PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Clinical risk score to predict in-hospital mortality in COVID-19 patients: a retrospective cohort study
AUTHORS	Fumagalli, Carlo; Rozzini, Renzo; Vannini, Matteo; Coccia, Flaminia; Cesaroni, Giulia; Mazzeo, Francesca; Cola, Maria; Bartoloni, Alessandro; Fontanari, Paolo; Lavorini, Federico; Marcucci, Rossella; Morettini, Alessandro; Nozzoli, Carlo; Peris, Adriano; Pieralli, Filippo; Pini, Riccardo; Poggesi, Loredana; Ungar, Andrea; Fumagalli, Stefano; MARCHIONNI, Niccolo'

VERSION 1 – REVIEW

REVIEWER	James Galloway King's College London
REVIEW RETURNED	03-Jun-2020

GENERAL COMMENTS	This is important work and adds to the scientific knowledge. The study population is clearly defined. I have a question about the outcome measure. Was out of hospital death (for discharged patients) captured? I am aware that the numbers of patients in Lombardy and Tuscany who died in the time frame were very high, and so these numbers suggest that the total hospital capacity is quite small. It might be helpful to state how many acute beds each hospital has, to provide context around these admission figures. It is difficult to externally validate these numbers from my perspective. How were data extracted? Do the hospitals use electronic records? Were paper notes hand searched? These details are helpful in appreciating the likelihood of different types of bias. I did not fully understand the CXR scoring system. Could this be clarified? The methods should discuss the survival model in more detail. What model diagnostics were performed on the models? How were deaths on day of admission handled? I would be keen to see some detail around the build of the risk model. The stepwise inclusion of terms is reasonable, although given the established understanding of the importance of age, did the authors consider including age/gender adjusted models for all variables? Were interaction terms or non-linear relationships explored?

	I am interested that neutrophil count was not included. Was there a reason for this? It is surprising that such a high % of patients were RTPCR positive. That is incongruous with other studies. Does this reflect case ascertainment methodology? Looking at the results - Table 1 and Table 2 make no mention of missing data. I find 0% missing data to be highlight unlikely, even in an era of electronic records. Table 4 legend is incomplete. I would like this table to be clearer about exactly what each model involved. What variables were included in the adjusted a model. A major concern I have with the discussion relates to comments about potential benefits of hydroxychloroquine or tocilizumab. The question of whether these drugs have benefit is of global importance. We have a responsibility in the scientific community to offer due diligence when reporting any results that suggest effects (beneficial or harmful). In my view the authors need much more clarity about the fact that their model was designed to predict mortality overall and not specific to these drugs. To unpick whether these drugs had an effect in an observational cohort, I would want to see baseline characteristics describing who had each drug. I would want to know the unadjusted associations of the drug exposures, age and gender adjusted hazards, and then a fully adjusted model capturing as much as possible in terms of confounders. A propensity model approach might be more appropriate given numbers of events / number of confounders. Either way - I think including a discussion comment suggesting a protective effect is not appropriate without conducting and presenting the full analyses. I always worry that some of the hidden biases cause major issues when looking at drugs. For example - ventilation associates with worse outcomes. Anyone ventilated shortly after arrival is excluded from kaletra therapy which cannot be administered via an NG tube. Extreme age is a major risk factor for death, and certainly in my experience, patients admitted who were very frail and clearly dying, were not commenced on any drugs (and definitely not monoclonals). Small numbers of patients like this could substantially impact on estimates of risk and interpretation. The hazard ratios for the different risk categories should be shown. Ideally 95% CI bands should be added to the KM plot. I am suspicious that the low / intermediate risk groups are not significantly different - and with thought, could the thresholds be improved to more clearly identify the intermediate risk group? Given the urgent nature of the COVID pandemic, I think there is an urgency to work such as this being published. The authors present a cohesive story, and the score itself has potential to support clinical decision making. Importantly, the score has face validity. I would encourage the authors to strengthen recommendations for validation of the score. Supplementing the publication with a link to an online calculator would increase impact.

