PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A Clinical Risk Score to Predict in-hospital Mortality in Critically Ill Patients with COVID-19: A Retrospective Cohort Study.
AUTHORS	AlKaabi, Salem; Alnuaimi, Asma; Harbi, Mariam Al; Amari, Mohammed; Ganapathy, Rajiv; Iqbal, Imran; Nauman, Javaid; Oulhaj , Abderrahim

VERSION 1 – REVIEW

REVIEWER	Altschul, David Montefiore Medical Center, Neurosurgery
REVIEW RETURNED	16-Jan-2021

GENERAL COMMENTS	The authors conduct a multi-center retrospective analysis of patients critically ill with COVID-19 attempting to create a risk score to predict mortality. The authors manuscript is well written, and seemingly very helpful. I have some questions regarding the methods. It is unclear when using the laboratory data whether this was their admission emergency data or labs once admitted to the ICU? Second, to my knowledge when creating a risk score the optimal approach is to have an observation or derivation cohort from which the score is generated and then a separate validation cohort from which the score can be tested against for accuracy and reliability. It is unclear from their "Validation of the risk prediction score" paragraphs if this was properly addressed. It appears to me they try to statistically explain away using the same dataset for both derivation and validation. I believe this paper would be much stronger if they validated this score using a different dataset. My third issue with the manuscript is the assumption that patients that were discharged from the hospital all survived. This assumption is fraught with potential bias and score failure. I believe the easiest approach to this issue is to reword the title and add inpatient mortality instead of just mortality, as this is what this score is truly predicting. Fourth their definition of ICU is very soft including patients that required any kind of oxygen therapy. In our experience this would include a majority of COVID patients admitted to the hospital. Last, in their conclusion they purport that this is the first score to predict mortality for COVID-19 in ICU patients. This is not entirely true especially given their soft definition of ICU patients. See references: 1. Altschul et al. A novel severity score to predict inpatient mortality in COVID-19 patients. Sci Rep. 2020 Oct 7;10(1):16726. doi: 10.1038/s41598-020-73962-9. 2. Knight et al. Risk stratification of patients admitted to hospital with

	covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score BMJ. 2020 Sep 9;370:m3339. doi: 10.1136/bmj.m3339. 3. Zhao et al. Prediction model and risk scores of ICU admission and mortality in COVID-19 PLoS One. 2020 Jul 30;15(7):e0236618. doi: 10.1371/journal.pone.0236618. eCollection 2020. The authors should discuss other scores utilized to predict mortality, how theirs differs and how their score adds to the published literature. In summary I think having an online prediction would be extremely useful to ICUs and hospitals. Some further clarification regarding the biomarkers and variables used and at what point during their admission they were obtained would be helpful, and expanding the discussion to include other covid severity scores that have been validated and published would also be helpful.

REVIEWER	Fumagalli, Carlo University of Florence, Department of Experimental and Clinical Medicine
REVIEW RETURNED	04-Feb-2021

GENERAL COMMENTS	In this multicenter study, Dr. Salem AlKaabi and Colleagues identify factors associated with COVID-19 related mortality in ICU patients and develop a clinical score to predict mortality in critically ill patients. Among several factors associated with the outcome, age, neutrophil percentage, lactate dehydrogenase, respiratory rate, creatinine, Glasgow Coma Scale, and oxygen saturation are included in the score. This is a very interesting and well written manuscript, offering an important epidemiological view of COVID-19 from Abu Dhabi, United Arab Emirates. Please find my comments below. Major comments -This is a multicenter study (Methods, Study design and Data sources paragraph). It would be very interesting to learn about the derivation of the cohort (e.g. number of centers or ICU/HDU that participated to the study; number of hospital beds available, to provide more information regarding hospital volumes during the pandemic). Furthermore, were there any exclusion criteria at the moment of data revision and inclusion in the study? How were missing data handled at the moment of data extraction? -Do the Authors have information regarding PaO2/FiO2 on admission for all patients? -The Introduction section presents several studies and evidence from previous reports on prevalence and predictors of mortality even for critical patients with COVID-19 in ICU. I wonder if some parts of the Introduction could be more suitable/useful for the Discussion section and if the Authors could explore more on the potential regional differences in the epidemiology of COVID-19 (vs European/US/Asian Cohorts). -Moreover, the Discussion section could be further expanded with the Authors' insights on the relevance of the score and potential implications in the ICU. Authors could speculate more on how ICUs (especially under strain) could benefit from the adoption of dedicated

	tools. -Table 1 describes the general characteristics of the study population, which appears to be younger than other ICU published cohorts. I would refer to this in the Discussion and in the limitations paragraph since it may influence external applicability on foreign cohorts. Minor comments -It would be helpful to include the study design in the title. -Table 1. The section dedicated to comorbidities occasionally is confusing, especially for counts and %s of certain items (e.g. respiratory diseases) where there is a mismatch (typos?).

