PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Clinical Characteristics of Hospitalized COVID-19 Patients and the Impact on Mortality: A Single-network, Retrospective Cohort Study from Pennsylvania State
AUTHORS	Gadhiya, Kinjal; Hansrivijit, Panupong; Gangireddy, Mounika; Goldman, John

VERSION 1 – REVIEW

REVIEWER	Giulio F Romiti Sapienza - University of Rome
REVIEW RETURNED	10-Aug-2020

GENERAL COMMENTS	In this manuscript from Gadhiya and colleagues, the authors reports the findings of a retrospective analysis, based in Pennsylvania, of 283 Covid-19 patients. The article is well-written and represent a potential valuable addition to the number of article reporting on clinical charateristics and course of patients with Covid-19 around the world. Major issues: 1) Aim of this article is clear: report charateristics and risk factors for in-hospital mortality from Covid-19 in a single-network retrospective cohorts based in Pennsylvania. However, the authors themselves stated that at least three reports from the US were already published. What is the novelty of their article? Is there any specific reason why clinical charateristics and course of Covid-19 in Pennsylvania may be different from those observed in Washington, California or New York states? I think that a brief explaination of what is the importance of their work may be useful to the readers (especially those in non-US countries) to understand the overall meaning of this manuscript. 2) How was the cut-off of laboratory variables (including D-Dimer, Ferritin, Troponin I, Procalcitonin etc.) chosen? Please state in the methods section. 3) Given the sample size and the total number of events, I am concerned about the number of independent variables inserted in the multivariate logistic regression models, which are not in accordance to the "one-in-ten rule". This may led to over-fitting and bias, especially on the strength of the associations. 4) As for model 2, I think that the meaning of the association which comes from the multivariate model are limited by potential confounding bias. In fact, the model was adjusted for age, sex, ethnicity and obesity; however, given the retrospective nature of the study, it is conceivable that most of these drugs were administered to patients which were more critical (or, at least, were perceived as "more critical" by their treating physicians". For example, it is possible that ascorbic acid, tocilizumab or convalescent plasma had

	been given to patients requiring intensive support, this in turn influencing the association between the treatments received and the outcomes (in-hospital mortality). I suggest the authors to repeat the analysis adjusting for comorbidities and variables which are at least "proxies" for severity (e.g. ICU admission, need for invasive ventilation etc.). This is extremely important to understand the true extent of the associations which come from multivariate models. This is also important for other models reported (especially, models 3, 4 and 5). 5) Caution is recommended in the discussion and conclusion when commenting on the results of the study. Page 17, "Hydroxychloroquine, ascorbic acid, zinc supplementation, and corticosteroids therapy were not recommended due to increased mortality risk" is way far from being demonstrated by a small retrospective study, given all the limitations. This sentence should be thoroughly amended to reflect the speculative nature of the findings presented. Similar edits are required in the discussion section (see for example page 16, "our cohort demonstrated that remdesivir and tocilizumab [...] and there was no significant imporevement in hospital length of stay [...]). Please try to use more caution, according to the nature of this study. 6) The authors may want to update their bibliographic search in order to reference some articles that were recently published and may deserve discussion, according to their results and in comparison to those reported in this article (see for example: https://www.nejm.org/doi/full/10.1056/NEJMoa2019014) Minor Issues: 1) Page 4: Limitation bullet points are maybe too scarce. Please try to expand these. 2) Page 5: is there any updated data on US mortality from Covid-19 and number of total cases in Pennsylvania? If so, please update in the manuscript. 3) Methods section (page 8): "An 95% CI greater than 1.0 or less than 1.0 is considered statistically significant". This sentence may be unclear to unexperienced reader. Please try to restate this definition in order to improve clarity. 4) What was the assay used for the Troponin I measurement? please report in the methods section. 5) Synthetic measures of illness severity (for example, SOFA) were not reported in the manuscript, and according to outcomes. Although not essentials, I think that they could be calculated according to data available. Reporting this measures would be useful for readers to understand difference of severity at presentation and their prognostic implications; also, this variables may be used as an adjustement in the multiavariate models. 6) Some small English edits are required (see for example page 16, "[...] hospital length of stay between in patients receivend these drugs" seems to me that "in" should be removed; etc.) 7) ref #43 should be updated to this: https://www.nejm.org/doi/full/10.1056/NEJMoa2021436

REVIEWER	Mark Rutherford University of Leicester, UK
REVIEW RETURNED	24-Aug-2020

GENERAL COMMENTS	This paper studies the patient characteristics of individuals hospitalised with COVID-19 and tries to evaluate predictors of mortality amongst those patients. The paper is observational in nature, covered a specific hospital network, and by its nature has a

