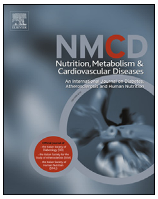
Nutrition, Metabolism & Cardiovascular Diseases (2020) **30**, 1899–1913

Available online at www.sciencedirect.com

Nutrition, Metabolism & Cardiovascular Diseases

journal homepage: www.elsevier.com/locate/nmcd



Common cardiovascular risk factors and in-hospital mortality in 3,894 patients with COVID-19: survival analysis and machine learning-based findings from the multicentre Italian CORIST Study

Augusto Di Castelnuovo [a], Marialaura Bonaccio [b], Simona Costanzo [b], Alessandro Gialluisi [b], Andrea Antinori [c], Nausicaa Berselli [d], Lorenzo Blandi [e], Raffaele Bruno [f,g], Roberto Cauda [h,i], Giovanni Guaraldi [j], Ilaria My [k], Lorenzo Menicanti [e], Giustino Parruti [l], Giuseppe Patti [m], Stefano Perlini [n,o], Francesca Santilli [p], Carlo Signorelli [q], Giulio G. Stefanini [k], Alessandra Vergori [r], Amina Abdeddaim [s], Walter Ageno [t], Antonella Agodi [u], Piergiuseppe Agostoni [v,w], Luca Aiello [x], Samir Al Moghazi [y], Filippo Aucella [z], Greta Barbieri [aa], Alessandro Bartoloni [ab], Carolina Bologna [ac], Paolo Bonfanti [ad,ae], Serena Brancati [af], Francesco Cacciatore [ag], Lucia Caiano [t], Francesco Cannata [k], Laura Carrozzi [ah], Antonio Cascio [ai], Antonella Cingolani [h,i], Francesco Cipollone [p], Claudia Colomba [ai], Annalisa Crisetti [z], Francesca Crosta [l], Gian B. Danzi [aj], Damiano D'Ardes [p], Katleen de Gaetano Donati [h], Francesco Di Gennaro [ak], Gisella Di Palma [al], Giuseppe Di Tano [aj], Massimo Fantoni [h,i], Tommaso Filippini [d], Paola Fioretto [am], Francesco M. Fusco [an], Ivan Gentile [ao], Leonardo Grisafi [m], Gabriella Guarnieri [ap], Francesco Landi [x], Giovanni Larizza [aq], Armando Leone [ar], Gloria Maccagni [aj], Sandro Maccarella [as], Massimo Mapelli [v,w], Riccardo Maragna [w], Rossella Marcucci [ab], Giulio Maresca [v,al], Claudia Marotta [ak], Lorenzo Marra [ar], Franco Mastroianni [aq], Alessandro Mengozzi [aa], Francesco Menichetti [aa], Jovana Milic [j], Rita Murri [h,i], Arturo Montineri [at], Roberta Mussinelli [o], Cristina Mussini [j], Maria Musso [au], Anna Odone [q], Marco Olivieri [av], Emanuela Pasi [aw], Francesco Petri [ad], Biagio Pinchera [ao], Carlo A. Pivato [k], Roberto Pizzi [t], Venerino Poletti [ax], Francesca Raffaelli [h], Claudia Ravaglia [ax], Giulia Righetti [aq], Andrea Rognoni [ay], Marco Rossato [am], Marianna Rossi [ad], Anna Sabena [n], Francesco Salinaro [n], Vincenzo Sangiovanni [an], Carlo Sanrocco [l], Antonio Scarafino [aq], Laura Scorzolini [az], Raffaella Sgariglia [as], Paola G. Simeone [l], Enrico Spinoni [m], Carlo Torti [ba], Enrico M. Trecarichi [ba], Francesca Vezzani [p], Giovanni Veronesi [t], Roberto Vettor [am], Andrea Vianello [ap], Marco Vinceti [d,bb], Raffaele De Caterina [ah], Licia Iacoviello [b,t,*], The COvid-19 RISk and Treatments (CORIST) collaboration

[a] *Mediterranea Cardiocentro, Napoli, Italy*
[b] *Department of Epidemiology and Prevention, IRCCS Neuromed, Pozzilli, IS, Italy*
[c] *UOC Immunodeficienze Virali, National Institute for Infectious Diseases "L. Spallanzani", IRCCS, Rome, Italy*
[d] *Section of Public Health, Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy*
[e] *IRCCS Policlinico San Donato, San Donato Milanese, Italy*
[f] *Division of Infectious Diseases I, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy*
[g] *Department of Clinical, Surgical, Diagnostic, and Paediatric Sciences, University of Pavia, Pavia, Italy*
[h] *Fondazione Policlinico Universitario A. Gemelli IRCCS, Roma, Italy*
[i] *Università Cattolica del Sacro Cuore- Dipartimento di Sicurezza e Bioetica Sede di Roma, Italy*
[j] *Infectious Disease Unit, Department of Surgical, Medical, Dental and Morphological Sciences, University of Modena and Reggio Emilia, Modena, Italy*

* Corresponding author. Department of Epidemiology and Prevention, IRCCS Neuromed, Via dell'Elettronica, 86077, Pozzilli, IS, Italy.
E-mail address: licia.iacoviello@moli-sani.org (L. Iacoviello).

https://doi.org/10.1016/j.numecd.2020.07.031

[k] Humanitas Clinical and Research Hospital IRCCS, Rozzano-Milano, Italy
[l] Department of Infectious Disease, Azienda Sanitaria Locale (AUSL) di Pescara, Pescara, Italy
[m] University of Eastern Piedmont, Maggiore della Carità Hospital, Novara, Italy
[n] Emergency Department, IRCCS Policlinico San Matteo Foundation, Pavia, Italy
[o] Department of Internal Medicine, University of Pavia, Pavia, Italy
[p] Department of Medicine and Aging, Clinica Medica, "SS. Annunziata" Hospital and University of Chieti, Chieti, Italy
[q] School of Medicine, Vita-Salute San Raffaele University, Milano, Italy
[r] HIV/AIDS Department, National Institute for Infectious Diseases "Lazzaro Spallanzani"-IRCCS, Roma, Italy
[s] UOC Malattie Infettive-Epatologia, National Institute for Infectious Diseases L. Spallanzani, IRCCS, Rome, Italy
[t] Department of Medicine and Surgery, University of Insubria, Varese, Italy
[u] Department of Medical and Surgical Sciences and Advanced Technologies "G.F. Ingrassia", University of Catania, AOU Policlinico "G. Rodolico - San Marco", Catania, Italy
[v] Centro Cardiologico Monzino IRCCS, Milano, Italy
[w] Department of Clinical Sciences and Community Health, Cardiovascular Section, University of Milano, Milano, Italy
[x] UOC Anestesia e Rianimazione. Dipartimento di Chirurgia Generale Ospedale Morgagni-Pierantoni Forlì, Italy
[y] UOC Infezioni Sistemiche dell'Immunodepresso, National Institute for Infectious Diseases L. Spallanzani, IRCCS, Rome, Italy
[z] Fondazione I.R.C.C.S "Casa Sollievo della Sofferenza", San Giovanni Rotondo, Foggia, Italy
[aa] Department of Clinical and Experimental Medicine, Azienda Ospedaliero-Universitaria Pisana, and University of Pisa, Pisa, Italy
[ab] Department of Experimental and Clinical Medicine, University of Florence, Firenze, Italy
[ac] Ospedale del Mare, ASL NA1, Naples, Italy
[ad] UOC Malattie Infettive, Ospedale San Gerardo, ASST Monza, Monza, Italy
[ae] School of Medicine and Surgery, University of Milano-Bicocca, Milano, Italy
[af] Department of General Surgery and Medical-Surgical Specialties, University of Catania, Catania, Italy
[ag] Department of Translational Medical Sciences. University of Naples, Federico II, Naples, Italy
[ah] Cardiovascular and Thoracic Department, Azienda Ospedaliero-Universitaria Pisana, and University of Pisa, Pisa, Italy
[ai] Infectious and Tropical Diseases Unit- Department of Health Promotion, Mother and Child Care, Internal Medicine and Medical Specialties (PROMISE) - University of Palermo, Palermo, Italy
[aj] Department of Cardiology, Ospedale di Cremona, Cremona, Italy
[ak] Medical Direction, IRCCS Neuromed, Pozzilli, IS, Italy
[al] UOC Medicina - PO S. Maria di Loreto Nuovo -ASL Napoli 1 Centro, Napoli, Italy
[am] Clinica Medica 3, Department of Medicine - DIMED, University hospital of Padova, Padova, Italy
[an] UOC Infezioni Sistemiche e dell'Immunodepresso, Azienda Ospedaliera dei Colli, Ospedale Cotugno, Napoli, Italy
[ao] Department of Clinical Medicine and Surgery, University of Naples "Federico II". Napoli, Italy
[ap] Respiratory Pathophysiology Division, Department of Cardiologic, Thoracic and Vascular Sciences, University of Padova, Padova, Italy
[aq] COVID-19 Unit. EE Ospedale Regionale F. Miulli, Acquaviva delle Fonti, BA, Italy
[ar] UOC di Pneumologia, P.O. San Giuseppe Moscati, Taranto, Italy
[as] ASST Milano Nord - Ospedale Edoardo Bassini, Cinisello Balsamo, Italy
[at] U.O. C. Malattie Infettive e Tropicali, P.O. "San Marco", AOU Policlinico "G. Rodolico - San Marco", Catania, Italy
[au] UOC Malattie Infettive-Apparato Respiratorio, National Institute for Infectious Diseases "L. Spallanzani", IRCCS, Rome, Italy
[av] Computer Service, University of Molise, Campobasso, Italy
[aw] Medicina Interna. Ospedale di Ravenna. AUSL della Romagna, Ravenna, Italy
[ax] UOC Pneumologia, Dipartimento di Malattie Apparato Respiratorio e Torace, Ospedale Morgagni-Pierantoni Forlì, Forlì, Italy
[ay] Coronary Care Unit and Catheterization Laboratory, A.O.U. Maggiore della Carità, Novara, Italy
[az] UOC Malattie Infettive ad Alta Intensità di Cura, National Institute for Infectious Diseases "L. Spallanzani", IRCCS, Rome, Italy
[ba] Infectious and Tropical Diseases Unit, Department of Medical and Surgical Sciences "Magna Graecia" University, Catanzaro, Italy
[bb] Department of Epidemiology, Boston University School of Public Health, Boston, USA