REVIEWER	Schwab, Camille Hôpital Saint-Antoine, Pharmacie
REVIEW RETURNED	09-Mar-2021

GENERAL COMMENTS	Thank you for the opportunity to review this paper. The methodology and the statistics used are appropriate but they are not described sufficiently. Thus, I have some remarks and questions: In the introduction section: The authors wrote: “to the best of our knowledge, no study has yet reported the prediction of death in patients admitted to ICU with confirmed COVID-19.” I don’t agree with the authors as Pan Pan et al., J Med Internet Res, Nov 2020, Ping Zhan et al. Clin Nutr, Feb 2021 and Xiaonan Li et al. Peer J, Nov 2020 have published about the prediction of death in patients admitted to ICU with confirmed COVID-19. In the methods section: The authors have not explicited if it was a prospective or a retrospective study. The inclusion criteria are not clear enough: patients admitted to ICU with confirmed COVID-19 between March, 1st 2020 and July, 22nd 2020. ICU being regular or related ICU. But what about “patients with available data on important clinical characteristics […] were included”. Does this mean that patient without data were excluded? I don’t understand the paragraph: “Patient and Public Involvement”. How can the public be involved in the design of a study? The authors detailed the statistical analysis prior to enumerate the potential predictive variables. For a better understanding of the study, I think that the paragraph “Potential Predictive Variables” should be placed at the top of the paragraph “Statistical analysis”. Furthermore, on what criteria were the potential predictive variables chosen? At the end of the methods section, the authors present their easy-to-use web-based risk calculator. Will the data given by the users be used to conduct an external validation of the score? What is the aim to present this computer tool in the method section as it’s not one of the objective. This article does not deal with the construction of the computer tool but with the risk score. Why did the authors not conduct a sensitivity analysis? Contrary to what is written on the STROBE Statement, the authors have not explain how the study size was arrived at and how missing data were addressed. To finish, the authors used the STROBE statement. However, as this study presents the development of a clinical risk score, they should have used the TRIPOD statement. In the results section:

	The authors wrote “among 221 women patients, 28 (12.7%) have died” but this percentage is different from the one in the table 1 (which is 14.3%) and “among 1321 men, 168 (12.7%) died” but the correct percentage is 85.7%. Table 1: the sum of the percentage of the ethnicities is not equal to 100 (24.2+73.3+1.8 = 99.3) for each of the 3 columns. The same for the variables “Oxygen therapy”, column 1, “Coexisting conditions”, column 2, Numbers are written on the line of the variable coexisting conditions (n (%)): 0.29 (±0.45) 0.38 (±0.49) 0.27 (±0.45), but they seem to be not percentage but means +/- sd. What are these numbers? The authors wrote that “Continuous variables were compared using the t-test or Wilcoxon-rank-sum test, while categorical variables were compared using the Chi-square or Fisher’s exact test.” But they have not detailed which variables were compared using Wilcoxon-rank-sum test. A p-value = 0.000 is commonly written <0.001. Table 3: for the Glasgow Coma Scale, is it the quantitative or the qualitative variable as both were presented in table 1? The authors wrote: “In case of continuous risk factors, we created a binary based on the median split”. Why was this split not made from the bivariate analyzes? Which allow to avoid the utilization of variables that do not follow a normal distribution.

REVIEWER	Sarker, Shah-Jalal University College London, Research Department of Medical Education
REVIEW RETURNED	13-Apr-2021

GENERAL COMMENTS	The authors presented a very useful timely article to identify risk factors and to develop risk prediction score for mortality of COVID-19 patients at ICU. The authors used the latest statistical techniques appropriately and the article will be very useful for readers. I recommend publishing it with minor revision as follows: 1. The article used data from March 2020 to July 2020 and the outcome time was from the date of ICU admission until the date of death. However, the risk prediction model is for 28-day mortality. This needs to be made clear in the article. If appropriate, 28-day (or 30-day) mortality needs to be added to the title. 2. Waterlow score is a well-known predictor for 30-day mortality (see, “Can Waterlow score predict 30-day mortality and length of stay in acutely admitted medical patients (aged ≥65 years)? Evidence from a single centre prospective cohort study”). Waterlow score data includes age, nutritional status, weight, mobility, gender, smoking status, comorbidities, use of medication and continence. Glasgow comma score has covered some of the severity factors including continence. However, a reference and some discussion need to be included based on Waterlow and other important predictors. 3. On page 12, “….The risk prediction score shows high accuracy in terms or discrimination (AUC = 88.1) and calibration (Bier score = 8.11)”. The word “or” should be “of”. 4. BMI or obesity is an important risk factor that should have been included in the model. 5. The article found no comorbidity but mostly laboratory findings

	were significant! Whereas, similar US studies (Ref 8) found significant comorbidities. These need to be discussed how these laboratory findings are likely to be related to the significant comorbidities found in other studies. 6. Possible interaction effect among risk factors need to be tried and included in the model? For example, age and GCS. 7. The last paragraph of the discussion is repeated in the conclusion. The conclusion can be much shorter.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 comments

Dr. David Altschul, Montefiore Medical Center

Comment 1. *The authors conduct a multi-center retrospective analysis of patients critically ill with COVID-19 attempting to create a risk score to predict mortality. The authors manuscript is well written, and seemingly very helpful.*
I have some questions regarding the methods. It is unclear when using the laboratory data whether this was their admission emergency data or labs once admitted to the ICU?

Our answer. We thank the reviewer for his kind words regarding the quality and the importance of the paper. We mentioned in section “potential predictive variables’ line 122-123 that all variables were measured at the time of ICU admission.

Comment 2. *Second, to my knowledge when creating a risk score the optimal approach is to have an observation or derivation cohort from which the score is generated and then a separate validation cohort from which the score can be tested against for accuracy and reliability. It is unclear from their "Validation of the risk prediction score" paragraphs if this was properly addressed. It appears to me they try to statistically explain away using the same dataset for both derivation and validation. I believe this paper would be much stronger if they validated this score using a different dataset.*

Our answer. There are two type of model validations indeed: Internal validation and External validation. The reviewer is referring in this comment to the external validation for which one needs an additional cohort to be used as the validation cohort. We have already acknowledged in the manuscript (section “Strengths and limitations of this study” line 60-61 and in the discussion section, line 249-252) that one limitation of this paper is the generalizability of risk prediction score in other settings, and that external validation of this risk prediction score in other populations is the next step. Hence, we agree with the reveiwer that this paper would be much stronger if we were able to validate this score using a different dataset. Unfortunately we don’t have this validation data set, however, we carried out internal validation using bootstrapping. We used the .632+ bootstrap method described in Tibshirani et al (Efron, Tibshirani (1997) Journal of the American Statistical Association 92, 548–560 Improvement On Cross-Validation: The .632+ Bootstrap Method).