	selective sample of individuals that required hospitalisation. Therefore, care should be made over generalising any inferences - there will be clearly be a selective population to start with. I have some comments on the manuscript. Major comment: 1. A lot of mortality percentages are given in the Introduction, but it's not clear what these values actually mean. You need to tie down what these values refer to. Who are the populations? What is the follow-up etc. etc.? It's not clear the numbers are fairly comparable as it stands. 2. How many people do each of the exclusions relate to? Are there not trial participants for treatments relating to COVID being excluded or not? Is not including trial patients dangerous - how many does this exclude? Some further details are needed. 3. Need to state the outcome clearly in the logistic regression. How did you handle the fact that some individuals had shorter follow-up and therefore less potential to contribute an event or was everyone followed up for a fixed minimum period? Are there important interaction effects that need to be considered in the models as opposed to just main effects? 4. What is the purpose of models 1-5? Need some more clarity on what the purpose of the models are. Is the plan to have one specific exposure of interest and then to adjust for confounders? Need some clarity over exactly what is of interest from each of the models to assess whether they are constructed and adjusted correctly. 5. What about the difference in characteristics between the ICU patients and not for the K-M plots? Obviously, this pattern is fairly expected given the severe cases are being separated off. 6. I don't follow the bootstrap analysis. More details are needed - do you just predict from your model and move the sample size up to a larger number? 7. You mention confounding a lot - but what's your main exposure of interest in that case? I still don't follow the purpose of the modelling in that sense. 8. "Multiple risk factors for in-hospital death were identified. Hydroxychloroquine, ascorbic acid, zinc supplementation, and corticosteroids therapy were not recommended due to increased mortality risk." - this is a fairly dangerous conclusion to be making. I'm not sure you can assume you are comparing like with like in any of these comparisons whether you think the model is helping in that regard or not. You are saying risk factors, rather than trying to say anything causal, but there's an implication in the conclusion here of causality. Some minor clarifications: "There was no significant difference in mean hospital stay between survivors and non-survivors." - is this a useful statistic? One of the groups is obviously curtailed by death. "Clinical risk factors that were significant from standard analyses " - what does this mean? "In our analysis, the cumulative survival declined with increasing length of hospital stay" - This has to stay constant or go down obviously - what is this sentence trying to say? Page 5. Line 1. What do you mean when you say "mortality of 5% in

	the US"? What does this number actually represent?

VERSION 1 – AUTHOR RESPONSE

Reviewer 1:
In this manuscript from Gadhiya and colleagues, the authors report the findings of a retrospective analysis, based in Pennsylvania, of 283 Covid-19 patients.
The article is well-written and represent a potential valuable addition to the number of articles reporting on clinical characteristics and course of patients with Covid-19 around the world.

Major issues:
1) Aim of this article is clear: report characteristics and risk factors for in-hospital mortality from Covid-19 in a single-network retrospective cohort based in Pennsylvania. However, the authors themselves stated that at least three reports from the US were already published. What is the novelty of their article? Is there any specific reason why clinical characteristics and course of Covid-19 in Pennsylvania may be different from those observed in Washington, California or New York states? I think that a brief explanation of what is the importance of their work may be useful to the readers (especially those in non-US countries) to understand the overall meaning of this manuscript.
A: Thank you for your comment. Our study is unique from others in several aspects. First, studies from Washington and California were reported early in the pandemic. The sample size was small in Washington study and novel therapies, such as remdesivir, tocilizumab, convalescent plasma, or steroids were not available during both studies. Second, although a landmark study from New York has its strength over these limitations, individual patient charts were not reviewed. Thus, multivariate analyses or survival analyses were not performed in this study. Moreover, our patient population is different from New York cohort, that is, our population is older and has more CKD and diabetes. We have also edited the manuscript by adding “Although the characteristics of hospitalized COVID-19 patients have been reported in other states, there are some limitations that preclude the generalization of the results toward our patient population. Studies from Washington and California were conducted in pre-remdesivir era and multivariate analysis was not performed in the New York City cohort. The association between clinical characteristics and in-hospital mortality in the U.S. population have not been clearly established.” in the Introduction. Moreover, we edited the last sentence of Introduction to expand the aim of study toward a global aspect, rather than just in the United States; “In this retrospective cohort, we aimed to demonstrate the clinical characteristics of COVID-19 patients, treatment outcomes and the impact on in-hospital mortality in an ethnically diverse population.”.

2) How was the cut-off of laboratory variables (including D-Dimer, Ferritin, Troponin I, Procalcitonin etc.) chosen? Please state in the methods section.
A: Due to a severe shortage of word count in the main text, the cutoff for each variable is provided in Supplemental Document 1. We added “Supplemental Documents 1 summarizes the description of each variable in our cohort as well as the cutoff values for each variable.” under Data Collection section for the readers. In brief, laboratory findings were recorded: leukocytosis (white blood cells count > 9,500/□L), leukopenia (white blood cells count < 3,900/□L), lymphocytopenia (absolute lymphocytes count < 600/□L), thrombocytosis (platelets > 400,000/□L), and thrombocytopenia (platelets < 140,000/□L), respiratory acidosis (arterial pH < 7.35 and arterial partial pressure of CO2 > 45 mmHg), transaminitis (alanine transaminase > 3 times of upper normal limit). Serum creatinine (mg/dL) was measured by enzymatic assay. Elevation of serum D-dimer (> 500 ng/mL), ferritin (> 336 ng/mL) lactate dehydrogenase (LDH; > 200 U/L), C-reactive protein (> 1 mg/dL), procalcitonin (> 0.25 ng/mL), and troponin I (> 0.03 ng/mL). Each cutoff is referenced from the National Reference Laboratory (LabCorp).