Received 14 July 2020; accepted 20 July 2020
Handling Editor: A. Siani
Available online 31 July 2020

KEYWORDS
COVID-19;
In-hospital mortality;
Risk factors;
Epidemiology

Abstract *Background and aims:* There is poor knowledge on characteristics, comorbidities and laboratory measures associated with risk for adverse outcomes and in-hospital mortality in European Countries. We aimed at identifying baseline characteristics predisposing COVID-19 patients to in-hospital death.

Methods and results: Retrospective observational study on 3894 patients with SARS-CoV-2 infection hospitalized from February 19th to May 23rd, 2020 and recruited in 30 clinical centres distributed throughout Italy. Machine learning (random forest)-based and Cox survival analysis. 61.7% of participants were men (median age 67 years), followed up for a median of 13 days. In-hospital mortality exhibited a geographical gradient, Northern Italian regions featuring more than twofold higher death rates as compared to Central/Southern areas (15.6% vs 6.4%, respectively). Machine learning analysis revealed that the most important features in death classification were impaired renal function, elevated C reactive protein and advanced age. These findings were confirmed by multivariable Cox survival analysis (hazard ratio (HR): 8.2; 95% confidence interval (CI) 4.6−14.7 for age ≥85 vs 18−44 y); HR = 4.7; 2.9−7.7 for estimated glomerular filtration rate levels <15 vs ≥ 90 mL/min/1.73 m^2; HR = 2.3; 1.5−3.6 for C-reactive protein levels ≥10 vs ≤ 3 mg/L). No relation was found with obesity, tobacco use, cardiovascular disease and related-comorbidities. The associations between these variables and mortality were substantially homogenous across all sub-groups analyses.

Conclusions: Impaired renal function, elevated C-reactive protein and advanced age were major predictors of in-hospital death in a large cohort of unselected patients with COVID-19, admitted to 30 different clinical centres all over Italy.

Introduction

As of July 10, 2020, there have been over 12 million of confirmed cases of COVID-19, with 549,247 deaths worldwide [1]. Robust knowledge of correlates and possible predictors of death among patients diagnosed with COVID-19 is crucial to target patients at highest risk through early and more intensive interventions. Since the COVID-19 pandemic outbreak, several studies have evaluated different factors that may predispose individuals to a higher risk of death from COVID-19, but evidence on this is still limited, especially from European countries. Recent meta-analyses [2,3] indicated comorbidities (hypertension, coronary heart disease and diabetes) as major predictors of higher mortality. Moreover, some laboratory parameters such as elevated levels of C-reactive protein (CRP), cardiac troponins and interleukin-6 resulted associated with a higher risk of death. However, those meta-analyses mainly relied on data from China and the United States. Subsequent studies have confirmed a role for age, male gender and the presence of comorbidities as major risk factors for mortality [4,5]; other investigators have also found higher death risk associated with smoking [6] and obesity [7–9].

Since the beginning of the pandemic, Italy was one of the most heavily hit countries, with 34,026 deaths recorded as of July 10, 2020, the largest part occurring in Northern regions [10]. Data from small studies conducted in Lombardy during the first wave (February–March 2020) of the Italian pandemic, showed high in-hospital mortality and higher rate of complications in cardiac patients [11], while older age, obesity and more advanced illness were major risk factors for 30-day mortality [9]. Data from 500 patients hospitalized in Milan, confirmed cardiovascular comorbidities as major predictors of death along with the chest X-ray quantitative radiographic assessment of lung oedema (RALE) score at admission [12]. No evidence to date is available from the Central/Southern regions of Italy, which were partially spared by the COVID-19-pandemic, with less than 20 percent of the total cases recorded nationwide [13]. Only recently, a larger study including also COVID-19 patients from Central-Southern regions of Italy provided insights into factors predisposing to in hospital death, however without a separate North-to-South analysis [14].

Low total number of cases was also documented in other countries bordering the Mediterranean Sea, with about 3000 confirmed cases and less than 200 deaths [15]. Given such broad geographic variability in the pandemic outbreak, risk factors for mortality might vary, in type and impact, across different geographical areas. Finally, most studies have mainly analysed data on patients hospitalized during the first wave of the pandemic, at the earliest stage of the outbreak, February–March 2020 [6,9,11,12,16,17].

To take into account potential non-linear relations of and interaction effects among clinical and sociodemographic risk factors for severity of COVID-19, other studies attempted to estimate mortality risk in COVID-19 patients using supervised machine learning algorithms, as recently reviewed [18]. Although these models were not always precisely described and thoroughly investigated and often only exploited limited sample sizes ($N < 1000$), these algorithms suggested a prominent effect of age, sex, comorbidities, like hypertension, cardiovascular and chronic respiratory disease, cancer and circulating biomarkers (CRP, lymphocyte count and lactate dehydrogenase) [19–25]. Moreover, most of these studies were based on Chinese patients and none of them included an Italian population of COVID-19 patients, which may be important to investigate whether there are differences in the prognostic values of the variables identified so far, based on the ethnicity of cases analysed.

Our report, therefore, aims at extending knowledge on factors predisposing to higher in-hospital death risk, based on data from the CORIST Collaboration [26], an observational multicentre study including 3894 patients with laboratory-confirmed SARS-CoV-2 infection, hospitalized in 30 Italian clinical centres, 41.5% of whom resident in Central/Southern Italian regions. This was accomplished through a composite approach involving the development of a machine learning algorithm to determine the predictive power of the features available and their relative importance in influencing mortality risk, followed by a classical statistical approach to analyse the directions of effects and the relation with time-to-events in a longitudinal setting. The CORIST cohort allowed to identify baseline factors associated with a higher risk of in-hospital death in COVID-19 patients and their efficacy as prognostic factors in a single machine learning algorithm. Moreover, the data analyzed here allowed to assess whether there was a geographic gradient in Italy in the association between predictors and in-hospital death risk and to evaluate whether there was a time-dependent pandemic wave-related difference in the type and/or strength of the associations of risk factors with in-hospital death.

Methods

Setting

This national retrospective observational study was conceived, coordinated and analysed within the CORIST Collaboration Project (ClinicalTrials.gov ID: NCT04318418). Initially focussed on the inhibitors of renin-angiotensin system, the CORIST Collaboration is a set of multicentre observational investigations launched in March 2020 and aimed at testing the association of risk factors and/or therapies with disease severity and mortality in COVID-19 hospitalised patients [26]. The study was approved by the institutional Ethics Board of the Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Neuromed, Pozzilli, and of all recruiting centres, including the IRCCS “Lazzaro Spallanzani”, Rome, which was the Coordinating Centre for clinical studies on COVID-19 in Italy. Data for the present analyses were provided by 30 hospitals distributed throughout Italy. Each hospital provided data from hospitalised adult ($\geq$18 years of age) patients who all had a positive test result for the SARS-CoV-2 virus at any time during their hospitalisation from February 19th to May 23rd, 2020. The follow-up continued through May 29th, 2020.

Data sources

We developed a cohort comprising 3971 patients with laboratory-confirmed SARS-CoV-2 infection in an inpatient setting. The SARS-CoV-2 status was declared on the basis of laboratory results (polymerase chain reaction on a nasopharyngeal swab) from each participating hospital. Clinical data were abstracted at one-time point from electronic medical records or charts, and collected using either a centrally-designed electronic worksheet or a centralized web-based database. Collected data included patients’ demographics, laboratory test results, medication administration, historical and current medication lists, historical and current diagnoses, and clinical notes. In addition, specific information on the most severe manifestations of COVID-19 that occurred during hospitalisation was retrospectively captured. Maximum clinical severity observed was classified as: either light-mild pneumonia; or severe pneumonia; or acute respiratory distress syndrome (ARDS) [27]. Specifically, we obtained the following information for each patient: hospital; date of admission and date of discharge or death; age; gender; the first recorded inpatient laboratory tests at entry (creatinine, CRP); past and current diagnoses of chronic degenerative disease or risk factors (myocardial infarction, heart failure, diabetes, hypertension, chronic pulmonary disease and cancer), and in-hospital drug therapies for COVID-19. Chronic kidney disease was classified as: stage 1: normal or increased glomerular filtration rate (eGFR) ($\geq$90 mL/min/1.73 m^2); stage 2: kidney damage with mild reduction in eGFR (60–89 mL/min/1.73 m^2); stage 3a: moderate reduction in eGFR (45–59 mL/min/1.73 m^2); stage 3 b: moderate reduction in eGFR (30–44 mL/min/1.73 m^2); stage 4: severe reduction in eGFR (15–29 mL/min/1.73 m^2); stage 5: kidney failure (eGFR <15 mL/min/1.73 m^2 or dialysis). eGFR was calculated by the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation. CRP was classified as $\leq$3, 3–10 and $\geq$ 10 mg/L.

