received treatment for these diseases
• 0	Absent
• 1	Only hypertension
• 2	Only dyslipemia
• 3	Only DM
• 4	Hypertension and dyslipidemia
• 5	Hypertension and DM
• 6	Dyslipidemia and DM
• 7	All factors (Hypertension and dyslipidemia and DM)
Heart failure and atrial fibrillation	Diagnosed according to clinical, biochemical, electrocardiographic, and echocardiographic criteria, as appropriate
• 0	Absent
• 1	Only atrial fibrillation
• 2	Only heart failure
• 3	Both, atrial fibrillation and heart failure
Vascular diseases	Only including coronary artery disease (CAD), peripheral artery disease (PAD), and stroke

• 0	Absent
• 1	Only CAD
• 2	Only PAD
• 3	Only stroke
• 4	CAD and PAD
• 5	CAD and stroke
• 6	PAD and stroke
• 7	All vascular diseases (CAD and PAD and stroke)
Digestive disorders	Only including liver diseases of any etiology and acid peptic disease
• 0	Absent
• 1	Peptic ulcer
• 2	Liver disease
• 3	Both, peptic ulcer and liver disease
Obesity, and obstructive sleep apnea (OSA)	Obesity defined as a body mass index greater than 30 kg/m^2. OSA: diagnosed via polysomnography
• 0	Absent
• 1	Only OSA
• 2	Only obesity
• 3	Obesity and OSA

TABLE 2: Characteristics of the clusters according to their comorbidities.

VARIABLE	TOTAL	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4	CLUSTER 5	CLUSTER 6	*p*
N	16455	8765	798	2651	3557	457	227	
NEURODEGENERATIVE DISEASES	1987 (12·1)	418 (4·8)	16 (2·01)	676 (25·5)	714 (20·1)	125 (27·3)	18 (16·7)	·00
KIDNEY DISEASES								
1-NO RENAL REPLACEMENT THERAPY	829 (5)	209 (2·4)	7 (0·9)	245 (9·2)	111 (3·1)	218 (47·7)	39 (17·8)	·00
2-RENAL REPLACEMENT THERAPY	199 (1·2)	51 (0·6)	2 (0·2)	71 (2·7)	11 (0·3)	51 (11·2)	13 (5·7)	
RHEUMATIC DISEASES	413 (2·5)	15 (0·2)	112 (14)	150 (5·7)	43 (1·2)	45 (9·8)	48 (21%)	·00
MENTAL HEALTH CONDITIONS								
1-DEPRESSION	1055 (6·4)	502 (5·7)	54 (6·8)	147 (5·5)	306 (8·6)	35 (7·7)	11 (4·8)	·00
2-GAD/PD	1274 (7·7)	857 (9·8)	40 (5)	19 (0·7)	286 (8)	56 (12·2)	16 (7)	
OBSTRUCTIVE LUNG DISEASES								
1-CHRONIC BRONCHITIS/ COPD	1422 (8·6)	696 (7·9)	84 (10·5)	87 (3·3)	459 (12·9)	81 (17·7)	15 (6·6)	·00
2-ASTHMA	1122 (7·1)	861 (9·8)	50 (6·3)	7 (0·3)	218 (6·1)	36 (7·9)	0	
HIV/AIDS								
1-AIDS	17 (0·1)	0	11 (1·4)	3 (0·1)	0	0	3 (1·3)	·00
2-HIV INFECTION/ SEROPOSITIV	101 (0.6)	0	68 (8·5)	16 (0·6)	0	2(0·4)	15 (6·6)	