3) Given the sample size and the total number of events, I am concerned about the number of independent variables inserted in the multivariate logistic regression models, which are not in accordance to the "one-in-ten rule". This may lead to over-fitting and bias, especially on the strength of the associations.
A: Thank you for comment. We understand that in some occasions, we have analyzed the data that were beyond the one-in-ten rule. However, several newer studies have started to report the associations when the number of events per variable is less than 10. We believe this conduct is as of result from McCulloch et al. (https://doi.org/10.1093/aje/kwk052) and Smeden et al. (https://dx.doi.org/10.1186%2Fs12874-016-0267-3). We apologize that we cannot re-analyze the whole results to fit under one-in-ten rule, however we have added a statement under Discussion to make the readers aware of this limitation, “Furthermore, our binary logistic regression analyses may not strictly follow the one-in-ten rule which may lead to over-fitting effect. However, our statistical rationale is supported by newer simulation studies by McCulloch et al. and Smeden et al.”

4) As for model 2, I think that the meaning of the association which comes from the multivariate model are limited by potential confounding bias. In fact, the model was adjusted for age, sex, ethnicity and obesity; however, given the retrospective nature of the study, it is conceivable that most of these drugs were administered to patients which were more critical (or, at least, were perceived as "more critical" by their treating physicians". For example, it is possible that ascorbic acid, tocilizumab or convalescent plasma had been given to patients requiring intensive support, this in turn influencing the association between the treatments received and the outcomes (in-hospital mortality).
I suggest the authors to repeat the analysis adjusting for comorbidities and variables which are at least "proxies" for severity (e.g. ICU admission, need for invasive ventilation etc.). This is extremely important to understand the true extent of the associations which come from multivariate models. This is also important for other models reported (especially, models 3, 4 and 5).
A: Thank you for your kind suggestion. We have adopted and applied your recommendations to the analyses. Model 2 to 5 were now adjusted for “ICU admission” and “the need for oxygen therapy”. The latter includes all modes of O2 delivery e.g. mechanical ventilation, NIPPV, high-flow nasal cannula, etc. We used these variables as surrogates for severity like you recommended. The results are quite interesting. We found like hydroxychloroquine, ascorbic acid, zinc, convalescent plasma, steroids, and arrhythmias were NO longer associated with mortality. This has made our Discussion easier. The manuscript was updated according to the revised results in Abstract, Results, Discussion, and Conclusion. We really appreciate your recommendations to improve the quality of our research.

5) Caution is recommended in the discussion and conclusion when commenting on the results of the study. Page 17, "Hydroxychloroquine, ascorbic acid, zinc supplementation, and corticosteroids therapy were not recommended due to increased mortality risk" is way far from being demonstrated by a small retrospective study, given all the limitations. This sentence should be thoroughly amended to reflect the speculative nature of the findings presented. Similar edits are required in the discussion section (see for example page 16, "our cohort demonstrated that remdesivir and tocilizumab [...] and there was no significant improvement in hospital length of stay [...]). Please try to use more caution, according to the nature of this study.
A: Thank you for your suggestions. We have deleted the old statement as our results were updated. The conclusions were changed to, “In conclusion, COVID-19 is a serious condition with a significant in-hospital mortality rate. Multiple risk factors for in-hospital death were identified. Increasing SOFA score, AKI and ARDS are significant risk factors for increased death in critically ill patients.” Similarly, we have made some changes to the statement on remdesivir and tocilizumab to soften the interpretation of results, “We have observed that remdesivir and tocilizumab were not associated with mortality and there was no significant improvement in hospital length of stay between in patients receiving these drugs. However, given the observational, non-randomized design of this study, it is difficult to determine the efficacy of such treatment.”

6) The authors may want to update their bibliographic search in order to reference some articles that were recently published and may deserve discussion, according to their results and in comparison to those reported in this article (see for example:
https://www.nejm.org/doi/full/10.1056/NEJMoa2019014)
A: Thank you for your suggestions. This article is fairly new and we appreciate your recommendation. We did not see this article when we wrote up the manuscript. The reference was added and discussed in the Discussion: “Later, the efficacy of hydroxychloroquine with or without azithromycin has been demonstrated in a recent multicenter, randomized, open-label, controlled trial in hospitalized patients with mild-to-moderate COVID-19. In this study, the authors found that the use of hydroxychloroquine, alone or with azithromycin did not improve the clinical outcome at 15 days compared to the standard treatment.”

Minor Issues:
1) Page 4: Limitation bullet points are maybe too scarce. Please try to expand these.
A: Thank you for your comment. The editor has suggested the same point. We have added more limitations into this section. However, please note that the editorial office has limited us to provide only 5 bullets to include both strengths and limitations to this section.

2) Page 5: is there any updated data on US mortality from Covid-19 and number of total cases in Pennsylvania? If so, please update in the manuscript.
A: We have updated the numbers as of August 29th, 2020 and I quote, “…rapidly escalated to 216 countries within five months with the highest number of infected cases in the United States. As of August 29th, 2020, the reported cumulated number of confirmed cases in the United States was close to 6 million with a mortality rate of 3.09%.” The data for the state of Pennsylvania is mentioned in the third paragraph of Introduction, “As of August 29th, 2020, the Pennsylvania Department of Health has announced more than 129,056 confirmed cases of COVID-19 leading to 7,671 deaths, making Pennsylvania the 13th state with the highest confirmed cases.”