The study index date was defined as the date of hospital admission. Index dates ranged from February 19th, 2020 to May 23rd, 2020. The study end point was the time from study index to death. The number of patients who either died or had been discharged alive, or were still admitted to hospital as of May 29th, 2020, were recorded, and hospital length of stay was determined. Patients alive had their data censored on the date of discharge or as the date of the respective clinical data collection.

Of the initial cohort of 3971 patients, 77 patients were excluded from the present analysis because of one or more missing data at baseline or during follow-up, including time to event (n = 59), outcome (death/alive, n = 8), COVID-19 severity (n = 4), age (n = 4 with missing data and n = 2 with age<18 years) or gender (n = 2).

At the end, the analysed cohort consisted of N = 3894 patients. Distribution of missing values for covariates is shown in Table 1.

Machine learning analysis

To take into account potential non-linear relations of and interaction effects among the investigated variables with the risk of death, we performed an exploratory machine learning analysis to compute the predictive power of a potential classification algorithm for the risk of death and to establish variable importance in more complex mortality prediction models. To this purpose, we trained a Random Forest (RF) algorithm in R [28], using age, gender, obesity and smoking status, chronic comorbidities (history of myocardial infarction, hypertension, diabetes, lung disease, heart failure and cancer), CRP and eGFR as input features. Random forest is an ensemble of decision trees which is often used in classification tasks, and represents one of the most used machine learning algorithms applied to risk prediction in COVID-19 patients [18]. Missing data were imputed through a k-nearest neighbour algorithm (*kNN()* function) of the *VIM* package [29], while continuous features (CRP, eGFR and age) underwent min–max normalization before analysis, with CRP transformed on the natural logarithm scale to attain normality. The resulting dataset (N = 3894) was divided in a random training and a test set (2725 and 1169 patients, respectively, 70:30 ratio). We then performed hyperparameter tuning through the *train()* function of the *caret* package [30], in a 10-fold cross validation setting, to optimize the algorithm over two varying parameters: the number of variables randomly sampled as candidate predictors at each node split in the decision tree (*mtry*, varying between 1 and 10), and the number of trees to grow in the random forest (*ntree* alternative values: 100, 500, 1000). Finally, we trained the optimized model (mtry = 3, ntree = 1000) within the training set, predicted the label (death/no death) in the independent test set, and performed a permutation feature importance (PFI) analysis to identify those variables showing the largest influence on the prediction of death. This implies shuffling measures of one marker at a time and then comparing the loss function (cross-entropy of the classification) of the perturbed RF model with that of

Table 1 Incidence rates and univariable hazard ratios for death in COVID-19 patients.

	Patient at risk N = 3894 (%)[a]	Death N = 712 (%)[b]	Person-days	Death Rate (×1,000 person-days)	Univariable HR (95% CI)
Gender					
Female	1491 (38.3)	243 (16.3)	24,327	10.0	-1-
Male	2403 (61.7)	469 (19.5)	39,559	11.9	1.18 (1.01–1.38)
Age, years					
18–44	348 (8.9)	6 (1.7)	4030	1.5	-1-
45–64	1413 (36.3)	75 (5.3)	22,387	3.4	2.40 (1.04–5.51)
65–74	808 (20.8)	145 (18.0)	15,437	9.4	7.12 (3.14–16.14)
75–84	849 (21.8)	266 (31.3)	14,462	18.4	13.56 (6.03–30.52)
≥85	476 (12.2)	220 (30.9)	7570	29.1	21.65 (9.60–48.82)
Hypertension					
No	1899 (48.8)	219 (11.5)	29,184	7.5	-1-
Yes	1943 (49.9)	461 (23.7)	33,994	13.6	1.85 (1.58–2.18)
Missing data	*52 (1.3)*	*32 (61.5)*			
Diabetes					
No	3103 (79.7)	481 (15.5)	49,303	9.8	-1-
Yes	739 (19.0)	203 (27.5)	13,855	14.7	1.56 (1.32–1.84)
Missing data	*52 (1.3)*	*28 (53.9)*			
Myocardial infarction					
No	3421 (87.9)	527 (15.4)	55,963	9.4	-1-
Yes	388 (10.0)	140 (36.1)	6641	21.1	2.31 (1.92–2.79)
Missing data	*85 (2.2)*	*45 (52.9)*			
Heart Failure					
No	3375 (86.7)	500 (14.8)	54,728	9.1	-1-
Yes	431 (11.1)	165 (38.3)	7710	21.4	2.44 (2.04–2.91)
Missing data	*88 (2.3)*	*47 (53.4)*			
Cancer					
No	3437 (88.3)	552 (16.1)	55,960	9.9	-1-
Yes	399 (10.3)	128 (32.1)	7074	18.1	1.88 (1.55–2.28)
Missing data	*58 (1.5)*	*32 (55.2)*			
Lung disease					
No	3269 (84.0)	513 (14.6)	53,097	9.7	-1-
Yes	557 (14.3)	167 (29.6)	9681	17.3	1.83 (1.53–2.17)
Missing data	*68 (1.8)*	*32 (47.1)*			
Obesity, BMI ≥30 kg/m^2					
No	2173 (55.8)	410 (18.9)	36,159	11.3	-1-
Yes	376 (9.7)	69 (18.4)	6683	10.3	0.92 (0.71–1.19)
Missing data	*1345 (34.5)*	*233 (17.3)*			
Smoking					
Non-smoker	2140 (55.0)	420 (19.6)	33,670	12.5	-1-
Current Smoker	319 (8.2)	59 (18.5)	5995	9.8	0.83 (0.63–1.09)
Missing data	*1435 (36.9)*	*233 (16.2)*			
CRP, mg/L					
≤3	884 (22.7)	50 (5.7)	12,906	3.9	-1-
3–10	835 (21.4)	165 (19.8)	12,384	13.3	3.39 (2.47–4.65)
≥10	1962 (50.4)	457 (23.3)	34,644	13.2	3.51 (2.62–4.71)
Missing data	*213 (5.5)*	*40 (18.8)*			
eGFR, CKD stage, mL/min/1.73 m^2					
≥90	1368 (35.1)	67 (4.9)	21,427	3.2	-1-
60–89	1409 (36.2)	205 (14.6)	23,950	8.6	2.75 (2.09–3.62)
45–59	433 (11.1)	131 (30.3)	7806	16.8	5.57 (4.15–7.48)
30–44	310 (8.0)	131 (42.3)	5317	24.6	8.17 (6.08–10.96)
15–29	197 (5.1)	110 (55.8)	2999	36.7	12.00 (8.85–16.26)
<15	87 (2.2)	47 (54.0)	1336	35.2	11.71 (8.06–17.00)
Missing data	*90 (2.3)*	*21 (23.3)*			

Abbreviations: BMI: body mass index; CKD: chronic kidney disease; CRP: C-reactive protein; eGFR: glomerular filtration rate.

[a] Column %.

[b] Row %: Death/Patient at risk.

the full model (i.e. with no permuted feature). This analysis was carried out through the *explain()* and the *variable_importance()* functions of the *DALEX* package [31].

Other statistical analyses

“Classical” (non-machine learning) statistical analyses were performed in SAS software for Windows, version 9.4.

Table 2 Subgroup analyses showing hazard ratio for mortality, according to gender. Multiple imputation analysis, N = 3894 patients and N = 712 deaths.

	Gender	
	Women	Men
Patient at risk	1491	2403
Death	243	469
Person-days	24,327	39,559
Death Rate, x1,000 PD	10.0	11.9
	HR (95% CI)	HR (95% CI)
Age, years		
18–44	-1-	-1-
45–64	1.77 (0.33–9.54)	1.77 (0.76–4.12)
65–74	3.37 (0.87–13.06)	4.06 (1.65–9.95)
75–84	6.24 (1.52–25.59)	6.18 (2.70–14.17)
≥85	8.82 (2.27–34.23)	8.06 (3.35–19.44)
Hypertension		
No	-1-	-1-
Yes	0.65 (0.51–0.82)	0.98 (0.79–1.22)
Diabetes		
No	-1-	-1-
Yes	1.05 (0.77–1.43)	0.99 (0.76–1.29)
Myocardial infarction		
No	-1-	-1-
Yes	1.27 (0.82–1.98)	1.16 (0.86–1.56)
Heart Failure		
No	-1-	-1-
Yes	1.02 (0.69–1.50)	1.09 (0.78–1.53)
Cancer		
No	-1-	-1-
Yes	1.33 (0.96–1.84)	1.41 (1.11–1.78)
Lung disease		
No	-1-	-1-
Yes	0.98 (0.64–1.52)	1.31 (1.01–1.69)
Obesity, BMI ≥30 kg/m^2		
No	-1-	-1-
Yes	1.39 (0.80–2.41)	1.15 (0.87–1.52)
Smoking		
Non-smoker	-1-	-1-
Current Smoker	0.77 (0.45–1.32)	0.98 (0.68–1.41)
CRP, mg/L		
≤3	-1-	-1-
3–10	2.53 (1.57–4.10)	2.77 (2.07–3.72)
≥10	2.27 (1.42–3.63)	2.31 (1.41–3.77)
eGFR, CKD stage, mL/min/1.73 m^2		
≥90	-1-	-1-
60–89	1.14 (0.60–2.20)	1.72 (1.20–2.49)
45–59	2.01 (1.04–3.88)	2.51 (1.64–3.85)
30–44	3.05 (1.57–5.96)	3.06 (1.82–5.16)
15–29	4.81 (2.69–8.61)	3.44 (2.37–4.99)
<15	7.93 (3.86–16.27)	3.47 (1.83–6.56)

Controlling for age, diabetes, hypertension, history of myocardial infarction, heart failure, chronic pulmonary disease, chronic kidney disease, CRP, obesity, smoking habit and anti-COVID19 drugs during hospitalization as fixed effects and repeated measures within hospital.
Abbreviations: BMI: body mass index; CKD: chronic kidney disease; CRP: C-reactive protein; eGFR: glomerular filtration rate; PD: person-days.