REVIEWER	Lee Wallis University of Cape Town South Africa
REVIEW RETURNED	28-Jun-2020

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript. The authors presents data of definite interest to the scientific community, but the paper requires substantial revision prior to consideration for publication. The overarching aim of the paper appears to be the development of the scoring system (based on the title, but in in the text, there are many unclear statements that confuse the aim); however, the majority of the methods, results and discussion focus on evaluating the utility of various features in predicting severe COVID-19 disease. These features and their potential for predicting severe illness have inherent value and are worthwhile publishing. My concern is that the manuscript focusses almost entirely on these features and does not provide an adequate description of the derivation of the scoring system itself. There is no explicit description of the derivation process in the methods – the only information provided is that a "survival analysis" was conducted. The results section suggests that the Kaplan-Meier analysis was used to derive the tool, but this is far from sufficient description. If developing the score is the primary objective of the paper, then the methods should be structured to reflect this: All methodological steps taken should logically advance towards the goal of deriving the score. The authors identified many potential determinants of mortality, as noted in Table 1. A small number were identified for inclusion in the two Cox multivariable regression models presented in Table 4. I am assuming that other determinants were eliminated during the backward stepwise elimination, though this is not clear. Only features were eligible for inclusion in their clinical risk score are shown in Table 4. It would be useful to include a table that documents all of the stepwise eliminations that occurred to get to the final model; this will allow for more transparent evaluation of the scoring system. I believe that there is value in the authors' data and their scoring system. In order for this manuscript to be effective, it needs to be reoriented to more logically and clearly describe the derivation portion. Alternatively, the content could be split into two manuscripts: One describing the predictive value of various demographic and clinical features (based on the publication of many similar studies, these data have standalone value) and a second that describes in greater depth the derivation and initial results of the scoring system. Some specific comments for the authors are below: General • The overarching aim of the paper appears to be the development of the scoring system; however, the majority of the methods, results and discussion focus on evaluating the utility of various features in predicting severe COVID-19 disease. These features and their potential for predicting severe illness have inherent value and are worthwhile of publishing. My concern is that the manuscript focusses almost entirely on these features and does not provide an adequate description of the derivation of the scoring system itself. • There are several grammar and spelling errors throughout the manuscript. I recommend it be reviewed by a native English speaker during revisions. • Appropriate statistical methods are employed in the study, though these methods are in need of more in-depth description. Background

	• The introduction is too brief; it does not make the argument for a need for a clinical risk prediction tool or give context to what other studies have considered for use as a tool. • There is information strewn throughout the paper (e.g. P13 L42) that could be more appropriately placed in the background section. • There are only two references cited in the entire background. This is insufficient in describing the situation in Italy and worldwide, and describing clinical risk prediction tools. • The study seems to describe two aims in its introduction: "defining the clinical and laboratory characteristics as assessed on hospital admission" and "to build…a novel risk scoring system." In the methods section, the primary outcome is noted as "In-hospital, all-cause death...". The ultimate aim of this study needs to be clarified. Methods • At present, the methods are quite confusing. Study design and population should be separated to provide more structure and clarity. • The description of statistical analyses needs strengthening: o Results mention Cox multivariable regressions, Kaplan-Meier survival analyses and ROC analyses, but only the Cox and a survival analysis (KM not specified) are noted in the methods section. o The analysis methods need to be presented in order, as they were used/presented in results section. o There is no explicit description of the derivation process in the methods – the only information provided is that a "survival analysis" was mentioned. The results section suggests that the Kaplan-Meier analysis was used to derive the tool, but this is far from sufficient description. If developing the score is the primary objective of the paper, than the methods should be structured as such: All steps taken should logically advance towards the goal of deriving the score. • If the goal was to create a scoring system that could use rapidly obtainable information/values, why were drugs and respiratory support (presumably provided post-admission) included in this evaluation? • P6: Results, such as availability of lab values and chest x-rays, should not be included in the methods section. • Patient and public involvement is discussed twice, on both P6 and P7. • P7 L15: Did the study authors make attempts to obtain retrospective informed consent or was it waived based on the conditions linked in the paper? I do not feel it is an issue if they did not, but the sentence is not clear. If attempts were made, they should be described. • The additional analysis described on P12 L18 is not described in the methods section. Such an analysis should have been set a priori, and it is concerning that this may not have been the case. • A serious limitation of this study was the exclusion of any variable that was not accounted for in all patients from the Cox analysis. This meant that a single missing lab value may have caused removal of what might have been a significant predictor from the scoring system. This is not well-described or motivated for in the methods (certainly there would be ways to account for missing values while maintaining inclusion of these potentially predictive features).