3) Methods section (page 8): "An 95% CI greater than 1.0 or less than 1.0 is considered statistically significant". This sentence may be unclear to unexperienced reader. Please try to restate this definition in order to improve clarity.
A: Thank you. We have modified the statement to make it more clear, “The analysis is considered statistically significant if the 95% CI crosses 1.0.” The reference was also added for readers who are interested in.

4) What was the assay used for the Troponin I measurement? please report in the methods section.
A: We have clarified our chemistry lab and we use Troponin I assay (the company name is Beckman Coulter DxI). We have added this information in Supplemental Document 1 as well. Please note that the cutoff value for each variable is determined by the National Reference Laboratory (LabCorp).

5) Synthetic measures of illness severity (for example, SOFA) were not reported in the manuscript, and according to outcomes. Although not essentials, I think that they could be calculated according to data available. Reporting this measure would be useful for readers to understand difference of severity at presentation and their prognostic implications; also, this variable may be used as an adjustment in the multivariate models.
A: Thank you for your recommendation. As we know that the SOFA score is associated with mortality in critically ill patients (JAMA, 2001), thus we went back and collected the SOFA score for our ICU patients. The SOFA score was collected on day 1 of ICU admission. The results were updated in Table 4. In brief, deceased patients had significantly higher SOFA score. We also used the SOFA score as a co-variate in the multivariate analysis. We found that AKI (OR 2.128, 95% CI 1.111-6.667; p = 0.034) and ARDS (OR 6.410; 95% CI 2.237-18.182; p = 0.023) are significantly predictive of in-hospital mortality among patients admitted the ICU after adjusted for age, sex, ethnicity, and SOFA

score. The SOFA score itself had an odds ratio of 1.544. The Results section was updated accordingly. We would like to thank you again for such a great advice.

6) Some small English edits are required (see for example page 16, "[...] hospital length of stay between in patients received these drugs" seems to me that "in" should be removed; etc.)
A: Thank you for noticing. We have corrected the typo. The correct statement is “remdesivir and tocilizumab were not associated with mortality and there was no significant improvement in hospital length of stay between patients receiving these drugs.”

7) ref #43 should be updated to this: https://www.nejm.org/doi/full/10.1056/NEJMoa2021436
A: Thank you so much for your suggestion. This article was not yet published when we submitted the manuscript. We have updated reference 43 to this one as suggested.

Reviewer 2:
This paper studies the patient characteristics of individuals hospitalised with COVID-19 and tries to evaluate predictors of mortality amongst those patients. The paper is observational in nature, covered a specific hospital network, and by its nature has a selective sample of individuals that required hospitalisation. Therefore, care should be made over generalising any inferences - there will be clearly be a selective population to start with.

I have some comments on the manuscript.

Major comment:
1. A lot of mortality percentages are given in the Introduction, but it's not clear what these values actually mean. You need to tie down what these values refer to. Who are the populations? What is the follow-up etc. etc.? It's not clear the numbers are fairly comparable as it stands.
A: Thank you for your comments. We rearranged the paragraphs in the Introduction. We started by describing the burden and magnitude of COVID-19 in the United States. We have adjusted the sentence to make it more comprehensible as follows; “As of August 29th, 2020, the reported cumulated number of confirmed cases in the United States was close to 6 million with a mortality rate of 3.09%.” Then we described the burden of disease in Pennsylvania and how our study differs from the previous cohorts. In brief, the associations between patient clinical characteristics and mortality have not been adequately demonstrated in US population. Overall, we are trying to prime the readers with the disease burden and why this study was conducted.

2. How many people do each of the exclusions relate to? Are there not trial participants for treatments relating to COVID being excluded or not? Is not including trial patients dangerous - how many does this exclude? Some further details are needed.
A: Thank you for your comment. We have revised Figure 1 to entail the excluded patients more clearly. The Results section was updated as well; “A total of 12,938 patients were identified during the study period. Thirty-nine patients were outpatient and did not require hospitalization. After excluding patients with negative PCR (n = 7,374), duplicate medical records (n = 145), pregnant woman (n = 3), and clinical trial patients (n = 2), 283 patients were included for further analysis. The flowchart of data selection from the UPMC Pinnacle COVID-19 registry is depicted in Figure 1.” To answer your question, No, we do not believe that by excluding patients who are enrolled in clinical trial would have a detrimental effect on our outcome. First, there were only 2 patients who enrolled in clinical trial (convalescent plasma trial with the Mayo Clinic), which is fairly little. Second, we cannot include these patients to our study as it would violate the IRB protocol.

3. Need to state the outcome clearly in the logistic regression. How did you handle the fact that some individuals had shorter follow-up and therefore less potential to contribute an event or was everyone

followed up for a fixed minimum period? Are there important interaction effects that need to be considered in the models as opposed to just main effects?
A: Thank you for your question. The outcome for logistic regression is clear. Only “death” was put in as dependent variable. For follow-up duration, we followed everyone while they remained in the hospital. Thus, the follow-up duration is the same as “hospital stay”, that is, 13.1 days. We did not have to worry about lost follow-up patients because there were none. If they were discharged within 1-2 days, that’s fine – the endpoint is clear, which is “no death”. Similarly, if they died within 1-2 days, we analyzed them as “death” and we also reviewed the patient chart for more details. For interaction that could interfere with the validity of our analysis is of course, it is the confounding effect. Although we adjusted several factors that were obvious confounders, there will be some unaware factors that could contribute to the outcome. We believe every retrospective cohort would face the same limitation, thus we also mentioned this to readers in the Limitations section of Discussion.

4. What is the purpose of models 1-5? Need some more clarity on what the purpose of the models are. Is the plan to have one specific exposure of interest and then to adjust for confounders? Need some clarity over exactly what is of interest from each of the models to assess whether they are constructed and adjusted correctly.
A: Thank you for your question. Model 1 is adjusted for the baseline co-morbidities. Model 2 is adjusted for the need of oxygen therapy because the interventions included in Model 2 (azithromycin, hydroxychloroquine, ascorbic acid, zinc, tocilizumab, convalescent plasma) were usually given in patients who required oxygen therapy. Thus, we held “the need for oxygen therapy” constant in this Model to determine if these interventions were associated with death. For Model 3, respiratory acidosis, and steroids therapy are usually seen in patients with asthma, COPD, critical illnesses and who required oxygen which are obvious confounders. Thus, we adjusted this Model for “asthma/COPD”; “need for oxygen therapy”; and “ICU admission”. These adjustments are in part of Reviewer 1’s recommendations to determine the true association between the variables and death. Model 4, “acute kidney injury” is defined by serum creatinine. Thus, patients with CKD would be a clear confounder. That is why we adjusted the Model for “CKD”; “need for O2 therapy”; and “ICU admission” as all of these factors may contribute to death. In Model 5, “arrhythmias” is a cardiac complication, hence we adjusted for every variable that could be the confounding factor, such as CAD, heart failure, history of arrhythmia/conduction disorder, and ICU admission. We understand that some readers may have the same question, so we included a statement and Supplemental Document 2 for clarification to the readers, “The rationale for each Model adjustment in multivariate analysis is available in Supplemental Document 2.”

5. What about the difference in characteristics between the ICU patients and not for the K-M plots? Obviously, this pattern is fairly expected given the severe cases are being separated off.
A: Thank you for your comment. You are correct. The Kaplan-Meier curves showed that ICU patients had less survival probability than non-ICU patients. This observation is straightforward. We have analyzed two more curves as presented in Figure 2. We compared the cumulative survival in ICU vs. ICU+AKI and ICU vs. ICU+ARDS. Please note that the p-values are not statistically significant because the outcome (death) was analyzed with the duration of stay.

6. I don't follow the bootstrap analysis. More details are needed - do you just predict from your model and move the sample size up to a larger number?
A: Thank you for your question. We used bootstrap function on SPSS. Basically, what we did was, when we had a significant association between variable X and death, we clicked bootstrapping. The software uses random sampling with replacement and expands the imputed same size to 1,000 (this number was pre-entered by us). Then, we determined if this variable X remained statistically significant. In this sense, bootstrapping was used as a form of sensitivity analysis.

7. You mention confounding a lot - but what's your main exposure of interest in that case? I still don't follow the purpose of the modelling in that sense.
A: Thank you for your comment. We are concerned about any variables that could potentially link to mortality. Our primary outcome is death and we know that several factors could result to this outcome. Thus, we developed the models of adjustments tailored for each variable. Some exposures were clear to link to mortality, so we made sure these exposures were adjusted for. We have provided the explanation of the rationale of each model for you above. Supplemental Document 2 is also available for the readers who might have the same question. We hope that you would find this document helpful as well.

8. "Multiple risk factors for in-hospital death were identified. Hydroxychloroquine, ascorbic acid, zinc supplementation, and corticosteroids therapy were not recommended due to increased mortality risk." - this is a fairly dangerous conclusion to be making. I'm not sure you can assume you are comparing like with like in any of these comparisons whether you think the model is helping in that regard or not. You are saying risk factors, rather than trying to say anything causal, but there's an implication in the conclusion here of causality.
A: Thank you for your suggestion. We have modified the conclusion as the results were revised as per Reviewer 1's suggestions. The new statement is, "In conclusion, COVID-19 is a serious condition with a significant in-hospital mortality rate. Multiple risk factors for in-hospital death were identified. Increasing SOFA score, AKI and ARDS are significant risk factors for increased death in critically ill patients."

Some minor clarifications:

"There was no significant difference in mean hospital stay between survivors and non-survivors." - is this a useful statistic? One of the groups is obviously curtailed by death.
A: Thank you for your suggestion. You are correct. This analysis is not so meaningful as one group is curtailed by death. Thus, we refrained from discussing on the length of stay under the Discussion part.

"Clinical risk factors that were significant from standard analyses " - what does this mean?
A: Thank you for your question. Here is our clarification, "Clinical risk factors that were significant from standard analyses (Pearson's □2 tests, Fisher-Exact tests, t-tests) were included in univariate binary logistic regression analysis."

"In our analysis, the cumulative survival declined with increasing length of hospital stay" - This has to stay constant or go down obviously - what is this sentence trying to say?
A: Thank you for your input. We provided that statement to describe the Kaplan-Meier curve for unexperienced readers. There are no other meanings that we were trying to relay.

Page 5. Line 1. What do you mean when you say, "mortality of 5% in the US"? What does this number actually represent?
A: Thank you for your question. The flow of Introduction was rearranged. We were trying to describe the burden and severity of COVID-19 in the United States for the readers and how important it is for this research to be conducted as it adds additional knowledge to what we already know. Then the second paragraph will mention the data from the state of Pennsylvania followed by the rationale of this study.

VERSION 2 – REVIEW

REVIEWER	Giulio Francesco Romiti Sapienza - University of Rome
REVIEW RETURNED	03-Sep-2020

GENERAL COMMENTS	Thank you for the opportunity to review the revised version of this manuscript. I think that the manuscript critically improved from the previous round of review, and I acknowledge the large efforts that the authors made to review their article. I still have some comments: 1) Abstract: the Results paragraph may need a double check. The long sentence starting with "in adjusted multivariate analysis [...]" does not come to an end, after the list of all the variables associated with mortality which finish with acute kidney injury. Also, I am aware that the abstract word count may represent a limit, but the significance of the relationship of these variables with the outcome should be clearly expressed in the abstract. If you are not able to report the p values/confidence intervals for each variable, please state at least that "all were significantly associated" (or similar), if applicable. 2) Row 301-305. The first sentence ("the efficacy of hydroxycholoroquine with or without azithromyicin has been demonstrated in a recent multicenter [...]") is quite misleading, since the authors themselves state that "in this study [...] the use of hydroxycholoroquine [...] did not improve the clinical outcome". At a first glance, readers may be confused by the first sentence, in which the authors wrote that "the efficacy has been demonstrated". Please change that. 3) Row 86: the term "pre-remdesivir era" may be changed to reflect the fact that remdesivir was already used in selected case at the moment in which those studies were conducted; in its current form, a reader may think that the introduction of remdesivir followed the publication of those articles. The authors may want to rephrase this passage, to reflect more precisely these temporal relationships.

REVIEWER	Mark Rutherford University of Leicester
REVIEW RETURNED	25-Sep-2020

GENERAL COMMENTS	The authors have yet to address some of my previous comments on the manuscript (see below). The authors have a useful data resource to describe the characteristics of hospital patients and look for some associations with mortality (but this study is observational in nature, so any comparisons across groups are made difficult). Also the number of variables simultaneously included with many interrelationships and common pathways makes it difficult to follow all the results. The clarity of the overall approach about what they are then trying to do needs firming up. I have some further comments: 1 To be a confounder a variable must be considered with it's relation to both the outcome and a main exposure - the authors still don't clearly state what the purpose of the models are in this regard(?) and therefore should be careful with the use of the word. Are they simply trying to make models looking at all risk factors possible? Some more careful thought about the relationships between the variables might be useful - some must be highly highly correlated and this could impact on the reported ORs. On that note, how does Table 3 work? Are the first set of variables also included in the subsequent models - surely the Odds Ratio values then change when the other variables are added? Why not give all the Odds

	Ratios - i.e. I can't see the one for obesity or sex. 2 The answer surrounding the bootstrap still doesn't make sense to me either. What's the purpose there, and what is this actually doing? "The association of these variables and overall mortality remained significant on bootstrap analysis when the sample size was imputed to 1,000." Are you just simulating from the model and increasing the sample size? This will only increase power, no? i don't see how this validates anything? 3 "In our analysis, the cumulative survival declined with increasing length of hospital stay." This is still not a great description of Figure 2. A survival curve has to go down or stay constant, by definition. Can more be added? 4 " However, one could argue that our study has significantly smaller sample size which might underestimate the actual mortality of COVID-19. " Why would this be necessarily true? 5 The reports of comparisons across lengths of stay by various population groups - these are all made difficult by the fact that non-survivors have a curtailed length of stay by definition. Be careful here. 6 The authors have added this sentence for clarity: "The analysis is considered statistically significant if the 95% CI crosses 1.0." But...this is the wrong way round.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Giulio Francesco Romiti

Institution and Country: Sapienza - University of Rome

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

Thank you for the opportunity to review the revised version of this manuscript. I think that the manuscript critically improved from the previous round of review, and I acknowledge the large efforts that the authors made to review their article.

A: Thank you Dr. Romiti for your time reviewing our manuscript again. It is clear to us that you have read through the entire manuscript again to provide us these helpful suggestions. We really appreciate your time and effort as well.

I still have some comments:

1) Abstract: the Results paragraph may need a double check. The long sentence starting with "in adjusted multivariate analysis [...]" does not come to an end, after the list of all the variables

associated with mortality which finish with acute kidney injury. Also, I am aware that the abstract word count may represent a limit, but the significance of the relationship of these variables with the outcome should be clearly expressed in the abstract. If you are not able to report the p values/confidence intervals for each variable, please state at least that "all were significantly associated" (or similar), if applicable.

A: Thank you Dr. Romiti for your recommendations. We have edited the Results part of the abstract to, “Several factors were identified from our adjusted multivariate analyses to be associated with in-hospital mortality: […]”. We also added the 95% confidence interval to each presented variable as the readers will likely want to see them as well. Your recommendation on this is in line with the editor’s recommendation as well. Thank you.

2) Row 301-305. The first sentence ("the efficacy of hydroxycholoroquine with or without azithromyicin has been demonstrated in a recent multicenter [...]") is quite misleading, since the authors themselves state that "in this study [...] the use of hydroxycholoroquine [...] did not improve the clinical outcome". At a first glance, readers may be confused by the first sentence, in which the authors wrote that "the efficacy has been demonstrated". Please change that.

A: Thank you for your suggestion. We have simplified the sentence to avoid misleading the readers. Thus, the sentence was changed to, “A recent multicenter, randomized, open-label, controlled trial in hospitalized patients with mild-to-moderate COVID-19 found that the use of hydroxychloroquine, alone or with azithromycin did not improve the clinical outcome at 15 days compared to the standard treatment.39 Thus, the recommendation for use of hydroxychloroquine was recalled by the Infectious Diseases Society of America (IDSA)”. This statement emphasizes our finding that hydroxychloroquine was not associated with improved mortality.

3) Row 86: the term "pre-remdesivir era" may be changed to reflect the fact that remdesivir was already used in selected case at the moment in which those studies were conducted; in its current form, a reader may think that the introduction of remdesivir followed the publication of those articles. The authors may want to rephrase this passage, to reflect more precisely these temporal relationships.

A: Thank you for your sophisticated comment again. We agree with you. Thus, we eliminated the term “pre-remdesivir” and expanded the sentence to, “Although the characteristics of hospitalized COVID-19 patients have been reported in other states, there are some limitations that preclude the generalization of the results toward our patient population. Studies from Washington and California were conducted and published during an early stage of the pandemic where treatment options, such as remdesivir or dexamethasone, were not recommended as the standard of care. In addition, multivariate analysis was not performed in the New York City cohort.”

Reviewer: 2

Reviewer Name: Mark Rutherford

Institution and Country: University of Leicester

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

The authors have yet to address some of my previous comments on the manuscript (see below). The authors have a useful data resource to describe the characteristics of hospital patients and look for some associations with mortality (but this study is observational in nature, so any comparisons across groups are made difficult). Also the number of variables simultaneously included with many interrelationships and common pathways makes it difficult to follow all the results. The clarity of the overall approach about what they are then trying to do needs firming up. I have some further comments:

1. To be a confounder a variable must be considered with it's relation to both the outcome and a main exposure - the authors still don't clearly state what the purpose of the models are in this regard(?) and therefore should be careful with the use of the word. Are they simply trying to make models looking at all risk factors possible? Some more careful thought about the relationships between the variables might be useful - some must be highly correlated and this could impact on the reported ORs. On that note, how does Table 3 work? Are the first set of variables also included in the subsequent models - surely the Odds Ratio values then change when the other variables are added? Why not give all the Odds Ratios - i.e. I can't see the one for obesity or sex.

A: Thank you for your comment. The reason why we adopted several models for adjustment is because there is NO single model where we can adjust for all the covariates and would be able to escape from the confounding factors. To explain this, take model 2 for example, we know that patients who were treated with azithromycin or hydroxychloroquine had to require oxygen therapy. It is an indication for treatment. Thus, if we only put age, sex, obesity as adjustments in our analysis – the result would not be accurate because we miss an important confounding factor; “the need for oxygen”. You are correct. The confounding factor is a variable that associates with BOTH independent variable and dependent variable. In this case, azithromycin or hydroxychloroquine were INdependent variables, and death is the dependent variable. Can “need for oxygen” affect the use of azithromycin or hydroxychloroquine? Yes, absolutely it is the indication for these treatments. Now, can “need for oxygen” affect death? A: Yes, absolutely. Patients with COVID-19 die from respiratory failure and other complications. Thus, we try to group the confounding factors the we know it will impact that variable together and add them in the analysis. The covariates that we chose are not random.

“Are they simply trying to make models looking at all risk factors possible?”

A: No, we are not trying to make models looking at all risk factors possible. If we want to identify as many risk factors as possible, we could have just included age, sex, obesity without considering for other confounding factors that might contribute to death as well – in this case we will have 30 or more significant variables. What we really want is the risk factors that are legit and may impact the clinical practice. That is why we included other factors in each different model. However, please note that basic patient characteristics, such as age, sex, ethnicity, and obesity were included in EVERY model. You can refer to the footnote of Table 3 for reference. If you would like us to use only one model that adjusted for only age, sex, ethnicity, obesity, we would be more than happy to do that if the reviewer 1 and the editor are agreeable with it.

“Some more careful thought about the relationships between the variables might be useful - some must be highly correlated and this could impact on the reported ORs.”

A: Yes, the development of each model to analyze the variables was derived from a consensus between the clinicians and our biostatistician in order to avoid potential confounding effect. We have invested a significant amount of time and thought in this process. Again, the variables included in each model are clinically relevant to each variable that we analyzed. Other studies have analyzed their data using different models too to make their findings as impactful as possible. We are not the only one. We apologize that we made you perceive that our choices of covariates included in the analysis were random and non-fundamental – but they are not.

"On that note, how does Table 3 work?"