For the purpose of the survival analyses, we used a multiple imputation technique (SAS PROC MI, followed by PROC MIANALYZE) to maximize data availability for all variables and get robust results over different simulations (n = 10 imputed datasets).

Cox proportional-hazards regression models were used to estimate the association between characteristics of patients at hospital admission and in-hospital mortality. Since the multiple imputation technique was applied, the final standard error was obtained using the Rubin's rule based on the robust variance estimator in Cox regression [32].

Multiple imputed (for all the variables) multivariable analysis controlling for gender, age, diabetes, hypertension, history of myocardial infarction, heart failure, chronic pulmonary disease, chronic kidney disease, CRP, obesity (body mass index (BMI)≥30 kg/m^2), smoking habit and in-hospital anti-COVID-19 therapy as fixed effects; accounting for clustering within hospital by using the robust sandwich estimator was considered as the main analysis. As secondary analyses, we considered a) multiple imputation for all the variables with the exception of smoking and obesity (which were the variables with the largest number of missing data); b) case complete analysis for all the variables except smoking and obesity (N = 3454 patients); c) case complete analysis for all variables (N = 1632 patients) (see Supplementary Table 1); d) accounting for clustering within hospital by inclusion of the hospital index as random effect or strata variable (Supplementary Table 2); e) censoring of follow up at 35 days (Supplementary Table 3). Data from 4 hospitals that provided less than 25 patients each were combined in a unique group for clustering. Sensitivity analyses were conducted according to temporal pandemic waves, geographical location of the hospitals, age or gender of patients or number of risk factors at baseline (Tables 2–5 *and* Supplementary Fig. 1).

For the latter we included the following: previous myocardial infarction, heart failure, hypertension, diabetes, cancer, chronic pulmonary disease, obesity and smoking. The presence of each of these conditions was summed up to obtain a score of risk factors potentially ranging from 0 to 8 (Supplementary Fig. 1).

Results

We included in the final analyses 3894 patients who were hospitalized with laboratory-confirmed SARS-CoV-2 infection at 30 clinical centres across Italy and either died or had been discharged or were still in hospital as of May 29, 2020. Baseline characteristics are shown in Table 1.

Among COVID-19 patients, there was a higher prevalence of men (61.7%) and elderly (54.8% of patients aged ≥65 years; median age 67 years, interquartile range (IQR): 55–79 years) (Table 1). Half of the patient sample reported a diagnosis of hypertension and about a fifth had diabetes. The prevalence of chronic degenerative diseases at admission (previous myocardial infarction, heart failure, cancer and lung disease) ranged from 10.0% to 14.3%.

At hospital admission, 71.8% of COVID-19 patients had high levels of CRP (>3 mg/L) and only 35.1% had normal eGFR (≥90 mL/min/1.73 m^2).

At the end of follow-up, out of 3894 patients, 712 died (18.3%), 2650 were discharged alive (68.0%) while 532

Table 3 Subgroup analyses showing hazard ratio for mortality, according to age classes. Multiple imputation analysis, N = 3894 patients and N = 712 deaths.

	Age classes		
	18–64 years	65–74 years	≥75 years
Patient at risk	1761	808	1315
Death	81	145	486
Person-days	26,417	15,437	22,032
Death Rate, x1,000 PD	3.1	9.4	22.1
	HR (95% CI)	HR (95% CI)	HR (95% CI)
Gender			
Female	-1-	-1-	-1-
Male	1.43 (0.69–2.96)	1.87 (1.33–2.64)	1.31 (1.07–1.61)
Hypertension			
No	-1-	-1-	-1-
Yes	0.85 (0.54–1.33)	0.92 (0.61–1.39)	0.87 (0.67–1.14)
Diabetes			
No	-1-	-1-	-1-
Yes	2.00 (1.15–3.50)	0.86 (0.59–1.25)	0.96 (0.78–1.19)
Myocardial infarction			
No	-1-	-1-	-1-
Yes	1.19 (0.43–3.28)	1.28 (0.75–2.20)	1.06 (0.82–1.39)
Heart Failure			
No	-1-	-1-	-1-
Yes	1.38 (0.56–3.37)	1.12 (0.66–1.92)	0.95 (0.74–1.23)
Cancer			
No	-1-	-1-	-1-
Yes	4.76 (2.46–9.21)	1.65 (0.95–2.86)	1.10 (0.86–1.40)
Lung disease			
No	-1-	-1-	-1-
Yes	1.09 (0.41–2.88)	1.76 (1.16–2.67)	1.11 (0.89–1.39)
Obesity, BMI ≥30 kg/m^2			
No	-1-	-1-	-1-
Yes	1.36 (0.75–2.46)	1.50 (0.92–2.45)	1.05 (0.73–1.52)
Smoking			
Non-smoker	-1-	-1-	-1-
Current Smoker	1.09 (0.47–2.49)	0.66 (0.39–1.11)	1.02 (0.71–1.46)
CRP, mg/L			
≤3	-1-	-1-	-1-
3–10	3.87 (1.74–8.60)	1.50 (0.81–2.78)	3.04 (2.19–4.22)
≥10	3.75 (1.92–7.34)	1.19 (0.68–2.08)	2.66 (1.62–4.35)
eGFR, CKD stage,mL/min/1.73 m^2			
≥90	-1-	-1-	-1-
60–89	1.76 (0.94–3.32)	1.49 (0.86–2.59)	1.10 (0.61–2.01)
45–59	3.30 (1.46–7.46)	1.93 (1.11–3.37)	1.84 (1.02–3.32)
30–44	3.24 (1.41–7.45)	3.33 (1.78–6.23)	2.36 (1.38–4.02)
15–29	13.89 (3.09–62.5)	4.86 (2.18–10.82)	3.00 (1.62–5.54)
<15	6.65 (1.66–26.73)	4.77 (1.63–13.9)	3.65 (1.92–6.96)

Controlling for gender, diabetes, hypertension, history of myocardial infarction, heart failure, chronic pulmonary disease, chronic kidney disease, CRP, obesity, smoking habit and anti-COVID19 drugs during hospitalization as fixed effects and repeated measures within hospital.
Abbreviation. BMI: body mass index; CKD: chronic kidney disease; CRP: C-reactive protein; eGFR: glomerular filtration rate; PD: person-days.

(13.7%) were still hospitalised. The median follow-up was 13 days (IQR: 7 to 22). Death rate (per 1000 person-days) and univariable hazard ratios for in-hospital death are shown in Table 1.

Main analyses

The trained Random Forest model proved to be quite accurate and robust in predicting survival in the test set (sensitivity 95.2%, specificity 30.8%, classification accuracy 83.4%, F1 value 90.4%). PFI analysis revealed that the most important features in classification were eGFR and CRP (average drop of loss function >5% in the perturbed models), followed by age (loss-drop >4%). All the other input features showed a lower influence within the model (Fig. 1).

Feature importance was in line with the multivariable hazard ratios for in-hospital mortality associated with demographic and clinical characteristics as obtained by survival analyses (Fig. 2). Indeed, men and the elderly COVID-19 patients had higher mortality as compared to their counterparts. In particular, patients aged ≥85 years had an 8-fold higher mortality as compared with the

Table 4 Subgroup analyses showing hazard ratio for mortality, according to geographical location of the hospitals. Multiple imputation analysis, N = 3894 patients and N = 712 deaths.