E								
HEMATOLOGIC MALIGNANCIES								
1-LEUKEMIA	188 (1·1)	11 (0·1)	73 (9·1)	25 (0·9)	18 (0·5)	10 (2·2)	51 (22·5)	·00
2-LYMPHOMA	218 (1·3)	2 (0·02)	95 (11·9)	28 (1·1)	12 (0·3)	8 (1·7)	73 (22·2)	
SOLID MALIGNANT NEOPLASMS								
1-NON-METASTATIC DISEASE	996 (6)	337 (3·8)	316 (39·6)	6 (0·23)	284 (7·9)	3 (0·7)	50 (22)	·00
2-METASTATIC DISEASE	335 (2)	55 (0·6)	173 (21·7)	0	91 (2·6)	0	16 (7)	
HEART FAILURE/ATRIAL FIBRILLATION								
1-AF	1247 (7·6)	381 (4·3)	27 (3·4)	356 (13·4)	380 (10·7)	72 (15·7)	31 (13·7)	·00
2-HF	595 (3·6)	190 (2·2)	7 (0·9)	148 (5·6)	180 (5·1)	55 (12·1)	15 (6·6)	
3-AF + HF	569 (3.5)	153 (1.7)	5 (0·6)	139 (5·2)	203 (5·7)	52 (11·4)	17 (7·5)	
DIGESTIVE DISORDERS								
1-PEPTIC ULCER DISEASE	370 (2·2)	228 (2·6)	20 (2·5)	8 (0·3)	33 (0·9)	76 (16·6)	5 (2·2)	·00
2-LIVER DISEASES	572 (3·5)	348 (3·9)	43 (5·4)	9 (0·3)	67 (1·9)	102 (22·3)	3 (1·3)	
3-PEPTIC ULCER + LIVER DISEASES	40 (0·2)	27 (0·3)	1 (0·13)	0	3 (0·1)	9 (2)	0	

OBESITY/OSA								
1-OSA	475 (2·9)	304 (3·5)	29 (3·6)	0	141 (3·9)	1 (0·2)	0	·00
2-OBESITY	2778 (16·9)	1896 (21·6)	119 (14·9)	0	763 (21·4)	0	0	
3-OSA + OBESITY	538 (3·3)	323 (3·7)	27 (3·4)	0	187 (5·6)	1 (0·2)	0	
CARDIOVASCULAR RISK FACTORS								
1-HYPERTENSION	2976 (18·1)	1578 (18)	157 (19)	379 (14·3)	705 (19·8)	101 (22·1)	56 (24·7)	·00
2-DYSLIPIDEMIA	1484 (9)	944 (10·8)	80 (10)	131 (4·9)	289 (8·1)	22 (4·8)	18 (7·9)	
3-DM	393 (2·4)	207 (2·4)	24 (3)	41 (1·5)	97 (2·7)	14 (3·1)	10 (4·4)	
4-HYPERTENSION + DYSLIPIDEMIA	2944 (17·9)	918 (10·5)	114 (14·3)	939 (35·4)	817 (22·9)	109 (23·8)	47 (20·7)	
5-HYPERTENSION + DM	756 (4·6)	229 (2·6)	39 (4·9)	210 (3·9)	217 (6·1)	48 (10·5)	13 (5·7)	
6-DYSLIPIDEMIA + DM	358 (2·2)	121 (1·4)	18 (2·3)	115 (4·3)	90 (2·5)	13 (2·8)	1 (0·4)	
7-HYPERTENSION + DYSLIPIDEMIA + DM	1710 (10·4)	572 (6·5)	45 (5·6)	465 (17·5)	505 (14·2)	96 (21)	27 (11·9)	
VASCULAR DISEASES								
1-CAD	973	368	3 (0·4)	232	304	46	20 (8·8)	·0

	(5·9)	(4·2)		(8·7)	(8·5)	(10·1)		0
2-PAD	458 (2·8)	152 (1·7)	1 (0·13)	122 (4·6)	137 (3·8)	38 (8·3)	8 (3·5)	
3-STROKE	880 (5·3)	259 (2·9)	0	285 (10·7)	273 (7·7)	46 (10·1)	17 (7·5)	
4-CAD + PAD	141 (0·9)	43 (0·5)	0	39 (1·5)	46 (1·3)	10 (2·2)	3 (1·3)	
5-CAD + STROKE	148 (0·9)	35 (0·4)	0	40 (1·5)	58 (1·6)	13 (2·8)	2 (0·9)	
6-PAD + STROKE	119 (0·7)	24 (0·3)	0	42 (1·6)	41 (1·1)	12 (2·6)	0	
7-CAD + PAD + STROKE	45 (0·3)	6 (0·1)	0	13 (0·5)	18 (0·5)	8 (1·7)	0	

Values are expressed as absolute values and percentages. The chi-square test was used to draw comparisons.

AF: Atrial Fibrillation; AIDS: Acquired Immune Deficiency Syndrome; CAD: coronary artery disease; COPD: Chronic Obstructive Pulmonary Disease; DM: Diabetes Mellitus; GAD: Generalized Anxiety Disorders; HF: Heart Failure; HIV: Human Immunodeficiency Virus; OSA: Obstructive Sleep Apnea; PAD: Peripheral Artery Disease; PD: Panic Disorder.

TABLE 3: Clinical and laboratory findings according to the clusters.

