Complement C3 identified as a unique Risk Factor for Disease Severity among Young COVID-19 Patients in Wuhan

Running title: Complement C3 is a unique risk factor for young COVID-19 patients

Weiting Cheng MD ¹, Roman Hornung PhD ², Kai Xu MD ^{3, #}, Jian Li PhD ²

¹ Oncology Department, Wuhan No.1 Hospital, Wuhan, 430022, China

² Institute of Medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilian-University Munich, Germany

³ Department of orthopedics, Tongji hospital, Huazhong University of Science and Technology, Wuhan, 430030, China.

Weiting Cheng: joycvt@126.com; Roman Hornung: hornung@ibe.med.uni-muenchen.de; Jian Li: lijian@ibe.med.uni-muenchen.de

Corresponding author: Kai Xu, Email: godocoto@163.com; Tel: +86-027-83665418
Tongji Hospital, Huazhong University of Science and Technology

Jiefang Avenue 1095

Wuhan 430030

Province Hubei
China

Words: 3000

number of references: 24

number of tables: 5

number of figures: 2

Key words COVID-2019; SARS-CoV-2; young patients; prognosis; severe disease course

Abstract

Background

Given that a substantial proportion of the subgroup of COVID-19 patients that face a severe disease course are younger than 60 years, it is critical to understand the disease-specific characteristics of young COVID-19 patients. Risk factors for a severe disease course for young COVID-19 patients and possibly non-linear influences remain unknown.

Methods

Data of COVID-19 patients with clinical outcome in a designated hospital in Wuhan, China, collected retrospectively from Jan 24th to Mar 27th, were analyzed. Clinical, demographic, treatment and laboratory data were collected from patients' medical records. Uni- and multivariable analysis using logistic regression and random forest, with the latter allowing the study of non-linear influences, were performed to investigate and exploit the clinical characteristics of a severe disease course.

Results

A total of 762 young patients (median age 47 years, interquartile ranges [IQR] 38 - 55, range 16 - 60; 55.9% female) were included, as well as 714 elderly patients as a comparison group. Among the young patients, 362 (47.5%) had a severe/critical disease course and the mean age was significantly higher in the severe subgroup than in the mild subgroup (59.3 vs. 56.0, Student's t-test: $p < 0.001$). The uni- and multivariable analysis suggested that several covariates such as elevated levels of ASS, CRP and LDH, and decreased lymphocyte counts are influential on disease severity independent of age. Elevated levels of complement C3 (odds ratio [OR] 15.6, 95% CI 2.41-122.3; $p=0.039$) are particularly associated with the risk for the development of severity specifically in young patients, where no such influence seems to exist for elderly patients. Additional analysis suggests that the influence of complement C3 in young patients is independent of age, gender, and comorbidities. Variable importance values and partial dependence plots obtained using random forests delivered additional insights, in particular indicating non-linear influences of risk factors on disease severity.

Conclusion

In young patients with COVID-19, the levels of complement C3 correlated with disease severity and tended to be a good predictor of adverse outcome.

Introduction

The pandemic caused by the coronavirus disease 2019 (COVID-19) that is associated with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has affected almost every corner of the world. As of June 30th, 2020, there have been more than 10 million confirmed cases, including more than 500,000 deaths according to a report by the World Health Organization (WHO)¹, where the numbers of cases and deaths are expected to continue to rise. The clinical spectrum of COVID-19 appears to be wide, encompassing asymptomatic infection, mild upper respiratory illness, neurological symptoms, renal and gastrointestinal complications, severe viral pneumonia with respiratory failure,

multiple organ failure and even death²⁻⁵. Approximately 20~25% of patients will have a severe disease course.

Despite numerous studies showing a higher risk of severe COVID-19 in elderly patients, a substantial proportion of young patients also have an increased risk of developing a severe course. According to a report of U.S. CDC, 47% of hospitalized patients are under the age of 65, and 48% of those admitted to ICU are under the age of 65⁶. Although potential risk factors for mortality were reported to include advanced age, male gender, presence of comorbidities, the development of a cytokine storm and an immunocompromised status⁸⁻⁹, the risk factors for the development of a severe course specific to young patients (≤ 60 years old) remain under investigation. Of note, to date, a large proportion of studies applied logistic regression to analyze risk factors related to COVID-19 infection, assuming an underlying causal linear influence on the log odds²⁻⁵. However, given the intricate complexity of COVID-19 infection, statistical analysis with the consideration of non-linear relationships might provide more insightful information on COVID-19 related potential risk factors.

In an effort to fill these gaps, an important aim of this single-center study was to analyze clinical, demographic and treatment data of patients sequentially admitted into the Wuhan No.1 hospital, in an attempt to elucidate risk factors and main causes among young COVID-19 patients for experiencing a severe disease course. A further aim was to use the large amount of data available in this study to foster the knowledge on general, age-independent, and non-linear relational risk factors for a severe disease course.