Comment 3. *My third issue with the manuscript is the assumption that patients that were discharged from the hospital all survived. This assumption is fraught with potential bias and score failure. I believe the easiest approach to this issue is to reword the title and add inpatient mortality instead of just mortality, as this is what this score is truly predicting.*

Our answer. We thank the reviewer for this very useful suggestion. We just wanted to emphasize the fact that being discharged alive from the hospital is a competing risk to mortality. The suggestion of

the reviewer is more accurate. We have reformulated the title of the paper as follows: “A Clinical Risk Score to Predict in-hospital Mortality in Critically Ill Patients with COVID-19: A Retrospective Cohort Study”.

Comment 4. *Fourth their definition of ICU is very soft including patients that required any kind of oxygen therapy. In our experience this would include a majority of COVID patients admitted to the hospital.*

Our answer. As it can be seen in Table 1 of the manuscript, 86.3% of the patients admitted to ICU had hypoxic respiratory failure requiring either supplemental oxygen, none-invasive mechanical ventilation or invasive mechanical ventilation. These figures are almost similar to the ones published in Table 1 in Gupta et al. JAMA Internal Medicine (2020). In their paper, they reported 87.1% of patients admitted to ICU to be on any form of oxygen therapy. Among their patients 67.4% were receiving invasive mechanical ventilatory support, 1.2% with noninvasive mechanical ventilation and 18.5% were receiving High-flow nasal cannula or nonrebreather mask. We agree that in some other countries or settings, patients admitted to ICU would all be on severe hypoxia requiring at least high-flow oxygen therapy.

Comment 5. *Last, in their conclusion they purport that this is the first score to predict mortality for COVID-19 in ICU patients. This is not entirely true especially given their soft definition of ICU patients. See references: 1. Altschul et al. A novel severity score to predict inpatient mortality in COVID-19 patients. Sci Rep. 2020 Oct 7;10(1):16726. doi: 10.1038/s41598-020-73962-9.*

2. Knight et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score BMJ. 2020 Sep 9;370:m3339. doi: 10.1136/bmj.m3339.

3. Zhao et al. Prediction model and risk scores of ICU admission and mortality in COVID-19 PLoS One. 2020 Jul 30;15(7):e0236618. doi: 10.1371/journal.pone.0236618. eCollection 2020.

The authors should discuss other scores utilized to predict mortality, how theirs differs and how their score adds to the published literature.

Our answer. We have now updated our reference list with the publications available so far, and have reformulated both the Introduction and Discussion sections (see line 69-97; 255-304).

Comment 6. *In summary I think having an online prediction would be extremely useful to ICUs and hospitals. Some further clarification regarding the biomarkers and variables used and at what point during their admission they were obtained would be helpful, and expanding the discussion to include other covid severity scores that have been validated and published would also be helpful.*

Our answer. We have expanded our discussion including the biomarkers/variables and discussing already published scores for COVID-19 inpatient mortality (see line 255-304).

Reviewer 2 comments

Dr. Carlo Fumagalli, University of Florence

Comment 1. *In this multicenter study, Dr. Salem AlKaabi and Colleagues identify factors associated with COVID-19 related mortality in ICU patients and develop a clinical score to predict mortality in critically ill patients. Among several factors associated with the outcome, age, neutrophil percentage,*

lactate dehydrogenase, respiratory rate, creatinine, Glasgow Coma Scale, and oxygen saturation are included in the score.
This is a very interesting and well written manuscript, offering an important epidemiological view of COVID-19 from Abu Dhabi, United Arab Emirates. Please find my comments below.

This is a multicenter study (Methods, Study design and Data sources paragraph). It would be very interesting to learn about the derivation of the cohort (e.g. number of centers or ICU/HDU that participated to the study; number of hospital beds available, to provide more information regarding hospital volumes during the pandemic).

Our answer. We thank the reviewer for his kind words regarding the quality and the importance of the paper. The data was collected from four major hospitals as well as newly developed field hospitals operating with some ICU bed capacity during the first wave of the COVID-19 between March, 1st 2020 and July, 22nd 2020. The estimated bed capacity for ICU and/or HDU was around 550 across the Emirate of Abu Dhabi.

We have added this information in the methods section (see line100-105).

Comment 2. *Furthermore, were there any exclusion criteria at the moment of data revision and inclusion in the study? How were missing data handled at the moment of data extraction?*

Our answer. The data set was extracted restrospectively from electronic medical records. Patients who were admitted to an ICU or HDU had their data extracted. Patients who were not admitted to ICU/HDU and who did not have any form of oxygen therapy were excluded. So the criteria for inclusion and exclusion were admission to ICU/HDU and oxygen therapy status. These two variables were not missing when the data was extracted.

Comment 3. *Do the Authors have information regarding PaO2/FiO2 on admission for all patients?*

Our answer. The laboratory and clinical data used in this analysis are measured at the time of admission to ICU. We were, unfortunately, not able to retrieve the levels of the ratio of PaO2/FiO2 for many of our patients.