	Results • A small number of potential determinants of mortality were identified for inclusion in the two Cox multivariable regression models presented in Table 4. It would be useful to include a table that documents all of the stepwise eliminations that occurred to get to the final model. Including these data will allow for more transparent evaluation of the scoring system. • ROCs should be provided. Discussion • In line with restructuring the paper to focus on the ultimate aim – derivation of the scoring system – the discussion needs to place more emphasis on the scoring system and its implications, instead of the predictive value of the wide range of features in Table 1. • The authors should consider including commentary about the results of the scoring system itself and next steps: How do the included variables align with other tools being developed? Where / at what point in the patient care journey could the tool be used? How will it be studied in the future. • The exclusion of any variable that was not accounted for in all patients from the Cox analysis is a serious limitation. This needs to be noted in the limitations portion of the discussion and any rationale or steps to mitigate the potential effects of this need to be described. • The concluding remarks are very short, lacking insight into the authors' ultimate conclusions and intended next steps.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

We thank the Reviewer for the interest and supportive comments.

The study population is clearly defined.

Q: I have a question about the outcome measure. Was out of hospital death (for discharged patients) captured? I am aware that the numbers of patients in Lombardy and Tuscany who died in the time frame were very high, and so these numbers suggest that the total hospital capacity is quite small. It might be helpful to state how many acute beds each hospital has, to provide context around these admission figures. It is difficult to externally validate these numbers from my perspective.

R: The purpose of the present study was to create a score for in-hospital mortality (now stated under "Study outcome", line 567 of the marked version). Mortality after discharge was not captured. Information on hospital capacity has now been provided in the Methods Section, Study Design (lines 426-429 of the marked version).

Q: How were data extracted? Do the hospitals use electronic records? Were paper notes hand searched? These details are helpful in appreciating the likelihood of different types of bias.

R: Both hospitals have electronic records. This information had been described in the original version of the manuscript. In accordance with the purpose of looking for early and readily available determinants of in-hospital mortality, we focused on variables collected during the triage phase at both hospitals. Such an aim has been now better highlighted in both the Introduction (lines 229 and

onwards of the marked version) and the Methods (lines 452-454 of the marked version) sections. Paper notes were never used or searched for. Accordingly, a number of variables that would be available for all patients upon admission (as part of routine work-up) were selected to be extracted from the electronic charts of the two hospitals and merged into a single database for SPSS software for statistical analysis.

Q: I did not fully understand the CXR scoring system. Could this be clarified?

R: Reading and interpretation of the main chest X-Ray features was performed according to 'Fleischner Society: Glossary of Terms for Thoracic Imaging' Guidelines. This reference has now been added to the Methods section, Study Population and data source (page 8, lines 462 and onwards of the marked version):

Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. Radiology. 2008;246(3):697-722. Doi:10.1148/radiol.2462070712

Q: The methods should discuss the survival model in more detail. What model diagnostics were performed on the models? How were deaths on day of admission handled?
I would be keen to see some detail around the build of the risk model. The stepwise inclusion of terms is reasonable, although given the established understanding of the importance of age, did the authors consider including age/gender adjusted models for all variables? Were interaction terms or non-linear relationships explored?

R: The methods section (as further detailed) has been extensively revised in order to explain the stepwise approach that lead to the development of the score. In particular, Scheonfeld residuals were used to check the proportional hazard assumption.
We have included a supplementary table to describe all variables that were considered to build the score at univariable analysis. Gender was not significantly associated with outcome and we decided not to consider gender adjusted analyses for the present manuscript.
Non-linear relationships were not explored since they generally require a very large number of cases.
Finally, no death occurred on day 0 of admission. This is now specified in the text (line 674 of the marked version).

Q: I am interested that neutrophil count was not included. Was there a reason for this?

R: Neutrophil count was available for 215 (41.7%) patients only. Therefore, we preferred to exclude this variable from descriptive statistics.

Q: It is surprising that such a high % of patients were RTPCR positive. That is incongruous with other studies. Does this reflect case ascertainment methodology?

R: Point well taken. Case ascertainment methodological bias, which may impact on patient selection and outcome, cannot be excluded as a partial explanation for the findings observed. Indeed, the vast majority of patients included in the present analysis had a positive RT-PCR on first testing and only in a minority of cases was sputum or bronchoalveolar lavage needed to confirm the infection (line 676-678 of the marked version). We are aware of this limitation and have now acknowledged it in the limitations paragraph of the revised manuscript (lines 1181 and onwards of the marked version).

Q: Looking at the results – Table 1 and Table 2 make no mention of missing data. I find 0% missing data to be highlight unlikely, even in an era of electronic records. Table 4 legend is incomplete. I would like this table to be clearer about exactly what each model involved.