Methods

Study design and participants

This retrospective, single-center cohort study involved adult patients who were diagnosed with COVID-19 pneumonia between Jan. 24th and Mar 27th, 2020, in the major government designated hospital in Wuhan: Wuhan No.1 Hospital. The date of the last follow-up was April 8th, 2020. The primary outcome was severity at the end of the study period. All patients were residents of Wuhan, and the diagnostic criteria of COVID-19 were based on the Diagnosis and Treatment Protocol for the 2019 Novel Coronavirus Pneumonia published by the National Health Commission of China.

The newly diagnosed patients were required to meet one of the following conditions: (1) positive signals of COVID-19 nucleic acids were found in the detection of fluorescent real-time RT-PCR; (2) the detection of viral gene sequencing showed a high degree of homology with the new coronavirus COVID-19. Patients with mild symptoms were required to meet the following conditions: (1) history of

epidemiology; (2) fever or other respiratory symptoms; (3) abnormalities of typical CT images of viral pneumonia. Patients with a severe condition met one of the following conditions: (1) shortness of breath, respiratory rate ≥ 30 times/min; (2) oxygen saturation (resting state) $\leq 93\%$; (3) $\text{PaO}_2/\text{FIO}_2 \leq 300 \text{ mm Hg}$. Critically ill patients were required to meet one of the following conditions: (1) respiratory failure, mechanical ventilation required; (2) shock; (3) combined with other organ failure, and ICU monitoring treatment required.

The following data were collected on admission: age, sex, symptoms from onset to hospital admission (fever, cough, dyspnea, myalgia, rhinorrhea, arthralgia, chest pain, headache, and vomiting), comorbidities (cardiovascular disease, chronic pulmonary disease, cerebrovascular disease and chronic neurological disorders, diabetes, malignancy, and smoking), vital signs (heart rate, respiratory rate, and blood pressure), laboratory values on admission (serum hemoglobin concentration, lymphocyte counts, platelet counts, diverse protein markers), treatment regime used for COVID-19 pneumonia (antiviral agents, antibacterial agents, and Chinese medicine), date of symptom onset, admission, virus testing, CT-scan, as well as condition improvement and living status. The study was approved by the Ethics Committee of Wuhan No.1 Hospital (No. 202008).

Treatment Protocol for SARS-CoV-2 Pneumonia

The treatment strategy for patients with COVID-19 pneumonia was based on the guidelines of the world health organization (WHO)¹⁰, which included symptom relief, treatment of underlying diseases, prevention of superimposed bacterial infections, active prevention of complications such as sepsis and ARDS and support organ vital function in a timely fashion. Oxygen supplementation was provided for patients with desaturation by means of high flow oxygen via nasal prong, non-invasive and invasive mechanical ventilation, or extracorporeal membrane oxygenation (ECMO) if required.

Statistical Considerations

The outcome in this study was whether or not the patients experienced a severe to critical course of the disease. This outcome will be denoted "severe vs. mild" in the following. All eligible variables were considered as covariates potentially influencing the outcome in the statistical analysis (supplement section "In-depth description of the statistical analysis flow"). Univariable analysis was performed using logistic regression analysis, where p-values were adjusted for multiple testing by means of the Benjamini-Hochberg procedure. The popular multiple imputation approach MICE¹¹ was employed to

deal with missing values in the multivariable analyses, where 20 imputed data sets were used in each analysis ($m = 20$). The ratio of CRP versus ALB and CRP, as well as WBC and ANC correlated very strongly ($\rho > 0.9$), which is why CRP and ANC were not considered in the multivariable analysis. The latter analysis was performed separately for the young patients and for all patients together 1) using logistic regression in combination with an automatic forward covariate selection procedure based on the Akaike information criterion (AIC)¹² applicable to multiply imputed data¹³ and 2) random forest¹⁴. As a sensitivity analysis¹⁵, the Bayesian information criterion (BIC)¹⁶ that tends to select fewer covariates than the AIC was also considered and backward selection was performed in addition, both, when using the AIC and the BIC. The prediction performance of the models was estimated using 20 times repeated stratified K-fold cross-validation ($K = 3,4,5$), repeating the whole model selection process on each training set in each cross-validation iteration, excluding the corresponding test set¹⁷. Multiple imputation was performed separately on training and test sets¹⁸. As prediction performance measures the AUC and the Brier score were used. Random forest was also used to rank the covariates with respect to their importance for prognosis via the AUC covariate importance values¹⁹ and to estimate the influence forms of the covariates using partial dependence plots (PDPs)²⁰. All statistical analyses were performed using the software R, version 3.6.3. All p-values smaller than 0.05 were considered as statistically significant. For further details and explanations, the interested reader is referred to the detailed description of the statistical analysis flow in the supplement (section "In-depth description of the statistical analysis flow").