Comment 4. *The Introduction section presents several studies and evidence from previous reports on prevalence and predictors of mortality even for critical patients with COVID-19 in ICU. I wonder if some parts of the Introduction could be more suitable/useful for the Discussion section and if the Authors could explore more on the potential regional differences in the epidemiology of COVID-19 (vs European/US/Asian Cohorts).*

Our answer. We have reformulated our Introduction section adding new publications related to the prognostic scores of COVID-19, and as suggested, moved some parts of Introduction to Discussion section. We have also updated epidemiology of COVID-19 in different regions as per the most recent WHO weekly update (4 May 2021). (line 305-316)

Comment 5. *Moreover, the Discussion section could be further expanded with the Authors' insights on the relevance of the score and potential implications in the ICU. Authors could speculate more on how ICUs (especially under strain) could benefit from the adoption of dedicated tools.*

Our answer. We have now reformulated the Discussion and the conclusion section to expand more on the relevance of the score and the potential implications in the ICU (line 318-327)

Comment 6. *Table 1 describes the general characteristics of the study population, which appears to be younger than other ICU published cohorts. I would refer to this in the Discussion and in the limitations paragraph since it may influence external applicability on foreign cohorts.*

Our answer. We have added the following sentence in the limitation (see line 252-254).

“Further, the participants included in this study were younger compared with other studies using the data at the time of hospital admission (Altschul DJ et al. Sci Rep 2020; 10(1): 16726, Zhao Z et al. PLoS One 2020;15(7): e0236618, Gupta S et al. JAMA Intern Med 2020, Pan P et al. J Med Internet Res 2020;22(11):e23128, Zhang P et al. Clin Nutr 2021;40(2):534-41), which may in turn limit the generalizability in older patients”

We also have included the following sentence in the Discussion (see line 280-283).

“Another plausible reason for this difference in the results is the younger age of the participants in our study compared to other studies (Altschul DJ et al. Sci Rep 2020; 10(1): 16726, Zhao Z et al. PLoS One 2020;15(7): e0236618, Gupta S et al. JAMA Intern Med 2020, Pan P et al. J Med Internet Res 2020;22(11):e23128, Zhang P et al. Clin Nutr 2021;40(2):534-41), which could likely influence the clinical features to be included in the risk prediction score.”

Comment 7. *Minor comments. It would be helpful to include the study design in the title.*

Our answer. The title of the manuscript has been updated accordingly. The new title is: “A Clinical Risk Score to Predict in-hospital Mortality in Critically Ill Patients with COVID-19: A Retrospective Cohort Study”

Comment 8. *Minor comments. Table 1. The section dedicated to comorbidities occasionally is confusing, especially for counts and %s of certain items (e.g. respiratory diseases) where there is a mismatch (typos?)..*

Our answer. The n (%) represents the number and percentage of cases. The percentage in Table 1 are column percentages. There is a typo for the total number of cases with respiratory diseases. The correct number is 181 instead of 81. We have fixed this issue in the main manuscript. We also removed the misplaced presentation in Coexisting conditions in Table 1.

Reviewer 3 comments

Dr. Camille Schwab, Hôpital Saint-Antoine, Institut Pierre Louis d'Epidemiologie et de Sante Publique

Comment 1. *Thank you for the opportunity to review this paper. The methodology and the statistics used are appropriate but they are not described sufficiently. Thus, I have some remarks and questions:*

Our answer. We thank the reviewer for her kind words.

Comment 2. *In the introduction section:The authors wrote: “to the best of our knowledge, no study has yet reported the prediction of death in patients admitted to ICU with confirmed COVID-19.” I don’t agree with the authors as Pan Pan et al., J Med Internet Res, Nov 2020, Ping Zhan et al. Clin Nutr, Feb 2021 and Xiaonan Li et al. Peer J, Nov 2020 have published about the prediction of death in patients admitted to ICU with confirmed COVID-19.*

Our answer. At the time of manuscript writing, some of these publications were not available. We have now updated our reference list with the publications available so far, and have reformulated both the Introduction and Discussion sections (see line 69-97; 255-304).

Comment 3. *In the methods section:*

The authors have not explicited if it was a prospective or a retrospective study.

The inclusion criteria are not clear enough: patients admitted to ICU with confirmed COVID-19 between March, 1st 2020 and July, 22nd 2020. ICU being regular or related ICU. But what about "patients with available data on important clinical characteristics […] were included". Does this mean that patient without data were excluded?

Our answer. This is a multicenter retrospective observational study. The design of the study is now reflected in the manuscript's title as: "A Clinical Risk Score to Predict in-hospital Mortality in Critically Ill Patients with COVID-19: A Retrospective Cohort Study"

The inclusion criteria for this study is: "All laboratory confirmed COVID-19 patients admitted to ICUs in the Emirate of Abu Dhabi, United Arab Emirates between March, 1st 2020 and July, 22nd 2020". The sentence about the availability of the data is not an inclusion or exclusion criteria. Patients with heavy missing data were not included in the final analysis. We preferred to use complete case analysis instead of using multiple imputations.

The inclusion criteria is now made clear in the main text. We removed the sentence on data availability from the section "Study design and Data sources" and we placed it, with different wording, in statistical analyses section (see line 134-135).

Comment 4. *I don't understand the paragraph: "Patient and Public Involvement". How can the public be involved in the design of a study?*

Our answer. This is a compulsory section for all research articles submitted to BMJ Open. We've been instructed by the editor not to remove it.

Comment 5. *The authors detailed the statistical analysis prior to enumerate the potential predictive variables. For a better understanding of the study, I think that the paragraph "Potential Predictive Variables" should be placed at the top of the paragraph "Statistical analysis". Furthermore, on what criteria were the potential predictive variables chosen?.*

Our answer. We followed the recommendation of the reviewer. The paragraph "Potential Predictive Variables" is now placed at the top of the section "Statistical analysis".

