

## 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

<b>TITLE (PROVISIONAL)</b>	Early prediction of in-hospital mortality in acute pancreatitis: A retrospective observational cohort study based on a large multicentre critical care database
<b>AUTHORS</b>	li, caifeng; Ren, Qian; Wang, Zhiqiang; wang, guolin

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Masayasu Horibe Keio university, Japan
<b>REVIEW RETURNED</b>	17-Jul-2020

<b>GENERAL COMMENTS</b>	<p>Authors showed that early prediction of in-hospital mortality in acute pancreatitis. Although the aim of this article is very interesting, authors should revisit following points.</p> <p>Major comments</p> <ol style="list-style-type: none"><li>1. More than half of the cases were excluded from creating a prognostic model because of lack of complete patient data on the variables included in the prognostic model. The prognostic models should be created using only those variables for which more than 80-90% of patients have been tested for to avoid selection bias. If the authors use only variables that majority of patients had, the number of excluded patients would be small and generalizability of prognostic model would improve.</li><li>2. The authors excluded 364 people due to insufficient data. Please add to Table 1 the number of patients, whose data is available for each variable. In the current table, it looks like all patients had been tested for all variables.</li><li>3. The number of variables that can be adjusted in multivariate analysis depends on the number of outcomes. In general, if the number of events is 30, the authors can adjust for 3 items. In this case, despite the fact that there were 33 death events, more than 10 variables were included in the model. This is statistically inappropriate.</li><li>4. How many deaths out of the 378 people were used to create the monograph?</li><li>5. The C-index for Apache IV is a result of 606 people. When the authors compare the ABL and Apache IV, the authors should limit those who were able to calculate both. Moreover, the author should use Delong's test method to compare two models. [Biometrics. 1988;44:837-45.]</li><li>6. How many cases were transferred to the ICU some time after the diagnosis of pancreatitis? The authors should exclude these patients to predict morbidity and mortality in directly transferred ICU patients.</li><li>7. The mortality rate for mild pancreatitis is less than 5% and is even higher in severe cases. Although this database is from for</li></ol>
-------------------------	--

	<p>patients admitted to the ICU, the mortality rate is 4.4% as low as mild pancreatitis. Why is this?  Minor comment.  There is a lot of overlapping content in the table and the text. Only the important results shown in the table should be included in the text and the rest should be removed.</p>
--	--

<b>REVIEWER</b>	Anirban Hom Choudhuri GB Pant Institute of Post Graduate Medical Education & Research New Delhi India
<b>REVIEW RETURNED</b>	20-Jul-2020

<b>GENERAL COMMENTS</b>	This is a retrospective study conducted from a large multicentric ICU database (eICU-CRD) by enrolling 378 patients to establish a nomogram model for predicting the in hospital mortality risk in acute pancreatitis and then internally validating the model. The approach has been scientific and the script lucidly explaining the facts. The research results are useful for internists, intensivists and gastroenterologists.
-------------------------	---

<b>REVIEWER</b>	Laszlo Czako Department of Medicine, University of Szeged
<b>REVIEW RETURNED</b>	25-Jul-2020

<b>GENERAL COMMENTS</b>	<p>The aim of this study was to develop and validate a prediction model in-hospital mortality in patients with acute pancreatitis in a retrospective multicenter study. Age, BUN, and lactate (ABL) were proved to be independent risk factors for in-hospital mortality. The generated nomogram ABL model was validated and proved to be as effective as APACHE IV.</p> <ol style="list-style-type: none"> <li>1. Why different patients' cohorts were included for the analysis of potential risk factors and in the nomogram model? Was the difference between the two cohorts examined? This should be added in Results. Which data were missing for the nomogram model?</li> <li>2. It seems that the APACHE IV score and the nomogram model ABL were not calculated in the same patients. Was the difference between the two cohorts examined? How this may have influenced the results?</li> <li>3. What was the CI for the C-index during internal validation (page 13, line 35)?</li> <li>4. It would elevate the quality of the manuscript, if the nomogram have had compared with more often used and specific prognostic scores such as BISAP, SIRS, SOFA.</li> </ol>
-------------------------	---