Results

Demographic and Clinical Features

The young cohort (age ≤ 60 years old) consisted of 762 patients (median age 47 years, interquartile ranges [IQR] 38 - 55, range 16 - 60; 55.9% female), who were admitted between Jan. 2020 and Mar. 2020 to Wuhan No.1 Hospitals. As shown in Table 1, 400 (52.5%) of the young patients had a mild condition during hospitalization (mild subgroup), while 362 (47.5%) developed a severe or critical disease course (severe subgroup). The mean age was significantly higher in the severe subgroup than in the mild subgroup (59.3 vs. 56.0, Student's t-test: $p < 0.001$). 145 (19.8%) patients were affected by underlying diseases, hypertension (14.3%) being the most common one (Table 1). The median body temperature on admission was 36.6°C (IQR 35.6 - 37.2), no difference of median temperature between both subgroups (mild vs. severe) was seen. The hospitalization in the severe subgroup was significantly longer than that in the mild subgroup (20.0 days [IQR 15.0 - 25.0] vs. 8.0 days [IQR 5.0 - 13.0]). The majority of patients received Chinese traditional medicine (93.4%), antiviral treatments (86.7%), and

antibiotics (75.3%). In this cohort, the most frequently applied oxygen therapy was UOC (78.2%). All key laboratory findings of this cohort are listed in Table 1. Additionally, an elderly cohort of patients (median age 69 years [IQR] 65 - 75, range 61 - 97, 52.8% female) was included, mainly for statistical comparisons with the young cohort.

Potential Risk Factors Associated with Disease Severity in the Young Patients

In the univariable analysis, elevated level of complement C3, SII, CRP, ASS, LDH, the ratio of CRP versus ALB, and the ratio of neutrophile versus lymphocyte were significantly associated with an increased risk for the development of a severe disease course of COVID-19 infection in the young patients (Table 2; Supplementary Table 1 contains the non-significant results in addition). In contrast, decreased levels of ALB and LYM were significantly correlated with the development of a severe disease course. ASS had the largest covariate importance value, both for young and elderly patients. The covariate importance values (Figure 1) and the PDPs (Figure 2, Supplementary Figures 1 to 4) obtained using random forest analysis suggest that complement C3 is only prognostic in young patients.

Generally, the covariate importance values reveal that, while the number of relevant risk factors is larger for the elderly patients, the difference in importance between the most important risk factors and the remaining risk factors is more pronounced for the young patients. Supplementary Figures 5 to 10 show the corresponding variable importance values and PDPs obtained for all patients, irrespective of age. Many of the PDPs indicate complex influence forms of the covariates. While ASS again is associated with the largest importance values here, the importance values of complement C3 are relatively small. The latter confirms that a higher level of complement C3 probably is a risk factor only associated with severity in young patients. After obtaining the latter result, additional analyses were performed in order to investigate, whether the influence of complement C3 is different for specific subgroups of young patients (see supplement section "Subgroup analysis of the influence of complement C3 in young patients" for details). The results of these analyses did not suggest any relevant dependence of the influence of complement C3 in young patients on age, gender, and comorbidities, indicating that this risk factor is relevant for young patients independent of their specific characteristics (Supplement Figures 11-13).

Multivariable Statistical Model-Analysis for Disease Severity

The results of the multivariable analysis using logistic regression in combination with the forward selection algorithm and the AIC criterion showed that the risk for the development of a severe disease

course for COVID-19 in young patients was higher for combinations of elevated levels of complement C3, increase of ASS and SII, and reduced levels of LYM, PLR and UA (Table 3). Applying the backward selection algorithm, UA was not selected in the multivariable model due to lack of importance, but gender, hypertension, thyroid related disease, and BUN were selected instead, indicating potential importance to the risk for development of a severe disease course in the young patients (Supplementary Table 2). However, a sensitivity analysis presented in the supplement (section "Model stability analysis") indicates that the results obtained using the forward selection algorithm (Table 3) may be more statistically stable. When using the BIC criterion instead of the AIC criterion, only complement C3 and ASS were shown to be relevant to the severity in these young patients (Supplementary Table 3).

The results of multivariable logistic regression for all patients (including young and elderly) using the AIC criterion and the forward selection algorithm differed partly from that obtained for the young patients. The risk for the development of a severe disease course in all patients was high for increased levels of complement C3, ASS, BUN, LDH and immunoglobulin G (IgG) and decreased levels of ALB, PLR and IgA (Table 4). Note that the fact that complement C3 was included in the forward selection for all patients is very likely only due to its importance within the young cohort: In the univariable analysis complement C3 was only significant for the young patients and, as seen in Figure 1, the covariate importance value of complement C3 is only large for the young patients. As revealed by the covariate importance values obtained through the random forest analysis (Figure 1), increased levels of ASS seem to be similarly associated with the development of a severe disease course in elderly patients as well as in young patients. When using the backward selection algorithm, LYM and blood platelet were selected in addition, indicating potential relevance associated with disease severity (Supplementary Table 4). Using the BIC criterion together with the forward selection algorithm, ASS, BUN, ALB, and LDH were selected, where the corresponding odds ratios (Supplementary Table 5) were very similar to those obtained in the model obtained using the AIC (Table 4). Applying the BIC criterion together with the forward selection algorithm delivered the same result.