About the criteria for choosing the potential predictive variables, we tried to include all risk factors that are known to be associated with the severity of COVID-19 and eventually with mortality due to COVID-19. These include demographics, clinical and past medical history. For the laboratory variables, we included all the routine labs.

Comment 6. *At the end of the methods section, the authors present their easy-to-use web-based risk calculator. Will the data given by the users be used to conduct an external validation of the score? What is the aim to present this computer tool in the method section as it's not one of the objective.This article does not deal with the construction of the computer tool but with the risk score. Why did the authors not conduct a sensitivity analysis?*

Our answer. We agree with the reviewer. The paragraph on the risk calculator is not fit for the method section. We removed it from the methods section and placed it at the end of the result's section (see line 227-233).

For the first part of the comment, the data given by the user might be, after obtaining consent, used for external validation indeed later on. We are currently in the process of obtaining required permission to implement this app into the CERNER (the electronic medical records system used in all

public hospitals in the UAE). Implementing the app into electronic medical records will provide us with a unique opportunity to externally validate our derived risk score.

Comment 7. *Contrary to what is written on the STROBE Statement, the authors have not explain how the study size was arrived at and how missing data were addressed.*

Our answer. As we explained earlier, in response to comment 3, all laboratory confirmed COVID-19 patients admitted to ICUs between March, 1st 2020 and July, 22nd 2020 were eligible for inclusion in this study. However, patients with missing data on the 36 potential predictive variables and/or on the main outcome were not included in the final analysis. We preferred to use complete case analysis instead of using multiple imputations especially with the LASSO procedure.

Comment 9. *To finish, the authors used the STROBE statement. However, as this study presents the development of a clinical risk score, they should have used the TRIPOD statement*

Our answer. We now have completed the TRIPOD statement. See our response to editor please.

Comment 10. *In the results section:*

The authors wrote "among 221 women patients, 28 (12.7%) have died" but this percentage is different from the one in the table 1 (which is 14.3%) and "among 1321 men, 168 (12.7%) died" but the correct percentage is 85.7%.

Our answer. The percentages presented in Table 1 are column percentages. The 14.3% mentioned by the reviewer represents the proportion of women among patients who died i.e. 28/196. The one reported in the main text is the proportion of death among women patients.

Comment 11. *Table 1: the sum of the percentage of the ethnicities is not equal to 100 (24.2+73.3+1.8 = 99.3) for each of the 3 columns. The same for the variables "Oxygen therapy", column 1, "Coexisting conditions", column 2, Numbers are written on the line of the variable coexisting conditions (n (%)): 0.29 (±0.45) 0.38 (±0.49) 0.27 (±0.45), but they seem to be not percentage but means +/- sd. What are these numbers?*

Our answer. We thank the reviewer for this comment. The percentages for the variable ethnicity do not sum up to 100% because there are missing values for ethnicity (0.7%). We have now added in the footnote of table 1 the following: *"*The percentages do not sum up to 100% because of the missing data."*

For the variable Oxygen therapy, there are 13.7% of patients without any oxygen therapy. This is why the percentage does not sum up to 100%. We have now added in the footnote of table 1 the following: *"** The percentages do not sum up to 100% because there are patients not requiring any form of oxygen therapy"*

For the variable "Coexisting conditions", there was a typo. We now have removed the misplaced presentation in in Table 1.

Comment 12. *The authors wrote that "Continuous variables were compared using the t-test or Wilcoxon-rank-sum test, while categorical variables were compared using the Chi-square or Fisher's exact test." But they have not detailed which variables were compared using Wilcoxon-rank-sum test.*

A p-value = 0.000 is commonly written <0.001.

Our answer. The p-values reported in Table 1 are parametric ones. For all variables, both the non-parametric Wilcoxon-rank-sum test and the parametric unpaired t-test (Welch) provided similar conclusions with comparable p-values. In addition, the normality assumption within each comparison group was satisfied for all variables.

The p-values throughout the manuscript are now presented as suggested by the reviewer.

Comment 13. *Table 3: for the Glasgow Coma Scale, is it the quantitative or the qualitative variable as both were presented in table 1?*

The authors wrote: "In case of continuous risk factors, we created a binary based on the median split". Why was this split not made from the bivariate analyzes? Which allow to avoid the utilization of variables that do not follow a normal distribution.

Our answer. For the first question, the Glasgow Coma Scale used in Table 3 is the quantitative one. We mentioned in this table that original quantitative variables were transformed using z-scores, and hazard ratios should be interpreted as 1 SD change in the values of the variables.

For the other comment, the median split was used only for graphical illustration purpose. In the LASSO, and Fine and Gray regression models, we did not transform or split by median any of the continuous variables. We have now made this point clear in the manuscript (see line 215-216).

Reviewer 4 comments

Dr. Shah-Jalal Sarker, University College London

Comment 1. *The authors presented a very useful timely article to identify risk factors and to develop risk prediction score for mortality of COVID-19 patients at ICU. The authors used the latest statistical techniques appropriately and the article will be very useful for readers. I recommend publishing it with minor revision as follows:*

Our answer. We thank the reviewer for his kind words regarding the quality and the importance of the paper.

Comment 2. *The article used data from March 2020 to July 2020 and the outcome time was from the date of ICU admission until the date of death. However, the risk prediction model is for 28-day mortality. This needs to be made clear in the article. If appropriate, 28-day (or 30-day) mortality needs to be added to the title.*

Our answer. As per the suggestions of other reviewers and the editor, we now have a new title emphasizing in-hospital mortality and the study design: "A Clinical Risk Score to Predict in-hospital Mortality in Critically Ill Patients with COVID-19: A Retrospective Cohort Study".

We have now clarified in the method section (line 164) that we derived the 28-day risk of in-hospital mortality.