	Geographical location of the hospitals	
	Northern regions	Central and Southern regions
Patient at risk	2278	1616
Death	513	199
Person-days	32,839	31,047
Death Rate, x1,000 PD	15.6	6.4
	HR (95% CI)	HR (95% CI)
Gender		
Female	-1-	-1-
Male	1.20 (0.95–1.52)	1.36 (0.97–1.91)
Age, years		
18–44	-1-	-1-
45–64	1.55 (0.67–3.57)	1.87 (0.45–7.71)
65–74	4.19 (1.87–9.36)	1.75 (0.67–4.61)
75–84	5.63 (2.48–12.78)	4.97 (1.90–13.04)
≥85	8.42 (3.87–18.31)	8.60 (3.54–20.94)
Hypertension		
No	-1-	-1-
Yes	1.00 (0.90–1.12)	0.63 (0.38–1.04)
Diabetes		
No	-1-	-1-
Yes	1.09 (0.85–1.38)	0.83 (0.62–1.12)
Myocardial infarction		
No	-1-	-1-
Yes	1.21 (0.96–1.51)	1.31 (0.81–2.12)
Heart Failure		
No	-1-	-1-
Yes	1.07 (0.83–1.38)	1.11 (0.77–1.60)
Cancer		
No	-1-	-1-
Yes	1.16 (0.92–1.48)	1.66 (1.21–2.27)
Lung disease		
No	-1-	-1-
Yes	1.13 (0.91–1.41)	1.70 (1.34–2.14)
Obesity, BMI ≥30 kg/m^2		
No	-1-	-1-
Yes	1.24 (0.94–1.62)	0.80 (0.45–1.41)
Smoking		
Non-smoker	-1-	-1-
Current Smoker	1.11 (0.82–1.50)	1.17 (0.76–1.81)
CRP, mg/L		
≤3	-1-	-1-
3–10	2.59 (1.77–3.80)	2.35 (1.40–3.93)
≥10	2.53 (1.71–3.74)	2.09 (0.90–4.86)
eGFR, CKD stage, mL/min/1.73 m^2		
≥90	-1-	-1-
60–89	1.45 (1.05–2.00)	1.42 (0.87–2.29)
45–59	2.06 (1.45–2.92)	2.87 (1.84–4.47)
30–44	3.11 (2.09–4.63)	2.99 (1.65–5.42)
15–29	3.29 (2.48–4.37)	4.89 (2.98–8.03)
<15	6.13 (3.62–10.38)	3.85 (2.00–7.41)

Controlling for gender, age, diabetes, hypertension, history of myocardial infarction, heart failure, chronic pulmonary disease, chronic kidney disease, CRP, obesity, smoking habit and anti-COVID19 drugs during hospitalization as fixed effects and repeated measures within hospital.

Abbreviation. BMI: body mass index; CKD: chronic kidney disease; CRP: C-reactive protein; eGFR: glomerular filtration rate; PD: person-days.

Northern regions: Novara, Milano, Cinisello Balsamo, Monza, Varese, Cremona, Padova, Forlì, Ravenna, Modena, Pavia.

Central-Southern regions: Firenze, Pisa, Roma, Chieti, Pescara, Napoli, Pozzilli, Acquaviva delle Fonti, San Giovanni Rotondo, Taranto, Catanzaro, Catania, Palermo.

Table 5 Subgroup analyses showing hazard ratio for mortality, according to pandemic wave. Multiple imputation analysis, N = 3894 patients and N = 712 deaths.

	Pandemic wave	
	From 19/02/2020 to 19/03/2020	From 20/03/2020 to 23/05/2020
Patient at risk	1712	2182
Death	324	388
Person-days	29,149	34,737
Death Rate, x1,000 PD	11.1	11.2
	HR (95% CI)	HR (95% CI)
Gender		
Female	-1-	-1-
Male	1.53 (1.27–1.85)	1.32 (1.04–1.68)
Age, years		
18–44	-1-	-1-
45–64	1.73 (0.55–5.49)	1.74 (0.58–5.24)
65–74	4.15 (1.51–11.42)	3.40 (1.24–9.29)
75–84	6.46 (2.17–19.25)	5.74 (1.94–17.02)
≥85	6.61 (2.29–19.02)	9.08 (3.38–24.37)
Hypertension		
No	-1-	-1-
Yes	1.06 (0.80–1.41)	0.73 (0.55–0.97)
Diabetes		
No	-1-	-1-
Yes	0.96 (0.68–1.36)	1.09 (0.87–1.38)
Myocardial infarction		
No	-1-	-1-
Yes	1.07 (0.77–1.48)	1.36 (0.98–1.89)
Heart Failure		
No	-1-	-1-
Yes	1.07 (0.73–1.58)	1.03 (0.77–1.39)
Cancer		
No	-1-	-1-
Yes	1.38 (0.92–2.07)	1.31 (1.01–1.68)
Lung disease		
No	-1-	-1-
Yes	1.30 (0.99–1.71)	1.18 (0.88–1.58)
Obesity, BMI ≥30 kg/m^2		
No	-1-	-1-
Yes	1.27 (0.92–1.75)	1.07 (0.70–1.64)
Smoking		
Non-smoker	-1-	-1-
Current Smoker	0.89 (0.54–1.48)	1.04 (0.74–1.46)
CRP, mg/L		
≤3	-1-	-1-
3–10	2.62 (1.58–4.36)	2.59 (1.74–3.86)
≥10	2.31 (1.40–3.82)	2.26 (1.29–3.97)
eGFR, CKD stage, mL/min/1.73 m^2		
≥90	-1-	-1-
60–89	1.46 (0.97–2.20)	1.48 (0.97–2.25)
45–59	2.23 (1.34–3.72)	2.25 (1.44–3.53)
30–44	2.45 (1.43–4.19)	3.42 (2.13–5.47)
15–29	2.99 (1.86–4.83)	4.71 (2.86–7.77)
<15	7.50 (4.32–13.01)	3.85 (2.01–7.38)

*Controlling for gender, age, diabetes, hypertension, history of myocardial infarction, heart failure, chronic pulmonary disease, chronic kidney disease, CRP, obesity, smoking habit and anti-COVID19 drugs during hospitalization as fixed effects and repeated measures within hospital.

Abbreviations: BMI: body mass index; CKD: chronic kidney disease; CRP: C-reactive protein; eGFR: glomerular filtration rate; PD: person-days.

youngest (18–64 years) patients. Reduced kidney function was associated with a proportionally higher increase of death risk. High levels of CRP, a well-known, sensitive marker of inflammation, were also associated with an increased mortality risk.

Among chronic degenerative diseases, history of or active cancer was associated with a high in-hospital mortality (HR: 1.33; 95% CI 1.09–1.63) and an upward trend of risk was found for previous myocardial infarction (HR: 1.20; 95% CI 0.93–1.55), chronic pulmonary disease (HR: 1.20; 95% CI 0.97–1.48) and obesity (HR: 1.21; 95% CI 0.91–1.61). On the contrary, current smoking, diabetes and hypertension were not associated with mortality in our COVID-19 patients. However, a score of risk factors ≥3 was associated with a 47% increase in mortality risk (Supplementary Fig. 1).

These results were confirmed in a number of secondary analyses that were run to test the robustness of the main findings by using both multiple imputation and complete-case analysis (Supplementary Table 1) or through accounting of clustering within hospitals (Supplementary Table 2) and censoring of follow up at 35 days (Supplementary Table 3).

Subgroup analyses

Subgroup analyses according to gender, age classes, geographical location of the hospitals (Northern, and Central/Southern regions) and temporal pandemic waves are presented in Tables 2–5 Additional information on age- and gender-adjusted prevalence of the here-studied determinants of in-hospital mortality according to gender, age classes, geographical location of the hospitals and pandemic wave are reported in Supplementary Table 4.

A higher death rate was observed in the elderly (age ≥ 75 years: 22.1 deaths x 1000 person-days) compared to younger patients (18–64 years: 3.1 deaths x 1000 person-days and 65–74 years: 9.4 deaths x 1000 person-days), as well as in patients admitted to hospitals of Northern regions (15.6 deaths x 1000 person-days) in comparison with those from Central/Southern regions (6.4 deaths x 1000 person-days) (Tables 3 and 4).

In general, subgroup analyses confirmed the findings reported in Fig. 1. However, some differences could be observed. A hypertensive status was associated with a reduction of mortality among women and those patients hospitalised during the second wave of the pandemic emergency (Tables 2 and 5). However, major risk factors were also confirmed among those patients with hypertension (Supplementary Table 5).

Diabetes turned out to be associated with a higher risk of in-hospital mortality in younger COVID-19 patients (18–64 years) (Table 3) The magnitude of the association between history of cancer and in-hospital mortality was stronger in younger than in older patients (Table 3).