VARIABLE	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4	CLUSTER 5	CLUSTER 6	*p*
N	8765	798	2651	3557	457	227	
AGE (Years)	59·8 (26·5)	65·2 (20·9)	73·2 (22·8)	74·9 (23·4)	79·1 (17·7)	77·5 (19·9)	·00
SEX (Female)	3867 (44·1)	359 (44·9)	1231 (46·4)	1296 (36·4)	174 (38·1)	86 (37·9)	·00
CCI	2 (3)	5 (3)	4 (3)	4 (3)	7 (3)	7 (4)	·00
SBP (mmHg)	127 (25)	126 (29)	130 (30)	128 (34)	126 (33)	122 (32)	·00
DBP (mmHg)	75 (16)	73 (16)	73 (17)	71 (19)	70 (20)	67 (15)	·00
P/F ratio	304·8 (98·1)	303·6 (95·6)	303·6 (94·1)	239·2 (114·5)	238·3 (122·7)	246 (106·9)	
Hemoglobin (g/dl)	14·1 (2·1)	13·3 (2·5)	13·5 (2·4)	13·7 (2·6)	12·5 (3·2)	12·4 (3)	·00
WBC (mm3)	4146 (3330)	4184 (4060)	4723 (3370)	7260 (5020)	6914 (5552)	6321 (5860)	·00
NLR	4·14 (4·04)	4·18 (5·09)	4·72 (4·74)	7·26 (8·75)	6·91 (9·16)	6·32 (8·83)	·00
PLR	190·1 (134·9)	209·3 (182·1)	197·2 (160·3)	240 (208·9)	217·3 (203·8)	211·7 (270·7)	·00
Prothrombin time	12·9 (2·1)	12·8 (2)	13 (3·2)	13·4 (3·5)	13·7 (5·7)	13·4 (3·2)	·00
Fibrinogen (g/L)	6·1 (2·2)	6·2 (2·4)	6 (2)	6·5 (2·5)	6·1 (2·5)	6·2 (2·6)	·00
D-dimer (ng/ml)	564·5 (659)	700 (934·5)	710 (972·7)	929 (1300)	1350 (1680)	1039 (1807)	·00
FBG (mg/dl)	108 (29)	109 (31·2)	114 (45)	126 (54)	131 (70)	118 (46·5)	·00
Creatinine (g/dl)	0·84 (0·33)	0·85 (0·35)	0·94 (0·52)	1·03 (0·58)	1·7 (1·5)	1·1 (0·94)	·00
Sodium	138 (5)	137 (5)	137 (5)	137 (6)	138 (7)	137 (6)	·0

(mEq/L)							0
Potassium (mEq/L)	4 (0·6)	4 (0·7)	4·1 (0·7)	4·1 (0·7)	4·4 (0·9)	4·2 (0·8)	·0 0
CRP (mg/dl)	49·8 (91)	47 (98·3)	51·1 (95·5)	107 (42·7)	91 (137·5)	118 (136·5)	·0 0
Serum Ferritin (ng/ml)	575 (901·5)	604 (859·5)	886·5 (711·3)	840 (1211·2)	875 (1381)	778 (1292)	·0 0
Serum Albumin (gr/dl)	3·9 (0·7)	3·8 (0·7)	3·7 (0·7)	3·5 (0·8)	3·4 (0·7)	3·5 (0·7)	·0 0

Variables are expressed as medians (interquartile range). Values were compared using the Kruskal-Wallis test or ANOVA with Welch's correction, depending on the similarity of variances.

CCI: age-adjusted Charlson Comorbidity Index; CRP: C-reactive protein; DBP: Diastolic blood pressure; FBG: Fasting blood glucose; NLR: Neutrophil-to-lymphocyte ratio; P/F ratio: PaO2/FiO2 ratio, the ratio of arterial oxygen partial pressure (PaO2 in mmHg) to fractional inspired oxygen; PLR: Platelet-to-lymphocyte ratio; SBP: Systolic blood pressure; WBC: White blood cells.

Supplementary table 1: Stratification of the outcomes of clusters by groups of age.