In summary, the estimated prediction performances of the multivariable logistic regression models for all patients were better than those of the models obtained specifically for the young patients (Table 5). The random forest, that considers also non-linear influences, performed best and the model selected using the BIC (that included less covariates) better than that selected using the AIC. Supplementary Table 6 provides an overview on which covariates were selected in each of the models obtained using the AIC and BIC criterion with forward and backward selection.

Discussions

COVID-19, caused by the SARS-CoV-2 virus, is a public health event that poses a serious threat to human health, and the number of COVID-19 cases in young adults is higher than expected: According to the US CDC report almost half of the patients are younger than 65, and young patients have a substantial risk for development of a severe disease course. Facing the rapidly evolving circumstances caused by the COVID-19 pandemic, it is essential to prioritize the medical resources by effectively conducting clinical stratification of COVID-19 young patients. Therefore, it is critical to early identify and investigate potential risk factors for the development of a severe/critical disease course during COVID-19 infection. In order to shed light on potential risk factors associated with a severe disease course in young patients, this study investigated clinical, demographic, treatment, and laboratory data from a group of COVID-19 patients using a set of comprehensive modern statistical methodologies. The univariable analysis suggested that potential risk factors for disease severity in young patients (age $\leq$ 60 years; n=762) are in part different from that in elderly patients (age $>$ 60 years; n=714). Specifically, elevated levels of complement C3 and ASS were only significantly associated with higher risks for a severe disease course in the young patients. In contrast, increased levels of ANC, AST, BUN, CK, CR, D-dimer, Myo-hemoglobin, and PLR, and decreased levels of RBC had a significant influence on this risk only for elderly patients. Even though ASS lacked a statistical significance in the elderly subgroup, the covariate importance values (Figure 1) suggest that ASS is an important risk factor irrespective of patients' age. While complement C4 missed significance in both age groups, the covariate importance values and PDPs suggest that this covariate, complement C3, could be a particularly strong risk factor in the young cohort, while its influence was not significant in elderly patients. Further analysis showed that the significance of complement C3 in young patients was independent of the common risk factors age, gender and the presence versus absence of comorbidities. Risk factors reported previously in COVID-19 infection such as CRP⁷, lactate dehydrogenase³ and decreased lymphocyte⁸ were validated both in young and elderly patient cohorts. Only the D-dimer and procalcitonin did not show significance in the young patient cohort. The PDPs confirmed the observed differences and revealed that the influences of many covariates on the log odds of disease severity of COVID-19 infection are strongly non-linear. However, in logistic regression it is assumed that these influences are linear, which is likely an important reason why the random forests outperformed the multivariable logistic regression models both for the cohort of young patients and for all patients taken together.

The complement family is an important integral component of the innate immune response to viruses, not only protects the body from infectious agents such as viruses and bacteria, but also plays a key role in promoting inflammatory processes triggering inflammatory cytokine storm²¹. The abnormal activation of various innate immune pathways, such as complement system, cytokines and thrombosis pathways, is considered as the driver of acute respiratory distress syndrome (ARDS) and may lead to multi-organ dysfunction^{22,23}. The activation of the complement system also can be found in patients infected with coronaviruses, such as MERs-COV, SARS-CoV-1 and SARS-CoV-2, which develop into ARDS²⁴. When mice with complement C3 deficiency were infected with SARS-COV, the infiltration of neutrophils and inflammatory monocytes in the lungs was significantly reduced, and the levels of cytokines and chemokines in the lungs and serum were decreased, as well as the incidence of respiratory failure. This suggests that the activation of the complement component C3 may aggravate the disease of SARS-COV-related ARDS²³. Young patients have fewer underlying diseases and are more immune-related compensated. This may be the reason why many indicators including D-dimer and procalcitonin may not change significantly even when they progress to severe COVID-19. However, young patients have more active immune function, which is why the serum complement C3 may be a potential indicator of the severity of COVID-19 patients.

This study possesses limitations aside from those inherent to all retrospective cohort studies such as their lack of causal inference. First, it is a single-center study featuring a limited number of cases. Second, the patient data were collected within 3 days after hospital admission, leading to missing data in a number of variables, which were imputed with a standard statistical approach.