Comment 3. *Waterlow score is a well-known predictor for 30-day mortality (see, "Can Waterlow score predict 30-day mortality and length of stay in acutely admitted medical patients (aged ≥65 years)? Evidence from a single centre prospective cohort study"). Waterlow score data includes age, nutritional status, weight, mobility, gender, smoking status, comorbidities, use of medication and continence. Glasgow comma score has covered some of the severity factors including continence. However, a reference and some discussion need to be included based on Waterlow and other important predictors.*

Our answer. We have now discussed the importance of Waterlow score for predicting 30-day mortality and length of hospital stay in admitted patients, and stress that application of this score in COVID-19 patients for mortality prediction requires future research. (see line 294-304).

Comment 4. *On page 12, "….The risk prediction score shows high accuracy in terms or discrimination (AUC = 88.1) and calibration (Bier score = 8.11)". The word "or" should be "of".*

Our answer. The typo is corrected now.

Comment 5. *BMI or obesity is an important risk factor that should have been included in the model.*

Our answer. We agree with the reviewers that BMI or obesity is an important factor. We managed to extract BMI but unfortunately, there are many missing values for BMI. One way to get around this is to use multiple imputations. However, we believe this will make the process complicated especially that we are using bootstrap for internal validation. Combining multiple imputation with bootstrapping is not straightforward.

Comment 6. *The article found no comorbidity but mostly laboratory findings were significant! Whereas, similar US studies (Ref 8) found significant comorbidities. These need to be discussed how these laboratory findings are likely to be related to the significant comorbidities found in other studies.*

Our answer. We have now discussed that inclusion of significant laboratory findings such as increased LDH and increased creatinine may represent underlying diseases such as liver disease, lung disease and kidney dysfunction (see line 290-293).

Comment 7. *Possible interaction effect among risk factors need to be tried and included in the model? For example, age and GCS.*

Our answer. We thank the reviewer for this interesting comment. Before the first submission of this manuscript, we investigated all statistical interactions between pairs of the retained predictors and found no statistically significant interaction terms.

We have now added a sentence in the method section regarding interaction terms (see line 161-162). We also added a sentence in the results section regarding the same topic (see line 212-213).

Comment 8. *The last paragraph of the discussion is repeated in the conclusion. The conclusion can be much shorter.*

Our answer. We have now reformulated the last paragraph of the discussion and shortened the conclusion section as suggested to avoid repetition (see line 316-327).

VERSION 2 – REVIEW

REVIEWER	Altschul, David Montefiore Medical Center, Neurosurgery
REVIEW RETURNED	27-May-2021

GENERAL COMMENTS	I think the authors have made all the appropriate revisions and responses to the initial manuscript. The paper is well written and ready for acceptance. The paper demonstrates that risk factors for mortality with COVID-19 changes depending on the population evaluated and the timing of the lab values evaluated in relation to the admission.

REVIEWER	Fumagalli, Carlo University of Florence, Department of Experimental and Clinical

	Medicine
REVIEW RETURNED	13-Jun-2021

GENERAL COMMENTS	I would like to thank and compliment Dr. Salem AlKaabi and Colleagues for responding to all comments raised in the first revision of the manuscript. All comments and observations that had been raised in the first round of revision have been replied to and, despite some limitations which have been acknowledged, the manuscript holds important epidemiological and clinical interest. I have one minor comment for the Discussion section regarding previously published severity scores developed to predict in-hospital mortality or disease progression in other clinical contexts (1. Clinical risk score to predict in-hospital mortality in COVID-19 patients: a retrospective cohort study – [PMID: 32978207 PMCID: PMC7520809] , and 2. A clinical risk score to identify patients with COVID-19 at high risk of critical care admission or death: An observational cohort study –[PMID: 32479771 PMCID: PMC7258846]): despite the younger age and epidemiological differences, some pathophysiological items seem to be in common and distinctive/specific to the disease. I would just add a comment in the Discussion. I have no further comments.

REVIEWER	Schwab, Camille Hôpital Saint-Antoine, Pharmacie
REVIEW RETURNED	31-May-2021

GENERAL COMMENTS	I thank the authors that have responded to the majority of my remarks. However, they have not corrected the table 1 where the percentages are incorrect. Most importantly, they have not describe how missing data were handled. Indeed, they wrote that "Patients with available data on these characteristics were included in the final analysis." but in the table 1, they mentionned that "The percentages do not sum up to 100% because of the missing data." Furthermore, they should explain why they have not conducted sensitivity analysis despite the missing data. To finish, they still have not explain how study size was calculated or why it has not been calculated.

REVIEWER	Sarker, Shah-Jalal University College London, Research Department of Medical Education
REVIEW RETURNED	07-Jun-2021

GENERAL COMMENTS	The authors have addressed all the comments. Strengths and limitations in the discussion section may be moved just before the conclusion.

VERSION 2 – AUTHOR RESPONSE

Reviewer Reports:

Reviewer: 1
Dr. David Altschul, Montefiore Medical Center

Comments to the Author:
I think the authors have made all the appropriate revisions and responses to the initial manuscript. The paper is well written and ready for acceptance. The paper demonstrates that risk factors for mortality with COVID-19 changes depending on the population evaluated and the timing of the lab values evaluated in relation to the admission.

Answer
We thank Dr. David Altschul for his constructive feedback and positive comments.

Reviewer: 2
Dr. Carlo Fumagalli, University of Florence

Comments to the Author:
I would like to thank and compliment Dr. Salem AlKaabi and Colleagues for responding to all comments raised in the first revision of the manuscript.
All comments and observations that had been raised in the first round of revision have been replied to and, despite some limitations, which have been acknowledged, the manuscript holds important epidemiological and clinical interest.

I have one minor comment for the Discussion section regarding previously published severity scores developed to predict in-hospital mortality or disease progression in other clinical contexts (1. Clinical risk score to predict in-hospital mortality in COVID-19 patients: a retrospective cohort study – [PMID: 32978207 PMCID: PMC7520809] , and 2. A clinical risk score to identify patients with COVID-19 at high risk of critical care admission or death: An observational cohort study –[PMID: 32479771 PMCID: PMC7258846]): despite the younger age and epidemiological differences, some pathophysiological items seem to be in common and distinctive/specific to the disease. I would just add a comment in the Discussion.

I have no further comments.

Answer
We thank Dr. Carlo Fumagalli for his constructive feedback and positive comments. We have now incorporated the suggested references and discuss it in the manuscript (see lines: 71-74, 258-260, 286-289).

Reviewer: 3
Dr. Camille Schwab, Hôpital Saint-Antoine, Institut Pierre Louis d'Epidemiologie et de Sante Publique

Comments to the Author:
I thank the authors that have responded to the majority of my remarks. However, they have not corrected the table 1 where the percentages are incorrect.
Most importantly, they have not describe how missing data were handled. Indeed, they wrote that "Patients with available data on these characteristics were included in the final analysis." but in the table 1, they mentioned that "The percentages do not sum up to 100% because of the missing data."
Furthermore, they should explain why they have not conducted sensitivity analysis despite the missing data.
To finish, they still have not explain how study size was calculated or why it has not been calculated.

Answer

We thank Dr. Camille Schwab for her constructive feedback and positive comments.
Regarding table 1 and the issue related to percentages, we realized that one of the variables described in table 1 (ethnicity) was not among the 36 potential predictors that we have chosen a priori for model selection and prediction. For consistency, we have now removed ethnicity from table 1 and described this variable in the text only (see lines 185-187). We checked table 1 and all the percentages sum up to 100. For example, for the variable oxygen saturation (SpO_2), 31.6% had SpO_2 < 90, 36.9% had SpO_2 between 90 and 94 and 31.5% had SpO_2 more or equal to 95. If the reviewer is still concerned regarding the percentages, please let us know exactly which variables are questionable.

Regarding the study size, a total of 1695 patients were eligible for the study entry among which 1542 had complete information on all the potential predictors. The characteristics of the 153 patients who were excluded due to heavy missing values were not different from those who were included in the current analysis (please see Table 1_supp below). Table 1_supp shows, in addition to the average values, the percentage of missing values for each variable. For instance, the average Lactate dehydrogenase was 409.32 IU/L in those included in the analysis (n = 1542) compared to 400 IU/L in those excluded from the analysis.
The average Lactate dehydrogenase calculations of 400 IU/L are based on only 20.92% of patients because 79.08 % of total patients who were excluded (i.e., n=153) had missing Lactate dehydrogenase values. In addition, the incidence of death in both samples were very similar. The manuscript has been updated accordingly (see lines 181:185).

Table 1 (suppl.). Demographics, commorbidity, vitals and laboratories for the whole dataset

Variables	Included 1,542 (90.97%)	Excluded 153 (9.03%)	P-value†
Age	48.94 (±12.66)	48.63 (±14.42)	0.656
Gender, n(%)			
Female	221 (14.33%)	25 (16.34%)	0.581
Male	1,321 (85.67%)	128 (83.66%)	
Diabetes, n(%)			
No	874 (56.68%)	77 (50.33%)	0.154
Yes	668 (43.32%)	76 (49.67%)	
Hypertension, n(%)			
No	854 (55.38%)	86 (56.21%)	0.912
Yes	688 (44.62%)	67 (43.79%)	
Respiratory disease, n(%)			
No	1,361 (88.26%)	134 (87.58%)	0.907
Yes	181 (11.74%)	19 (12.42%)	
CVD, n(%)			
No	1,053 (68.29%)	103 (67.32%)	0.877
Yes	489 (31.71%)	50 (32.68%)	
CKD, n(%)			
No	1,397 (90.60%)	140 (91.50%)	0.824
Yes	145 (9.40%)	13 (8.50%)	
Cancer, n(%)			
No	1,479 (95.91%)	147 (96.08%)	1.000
Yes	63 (4.09%)	6 (3.92%)	
Liver disease, n(%)			
No	1,437 (93.19%)	143 (93.46%)	1.000
Yes	105 (6.81%)	10 (6.54%)	
Coexisting conditions, n(%)			
0	498 (32.30%)	56 (36.60%)	0.05
1	403 (26.13%)	26 (16.99%)	
≥ 2	641 (41.57%)	71 (46.41%)	
Systolic BP, Mean (±SD)			
sbp	126.16 (±17.37)	124.86 (±18.19)	0.137

Table 1 (suppl.). Demographics, commorbidity, vitals and laboratories for the whole dataset

Variables	Included 1,542 (90.97%)	Excluded 153 (9.03%)	P-value†
Missing	0 (0%)	5 (3.27%)	
Diastolic BP, Mean (±SD)			
dbp	75.54 (±12.12)	73.66 (±9.86)	0.070
Missing	0 (0%)	5 (3.27%)	
Respiratory rate, Mean (±SD)	23.26 (±6.70)	22.32 (±6.62)	0.097
Glasgow Coma Scale, Mean (±SD)			
gcs	13.96 (±3.25)	13.92 (±3.25)	0.635
Missing	0 (0%)	28 (18.30%)	
Glasgow Coma Scale, n(%)			
Mild	1,391 (90.21%)	112 (73.20%)	0.639
Moderate	21 (1.36%)	3 (1.96%)	
Severe	130 (8.43%)	10 (6.54%)	
Missing	0 (0.00%)	28 (18.30%)	
Chloride, Mean (±SD)			
Chloride	99.25 (±4.50)	99.42 (±4.97)	0.675
Missing	0 (0%)	61 (39.87%)	
Bicarbonates, Mean (±SD)			
hco3	22.80 (±3.28)	22.62 (±3.54)	0.210
Missing	0 (0%)	61 (39.87%)	
Hemoglobin, Mean (±SD)			
hgb	131.90 (±18.24)	130.04 (±21.45)	0.814
Missing	0 (0%)	68 (44.44%)	
Monocytes percent, Mean (±SD)			
mono_perc	6.93 (±3.67)	6.63 (±3.31)	0.474
Missing	0 (0%)	69 (45.10%)	
Neutrophil percent, Mean (±SD)			
neutro_perc	73.04 (±13.28)	72.58 (±14.16)	0.865
Missing	0 (0%)	69 (45.10%)	
Platelets count, Mean (±SD)			
platelets	263.57 (±107.94)	257.51 (±118.38)	0.359
Missing	0 (0%)	69 (45.10%)	
Potasium, Mean (±SD)			
potas	4.05 (±0.55)	4.05 (±0.55)	0.980
Missing	0 (0%)	61 (39.87%)	
Red blood cell distribution width, Mean (±SD)			
rdw	13.49 (±1.60)	13.74 (±2.27)	0.829
Missing	0 (0%)	67 (43.79%)	
White blood cells, Mean (±SD)			
wbc	7.89 (±3.92)	9.21 (±6.08)	0.257
Missing	0 (0%)	68 (44.44%)	
Creatinine, Mean (±SD)			
creat	97.76 (±111.80)	95.09 (±71.56)	0.300
Missing	0 (0%)	59 (38.56%)	
Lactate dehydrogenase, Mean (±SD)			
ldh	409.32 (±223.16)	400.00 (±125.96)	0.431
Missing	0 (0%)	121 (79.08%)	
Ferritin, Mean (±SD)			
ferritin	1209.37 (±1374.60)	1031.91 (±996.80)	0.400
Missing	0 (0%)	116 (75.82%)	
C-reactive protein level, Mean (±SD)			
crp	102.13 (±94.00)	88.42 (±94.52)	0.057
Missing	0 (0%)	76 (49.67%)	
Sodium, Mean (±SD)			
sodium	136.98 (±4.36)	136.87 (±4.76)	0.446
Missing	0 (0%)	61 (39.87%)	
Haematocrit, Mean (±SD)			
v_hct	0.39 (±0.05)	0.39 (±0.06)	0.462

Table 1 (suppl.). Demographics, commorbidity, vitals and laboratories for the whole dataset

Variables	Included 1,542 (90.97%)	Excluded 153 (9.03%)	P-value†
Missing	0 (0%)	67 (43.79%)	
Red blood cells, Mean (±SD)			
rbc	4.75 (±0.71)	4.64 (±0.73)	0.542
Missing	0 (0%)	67 (43.79%)	
Lymphocytes count, Mean (±SD)			
Lymph_n	1.28 (±0.71)	1.38 (±0.90)	0.440
Missing	0 (0%)	69 (45.10%)	
Neutrophils count, Mean (±SD)			
neutro_n	6.00 (±3.77)	6.94 (±4.87)	0.376
Missing	0 (0%)	69 (45.10%)	
Monocyte count, Mean (±SD)			
mono_n	0.51 (±0.36)	0.54 (±0.33)	0.514
Missing	0 (0%)	69 (45.10%)	
Lymphocyte percent, Mean (±SD)			
lymph_perc	18.76 (±10.74)	18.72 (±11.70)	0.824
Missing	0 (0%)	69 (45.10%)	
Neutrophil-lymphocyte percent ratio, Mean (±SD)			
neutro_lympho_perc_ratio	6.82 (±9.46)	8.09 (±11.18)	0.828
Missing	0 (0%)	69 (45.10%)	
Neutrophil-lymphocyte count ratio, Mean (±SD)			
neutro_lympho_n_ratio	6.81 (±9.39)	8.09 (±11.18)	0.830
Missing	0 (0%)	69 (45.10%)	

† Continuous variables were compared using the t-test Wilcoxon-rank-sum test, while discrete variables were compared using the Chi-square test Fisher's exact test .

Regarding the sensitivity analysis, we did not do it for two main reasons:

The first reason is because of the heavy percentage of missing values in many potential predictors (see table1_supp). It is well known that multiple imputation methods in these scenarios might be unstable.

The second reason is the complexity of our procedure. To come up with the final model, we proceeded as follows: (1) we selected the set of important variables using LASSO. (2) the model chosen was internally validated using the .632+ Bootstrap method. To our knowledge, there is no statistical theory providing a consistent solution for using multiple imputation in this framework. The pooling of the parameter estimates in multiple imputation procedure can also be problematic when using the competing risk model. The scope of sensitivity analysis using multiple imputation or similar techniques in this framework are beyond the scope of the current study.

However, despite we have not done the sensitivity analysis, we believe that our results are consistent because the distribution of the variables in those included compared to those excluded are similar. Hence, one can assume that excluded patient are missing completely at random.

Reviewer: 4
Dr. Shah-Jalal Sarker, University College London

Comments to the Author:
The authors have addressed all the comments. Strengths and limitations in the discussion section may be moved just before the conclusion.

Answer
We thank Dr. Shah-Jalal Sarker for his constructive feedback and positive comments. As per the BMJ Open instructions to authors, the strengths and limitations should be placed just after the statement of the principal findings.

We are willing to move this section just before the conclusion, as suggested by the reviewer, if the editor approves it.