VARIABLE	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4	CLUSTER 5	CLUSTER 6	P
AGE < 50	N=2112	N=100	N=130	N=218	N=11	N=7	
AGE (YEARS)	42.2 (10.6)	43.6 (9.3)	43.7 (8.8)	42.9 (10.1)	44.4 (9.1)	41 (8.2)	0.24
SEX (FEMALE)	867 (41)	51 (51)	46 (35.4)	69 (31.6)	2 (18.2)	0	<0.0
CHARLSON COMORBIDITY INDEX	0 (8)	2 (13)	1 (12)	0 (6)	3 (4)	2 (7)	<0.0
HEART FAILURE	9 (0.43)	0	4 (3.1)	9 (4.1)	0	2 (28.6)	<0.0
ACUTE LUNG INJURY	235 (11.1)	15 (15)	16 (12.3)	188 (86.2)	8 (72.7)	6 (85.7)	<0.0
ACUTE KIDNEY INJURY	29 (1.4)	2 (2)	17 (13.1)	39 (17.9)	3 (27.3)	1 (14.3)	<0.0
DISSEMINATED INTRAVASCULAR COAGULATION	2 (0.1)	0	1 (0.8)	6 (2.7)	1 (9.1)	1 (14.3)	<0.0
THROMBOSIS							
• **DEEP VENOUS THROMBOSIS (DVT)**	4 (0.2)	4 (4)	1 (0.77)	6 (2.7)	0	1 (14.3)	
• **PULMONARY EMBOLISM (PE)**	18 (0.8)	0	0	16 (7.3)	0	0	<0.0
• **DVT+PE**	1 (0.05)	0	0	1 (0.46)	0	0	
ADMISSION TO INTENSIVE	0	0	0	197 (90.4)	8 (72.7)	6 (85.7)	<0.0

CARE UNIT							
MORTALITY	0	0	0	67 (30.7)	5 (45.4)	6 (85.7)	<0.0
AGE 50-60	N=1879	N=156	N=278	N=315	N=20	N=16	
AGE (YEARS)	55.4 (5)	56.1 (4.9)	55.7 (4.2)	55.7 (4.7)	57.9 (3.9)	56.9 (4.1)	<0.0
SEX (FEMALE)	769 (40.9)	82 (52.6)	89 (32)	94 (29.8)	5 (25)	6 (37.5)	<0.0
CHARLSON COMORBIDITY INDEX	1 (1)	3 (4)	2 (1)	1 (1)	3.5 (1.5)	3 (1)	<0.0
HEART FAILURE	12 (0.64)	2 (1.3)	6 (2.2)	26 (8.2)	2 (10)	0	<0.0
ACUTE LUNG INJURY	307 (16.3)	20 (12.8)	47 (16.9)	286 (9.8)	19 (95)	15 (93.7)	<0.0
ACUTE KIDNEY INJURY	47 (2.5)	7 (4.5)	21 (7.5)	75 (23.8)	4 (20)	4 (25)	<0.0
DISSEMINATED INTRAVASCULAR COAGULATION	7 (0.4)	0	1 (0.4)	12 (13.8)	2 (10)	0	<0.0
THROMBOSIS							
• **DEEP VENOUS THROMBOSIS (DVT)**	4 (0.2)	1 (0.6)	1 (0.4)	6 (1.9)	0	1 (6.2)	
• **PULMONARY EMBOLISM (PE)**	16 (0.8)	4 (2.6)	2 (0.7)	13 (4.1)	0	2 (12.5)	<0.0
• **DVT+PE**	4 (0.21)	0	0	1 (0.3)	0	0	
ADMISSION TO INTENSIVE CARE UNIT	0	0	0	271 (86)	11 (55)	11 (68.7)	<0.0
MORTALITY	0	0	0	147	11 (55)	11 (68.7)	<0.0

				(46.7)			
AGE 60-70	N=1916	N=202	N=484	N=581	N=42	N=27	
AGE (YEARS)	64.7 (5.2)	65.2 (5.4)	65.7 (5)	65.4 (4.8)	65.2 (4.1)	65.7 (4)	0.02
SEX (FEMALE)	861 (44.9)	100 (49.5)	175 (36.2)	199 (34.2)	8 (19)	6 (22.2)	<0.0
CHARLSON COMORBIDITY INDEX	2 (1)	4 (3)	3 (1)	3 (2)	5 (4)	5 (3)	<0.0
HEART FAILURE	35 (1.8)	5 (2.5)	8 (1.6)	65 (11.2)	6 (14.3)	3 (11.1)	<0.0
ACUTE LUNG INJURY	376 (19.2)	49 (24.3)	101 (20.9)	535 (92.1)	35 (83.3)	24 (88.9)	<0.0
ACUTE KIDNEY INJURY	113 (5.9)	10 (4.9)	35 (7.2)	197 (33.9)	22 (52.4)	7 (25.9)	<0.0 0
DISSEMINATED INTRAVASCULAR COAGULATION	5 (0.26)	3 (1.5)	2 (0.4)	29 (4.9)	4 (9.5)	4 (14.8)	<0.0
THROMBOSIS							
• **DEEP VENOUS THROMBOSIS (DVT)**	13 (0.7)	4 (1.9)	2 (0.4)	7 (1.2)	0	0	
• **PULMONARY EMBOLISM (PE)**	24 (1.2)	9 (4.5)	6 (1.2)	22 (3.8)	3 (7.1)	0	<0.0
• **DVT+PE**	3 (0.2)	0	1 (0.2)	5 (0.9)	0	0	
ADMISSION TO INTENSIVE CARE UNIT	0	0	0	448 (77.1)	22 (52.4)	15 (55.6)	<0.0
MORTALITY	110 (5.7)	8 (3.9)	40 (8.3)	349 (60.1)	33 (78.6)	19 (70.4)	0.00 1