In summary, this study conducted a comprehensive statistical analysis with a focus on non-linear relationships to identify risk factors and possible pathogenesis for the development of a severe disease course during COVID-19 infection in young patients. However, large-scale and multi-center analysis is needed to foster the obtained knowledge.

Funding

This study was funded by the Natural Science Foundation of Hubei Province (No. 2019CFB641) and by the German Science Foundation (DFG-Einzelförderung HO6422/1-2 to RH).

Role of the Funder/Sponsor

The funder had no role in any activities of this study aside providing financial support.

Contributors

Acquisition, analysis, or interpretation of data: WC, KX, RH, JL;
 Drafting of the manuscript: WC, RH, JL;
 Statistical analysis: RH, JL;
 Critical revision of the manuscript for important intellectual content: KH, KX;
 Obtained funding: WC, KX, RH;
 Administrative, technical, or material support: WC, KX;
 Conception design: WC, KX, JL;
 Supervision: WC, KX, JL.

Competing interests

All authors declare that there is no conflict of interest to report.

Ethics, consent and permissions

The study was approved by the Ethics Committee of Wuhan No.1 Hospital (No. 202008). All methods were carried out in accordance with relevant guidelines and regulations.

Consent to publish

For this study, informed consent was obtained from all subjects or, if subjects are under 18, from a parent and/or legal guardian.

References

1. World Health Organization. Novel coronavirus (2019-nCoV) situation reports. Situation report - 162. June 30, 2020 https://www.who.int/docs/default-source/coronaviruse/20200630-covid-19-sitrep-162.pdf?sfvrsn=e00a5466_2
2. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet Lond. Engl.* (2020) doi:10.1016/S0140-6736(20)30183-5.
3. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* **0**, (2020).
4. Chen, N. *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* **0**, (2020).
5. Richardson, S. *et al.* Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* (2020) doi:10.1001/jama.2020.6775.
6. US. CDC COVID-19 response team. Severe outcomes among patients with coronavirus disease 2019 (COVID-19) -- United States. <https://www.cdc.gov/mmwr/volumes/69/wr/mm6912e2.htm> (2020).
7. Wu, C. *et al.* Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern. Med.* (2020) doi:10.1001/jamainternmed.2020.0994.
8. Ruan, Q., *et al.* Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med.* (2020) doi:10.1007/s00134-020-05991-x.
9. Liang, W. *et al.* Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. *JAMA Intern. Med.* (2020) doi:10.1001/jamainternmed.2020.2033.
10. Clinical management of severe acute respiratory infection when novel coronavirus (nCoV) infection is suspected. [https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novel-coronavirus-\(ncov\)-infection-is-suspected](https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novel-coronavirus-(ncov)-infection-is-suspected).
11. Buuren, S. van, *et al.* Multivariate Imputation by Chained Equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).
12. Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. in *Selected Papers of Hirotugu Akaike* (eds. Parzen, E., Tanabe, K. & Kitagawa, G.) 199–213 (Springer, 1998). doi:10.1007/978-1-4612-1694-0_15.

13. Wood, A. M., *et al.* How should variable selection be performed with multiply imputed data? *Stat. Med.* **27**, 3227–3246 (2008).
14. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
15. Thabane, L. *et al.* A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med. Res. Methodol.* **13**, 92 (2013).
16. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **6**, 461–464 (1978).
17. Hornung, R. *et al.* A measure of the impact of CV incompleteness on prediction error estimation with application to PCA and normalization. *BMC Med. Res. Methodol.* **15**, (2015).
18. Wahl, S., *et al.* Assessment of predictive performance in incomplete data by combining internal validation and multiple imputation. *BMC Med. Res. Methodol.* **16**, (2016).
19. Janitza, S., Strobl, C. & Boulesteix, A.-L. An AUC-based permutation variable importance measure for random forests. *BMC Bioinformatics* **14**, 119 (2013).
20. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
21. Li, G. *et al.* Coronavirus infections and immune responses. *J. Med. Virol.* **92**, 424–432 (2020).
22. C.M. Campbell, *et al.* Will complement inhibition be the new target in treating COVID-19 related systemic thrombosis? *Circulation* (2020)
23. Ciceri, F. *et al.* Microvascular COVID-19 lung vessels obstructive thromboinflammatory syndrome (MicroCLOTS): an atypical acute respiratory distress syndrome working hypothesis, *Crit. Care Resusc.* (2020)
24. Risitano, A. M. *et al.* Complement as a target in COVID-19? *Nat. Rev. Immunol.* **20**, 343–344 (2020).

