Journal of Clinical and Translational Research
Peer review process file 07.202103.010

Hospital length-of-stay among COVID-19 positive patients

Tze Chiam, Keshab Subedi, David Chen, Eric Best, Federica B. Bianco, Gregory Dobler, Mia Papas

Corresponding author
Keshab Subedi
Value Institute, ChristianaCare, 4755 Ogletown-Stanton Rd, Newark, DE 19718

Handling editor:
Michal Heger
Department of Pharmaceutics, Utrecht University, the Netherlands
Department of Pharmaceutics, Jiaxing University Medical College, Zhejiang, China

Review timeline:

Received: 12 March, 2021
Editorial decision: 15 April, 2021
Revision received: 12 May, 2021
Editorial decision: 13 May, 2021
Published online: 5 June, 2021

Ref.: Ms. No. JCTRes-D-21-00036
Hospital length-of-stay among COVID-19 positive patients
Journal of Clinical and Translational Research

Dear Mr Subedi,

Reviewers have now commented on your paper. You will see that they are advising that you revise your manuscript. If you are prepared to undertake the work required, I would be pleased to reconsider my decision.

For your guidance, reviewers' comments are appended below.

If you decide to revise the work, please submit a list of changes or a rebuttal against each point which is being raised when you submit the revised manuscript. Also, please ensure that the track changes function is switched on when implementing the revisions. This enables the reviewers to rapidly verify all changes made.

Your revision is due by May 15, 2021.

To submit a revision, go to https://www.editorialmanager.com/jctres/ and log in as an Author. You will see a menu item call Submission Needing Revision. You will find your submission record there.

Yours sincerely

Michal Heger
Editor-in-Chief

Journal of Clinical and Translational Research

Reviewers' comments:

Editor: Dear authors, thank you for submitting your work to JCTR. We have fast-tracked your paper because of its importance. References to sections in your paper in this review have the following format: page # line #s (P#L#).

- P5L41: there are no treatments for SARS-CoV-2, and none have become available during the course of the last year. Vaccines are not considered treatment. Please rephrase to e.g., “pharmacological interventions and clinical management strategies” to better reflect the status quo.

- P6L37-54: I do not understand why you chose to exclude patients who died in the hospital. It seems to me that in your multivariate analysis a lot of critical data could be obtained on the possible factors that may have resulted in lethality. These are equally important in the clinical management of COVID-19 patients, in addition to providing clues on prognosis regarding SARS-CoV-2 infection. The same applies to patients who had an LOS of > 40 d. I can imagine that the sample size was ultimately too small for valuable analytics (N = 98 deaths and N = 26 LOS > 40 d), but perhaps denoting that premise for exclusion in the text provides the readers with the proper context for the exclusions.

- P7: please describe which statistical test was conducted to ascertain a normal distribution of the data set to justify subsequent statistical methods that rely on a Gaussian population.

- P8L4-40: Figure 1 is actually a result and should be placed in the respective section.

Reviewer #1: This paper is well-written and timely. It addresses the current literature gap in understanding the contributing factors of LOS for COVID-19 patients in a US setting. The study is thorough and the statistical analysis is sound. Here are some comments.
o There is 16% of Other Race and I wonder which racial groups make up this Other Race. Are there any insights in the data that can be shared?
o Please clarify how you count the patients who have double eligibility (Medicaid and Medicare).
o I wonder if Model 3 results will change if both arrival day and arrival shift are removed.
o The study excluded all the patients who died (98) in the hospitals. The number is significant compared to the ICU patients in this study (146). However, those patients likely also need ICU resources. Authors may want to bring this point to the overall discussion, especially the conclusion of the paper focusing on improving the prediction accuracy of COVID-19 patient census in hospitals for resource planning and care delivery.

Reviewer #2: 1. The stats don't make sense. They say "We excluded i) patients who died in the hospital, ii) patients who had hospital LOS of more than 40 days, and iii) patients who were not discharged as of August 11, 2020" So they exclude the sickest patients and those who stay in the hospital the longest? Survival analyses are designed to deal with "censored" data -- you know someone "survived" (stayed in hospital in this case) up to time X but don't

know after that (because they left the study or are still 'alive" at time of data collection. Since the authors didn't do the standard survival analysis, I wouldn't believe the results.