AGE 70-80	N=1741	N=229	N=793	N=1021	N=135	N=65	
AGE (YEARS)	74.7 (4.9)	74.2 (4.1)	74.8 (4.9)	75.5 (4.8)	75.7 (5.1)	76.1 (4.4)	<0.0
SEX (FEMALE)	807 (46.3)	79 (34.5)	357 (45.2)	292 (28.6)	43 (31.8)	29 (44.6)	<0.0
CHARLSON COMORBIDITY INDEX	4 (2)	6 (3)	4 (2)	4 (3)	6.5 (3)	6 (2)	<0.0
HEART FAILURE	63 (3.6)	7 (3.1)	27 (3.4)	115 (11.3)	24 (17.8)	9 (13.8)	<0.0
ACUTE LUNG INJURY	353 (20.3)	52 (22.7)	157 (19.8)	864 (84.6)	112 (82.9)	55 (84.6)	<0.0
ACUTE KIDNEY INJURY	152 (8.7)	22 (9.6)	90 (11.3)	360 (35.3)	61 (45.2)	17 (26.1)	<0.0
DISSEMINATED INTRAVASCULAR COAGULATION	8 (0.5)	2 (0.9)	2 (0.2)	32 (3.1)	6 (4.4)	2 (3.1)	<0.0
THROMBOSIS							
• **DEEP VENOUS THROMBOSIS (DVT)**	9 (0.5)	2 (0.9)	3 (0.4)	13 (1.3)	0	1 (1.5)	
• **PULMONARY EMBOLISM (PE)**	23 (1.3)	8 (3.5)	11 (1.4)	27 (2.6)	3 (2.2)	2 (3.1)	<0.05
• **DVT+PE**	3 (0.2)	0	1 (0.1)	5 (0.5)	0	0	
ADMISSION TO INTENSIVE CARE UNIT	0	0	0	386 (37.8)	32 (23.7)	16 (24.6)	<0.0
MORTALITY	0	0	0	860 (84.2)	125 (92.6)	50 (90.8)	<0.0
AGE > 80	N=1117	N=111	N=966	N=1422	N=249	N=112	

AGE (YEARS)	85.5 (6.2)	85.5 (6.3)	86.6 (6.4)	86.3 (6.5)	86.7 (6.6)	86.1 (5.2)	<0.0
SEX (FEMALE)	563 (50.4)	47 (42.3)	564 (58.4)	642 (45.1)	116 (46.6)	45 (40.2)	<0.0
CHARLSON COMORBIDITY INDEX	5 (3)	7 (4)	6 (2)	6 (2)	8 (2)	7 (2)	<0.0
HEART FAILURE	98 (8.8)	7 (6.3)	101 (10.5)	226 (15.9)	52 (20.9)	23 (20.5)	<0.0
ACUTE LUNG INJURY	237 (21.1)	14 (12.6)	146 (15.1)	1033 (79.9)	174 (69.9)	87 (77.7)	<0.0
ACUTE KIDNEY INJURY	183 (16.4)	13 (11.7)	165 (17.1)	435 (30.6)	107 (42.9)	46 (41.2)	<0.0
DISSEMINATED INTRAVASCULAR COAGULATION	3 (0.3)	1 (0.9)	6 (0.6)	26 (1.8)	7 (2.8)	5 (4.5)	<0.0
THROMBOSIS							
• **DEEP VENOUS THROMBOSIS (DVT)**	10 (0.9)	3 (2.7)	5 (0.5)	2 (0.1)	2 (0.8)	0	
• **PULMONARY EMBOLISM (PE)**	16 (1.4)	4 (3.6)	12 (1.2)	13 (0.9)	1 (0.4)	1 (0.9)	<0.0
• **DVT+PE**	1 (0.1)	0	0	1 (0.1)	0	1 (0.9)	
ADMISSION TO INTENSIVE CARE UNIT	0	0	0	46 (3.2)	3 (1.2)	1 (0.9)	
MORTALITY	0	0	0	1048 (99)	249 (100)	112 (100)	<0.0

Figure 1: Variable importance according to the random forest algorithm.