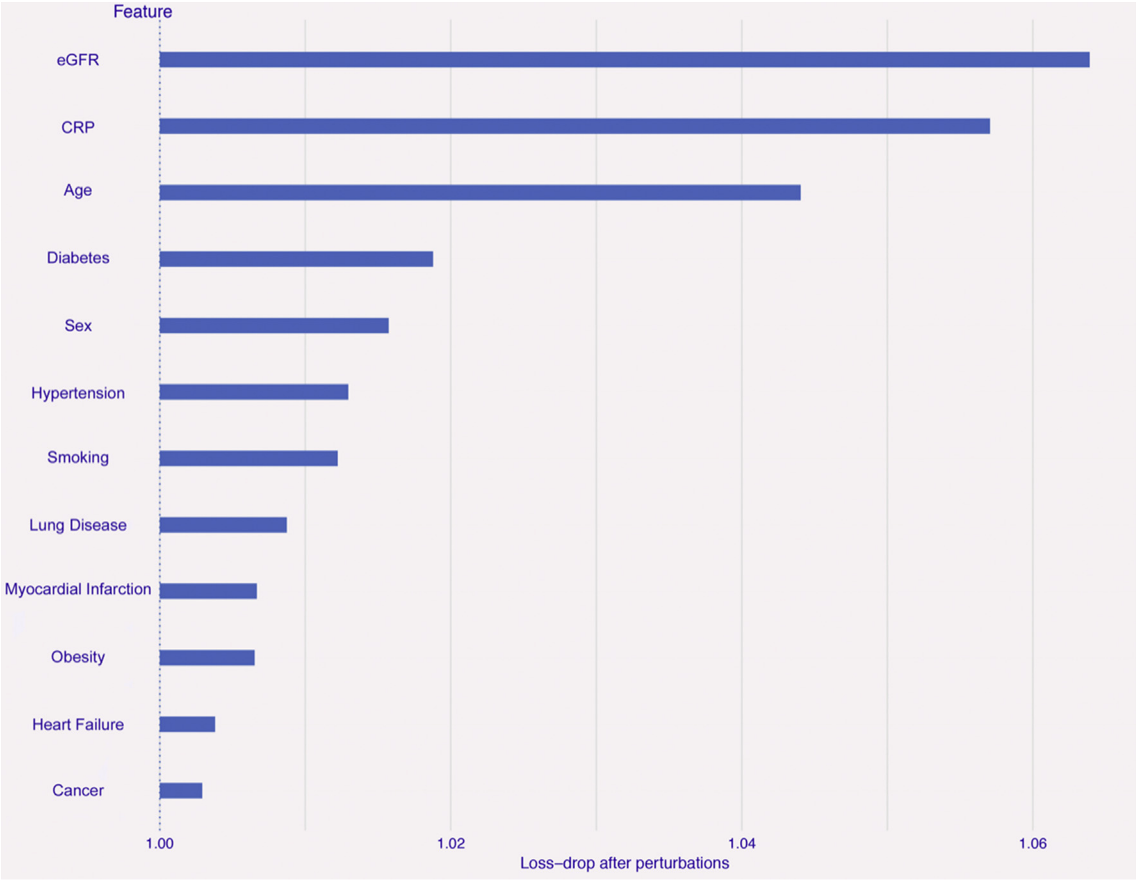


Figure 1 Permutation Feature Importance analysis for the prediction of mortality. Bars indicate the importance of each feature used for the prediction of death in the Random Forest algorithm within the training set (N = 2,725, 70% of the total imputed sample available), based on the ratio between loss function in the perturbed model (i.e., after permutation of a given variable) and loss function in the full model (with no permuted variable). The higher the ratio, the more the perturbed model is altered and the more important is the permuted feature. Abbreviations. CRP: C-reactive protein; eGFR: glomerular filtration rate.

Chronic pulmonary disease was associated with in-hospital mortality both in men and in patients aged 65–74 years, as well as in those admitted to hospitals of Central-Southern regions (Tables 2–4) Additionally, non-hypertensive patients with chronic pulmonary diseases showed a 2-fold higher risk as compared with those without lung disease (Supplementary Table 5).

We found no association between obesity and in-hospital mortality, data confirmed by complete-case analysis restricted to 1517 patients without missing data on BMI and by distinguishing obesity as stage 1 (BMI = 30–34.9 kg/m^2) and stage 2 (BMI ≥35 kg/m^2) (Supplementary Table 6).

Discussion

We analysed data from a large cohort of patients hospitalised between February 19 and May 23, 2020 with laboratory-confirmed SARS-CoV-2 infection, with the aim of defining an algorithm to classify death risk and identifying major predictors of in-hospital death.

Our sample had a higher prevalence of men (61.7%) and elderly aged ≥65 years (54.8%) as compared to a large study on COVID-19-patients in New York City (49.5% of men and less than 30% of elderly individuals [4]), while in our cohort obese subjects were under-represented (9.7% vs 35% in the US cohort).

Through a supervised machine learning approach, we built a random forest algorithm to classify death risk in our cohorts, also identifying the most influential predictors of this risk and taking into account potential non-linear relations and interaction effects among them. This algorithm reached a good (>83%) accuracy, with a high (>90%) sensitivity for classification of survivors (>95%), probably due to the fact that the majority of our patients did not die, which made the model more prone and fitted to recognise these outcomes. Moreover, this approach revealed that age, renal function (eGFR) and circulating inflammation (hs-CRP) were the most important features predicting survival. While CRP and age have been already identified as top influential features in previous machine learning algorithms predicting mortality risk [21,22,25], we are not aware of any comparable approach identifying renal function among top features. This may be due to the huge heterogeneity of settings and of clinical variables available from COVID-19 patients in different studies, which make it hard to compare the studies among themselves. Further

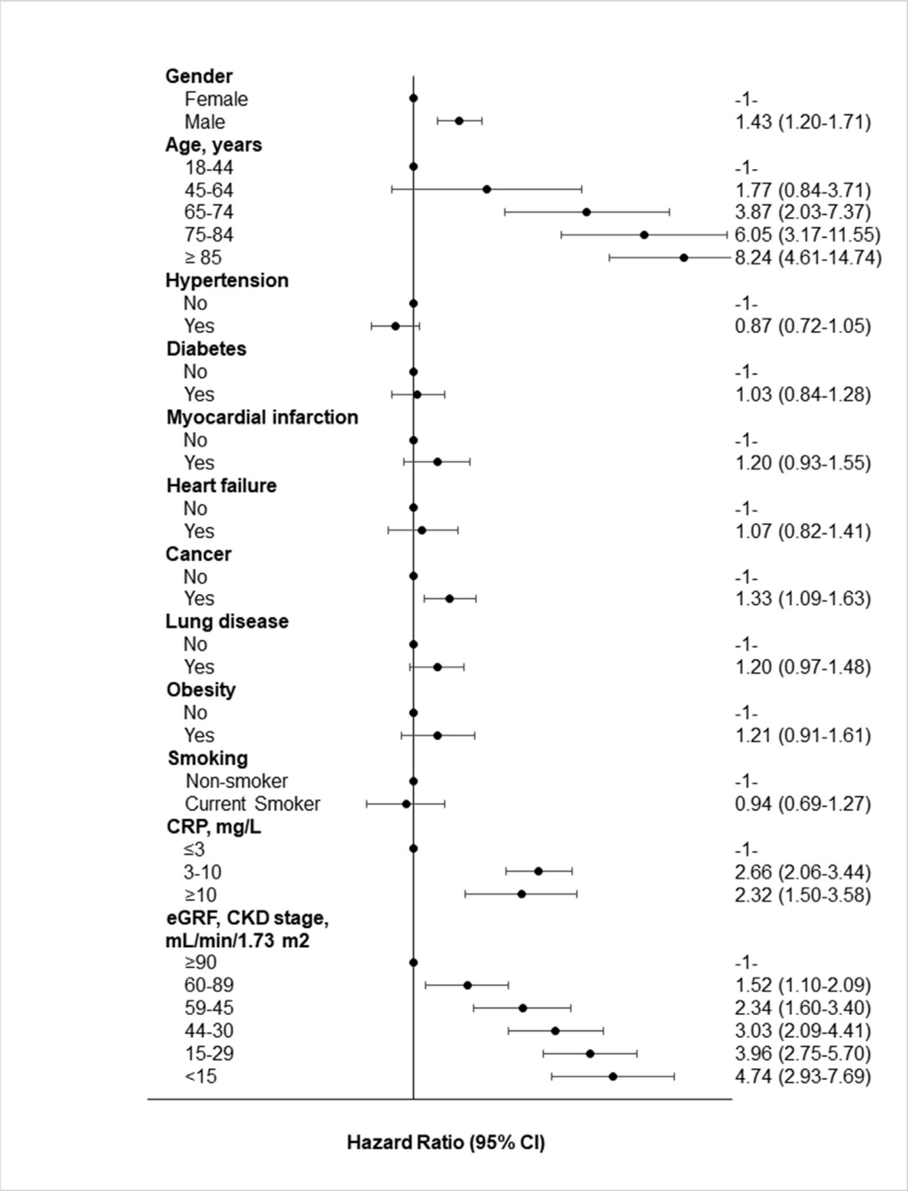


Figure 2 Multivariable hazard ratios for in-hospital mortality for different characteristics of patients. Multiple imputation analysis, N = 3894 patients and N = 712 deaths. Controlling for gender, age, diabetes, hypertension, history of myocardial infarction, heart failure, chronic pulmonary disease, chronic kidney disease, CRP, obesity, smoking habit and anti-COVID19 drugs during hospitalization as fixed effects and repeated measures within hospital. Abbreviations. BMI: body mass index; CKD: chronic kidney disease; CRP: C-reactive protein; eGFR: glomerular filtration rate.

investigations with more homogeneous settings and clinical variables are needed to further substantiate these findings. However, "classical" (non-machine learning based) statistical analyses (see below) revealed data in line with the results of the permutation feature importance analysis applied to the random forest algorithm, providing robustness to our findings. Indeed, Cox survival regressions revealed that advanced age, impaired renal function and elevated levels of CRP, as well as male gender and cancer, were independent predictors of in-hospital death.

These data are in line with prior evidence on a small sample of Italian patients admitted to hospital during the first wave of the COVID-19 pandemic, documenting increased risk of death associated with older age, critical disease and high levels of CRP [9]. Accordingly, investigations from the U.S. and China also identified advanced age [2,4,5], cancer [4] and inflammation [2,4,5] as major predictors of death.

At variance from other studies [4,7–9], however, we failed to find any association between obesity and smoking on admission with the risk of in-hospital death, although findings on tobacco use were in accordance with what seen in a U.S. cohort [4]. Obesity was not a risk factor in a large U.S. cohort [5] and in another Italian series of COVID-19 patients [33].

Our findings do not support a pivotal role for mortality of previous known cardiovascular disease and other comorbidities, data apparently contrasting from those of previous studies suggesting a major association of pre-existing cardiovascular conditions with higher mortality [5,12] and poor prognosis [34].