2. What's the point? It shows older and sicker patients stay in the hospital longer. Who would have guessed? I suppose the goal must be to predict hospital stay (which could be useful for planning). If so, they need to test their predictive model on new patients. Just describing one set of patients is not enough. Models always over predict, and need a reality check with new data.

Reviewer #3: In general, the statistical analysis is sound. However I would advise the following changes:
1. For the log-rank tests correct for multiple testing to better control the type I error rate
2. Include non-ICU/ICU as covariate in the accelerated failure time models corresponding to Table 2.
3. Extend the section on limitations and also mention:
a. retrospective nature of the study
b. relatively modest sample size
c. the fact that hospital that patients who had an LOS of more than 40 days and patients who were not discharged as of August 11, 2020 were excluded. These could have been included and, for those not discharged, been handled using right censoring

In terms of presentation, I suggest:
1. Fig. 1: move to the supplementary information
2. Table 1
a. Explain NHL and HL in the legend
b. Replace 'gender' by 'sex' (also in the corresponding Kaplan-Meier curve)
c. Include information for all variables that are used in the survival analysis, for example the different comorbidities
3. Fig. 2: leave out (since this information is already included in Appendix 1) and include Bonferroni-adjusted p-values for the log-rank tests as an additional column in Table 1
4. In the text (p.12, l.10) there is a reference to Table 3. This table doesn't exist. Do the authors mean Appendix 2? Please correct.
5. The discussion should be restructured to more clearly focus on the most important findings. Furthermore:
a. p.12-13: "Analysis also shows that the age of hospitalized COVID-19 patients decreases since the pandemic started. Hence, the decrease of LOS is also consistent with the age association with LOS discussed above." Add a remark that the accelerated failure time models (Table 2, Appendix 2) indicate that the associations with age and week of onset of the pandemic are independently associated with LOS.
b. On p.13, l.27 there is a reference to Richardson et al. 2020, which is not included in the bibliography. Please add.
c. Related to the discussion on association of gender with LOS (p.13, l.36-56) the results (Table 2, Appendix 2) suggest a possible interaction between gender and whether a patient had to be sent to the ICU or not. I therefore advise to include non-ICU/ICU as covariate in the accelerated failure time models corresponding to Table 2 and discuss these results.
6. Appendix 1. Put Kaplan-Meier curves in the same order as the clinical variables in Table 1

Authors' response

Reviewers' comments:

Editor: Dear authors, thank you for submitting your work to JCTR. We have fast-tracked your paper because of its importance. References to sections in your paper in this review have the following format: page # line #s (P#L#).

- P5L41: there are no treatments for SARS-CoV-2, and none have become available during the course of the last year. Vaccines are not considered treatment. Please rephrase to e.g., "pharmacological interventions and clinical management strategies" to better reflect the status quo.

Agreed and updated

- P6L37-54: I do not understand why you chose to exclude patients who died in the hospital. It seems to me that in your multivariate analysis a lot of critical data could be obtained on the possible factors that may have resulted in lethality. These are equally important in the clinical management of COVID-19 patients, in addition to providing clues on prognosis regarding SARS-CoV-2 infection. The same applies to patients who had an LOS of > 40 d. I can imagine that the sample size was ultimately too small for valuable analytics (N = 98 deaths and N = 26 LOS > 40 d), but perhaps denoting that premise for exclusion in the text provides the readers with the proper context for the exclusions.

Thank you for this critical observation. We sourced the data from the electronic health records and computed the length of stay as time from admission to discharge or death. However, the death time or discharge time are not accurately captured in many instances for patients who died in hospital. For that reason, we opted to not included patients who died. We agree the data has potential to be use for mortality related outcomes, however, we did not include mortality related analysis in this paper to remain focused on analysis of hospital length-of-stay. We excluded patients with LOS>40 d and who were not discharged as of August 11, 2020 to make sure everyone has at least a 40-day of follow up time. We have modified manuscript text related to inclusion/exclusion criteria to better explain this. Now the text reads-