Figure 2: Conformation of clusters according the comorbidities included.

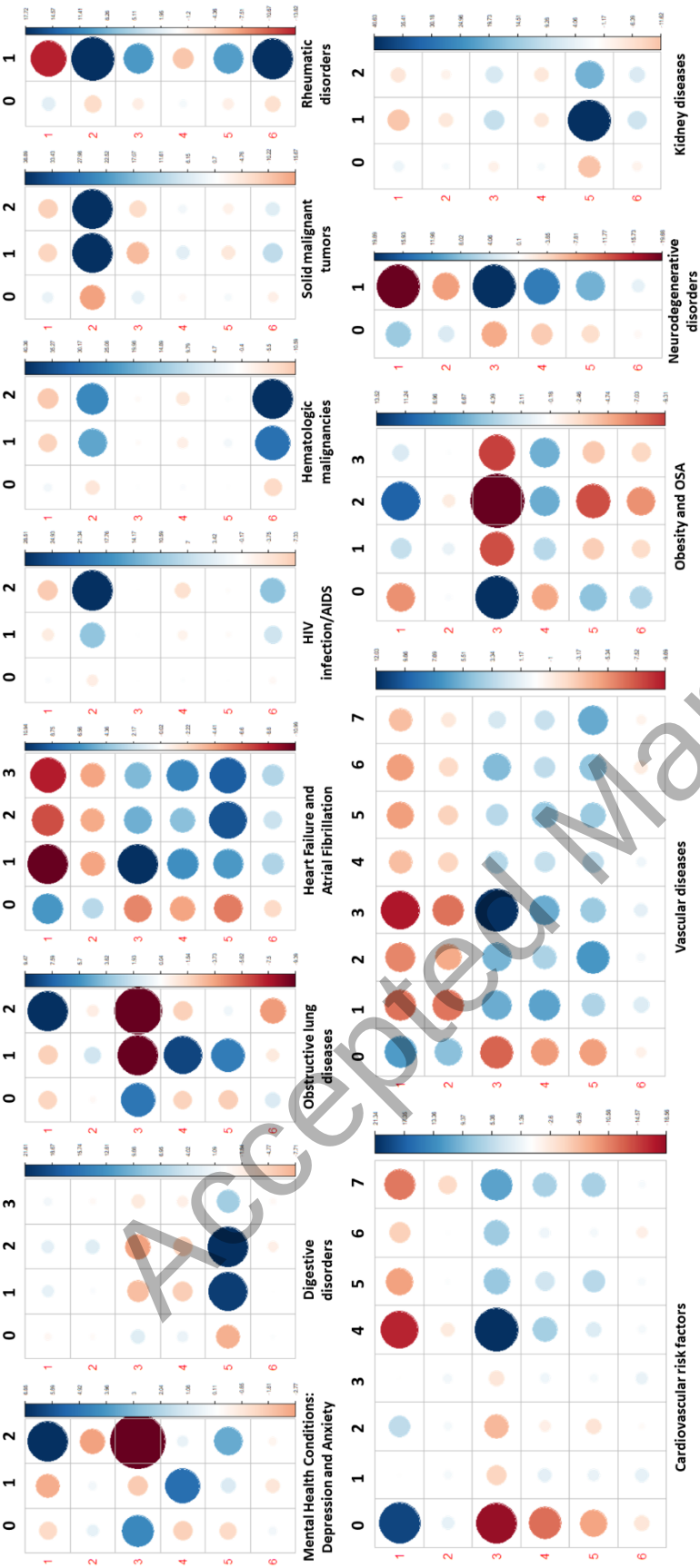
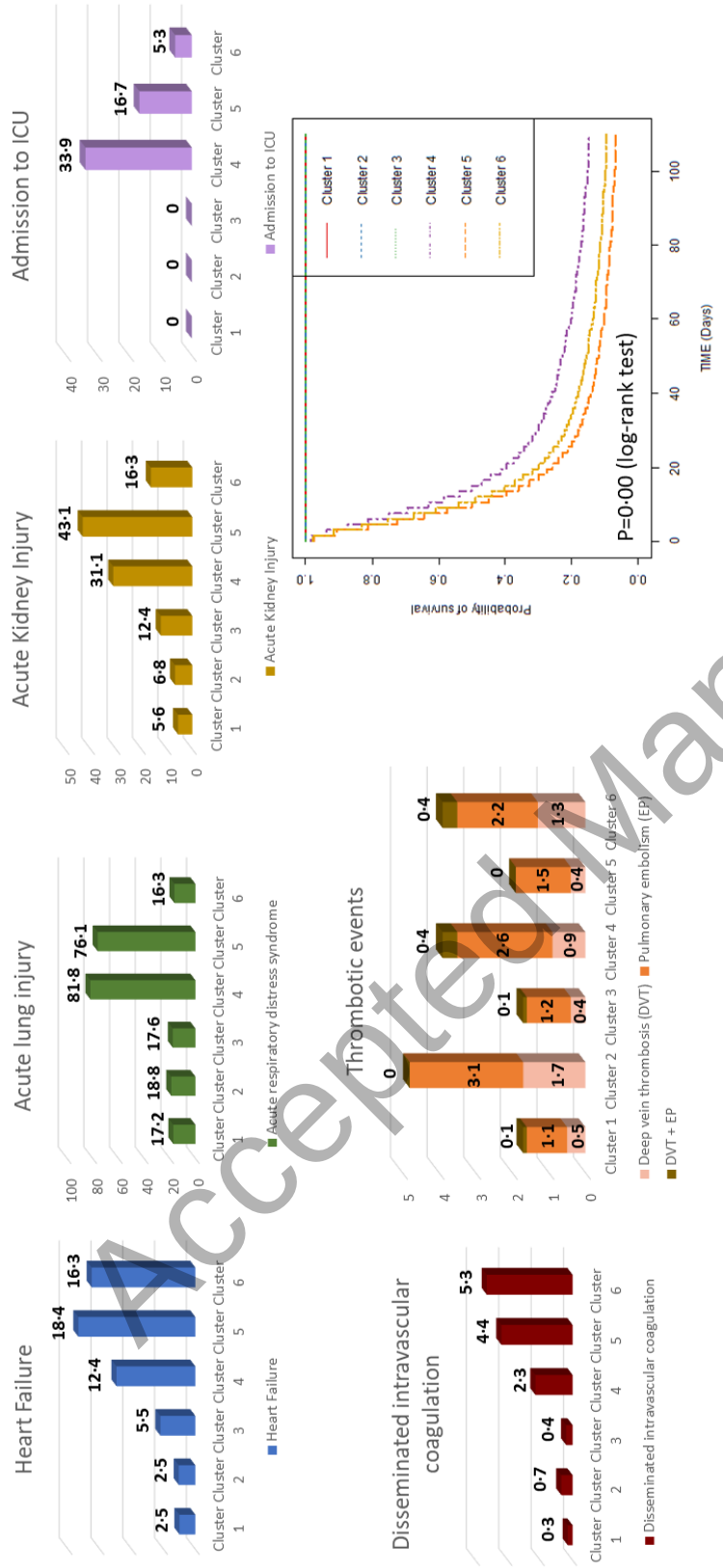