However, neither the study by Cummings [5] nor that by Ciceri [12] did account for smoking status in their multivariable analysis, and this may have biased their results; moreover, the Italian analysis [12] was conducted on a small sample of subjects from one single centre, as in the study by Colaneri and colleagues [35]. In another setting, the associations between cardiovascular disease and related comorbidities were mitigated after multivariable adjustments [36]. In addition, studies with a larger sample size, similar to ours, failed to find any association between cardiovascular comorbidities and in-hospital mortality after multivariable adjustments [4].

Age is one of the strongest predictors of mortality in COVID-19 patients, and the prevalence of cardiovascular comorbidities obviously increases with age. In univariable analysis, all cardiovascular risk factors are associated with a higher mortality. Adjustment for age strongly reduced almost all associations, and further adjustment for kidney impairment completely eliminated any residual association, suggesting that the latter were essentially driven by these two main factors (Supplementary Table 7), as also discussed previously [37,38].

The importance of renal function is reinforced by the observation that its deterioration was predictive of mortality across all sub-groups analysed in this study, consistent with data from 701 Chinese patients with COVID-19 showing that kidney disease was associated with a higher risk of in-hospital death [39]. We acknowledge that our analysis is based on one-time measurement of eGFR values at the hospital entry that does not allow to discern infection-induced acute kidney injury (AKI). It has been recently reported that patients with hospital-acquired AKI as opposed to those with community-acquired AKI had higher rates of in-hospital death [40].

Moreover, in a small sample of Italian patients, low levels of albumin were associated with higher risk [41]. It is known that SARS-CoV-2 can infect podocytes and tubular epithelial cells [42]. Thus, viral infection may be the cause of renal abnormalities, partially explaining the early involvement of kidney impairment in the severity of COVID-19. Alternatively – or possibly complementary to this – renal dysfunction may predispose affected individuals to a more rapid deterioration and death.

We carried out a number of subgroup analyses in order to test whether the association between variables found to predict in-hospital death was likely to vary according to age, gender, geographical area and temporal pandemic waves. Our results indicate a similar magnitude of the association between selected determinants and the outcome, although the strength of the relation appeared to differ for some baseline characteristics.

First, in-hospital mortality was likely to follow a geographical gradient, death rates being more than double in hospitals located in Northern regions as compared to Central-Southern centres. These inter-regional disparities in death rates are likely linked to the timing of the COVID-19 outbreak that first hit Northern regions, forced to face an almost unknown and unexpected infectious agent that only later spread to Central-Southern areas of the country. Moreover, evidence-based management guidelines for patients diagnosed with COVID-19 were not available during the early stages of the pandemic. In addition, a viral load in Northern Italian coronavirus infections higher than in other parts of the country may have contributed to the different case-fatality ratio.

Analyses between temporal pandemic waves revealed similar death rates in the first and second wave, although the impact of hypertension on mortality was likely mitigated during the second wave (end of March-end of May), as compared to the first pandemic wave (mid-February—mid-March), while the other predictors were substantially homogenous. These data indicate that the risk factors analysed here were basically independent from the advancement in therapies against COVID-19 that occurred over time, and highlight the importance of prevention strategies targeting high risk individuals more aggressively. Finally, our data are unlikely to support the assumption of a reduced viral load over time.

Risk predictors slightly varied between genders. Gender-related differences are well established both for infectious and non-communicable diseases [43,44]; more recently, testosterone has been suspected as playing a critical role in driving the excess of risk observed among men tested positive for COVID-19 [45].

Among subjects under 65 years, major risk factors were kidney disease and inflammation, as well as cancer and diabetes, while for those aged ≥75, male gender, kidney disease and inflammation were the only factors independently associated with fatality.

Overall, our data suggest that among young patients pre-existing diseases represent a major risk factor for a clinical outcome such as mortality, while they have a marginal role in the elderly.

A laboratory predictor, such as CRP levels, was consistently and independently associated with death risk, suggesting that pre-hospitalisation inflammation (possibly in response to viral infection) might have been even more important than baseline personal characteristics and comorbidities. Our data on excess of death risk associated with inflammation support the hypothesis of a 'cytokine storm' caused by SARS-CoV-2 being a main player in the high mortality of COVID-19 pneumonia and ARDS [46].

Strengths and limitations

To the best of our knowledge, this is the largest study on COVID-19 patients actually available in Italy and one of the largest multicentre studies of patients with COVID-19 in Europe. Moreover, we are not aware of other larger studies attempting to classify death risk in COVID-19 patients through a machine learning approach, in a scenario where most studies show a sample size <1000 [17,32]. Indeed, an unselected patient sample from 30 hospitals, covering the entire Italian territory and all the overt epidemic period in Italy, was collected and analysed.

A second strength is the use of a multifaceted approach to estimate the risk and the most influential predictors of in-hospital death, which implied i) a supervised machine learning algorithm to classify death and survival and ii) multivariable, time-varying analyses to describe in detail the relationship of independent risk factors with in-hospital death. This allowed to also take into account potential non-linear and synergistic relationships among the risk factors. Moreover, several statistical approaches were used to test the robustness of the associations and overcome biases due to the observational nature of the study.

On the other hand, the results of this study should be interpreted in the light of several limitations. First, results may not be generalized to other populations with different geographical and socioeconomic conditions, differences in national health service or insurance-based health expenses, and in the natural history of COVID-19. Due to the retrospective nature of our study, some information (such as smoking and obesity) was not available in all patients. Lack of data on vital signs (e.g. respiratory rate or oxygen saturation) represents another limitation of this study although CRP levels and renal function were likely to be proxies of patients' severity upon admission.

Finally, the possibility of unmeasured residual confounders cannot be ruled-out due to the observational nature of our study.

Conclusions

Impaired renal function, elevated levels of CRP and advanced age at hospital admission were powerful predictors of higher in-hospital death in a large cohort of unselected patients with COVID-19 admitted to 30 different clinical centres all over Italy.

Some of these risk factors are likely to vary somewhat according to gender, age and geographical location for reasons that are not fully understood, although being possibly associated, at least in part, with timing of the pandemic outbreak and the local management modalities of patients. Along with epidemiological, pathophysiologic and possibly epigenetic mechanisms underlying such differences need to be elucidated.

Our data also support the importance of measuring inflammatory markers other than CRP at admission and during hospital stay in patients with COVID-19.

Finally, our findings contribute to identify some potential effect modifiers in the association between risk factors and in-hospital mortality among COVID-19 diagnosed patients. Future comparison with other European and Mediterranean Countries, possibly including assessments of other factors such as socioeconomic status [47], will be useful to elucidate possible differences in risk factors and outcomes observed across different geographical areas.

Authors' contribution

ADC, LI, RDC designed the study. ADC and LI supervised research and led the data collection. All Authors contributed to collect, re-analyse, check and pool data, critically reviewed and approved the manuscript. SC and AG analysed the data and prepared the results. GV advised statistical analyses. MO managed the web-based database. MB and SC wrote the first draft of the manuscript with input from RDC, KdD, ADC, AG, LI, and the other Authors.

Source of funding

None.

Additional information

Correspondence and other requests should be addressed to LI.

Declaration of Competing Interest

All Authors declare no competing interests.

Acknowledgments

We thank the 30 participating clinical Centres included in this cohort. This Article is dedicated to all patients who suffered or died, often in solitude, due to COVID-19; their tragic fate gave us moral strength to initiate and complete this research.

The Authors alone are responsible for the views expressed in this article. They do not necessarily represent the views, decisions, or policies of the Institutions with which they are affiliated. AG was supported by Fondazione Umberto Veronesi.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.numecd.2020.07.031.

References

[1] https://covid19.who.int. [Accessed 7 August 2020].

[2] Tian W, Jiang W, Yao J, Nicholson CJ, Li RH, Sigurslid HH, et al. Predictors of mortality in hospitalized COVID-19 patients: a systematic review and meta-analysis. J Med Virol 2020 May 22. https://doi.org/10.1002/jmv.26050.

[3] Mantovani A, Byrne CD, Zheng MH, Targher G. Diabetes as a risk factor for greater COVID-19 severity and in-hospital death: a meta-analysis of observational studies [published online ahead of print,

2020 May 29] Nutr Metabol Cardiovasc Dis 2020;30(8):1236–48. https://doi.org/10.1016/j.numecd.2020.05.014.

[4] Petrilli CM, Jones SA, Yang J, Rajagopalan H, O'Donnell L, Chernyak Y, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. BMJ 2020 May 22;369: m1966. https://doi.org/10.1136/bmj.m1966.

[5] Cummings MJ, Baldwin MR, Abrams D, Jacobson SD, Meyer BJ, Balough EM, et al. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. Lancet 2020;395(10239):1763–70. https://doi.org/10.1016/S0140-6736(20)31189-2.

[6] Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, et al. Risk factors of critical & mortal COVID-19 cases: a systematic literature review and meta-analysis. J Infect 2020 Apr 23. https://doi.org/10.1016/j.jinf.2020.04.021. S0163-4453(20)30234-6.

[7] Klang E, Kassim G, Soffer S, Freeman R, Levin MA, Reich DL. Morbid obesity as an independent risk factor for COVID-19 mortality in hospitalized patients younger than 50. Obesity 2020 May 23. https://doi.org/10.1002/oby.22913.

[8] Tamara A, Tahapary DL. Obesity as a predictor for a poor prognosis of COVID-19: a systematic review. Diabetes Metab Syndrome 2020 May 12;14(4):655–9. https://doi.org/10.1016/j.dsx.2020.05.020.