Figure Legends

Figure 1: AUC variable importance with respect to predicting the outcome "severe vs. mild" for young and elderly patients calculated using random forests. The larger the importance value of a covariate is, the greater the improvement of prediction performance by including this covariate in prognosis tends to be. Complement C3, C4, LYM, the ratio of CRP versus ALB, and UA seem to influence the prognosis

of the development of severity mostly in young patients. The bars show the medians of the 20 importance values calculated using the 20 imputed data sets from the multiple imputation. The error bars illustrate the variabilities of the importance values: The lower / upper ends show the first / third quartiles of the 20 importance values, that is, 25% percent of the importance values lie below / above these values. To make the raw importance values comparable between young and elderly patients, both for the young and for the elderly patients, the raw importance values were divided by the means of all importance values with positive sign.

Figure 2: Partial dependence plots (PDPs) for young and elderly patients calculated using random forests. In simplified terms, a PDP shows the influence of a covariate on the outcome after adjusting for the influences of the other covariates. The PDPs for the five variables with largest AUC importance values in young patients and that for 'Age' are shown. The light lines show the 20 individual PDPs calculated using the imputed data sets from the multiple imputation. The bold lines show averages over the 20 individual PDPs.

	All patients	Young patients (<= 60 years)	Elderly patients (> 60 years)	p-value	
Age	60.00 (46.00-69.00)	47.00 (38.00-55.00)	69.00 (65.00-75.00)		
Gender: female				0.2500	
Yes	803 (54.4)	426 (55.9)	377 (52.8)		
No	673 (45.6)	336 (44.1)	337 (47.2)		
Mild patients	718(48.6)	400(55.7)	318(44.3)		
Severe patients	758(51.4)	362(47.8)	396(52.2)		
Underlying diseases					
Coronary heart disease				<0.0001	
Yes	124 (8.4)	19 (2.5)	105 (14.7)		
No	1352 (91.6)	743 (97.5)	609 (85.3)		
Diabetes				<0.0001	
Yes	200 (13.6)	51 (6.7)	149 (20.9)		
No	1276 (86.4)	711 (93.3)	565 (79.1)		
Hypertension				<0.0001	
Yes	441 (29.9)	109 (14.3)	332 (46.5)		
No	1035 (70.1)	653 (85.7)	382 (53.5)		
Thyroid related diseases				0.4751	
Yes	32 (2.2)	19 (2.5)	13 (1.8)		
No	1444 (97.8)	743 (97.5)	701 (98.2)		
Initial symptoms					
Fever	973(65.9)	529(54.4)	444(45.6)		
cough	932(63.1)	480(51.5)	452(48.5)		
dyspnea	519(35.2)	285(54.9)	234(45.1)		
Therapy					
Antiviral treatments	1232(83.5)	658(53.4)	574(46.6)		
Chinese traditional medicine	1150(77.9)	618(53.7)	532(46.3)		
Antibiotics	1071(72.6)	543(50.7)	528(49.3)		

Oxygen therapy	1199(81.2)	596(49.7)	603(50.3)		
Accu Troponin	0.0050 (0.0020-0.0120)	0.0020 (0.0010-0.0050)	0.0080 (0.0030-0.0190)	<0.0001	
Key laboratory findings					
ALB	35.90 (32.50-39.00)	37.80 (34.90-40.20)	33.70 (30.70-36.60)	<0.0001	
Complement C3	1.07 (0.92-1.23)	1.12 (0.98-1.26)	0.98 (0.88-1.18)	0.0076	
Complement C4	0.27 (0.21-0.34)	0.29 (0.23-0.35)	0.24 (0.18-0.32)	0.0185	
ALT	22.00 (14.00-35.00)	22.00 (14.00-37.00)	22.00 (15.00-33.00)	0.5136	
ANC	3.30 (2.41-4.43)	3.07 (2.25-3.94)	3.62 (2.71-5.02)	<0.0001	
ASS	58.40 (6.20-133.35)	38.75 (5.55-126.50)	76.55 (10.57-139.35)	0.0874	
AST	24.00 (19.00-34.00)	23.00 (18.00-32.00)	25.50 (20.00-36.00)	<0.0001	
Blood glucose	5.40 (4.90-6.50)	5.10 (4.70-5.80)	5.90 (5.00-7.75)	<0.0001	
Blood platelet	218.00 (170.00-280.75)	216.00 (171.00-275.00)	219.00 (168.00-286.00)	0.3292	
BUN	4.10 (3.30-5.30)	3.80 (3.10-4.50)	4.70 (3.70-6.40)	<0.0001	
CK	58.00 (40.00-90.00)	57.00 (40.00-86.00)	59.00 (41.00-97.00)	0.3034	
CK MB	7.00 (6.00-9.00)	7.00 (6.00-9.00)	8.00 (6.00-10.00)	0.0075	
Cr	62.00 (53.00-76.00)	60.00 (52.00-73.00)	65.00 (55.00-81.25)	<0.0001	
CRP	3.87 (3.00-21.80)	3.00 (3.00-12.10)	7.59 (3.00-39.10)	<0.0001	
D Dimer	0.47 (0.23-1.10)	0.29 (0.18-0.55)	0.76 (0.37-2.02)	<0.0001	
Erythrocyte sedimentation rate	27.00 (15.00-48.00)	21.00 (12.00-36.00)	36.00 (19.00-59.00)	<0.0001	
Hemoglobin	128.00 (118.00-139.00)	132.00 (123.00-143.00)	124.00 (114.00-135.00)	<0.0001	
Immunoglobulin IgA	2.23 (1.61-2.82)	2.09 (1.58-2.62)	2.30 (1.78-3.06)	0.0761	
Immunoglobulin IgG	9.84 (8.21-12.00)	9.88 (8.17-11.83)	9.84 (8.23-12.50)	0.6426	
Immunoglobulin IgM	1.00 (0.79-1.37)	1.04 (0.81-1.37)	0.97 (0.74-1.33)	0.3179	
LDH	194.00 (158.00-262.00)	178.00 (150.00-	216.00 (170.50-	<0.0001	