A: To help you understand Table 3, let's take model 3 for example. We put "respiratory acidosis" and "steroids therapy" on there. It means that these variables were analyzed when we included age, sex, ethnicity, obesity, asthma/COPD, and need for oxygen therapy as covariates in the logistic regress analysis. Why did we include asthma/COPD and need for oxygen therapy? Because "respiratory acidosis" is an extremely common presentation of asthma/COPD patients and usually "steroids" are used in these patients. These patients tend to require oxygen – that's why we included "need for oxygen" in the model. Now, take model 4, for example, "CKD" which stands for chronic kidney disease, and "need for oxygen therapy" were the covariates for the analysis of the variable, "acute kidney injury". CKD has a known risk factor to both acute kidney injury (independent variable) and death (outcome). "The need for oxygen therapy" can also impact both acute kidney injury (independent variable) and death (outcome) as well. Thus, including these variables, CKD and need for oxygen is reasonable in the analysis. I am certain that our readers, who are mostly physicians that have clinical background would easily understand the findings demonstrated in Table 3.

"Are the first set of variables also included in the subsequent models"

A: You misunderstood. The model 1-5 are not the stepwise analysis. Each model had a unique set of variables included as adjustments. It is possible that some models may share the same set of covariables – but it is NOT a stepwise thing. For example, note that all models had "age, sex, ethnicity, obesity" as baseline adjustments because these are patient's baseline characteristics that may associate with death. Thus, we think it is reasonable to include these 'basic' variables as adjusting covariates.

"Why not give all the Odds Ratios - i.e. I can't see the one for obesity or sex."

A: Every model is adjusted for at least "age, sex, ethnicity, obesity". Please refer to the footnote of Table 3 for reference. We have provided all the odds ratios in Table 3 already regardless of their statistical significance. If there is a specific way that you would like it to be. Please feel free to let us know. Do you mean why you do not see obesity or sex as 'variables' in Table 3? Because they are not statistically significant on univariate analysis (Table 1). We do not perform multivariate analysis on variables that are not significant from the basic Chi square test.

To facilitate our mutual understanding, if you have a certain way that we could do to satisfy you. Please kindly let us know. It is quite broad to debate about this through the review as we are not sure what do you exactly want.

2. The answer surrounding the bootstrap still doesn't make sense to me either. What's the purpose there, and what is this actually doing? "The association of these variables and overall mortality remained significant on bootstrap analysis when the sample size was imputed to 1,000." Are you just simulating from the model and increasing the sample size? This will only increase power, no? i don't see how this validates anything?

A: Thank you for your comment. To prevent miscommunication and disagreement, we have deleted the section for sensitivity analysis and bootstrapping. All sentences that mentioned bootstrapping analysis were deleted. Please note that these changes will not affect the findings and results of our study. Thank you.

3. "In our analysis, the cumulative survival declined with increasing length of hospital stay." This is still not a great description of Figure 2. A survival curve has to go down or stay constant, by definition. Can more be added?

A: Thank you for your comment. We have deleted that sentence to avoid confusion to the readers. The remaining paragraph is, “Survival analyses were evaluated using Kaplan-Meier curve of all patients were presented in Figure 2A. The median survival time was 25.0 days with standard error of 7.0. The Kaplan-Meier curves for ICU and no ICU patients were illustrated in Figure 2B.”

4. " However, one could argue that our study has significantly smaller sample size which might underestimate the actual mortality of COVID-19. " Why would this be necessarily true?

A: Thank you for your comment. In this context, we are comparing the mortality rate from our study and the study from New York. In New York study, the mortality from COVID is higher. However, to avoid confusion, we have adjusted the sentence to, “Richardson et al found that the overall mortality of 5,700 hospitalized COVID-19 patients from 12 hospitals in New York City was 21%. However, one could argue that our study has a significantly smaller sample size. Our data need confirmation from other studies with a larger sample size.”

5. The reports of comparisons across lengths of stay by various population groups - these are all made difficult by the fact that non-survivors have a curtailed length of stay by definition. Be careful here.

A: Thank you for your warning. We understand that the length of stay may be curtailed by non-survivor groups. That is why we did not discuss Figure 2C-D in the Discussion. We also included the following sentence in the Limitation, “The length of stay was computed in the Kaplan-Meier analysis to represent the time to death. It is worth noting that the non-survivors had a curtailed length of stay.”

6. The authors have added this sentence for clarity: "The analysis is considered statistically significant if the 95% CI crosses 1.0." But...this is the wrong way round.

A: Thank you for your comment. We have modified the statement to, “A 95% CI that crosses 1.0 and a p-value of less than 0.05 is considered statistically significant.” Is there a specific way that you would like it this sentence to be?

Thank you, Dr. Rutherford for your time and effort in reviewing the revised manuscript and providing us with your thoughtful suggestions to improve our manuscript. We hope that our responses have satisfied you. We strongly believe that our research will be helpful to the readers in understanding the nature of COVID-19 in order to improve the quality of care to the patients. Should you have any questions or additional recommended guidance, we are delighted to get back to you.

VERSION 3 – REVIEW

REVIEWER	Giulio Francesco Romiti Sapienza - University of Rome, Rome, Italy
REVIEW RETURNED	07-Oct-2020

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript again. The authors sufficiently addressed all the issues that I have raised in previous round of review, and I think that the manuscript clearly improved. I have no further comments.