[9] Giacomelli A, Ridolfo AL, Milazzo L, Oreni L, Bernacchia D, Siano M, et al. 30-day mortality in patients hospitalized with COVID-19 during the first wave of the Italian epidemic: a prospective cohort study. Pharmacol Res 2020 May 22;158:104931. doi: 10.1016/j.phrs.2020.104931.

[10] https://www.epicentro.iss.it/en/coronavirus/sars-cov-2-dashboard. [Accessed 7 August 2020].

[11] 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(19):1821–9. https://doi.org/10.1093/eurheartj/ehaa388.

[12] Ciceri F, Castagna A, Rovere-Querini P, De Cobelli F, Ruggeri A, Galli L, et al. Early predictors of clinical outcomes of COVID-19 outbreak in Milan, Italy [published online ahead of print, 2020 Jun 11] Clin Immunol 2020:108509. doi:10.1016/j.clim.2020.108509.

[13] http://opendatadpc.maps.arcgis.com/apps/opsdashboard/index.html#/b0c68bce2cce478eaac82fe38d4138b1. [Accessed 7 August 2020].

[14] Iaccarino G, Grassi G, Borghi C, Ferri C, Salvetti M, Volpe M, et al. Age and multimorbidity predict death among COVID-19 patients: results of the SARS-RAS study of the Italian society of hypertension. 2020. https://doi.org/10.1161/HYPERTENSIONAHA.120.15324 [published online ahead of print, 2020 Jun 22]. Hypertension HYPERTENSIONAHA12015324.

[15] https://coronavirus.jhu.edu/map.html [Accessed August 7, 2020].

[16] 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 Mar 28; 395(10229):1054–62. https://doi.org/10.1016/S0140-6736(20)30566-3.

[17] Shi S, Qin M, Cai Y, Liu T, Shen B, Yang F, et al. Characteristics and clinical significance of myocardial injury in patients with severe coronavirus disease 2019. Eur Heart J 2020 May 11:ehaa408. https://doi.org/10.1093/eurheartj/ehaa408.

[18] Laure Wynants, Ben Van Calster, Collins Gary S, Riley Richard D, Georg Heinze, Ewoud Schuit, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020;369:m1328.

[19] Xie J, Hungerford D, Chen H, Abrams ST, Li S, Wang G, et al. Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19. medRxiv [Preprint]; 2020. https://doi.org/10.1101/2020.03.28.20045997.

[20] Caramelo F, Ferreira N, Oliveiros B. Estimation of risk factors for COVID-19 mortality - preliminary results. medRxiv [Preprint]; 2020. https://doi.org/10.1101/2020.02.24.20027268.

[21] Lu J, Hu S, Fan R, Liu Z, Yin X, Wang Q, et al. ACP risk grade: a simple mortality index for patients with confirmed or suspected severe acute respiratory syndrome coronavirus 2 disease (COVID-19) during the early stage of outbreak in Wuhan. China: medRxiv [Preprint; 2020. https://doi.org/10.1101/2020.02.20.20025510.

[22] Yan L, Zhang H-T, Xiao Y, Wang M, Sun C, Liang J, et al. Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. medRxiv [Preprint]; 2020. https://doi.org/10.1101/2020.02.27.20028027.

[23] Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. Crit Care 2020;24:108. https://doi.org/10.1186/s13054-020-2833-7. pmid:32188484.

[24] Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with covid-19 using artificial intelligence to help medical decision-making. medRxiv [Preprint]; 2020. https://doi.org/10.1101/2020.03.30.20047308.

[25] Yan L, Zhang H, Goncalves J, Xiao J, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell 2020;2:283–8. https://doi.org/10.1038/s42256-020-0180-7.

[26] Di Castelnuovo A, De Caterina R, de Gaetano G, Iacoviello L. Controversial relationship between renin-angiotensin system inhibitors and severity of COVID-19: announcing a large multicentre case-control study in Italy. 2020. https://doi.org/10.1161/HYPERTENSIONAHA.120.15370 [published online ahead of print, 2020 May 8]. Hypertension 10.1161/HYPERTENSIONAHA.120.15370.

[27] Clinical management of severe acute respiratory infection when novel coronavirus (2019-nCoV) infection is suspected: interim guidance. 28 January 2020. Available at: https://apps.who.int/iris/bitstream/handle/10665/330893/WHO-nCoV-Clinical-2020.3-eng.pdf?sequence=1&isAllowed=y. [Accessed 31 May 2020].

[28] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. http://www.r-project.org/.

[29] Kowarik A, Templ M. Imputation with the R package VIM. J Stat Software 2016;74:1–16. https://doi.org/10.18637/jss.v074.i07.

[30] https://CRAN.R-project.org/package=caret. [Accessed 7 August 2020].

[31] Biecek P. DALEX: explainers for complex predictive models in R. J Mach Learn Res 2018;19:1–5. http://jmlr.org/papers/v19/18-416.html. [Accessed 10 April 2020].

[32] Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley; 1987.

[33] Moriconi D, Masi S, Rebelos E, Virdis A, Manca ML, De Marco S, et al. Obesity prolongs the hospital stay in patients affected by COVID-19, and may impact on SARS-COV-2 shedding [published online ahead of print, 2020 Jun 4] Obes Res Clin Pract 2020; S1871–403X(20):30401–4. https://doi.org/10.1016/j.orcp.2020.05.009.

[34] Li M, Dong Y, Wang H, Guo W, Zhou H, Zhang Z, et al. Cardiovascular disease potentially contributes to the progression and poor prognosis of COVID-19. Nutr Metabol Cardiovasc Dis 2020; 30(7):1061–7. https://doi.org/10.1016/j.numecd.2020.04.013.

[35] Colaneri M, Sacchi P, Zuccaro V, Biscarini S, Sachs M, Roda S, et al. Clinical characteristics of coronavirus disease (COVID-19) early findings from a teaching hospital in Pavia, North Italy, 21 to 28 February 2020. Euro Surveill 2020;25(16):2000460. https://doi.org/10.2807/1560-7917.ES.2020.25.16.2000460.

[36] Palaiodimos L, Kokkinidis DG, Li W, Karamanis D, Ognibene J, Arora S, et al. Severe obesity, increasing age and male sex are independently associated with worse in-hospital outcomes, and higher in-hospital mortality, in a cohort of patients with COVID-19 in the Bronx, New York [published online ahead of print, 2020 May 16] Metabolism 2020;108:154262. https://doi.org/10.1016/j.metabol.2020.154262.

[37] Cappuccio FP, Siani A. Covid-19 and cardiovascular risk: susceptibility to infection to SARS-CoV-2, severity and prognosis of Covid-19 and blockade of the renin-angiotensin-aldosterone system. An evidence-based viewpoint [published online ahead of print, 2020 May 29] Nutr Metabol Cardiovasc Dis 2020. https://doi.org/10.1016/j.numecd.2020.05.013. S0939-4753(20)30206-4.

[38] Li G, Hu R, Gu X. A close-up on COVID-19 and cardiovascular diseases. Nutr Metabol Cardiovasc Dis 2020;30(7):1057–60. https://doi.org/10.1016/j.numecd.2020.04.001.

[39] Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. Kidney Int 2020;97(5):829–38. https://doi.org/10.1016/j.kint.2020.03.005.

[40] Pelayo J, Lo KB, Bhargav R, Gul F, Peterson E, De Joy III R, et al. Clinical characteristics and outcomes of community- and hospital-acquired acute kidney injury with COVID-19 in a US inner city hospital system [published online ahead of print, 2020 Jun 18] Cardiorenal Med 2020:1–9. https://doi.org/10.1159/000509182.

[41] Violi F, Cangemi R, Romiti GF, Ceccarelli G, Oliva A, Alessandri F, et al. IS albumin predictor OF mortality IN COVID-19? [published online ahead of print, 2020 Jun 11] Antioxidants Redox Signal 2020. https://doi.org/10.1089/ars.2020.8142. 10.1089/ars.2020.8142.

[42] Martinez-Rojas MA, Vega-Vega O, Bobadilla NA. Is the kidney a target of SARS-CoV-2? Am J Physiol Ren Physiol 2020;318(6): F1454–62. https://doi.org/10.1152/ajprenal.00160.2020.

[43] Van Lunzen J, Altfed M. Sex differences in infectious diseases-common but neglected. J Infect Dis 2014;209(Suppl 3):S79–80. https://doi.org/10.1093/infdis/jiu159.

[44] Vlassoff C. Gender differences in determinants and consequences of health and illness. J Health Popul Nutr 2007 Mar;25(1):47–61.

[45] Giagulli VA, Guastamacchia E, Magrone T, Jirillo E, Lisco G, De Pergola G, et al. Worse progression of COVID-19 in men: is Testosterone a key factor? [published online ahead of print, 2020 Jun 11] Andrology 2020. https://doi.org/10.1111/andr.12836. 10.1111/andr.12836.

[46] Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ, et al. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet 2020;395(10229):1033–4. https://doi.org/10.1016/S0140-6736(20)30628-0.

[47] Bonaccio M, Iacoviello L, Donati MB, de Gaetano G. A socioeconomic paradox in the COVID-19 pandemic in Italy: a call to study determinants of disease severity in high and low-income Countries. Mediterr J Hematol Infect Dis 2020;12(1):e2020051. https://doi.org/10.4084/MJHID.2020.051.