### VERSION 1 – AUTHOR RESPONSE

#### Major comments

1. More than half of the cases were excluded from creating a prognostic model because of lack of complete patient data on the variables included in the prognostic model. The prognostic models should be created using only those variables for which more than 80-90% of patients have been tested for to avoid selection bias. If the authors use only variables that majority of patients had, the

number of excluded patients would be small and generalizability of prognostic model would improve. Thank you very much for pointing out the shortcomings of our study. Indeed, an ideal prognostic model should be reliable and easy-to-use at the early stage of the disease, using readily available variables. Actually, we tried our best to screen such easy-obtainable and popular variables to establish a predictive model. Univariate and multivariate logistic regression analysis showed that Age, BUN, Lactate were independent risk factor for predicting in hospital mortality in AP. Due to the incomplete data on independent risk factors (Age, BUN and Lactate), many cases were excluded from the establishment of prognostic models, which may lead to selection bias. Therefore, we examined the difference of baseline characteristics between included and excluded patients (Table s). The only significant difference between the two cohorts was weight, providing evidence for only limited potential selection bias despite a large number of excluded patients. Due to the small sample size and small number of events, large scale clinical studies are still needed to externally validate the model in the future. Thanks!

2. The authors excluded 364 people due to insufficient data. Please add to Table 1 the number of patients, whose data is available for each variable. In the current table, it looks like all patients had been tested for all variables.

Thank you very much for your kind suggestion. We have added the information required to Table 1 (page 23-24, Table 1). Thanks!

3. The number of variables that can be adjusted in multivariate analysis depends on the number of outcomes. In general, if the number of events is 30, the authors can adjust for 3 items. In this case, despite the fact that there were 33 death events, more than 10 variables were included in the model. This is statistically inappropriate.

It is really a good question. Thanks a lot for pointing out the shortcomings of this study (page 16, line 12-15). As you mentioned, the number of variables that can be adjusted in multivariate analysis depends on the number of outcomes. Due to relatively small sample size and small number of events, the statistical power in the analyses remained low, large-scale prospective cohorts are needed to verify our results. Thanks!

4. How many deaths out of the 378 people were used to create the monograph?

Thank you for your efforts and comments on our manuscript. There are 21 deaths out of the 378 patients to establish the monograph. Thanks!

5. The C-index for Apache IV is a result of 606 people. When the authors compare the ABL and Apache IV, the authors should limit those who were able to calculate both. Moreover, the author should use Delong's test method to compare two models. [Biometrics. 1988;44:837-45.]

Thanks for your nice comments. The paper you provide helps me a lot. In the revised manuscript, we compare the ABL and Apache IV in patients who were able to calculate both. (page 11, line 11-17; Figure 5) Thanks!

6. How many cases were transferred to the ICU some time after the diagnosis of pancreatitis? The authors should exclude these patients to predict morbidity and mortality in directly transferred ICU patients.

Thank you very much for your constructive suggestion. Indeed, these two types of patients should be distinguished to reduce the heterogeneity of the study. I am so sorry that the database does not provide information on the duration of disease before admission to ICU, it is also a shortcoming of this study. (page 16, line 18-19). We hope that there will be large-scale studies in the future to prove this viewpoint. Thanks!

7. The mortality rate for mild pancreatitis is less than 5% and is even higher in severe cases. Although this database is from for patients admitted to the ICU, the mortality rate is 4.4% as low as mild pancreatitis. Why is this?