Figure 3: Complications according to the clusters.



Supplementary figure 1

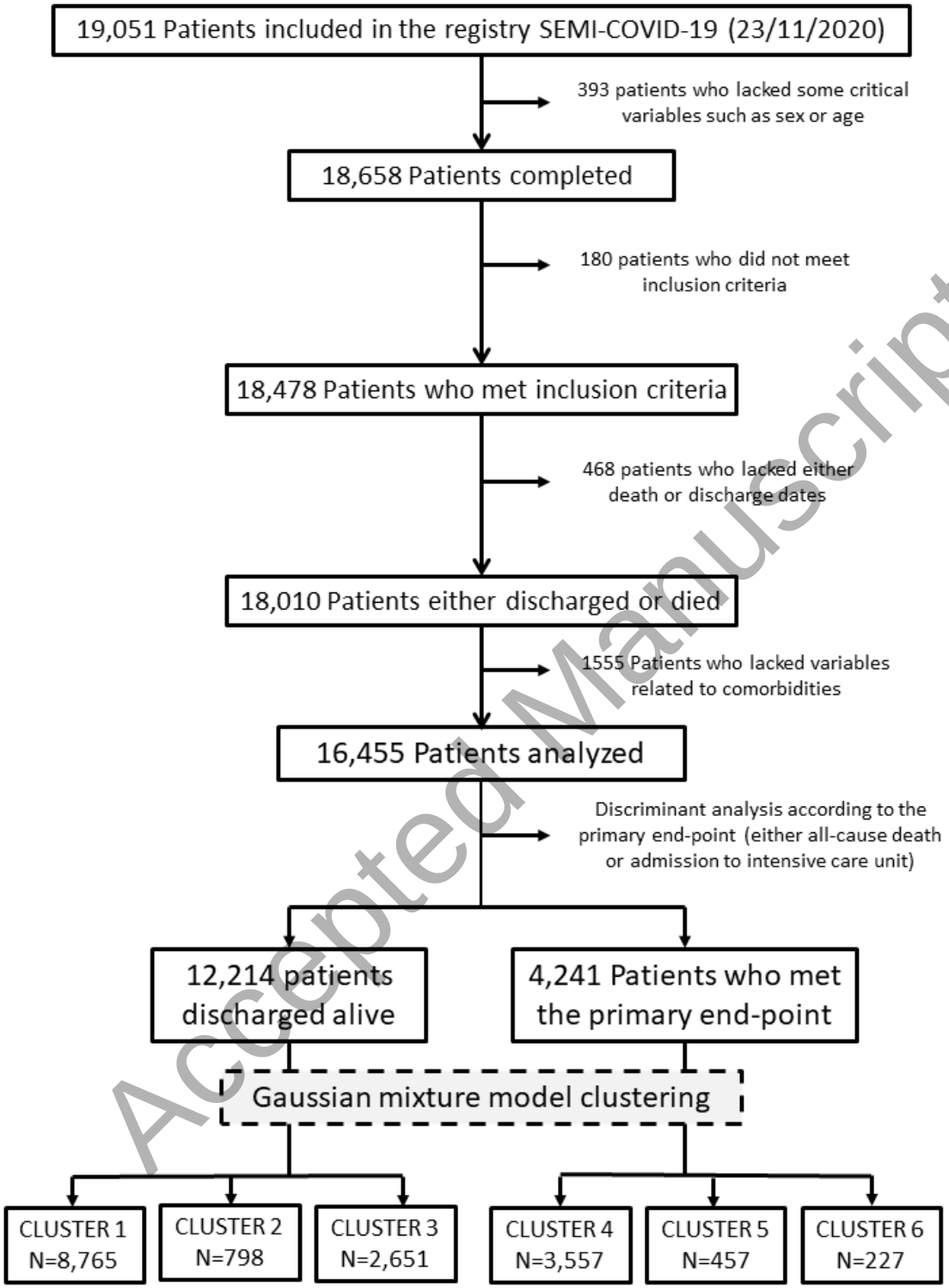


Figure captions

Figure 1: Variable importance according to the random forest algorithm. The unscaled permutation-based importance measure is based on how much the accuracy decreases when the variable is excluded. Mean decrease of Gini impurity based on the decrease of Gini impurity (Gini Impurity is a measurement of the likelihood of an improper classification of a new instance of a random variable).

CVRF: cardiovascular risk factors; CKD: Chronic kidney disease; HF/AF: Heart failure/Atrial Fibrillation; HIV/AIDS: Human Immunodeficiency Virus infection/Acquired Immune Deficiency Syndrome; OSA: obstructive sleep apnea.

Figure 2: Composition of clusters according the comorbidities included. The circles correspond with the residuals on the chi-square test. Positive residuals are indicated in blue. Positive values in cells specify an attraction (positive association) between the corresponding row and column variables. Negative residuals are indicated in red. This implies a repulsion (negative association) between the corresponding row and column variables. Row number corresponds to number of clusters. Numbers in columns correspond to comorbidities as described in methods and in table 1.

HIV/AIDS: Human Immunodeficiency Virus infection/Acquired Immune Deficiency Syndrome; OSA: obstructive sleep apnea.

Figure 3: Complications according to the clusters. The bars show the percentage of patients that experienced the complication. All differences were significant ($p = .00$) as per the chi-square test. The Kaplan-Meier curve for all-cause mortality appears in the lower right corner of the figure.

Supplementary Figure 1: Patient inclusion flowchart.