		231.00)	300.50)		
LYM	1.45 (1.01-1.92)	1.58 (1.15-2.00)	1.34 (0.91-1.74)	<0.0001	
MONO	0.52 (0.40-0.68)	0.51 (0.39-0.66)	0.54 (0.41-0.70)	0.0045	
Myohemoglobin	37.15 (26.20-65.92)	27.60 (22.60-33.90)	49.20 (33.30-91.10)	<0.0001	
PLR	150.44 (111.06-213.47)	137.41 (104.98-191.12)	169.93 (119.89-242.33)	<0.0001	
Procalcitonin	0.05 (0.05-0.05)	0.05 (0.05-0.05)	0.05 (0.05-0.06)	<0.0001	
RBC	4.20 (3.87-4.58)	4.36 (4.02-4.67)	4.10 (3.72-4.41)	<0.0001	
SII	481.19 (308.34-842.83)	389.75 (271.34-653.01)	590.55 (367.77-1069.11)	<0.0001	
UA	286.00 (233.00-355.00)	285.00 (233.00-348.00)	287.50 (233.00-366.00)	0.5587	
WBC	5.58 (4.47-7.04)	5.38 (4.37-6.66)	5.86 (4.68-7.60)	<0.0001	
Ratios					
Ratio of CRP versus ALB	0.09 (0.08-0.50)	0.08 (0.07-0.25)	0.15 (0.08-1.00)	<0.0001	
Ratio of neutrophile versus lymphocyte	2.19 (1.48-3.44)	1.88 (1.35-2.66)	2.68 (1.77-4.54)	<0.0001	
Ratio of CRP versus ALB	0.09 (0.08-0.50)	0.08 (0.07-0.25)	0.15 (0.08-1.00)	<0.0001	
CT-scan					
Pulmonary consolidation				0.7776	
Yes	52 (3.7)	26 (3.5)	26 (3.9)		
No	1359 (96.3)	716 (96.5)	643 (96.1)		
Ground glass opacity				0.8334	
Yes	247 (17.5)	128 (17.3)	119 (17.8)		
No	1164 (82.5)	614 (82.7)	550 (82.2)		
Multiple patchy shadows				0.0219	
Yes	1176 (83.3)	602 (81.1)	574 (85.8)		
No	235 (16.7)	140 (18.9)	95 (14.2)		

Table 1 : Clinical and demographic characteristics, treatment and key laboratory findings. Metric covariates are reported as medians with interquartile ranges in the following form: 'Median (First Quartile-Third Quartile)'. Categorical covariates are reported as percentages in the following form: 'Absolute number (Percentage)'. Differences between the young and elderly patients were tested using the Wilcoxon test and Fisher's exact test for metric covariates and categorical covariates, respectively.