R: We confirm that variables reported in Table 1 (Demographic and clinical characteristics on hospital admission) were available in 100% of patients. Numbers in parentheses of variables included in the original version of Table 2 and a subtitle 'Variables not available in all patients' indicated the numbers and proportions of missing variables. However, these variables (e.g. IL-6, TNF-a, etc.) are usually available many hours after hospitalization and therefore are not useful to build a rapid risk score in the triage phase, which was the purpose of our study. Hence, in this perspective, these variables (which were not included in any multivariate model in the original version of the manuscript) have been removed throughout the revised version of the manuscript.

Q: What variables were included in the adjusted a model. A major concern I have with the discussion relates to comments about potential benefits of hydroxychloroquine or tocilizumab. The question of whether these drugs have benefit is of global importance. We have a responsibility in the scientific community to offer due diligence when reporting any results that suggest effects (beneficial or harmful). In my view the authors need much more clarity about the fact that their model was designed to predict mortality overall and not specific to these drugs. To unpick whether these drugs had an effect in an observational cohort, I would want to see baseline characteristics describing who had each drug. I would want to know the unadjusted associations of the drug exposures, age and gender adjusted hazards, and then a fully adjusted model capturing as much as possible in terms of confounders. A propensity model approach might be more appropriate given numbers of events / number of confounders. Either way – I think including a discussion comment suggesting a protective effect is not appropriate without conducting and presenting the full analyses.

I always worry that some of the hidden biases cause major issues when looking at drugs. For example – ventilation associates with worse outcomes. Anyone ventilated shortly after arrival is excluded from kaletra therapy which cannot be administered via an NG tube. Extreme age is a major risk factor for death, and certainly in my experience, patients admitted who were very frail and clearly dying, were not commenced on any drugs (and definitely not monoclonals). Small numbers of patients like this could substantially impact on estimates of risk and interpretation.

R: We agree that it may be difficult to draw hard conclusions regarding treatment efficacy, given the lack of propensity score matched populations and the substantial diversity of patients admitted to the ER. Moreover, the Reviewer's comment helped us in realizing that taking into account the potentially protective effect of whatever therapy patients were receiving was clearly beyond the scope of our work and might have a confounding effect, without adding any valuable information to our main scope, which was to design a rapid risk score. Hence, we have now removed Model 2 (which included pharmacological treatments) from the manuscript and from Table 4 (Cox Multivariate Analysis). However, we have decided to leave treatment strategies summarized in Table 3 to provide readers some insight into how patients with COVID-19 were managed. The 'Predictors of mortality and development of the mortality risk score' Paragraph in the Results section has now been re-written accordingly (lines 695 and onwards of the marked version).

Q: The hazard ratios for the different risk categories should be shown. Ideally 95% CI bands should be added to the KM plot. I am suspicious that the low / intermediate risk groups are not significantly different – and with thought, could the thresholds be improved to more clearly identify the intermediate risk group?

R: We have now added hazard ratios of the risk categories in the text ('Predictors of mortality and development of the mortality risk score' paragraph, Results section) and added 95% CI (represented as shaded areas) in the KM plot.
The low- and intermediate-risk groups were significantly different, and this is now specified in the text (lines 832-835).

Q: Given the urgent nature of the COVID pandemic, I think there is an urgency to work such as this being published. The authors present a cohesive story, and the score itself has potential to support clinical decision making. Importantly, the score has face validity.
I would encourage the authors to strengthen recommendations for validation of the score. Supplementing the publication with a link to an online calculator would increase impact.

R: We thank the reviewer for the encouraging and supportive comment. A validation project is ongoing using a large case series admitted to another Italian Hospital (Milan). Results will be object of a further work that, in case of positive confirmation, will include the proposal of an online calculator.

Reviewer 2

We thank the Reviewer for the precious comments.