"*We excluded i) patients who died in the hospital, ii) patients who had hospital LOS of more than 40 days, and iii) patients who were not discharged as of August 11, 2020. This cut-off date was selected to provide a 40-day follow-up period for patients admitted on the last day of the study period. For patients with multiple hospitalizations related to COVID-19 within the study period, only the first visit was included.*"

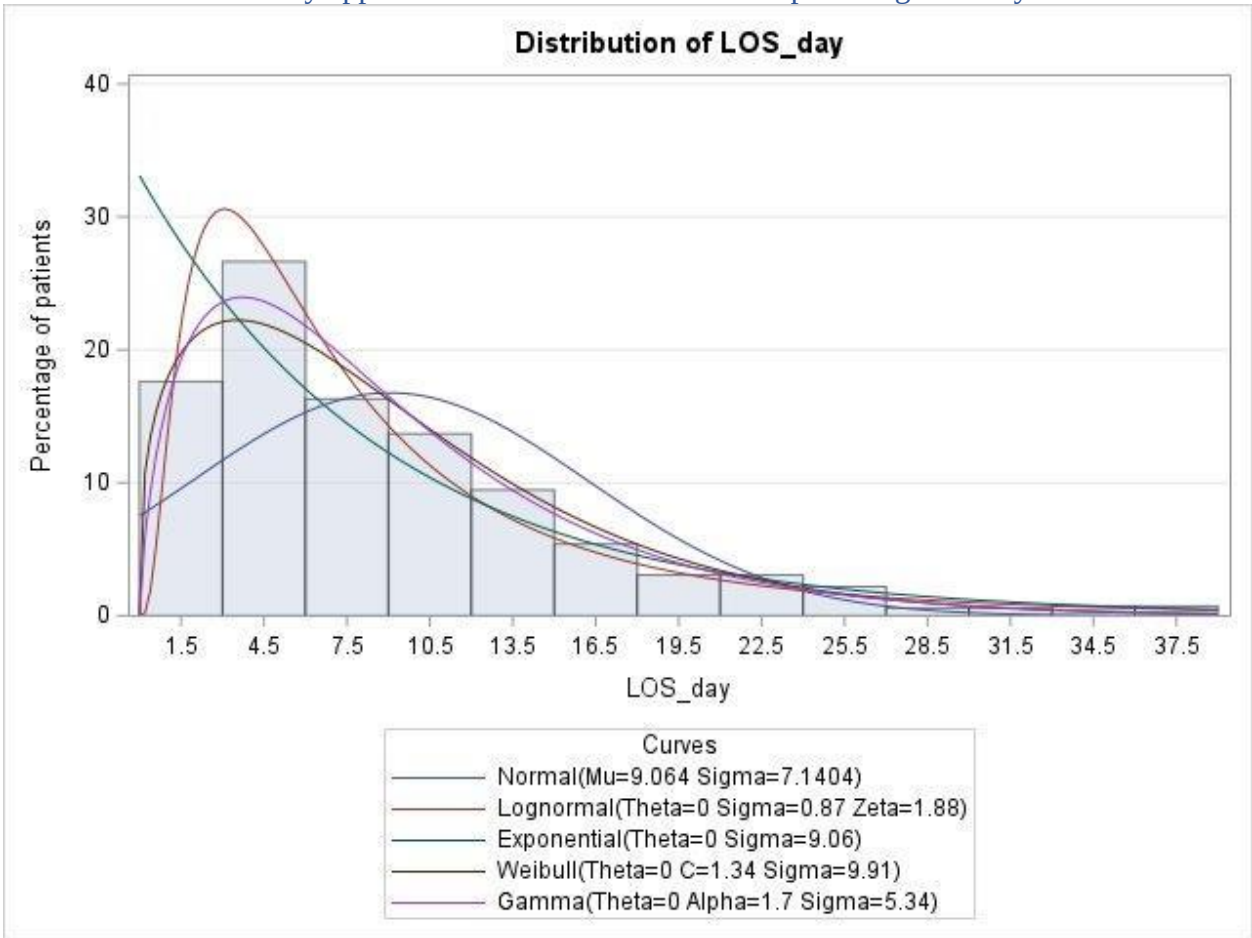
- P7: please describe which statistical test was conducted to ascertain a normal distribution of the data set to justify subsequent statistical methods that rely on a Gaussian population.