variable	All patients		Young patients (<= 60 years)		Elderly patients (> 60 years)	
	Odds ratio [CI]	adj. p	Odds ratio [CI]	adj. p	Odds ratio [CI]	adj. p
ALB	0.9320 [0.9109, 0.9530]	<0.0001	0.9501 [0.9170, 0.9831]	0.0317	0.9211 [0.8900, 0.9521]	<0.0001
ANC	1.1055 [1.0549, 1.1611]	0.0002	1.0794 [0.9937, 1.1773]	0.2062	1.1032 [1.0408, 1.1741]	0.0062
ASS	1.0065 [1.0032, 1.0100]	0.0006	1.0066 [1.0023, 1.0112]	0.0317	1.0059 [1.0008, 1.0114]	0.0663
AST	1.0100 [1.0048, 1.0158]	0.0011	1.0062 [1.0005, 1.0130]	0.1762	1.0153 [1.0066, 1.0251]	0.0059
BUN	1.1116 [1.0658, 1.1661]	<0.0001	1.0461 [0.9923, 1.1270]	0.3516	1.1324 [1.0730, 1.2051]	0.0002
CK	1.0023 [1.0011, 1.0036]	0.0011	1.0021 [1.0005, 1.0042]	0.0926	1.0023 [1.0007, 1.0041]	0.0196
Complement C3	2.0055 [0.6022, 6.9684]	0.3699	15.5808 [2.4111, 122.2841]	0.0392	0.3323 [0.0478, 1.9321]	0.3700
Cr	1.0076 [1.0035, 1.0122]	0.0017	1.0046 [0.9979, 1.0121]	0.3963	1.0083 [1.0030, 1.0143]	0.0114
CRP	1.0119 [1.0079, 1.0161]	<0.0001	1.0122 [1.0043, 1.0209]	0.0317	1.0110 [1.0064, 1.0161]	0.0001
D-Dimer	1.0808 [1.0386, 1.1358]	0.0017	1.0580 [0.9368, 1.2306]	0.6455	1.0770 [1.0319, 1.1374]	0.0081
LDH	1.0027 [1.0018, 1.0038]	<0.0001	1.0026 [1.0011, 1.0043]	0.0317	1.0027 [1.0015, 1.0040]	0.0002
LYM	0.5842 [0.4926, 0.6908]	<0.0001	0.5947 [0.4664, 0.7536]	0.0009	0.6016 [0.4698, 0.7662]	0.0003
Myohemoglobin	1.0054 [1.0027, 1.0091]	0.0024	1.0022 [0.9918, 1.0134]	0.7994	1.0055 [1.0025, 1.0096]	0.0081
PLR	1.0015 [1.0006,	0.0040	1.0008 [0.9996,	0.3963	1.0017 [1.0005,	0.0196

		1.0024]		1.0022]		1.0031]	
Ratio of CRP	1.3836	<0.0001	1.3994	0.0430	1.3393	0.0002	
versus ALB	[1.2341,		[1.0999,		[1.1759,		
	1.5667]		1.8321]		1.5467]		
Ratio of	1.1061	<0.0001	1.0835	0.0416	1.1105	0.0001	
neutrophile versus	[1.0687,		[1.0256,		[1.0629,		
lymphocyte	1.1485]		1.1539]		1.1666]		
RBC	0.7775	0.0140	0.9856	0.9337	0.6592	0.0081	
	[0.6493,		[0.7644,		[0.5018,		
	0.9287]		1.2707]		0.8600]		
SII	1.0003	<0.0001	1.0003	0.0430	1.0003	0.0063	
	[1.0002,		[1.0001,		[1.0001,		
	1.0004]		1.0006]		1.0004]		

Table 2: Univariable logistic regression for the outcome "severe vs. mild". The p-values were adjusted for multiple testing separately for the analysis of all patients, young patients, and elderly patients. Increased levels of complement C3 and ASS were associated with an increased risk for severity only in the young patient cohort. Supplementary Table 1 contains the results for all covariates.

	Regression coefficient	Odds ratio
Intercept	-0.604671	-
Complement C3	1.621768	5.0620
ASS	0.005154	1.0052
LYM	-0.477558	0.6203
PLR	-0.003868	0.9961
SII	0.000401	1.0004
UA	-0.001553	0.9984

Table 3: Multivariable logistic regression models for the outcome "severe vs. mild" in young patients selected using the AIC criterion and forward selection.

	Regression coefficient	Odds ratio
Intercept	0.191987	-
Complement C3	0.506456	1.6594
Immunoglobulin IgA	-0.134694	0.8740
Immunoglobulin IgG	0.035108	1.0357
ALB	-0.044892	0.9561
ASS	0.00574	1.0058
BUN	0.06745	1.0698
LDH	0.001668	1.0017
PLR	-0.000795	0.9992

Table 4: Multivariable logistic regression models for the outcome "severe vs. mild" in all patients irrespective of age selected using the AIC criterion and forward selection.

	K in cross-validation	AUC			Brier score		
		Logistic regression - AIC	Logistic regression - BIC	Random forest	Logistic regression - AIC	Logistic regression - BIC	Random forest
Young patients	3	0.5512	0.5542	0.5962	0.2740	0.2584	0.2438
	4	0.5405	0.5454	0.5847	0.2764	0.2606	0.2462
	5	0.5327	0.5344	0.5838	0.2772	0.2640	0.2461
All patients	3	0.6069	0.6186	0.6310	0.2502	0.2425	0.2369
	4	0.6027	0.6140	0.6271	0.2525	0.2444	0.2378
	5	0.6059	0.6156	0.6276	0.2500	0.2433	0.2378

Table 5: Performances of the models measured using stratified K-fold cross-validation. For each value of K, the stratified K-fold cross-validation was repeated twenty times and the results averaged. Higher values of the AUC and smaller values of the Brier score are preferable.