Thank you for your kind suggestion. The mortality rate of mild pancreatitis was about 5% in the past. With the increased understanding of the pathogenesis and pathophysiology of pancreatitis and the development of medical technology in recent years, especially critical care medicine, AP patients benefit from closer monitoring, early aggressive fluid resuscitation, rational use of antibiotics, and timely invasive strategies, the mortality rate of pancreatitis has gradually declined. That may be why

the mortality rate is as low as mild pancreatitis. Thanks!

Minor comment.

There is a lot of overlapping content in the table and the text. Only the important results shown in the table should be included in the text and the rest should be removed.

Thank you very much for your kind advice. The overlapping content in the text has been removed as suggested. (page 8, line 11-14). We really appreciate your kindness. Thanks!

COMMENTS FROM Reviewer 2:

This is a retrospective study conducted from a large multicentric ICU database (eICU-CRD) by enrolling 378 patients to establish a nomogram model for predicting the in hospital mortality risk in acute pancreatitis and then internally validating the model. The approach has been scientific and the script lucidly explaining the facts. The research results are useful for internists, intensivists and gastroenterologists.

Dear professor Anirban Hom Choudhuri,

Thank you very much for your kindness. We really appreciate for your nice comments. Thanks!

COMMENTS FROM Reviewer 3:

Dear professor Laszlo Czako,

Thank you very much for your valuable and helpful suggestions and comments. We have revised our manuscript very carefully. Hope the revised manuscript is acceptable for publication.

1. Why different patients' cohorts were included for the analysis of potential risk factors and in the nomogram model? Was the difference between the two cohorts examined? This should be added in Results. Which data were missing for the nomogram model?

Thank you very much for your helpful advice. The patients' cohorts were different in risk factor selection and in the nomogram development, there are 742 patients included for risk factor selection, while 378 eligible patients were included for nomogram development, with 364 patients excluded due to incomplete data on Age, BUN, Lactate. We examined the difference of baseline characteristics between included and excluded patients as suggested (Table s). The only significant difference between the two cohorts was weight, providing evidence for only limited potential selection bias despite a large number of excluded patients. Thanks!

2. It seems that the APACHE IV score and the nomogram model ABL were not calculated in the same patients. Was the difference between the two cohorts examined? How this may have influenced the results?

Thank you very much for your constructive suggestion. In previous manuscript, the APACHE IV score and the nomogram model ABL were not calculated in the same patients. There was no statistically significant difference between the C-index of the nomogram model ABL and that of APACHE IV ( $P = 0.499$ ). In the revised manuscript, we compared the predictive accuracy for in-hospital mortality between the nomogram model ABL and APACHE IV in the same cohort, There was also no statistical difference between nomogram model ABL and APACHE IV in the C-index of ( $P = 0.086$ ), indicating the nomogram model ABL had a similar predictive accuracy with APACHE IV. (page 11, line 11-17).

Thanks!

3. What was the CI for the C-index during internal validation (page 13, line 35)?

Thank you very much for your kind suggestion. The C-index during internal validation was 0.892(95%CI: 0.822 to 0.962). The manuscript has been revised as suggested. (page 10, line 18). Thanks!

4. It would elevate the quality of the manuscript, if the nomogram have had compared with more often used and specific prognostic scores such as BISAP, SIRS, SOFA.

Thank you very much for your kind comments. We believe that comparing nomogram ABL with more commonly used and more specific prognostic scores would elevate the quality of the manuscript. However, there were no sufficient data to calculate these prognostic scores. We hope there will be large-scale clinical studies to verify it in the future. Thanks!

## VERSION 2 – REVIEW

<b>REVIEWER</b>	Masayasu Horibe Mayo Clinic, USA
<b>REVIEW RETURNED</b>	23-Sep-2020

<b>GENERAL COMMENTS</b>	I believe that the matters I have pointed out have been appropriately addressed.
-------------------------	--

<b>REVIEWER</b>	Laszlo Czako University of Szeged, Hungary
<b>REVIEW RETURNED</b>	23-Sep-2020

<b>GENERAL COMMENTS</b>	none
-------------------------	------