The models used do not depend on the assumption of normal distribution of outcome. The final models are based on Gamma distribution. We selected Gamma distribution among other

distribution within exponential family of distribution based on model fit indices -log likelihood and AIC. This is explained in methods as:
"To evaluate the effect of patients' demographic and clinical characteristics on the LOS we utilized a multivariable accelerated failure time (AFT) models with exponential, log-normal, loglogistic, Weibull and Gamma distribution. The best fitting model was selected based on loglikelihood statistics."

In the Result section we mentioned that the Gamma model were found to be best fitting based on log-likelihood. We also updated the Analysis section to emphasize the Gamma model to provide further clarity.

Also, visual evaluation of distribution of hospital length of stay supports that gamma distribution most closely approximates the distribution of hospital length of stay.



- P8L4-40: Figure 1 is actually a result and should be placed in the respective section.
Agreed. Figure 1 is now moved to the result section.

Reviewer #1: This paper is well-written and timely. It addresses the current literature gap in understanding the contributing factors of LOS for COVID-19 patients in a US setting. The study is thorough, and the statistical analysis is sound. Here are some comments.

o There is 16% of Other Race and I wonder which racial groups make up this Other Race. Are there any insights in the data that can be shared?

We grouped race into three categories -Black, White and the other. The other categories include American Indian, Asian, Pacific Islander, Other races, and Unknown races. Some of those categories have few patients (<1%), so both for data privacy purpose and ease of interpretation we grouped all these together in the Other race group. We have also updated the manuscript with this explanation as a footnote to Table 1

o Please clarify how you count the patients who have double eligibility (Medicaid and Medicare).

For people with multiple insurance, we used the one which was registered as primary insurance during the hospital admission (updated in the manuscript).

o I wonder if Model 3 results will change if both arrival day and arrival shift are removed.
If we remove arrival day and arrival shift from the model three, the direction and significance of the estimates for rest of the parameters remain same, though the point estimate slightly change.
The arrival day and shift are documented to be important confounder for length of stay outcomes, so we want to keep them as confounder, though they are not significant in our model.

o The study excluded all the patients who died (98) in the hospitals. The number is significant compared to the ICU patients in this study (146). However, those patients likely also need ICU resources. Authors may want to bring this point to the overall discussion, especially the conclusion of the paper focusing on improving the prediction accuracy of COVID-19 patient census in hospitals for resource planning and care delivery.

See the previous response to the editor comment 2.

Reviewer #2: 1. The stats don't make sense. They say "We excluded i) patients who died in the hospital, ii) patients who had hospital LOS of more than 40 days, and iii) patients who were not discharged as of August 11, 2020" So they exclude the sickest patients and those who stay in the hospital the longest? Survival analyses are designed to deal with "censored" data -- you know someone "survived" (stayed in hospital in this case) up to time X but don't know after that (because they left the study or are still 'alive" at time of data collection. Since the authors didn't do the standard survival analysis, I wouldn't believe the results.

Having the LOS of 40 days as cut off and excluding individuals not discharged as of Aug 11 2020 provides a follow up time of at least 40 days for all the patients (as we included patients who were admitted up to June 30, 2020). We explained our rationale for excluding the patients who died in earlier responses. The inclusion/exclusion framework does not exclude the sickest patients. Rather, by defining the inclusion criteria as: patients admitted up to June 30, 2020; and exclusion criteria as i) LOS >40, ii) not discharged as of August 11, 2020 we made sure that everyone got at least 40 days of follow up time after their admission. This actually prevents removing the sickest patients in contrary to what the reviewer has pointed.

2. What's the point? It shows older and sicker patients stay in the hospital longer. Who would have guessed? I suppose the goal must be to predict hospital stay (which could be useful for

planning). If so, they need to test their predictive model on new patients. Just describing one set of patients is not enough. Models always over predict, and need a reality check with new data.

It is not our intention to provide a predictive model in this paper. Hence we did not do a validation study (the "reality check" mentioned by the reviewer). Our objective in this paper is to study the association of clinical and demographics factors with hospital LOS retrospectively, not to predict.