Thank you for the opportunity to review this manuscript. The authors presents data of definite interest to the scientific community, but the paper requires substantial revision prior to consideration for publication.
The overarching aim of the paper appears to be the development of the scoring system (based on the title, but in in the text, there are many unclear statements that confuse the aim); however, the majority of the methods, results and discussion focus on evaluating the utility of various features in predicting severe COVID-19 disease. These features and their potential for predicting severe illness have inherent value and are worthwhile publishing. My concern is that the manuscript focusses almost entirely on these features and does not provide an adequate description of the derivation of the scoring system itself. There is no explicit description of the derivation process in the methods – the only information provided is that a “survival analysis” was conducted. The results section suggests that the Kaplan-Meier analysis was used to derive the tool, but this is far from sufficient description. If developing the score is the primary objective of the paper, then the methods should be structured to reflect this: All methodological steps taken should logically advance towards the goal of deriving the score.
Q: The authors identified many potential determinants of mortality, as noted in Table 1. A small number were identified for inclusion in the two Cox multivariable regression models presented in Table 4. I am assuming that other determinants were eliminated during the backward stepwise elimination, though this is not clear. Only features were eligible for inclusion in their clinical risk score are shown in Table 4. It would be useful to include a table that documents all of the stepwise eliminations that occurred to get to the final model; this will allow for more transparent evaluation of the scoring system.
R: The reviewer makes several important observations. Of the potential candidate determinants of in-hospital mortality, only 6 were included as they were retained in the final multivariate model. Highly significant associations with biomarkers that were not available for all patients (Table 2) as they are not readily available soon after triage nor are they routinely measured in all patients, were not entered into the model. This is now specified in the Methods section (Introduction – lines 201 and onwards; Methods – lines 425 and onwards of the marked version).
I believe that there is value in the authors’ data and their scoring system. In order for this manuscript to be effective, it needs to be reoriented to more logically and clearly describe the derivation portion. Alternatively, the content could be split into two manuscripts: One describing the predictive value of various demographic and clinical features (based on the publication of many similar studies, these data have standalone value) and a second that describes in greater depth the derivation and initial results of the scoring system.
Some specific comments for the authors are below:

General
Q: The overarching aim of the paper appears to be the development of the scoring system; however, the majority of the methods, results and discussion focus on evaluating the utility of various features in predicting severe COVID-19 disease. These features and their potential for predicting severe illness have inherent value and are worthwhile of publishing. My concern is that the manuscript focusses almost entirely on these features and does not provide an adequate description of the derivation of the scoring system itself.
R: We appreciate the reviewer comments. We have restructured the Introduction, Methods and Results section in order to reflect the objective of the paper. The Discussion now focuses mainly on the score and its clinical relevance.
Q: There are several grammar and spelling errors throughout the manuscript. I recommend it be reviewed by a native English speaker during revisions.
R: An in-depth revision by a native English speaker has now been performed. This has now been acknowledged in the “Acknowledgements” (lines 1377 of the marked version).
Q: Appropriate statistical methods are employed in the study, though these methods are in need of more in-depth description.
R: The Statistical Analysis Paragraph has been re-written to better describe the step-by-step derivation of the score. Furthermore, as suggested by the Reviewer 1, model diagnostics on Cox analysis have been acknowledged.
Background
Q1: The introduction is too brief; it does not make the argument for a need for a clinical risk prediction tool or give context to what other studies have considered for use as a tool.
Q2: There is information strewn throughout the paper (e.g. P13 L42) that could be more appropriately placed in the background section.
Q3: There are only two references cited in the entire background. This is insufficient in describing the situation in Italy and worldwide, and describing clinical risk prediction tools.
R1-R3: We have restructured and increased the length of the Introduction as suggested by the Reviewer (from 201 to 350 words). We have stressed the importance of having a prompt clinical assessment score for mortality directly in the Emergency Room with potentially associated benefits. Furthermore, we have added 10 new references regarding the situation in Italy and worldwide.
Q: The study seems to describe two aims in its introduction: “defining the clinical and laboratory characteristics as assessed on hospital admission” and “to build...a novel risk scoring system.” In the methods section, the primary outcome is noted as “In-hospital, all-cause death...”. The ultimate aim of this study needs to be clarified.
R: The aim of the paper (development of a mortality risk score) is now clearly reported in the last paragraph of the Introduction (lines 201 and onwards of the marked version).
Methods
Q: At present, the methods are quite confusing. Study design and population should be separated to provide more structure and clarity. The description of statistical analyses needs strengthening:
R: As further detailed in the Point-by-point letter (see below) the Methods have been rewritten. In particular, sections describing that Study Design and Population have now been separated as two paragraphs of the Methods section. The paragraphs are now: Study design, Study population data source, Study Outcome, Patient and public involvement and Statistical Analysis and score development.
Q: Results mention Cox multivariable regressions, Kaplan-Meier survival analyses and ROC analyses, but only the Cox and a survival analysis (KM not specified) are noted in the methods section.
The analysis methods need to be presented in order, as they were used/presented in results section. There is no explicit description of the derivation process in the methods – the only information provided is that a “survival analysis” was mentioned. The results section suggests that the Kaplan-Meier analysis was used to derive the tool, but this is far from sufficient description. If developing the score is the primary objective of the paper, than the methods should be structured as such: All steps

taken should logically advance towards the goal of deriving the score.

R: We appreciate the Reviewer's comment. The description of the Statistical Analysis Paragraph has been strengthened and presented in a step-by-step fashion to advance towards score derivation: from Descriptive Characteristics, to Cox multivariable model derivation, to Kaplan Meier and ROC analyses (with AUC).
Q: If the goal was to create a scoring system that could use rapidly obtainable information/values, why were drugs and respiratory support (presumably provided post- admission) included in this evaluation?
R: The Reviewer's concern is understandable. Given that an analysis on management was beyond the scope of the present study, we have now removed all paragraphs related to treatment throughout the manuscript.
Q: Results, such as availability of lab values and chest x-rays, should not be included in the methods section.
R: These results have now been removed.
Q: Patient and public involvement is discussed twice, on both P6 and P7.
R: We have now left a dedicated paragraph entitled 'Patient and public involvement' (lines 493 and onwards of the marked version).
Q: Did the study authors make attempts to obtain retrospective informed consent or was it waived based on the conditions linked in the paper? I do not feel it is an issue if they did not, but the sentence is not clear. If attempts were made, they should be described.
R: Ethical Committees of both hospitals approved data collection and granted a waiver of informed consent from study participants. In all instances, patients' identity was anonymized, and information protected by password (Lines 481-486).
Q: The additional analysis described on P12 L18 is not described in the methods section. Such an analysis should have been set a priori, and it is concerning that this may not have been the case.
R: As mentioned previously in response to other comments, all analyses on treatment management have now been removed from the present paper.
Q: A serious limitation of this study was the exclusion of any variable that was not accounted for in all patients from the Cox analysis. This meant that a single missing lab value may have caused removal of what might have been a significant predictor from the scoring system. This is not well-described or motivated for in the methods (certainly there would be ways to account for missing values while maintaining inclusion of these potentially predictive features).
R: The reviewer is correct. All variables that were not available for all patients (e.g. neutrophil count, cytokines etc.) have been removed from the Results section and Tables and this has been acknowledge in the limitations Paragraph (lines 1162 and onwards of the marked version).
Results
Q: A small number of potential determinants of mortality were identified for inclusion in the two Cox multivariable regression models presented in Table 4. It would be useful to include a table that documents all of the stepwise eliminations that occurred to get to the final model. Including these data will allow for more transparent evaluation of the scoring system.
R: We have added a Supplementary Table 1 which documents the stepwise eliminations that occurred to obtain to the final model.
Q: ROCs should be provided.
R: A ROC curve of the score has now been included in Supplementary Figure 1.

Discussion
Q: In line with restructuring the paper to focus on the ultimate aim – derivation of the scoring system – the discussion needs to place more emphasis on the scoring system and its implications, instead of the predictive value of the wide range of features in Table 1.
The authors should consider including commentary about the results of the scoring system itself and next steps: How do the included variables align with other tools being developed? Where / at what

point in the patient care journey could the tool be used? How will it be studied in the future.
R: The discussion paragraph has now been rewritten to comply with the Reviewers' observation. It now focuses more on the development and clinical implications of the scoring system and its applicability in the emergency department.
Q: The exclusion of any variable that was not accounted for in all patients from the Cox analysis is a serious limitation. This needs to be noted in the limitations portion of the discussion and any rationale or steps to mitigate the potential effects of this need to be described.
R: As mentioned previously in response to another point by this reviewer (see page 8), we have now acknowledged this in the limitations paragraph.
Q: The concluding remarks are very short, lacking insight into the authors' ultimate conclusions and intended next steps.
R: Concluding remarks have now also been expanded to address the need for future validation studies.

VERSION 2 – REVIEW

REVIEWER	James Galloway King's College London, UK
REVIEW RETURNED	25-Aug-2020

GENERAL COMMENTS	Thank you for responding to my comments. You have provided clear answers, and I believe the revisions strengthen the manuscript substantially. No further comments.

REVIEWER	Lee Wallis University of Cape Town South Africa
REVIEW RETURNED	11-Aug-2020

GENERAL COMMENTS	Thank you. the paper reads much better now