Reviewer #3: In general, the statistical analysis is sound. However I would advise the following changes:
1. For the log-rank tests correct for multiple testing to better control the type I error rate

The Kaplan-Meir curves and corresponding log rank test were provided as a part of exploratory analysis and not for definitive inference.
2. Include non-ICU/ICU as covariate in the accelerated failure time models corresponding to Table 2.
This is a very important point however we decided not include ICU/NON-ICU as covariate for the following reason. 1. Simply having icu/non-ICU is not meaningful as we do not know the ICU/NON ICU at the time of admission and the estimate for all the parameters would depend upon some future value Second, if icu/non-icu is mediating some of the relationships we showed then there is still value in showing the estimates of those relationships without controlling for the mediator. However, we agree that there is scope of separate mediation/pathway analysis to understand the length-of-stay.

3. Extend the section on limitations and mention:
a. retrospective nature of the study Updated
b. relatively modest sample size Updated
c. the fact that hospital that patients who had an LOS of more than 40 days and patients who were not discharged as of August 11, 2020 were excluded. These could have been included and, for those not discharged, been handled using right censoring

We used that approach as well. However, due to data quality issue related to length of stay of people who died (due to new hospital process of handling deceased patients), we only included results from the analysis where we excluded the patients who died.

In terms of presentation, I suggest:
1. Fig. 1: move to the supplementary information Updated – This suggestion differs with the editor's suggestion and we have decided to follow the editor's suggestion of moving it to the Results section
2. Table 1
a. Explain NHL and HL in the legend Updated
b. Replace 'gender' by 'sex' (also in the corresponding Kaplan-Meier curve) Updated
c. Include information for all variables that are used in the survival analysis, for example the different comorbidities The reference to Elixhauser is provided in the manuscript (reference #20) that includes all the standard definitions of comorbidities as part of the Elixhauser algorithm.

3. Fig. 2: leave out (since this information is already included in Appendix 1)
Updated and include Bonferroni-adjusted p-values for the log-rank tests as an additional column in Table 1
Done.

4. In the text (p.12, l.10) there is a reference to Table 3. This table doesn't exist. Do the authors mean Appendix 2? Please correct. Updated
5. The discussion should be restructured to more clearly focus on the most important findings.
We moved the discussion on comorbidities and demographics to the first paragraph as they were likely of highest interests to the audience, and removed the discussion on random forest and bootstrapping to hopefully provide a more focused discussion on the work performed.
Furthermore:
a. p.12-13: "Analysis also shows that the age of hospitalized COVID-19 patients decreases since the pandemic started. Hence, the decrease of LOS is also consistent with the age association with LOS discussed above." Add a remark that the accelerated failure time models (Table 2, Appendix 2) indicate that the associations with age and week of onset of the pandemic are independently associated with LOS. Updated the manuscript with "… LOS discussed above, assuming no interactions among factors used in the study".
b. On p.13, l.27 there is a reference to Richardson et al. 2020, which is not included in the bibliography. Please add. Updated
c. Related to the discussion on association of gender with LOS (p.13, l.36-56) the results (Table 2, Appendix 2) suggest a possible interaction between gender and whether a patient had to be sent to the ICU or not. I therefore advise to include non-ICU/ICU as covariate in the accelerated failure time models corresponding to Table 2 and discuss these results. (Please see response to comment #2 above)
6. Appendix 1. Put Kaplan-Meier curves in the same order as the clinical variables in Table 1
Done

2nd Editorial decision
13-May-2021

Ref.: Ms. No. JCTRes-D-21-00036R1
Hospital length-of-stay among COVID-19 positive patients
Journal of Clinical and Translational Research

Dear authors,

I am pleased to inform you that your manuscript has been accepted for publication in the Journal of Clinical and Translational Research.

You will receive the proofs of your article shortly, which we kindly ask you to thoroughly review for any errors.

Thank you for submitting your work to JCTR.

Kindest regards,

Michal Heger
Editor-in-Chief

Journal of Clinical and Translational Research

Comments from the editors and reviewers: