July 25, 2022

Lian-Sheng Ma

Editorial Office Director, Company Editor-in-Chief

World Journal of Gastroenterology

Dear Editor:

We wish to re-submit the manuscript titled "**Online Calculator for Predicting the Risk of Malignancy in Patients with Pancreatic Cystic Neoplasms: A Multicenter, Retrospective Study**". The manuscript was originally titled "**Clinical Online Calculator for Predicting the Individual Risk of Malignancy in Patients with Pancreatic Cystic Neoplasms: A Multicenter, Retrospective Study**". The manuscript number is 78250.

We thank you and the reviewers for your thoughtful suggestions and insights. The manuscript has benefited from these insightful suggestions. Moreover, the comments provided significant guidance to our study. I look forward to working with you and the reviewers to move this manuscript closer to publication in the *World Journal of Gastroenterology.*

The manuscript has been rechecked and the necessary changes have been made in accordance with the reviewers' suggestions. The revised sections are marked in boldface letters or strikethrough in the manuscript. The responses to all comments have been prepared and attached herewith.

Reviewer #1:

Scientific Quality: Grade D (Fair)

Language Quality: Grade A (Priority publishing)

Conclusion: Major revision

Specific Comments to Authors:

1.Methodology: authors excluded some other types of cystic lesions, How can clinicians decided to use/ or not use the proposed calculator during clinical practice?

Response: Dear reviewer, many thanks for taking your precious time to review our work. Thank you for the pertinent question. In this article, we developed a nomogram-based online calculator which provided important information for surgeons and patients in assessing an individual's risk for malignant lesions. We did exclude some rare cystic lesions of the pancreas such as solid pseudopapillary tumors (SPTs), cystic neuroendocrine tumors (cNET) and other undefined cystic tumors of the pancreas. Because these cystic lesions account for less than 10% of pancreatic cystic neoplasms[1], and most of these lesions are with typical clinical and radiological features so that they can be easily identified[2, 3]. During clinical practice, clinicians were suggested to use the calculator when they are unable to confirm the nature of the PCNs included in the study or when the patient and physician are unsure about whether surgical operation should be taken. The cystic lesions we excluded were not involved in the development of the model, so the performance of the model won't be affected by the exclusion. The study followed clear and strict inclusion and exclusion criteria ,the excluded cyst lesions were not in the scope of the study. Therefore, exclusion of these lesions didn't cause obvious selection bias. We are honored to work with you to raise the article to a higher level. Our deepest respect for you.

2.Methodology: no details about steps of creating the online calculator

Response: Thank you very much for your comments. The steps of creating the online calculator are as follows. Firstly, the package named DynNom was invoked in the RStudio to build the regression equation. Then the link to the associated terminal code was then generated for public use though the website shinyapps.io. Meanwhile, we have added the relevant content to the method. (The part "***Statistical analysis***" of the manuscript, rear position, marked in boldface letters)

3.Methodology: Why did not you use machine learning algorisms rather than simple multivariate nomogram model

Response: Thank you for your helpful comments. This is indeed a good proposal; we express our admiration for your profound insights. Machine learning algorisms has many advantages. For example, advantages of machine learning include flexibility and scalability compared with traditional statistical methods, which makes it deployable for many tasks, such as risk stratification, diagnosis and classification, and survival predictions. Another advantage of machine learning algorithms is the ability to analyze diverse data types (e.g., demographic data, laboratory findings, imaging data, and doctors' free-text notes) and incorporate them into predictions for disease risk, diagnosis, prognosis, and appropriate treatments. Despite these advantages, the application of machine learning in health-care delivery also presents unique challenges that require data pre-processing, model training, and refinement of the system with respect to the actual clinical problem. Also crucial are ethical considerations, which include medico-legal implications, doctors' understanding of machine learning tools, and data privacy and security. Unfortunately, we have not mastered the skill so that we are unable to develop the model by the method you suggested. However, your helpful suggestions broaden our horizon, makes us realize our own shortcomings, and also provides valuable reference for the direction of our future efforts.

4.Methodology: The study focusses on the developing and validation a nomogram, why to add preoperative, surgical and postoperative details? "Please move it to a supplementary document"

Response: Thank you for your recommendations. We have carefully considered your opinion. That's a reasonable advice. After reflection and discussion, we think that the part "preoperative evaluation" helps us learn about the general information of the patients and the various examinations performed before surgery. Meanwhile, this part also reveals the main factors studied in this paper. Therefore, we would prefer to retain most of "preoperative evaluation". As for surgical and postoperative details, we quiet agree with your opinion, in order to maintain the structural integrity of the article and the therapeutic process. With the advices and guidance of the reviewers,

we have streamlined the contents of these two parts and the unnecessary or controversial expressions were deleted. Finally, in order to make the focus of the article more prominent, we have integrated preoperative, surgical and postoperative details into perioperative management. We sincerely hope that the revision may gain your understanding and support. (The part "***Perioperative Management***" of the manuscript, marked in strikethrough)

5.Methodology: AUC is not enough for discriminatory performance assessment. Please add other more in-depth statistical techniques.

Response: Thank you very much for your suggestions. We quite agree with your viewpoint about that only AUC is not enough for discriminatory performance assessment. So in the study, we also applied other quality indices to assess the performance of the model such as calibration plots with bootstrap samples, decision curve analysis (DCA), clinical impact curves. Besides, we have also added the description of the sensitivity, specificity, accuracy, positive predictive value, and negative predictive value with the associated 95%CI of the model performance in the training and validation cohorts, respectively. We sincerely hope that after adding these indicators, the assessment of performance can meet your expectations. We have added the relevant content to the method. (The part "***Statistical analysis***" and "***Comparison of the performance of the nomogram and the risk factors identified in the relevant guidelines***" of the manuscript, middle position, marked in boldface letters)

6.Results: the authors stated" In the training cohort, the nomogram achieved a C-index of 0.824 for predicting the risk of malignancy. The predictive ability of the model was validated in an external cohort (C-index: 0.893)". Is it logic that model performance on the external validation dataset surpassed the model performance during training?? Please this point must be explained in detail.

Response: Thank you very much for your comments. The question you raised is very profound. Firstly, our study applied independent external validation. External
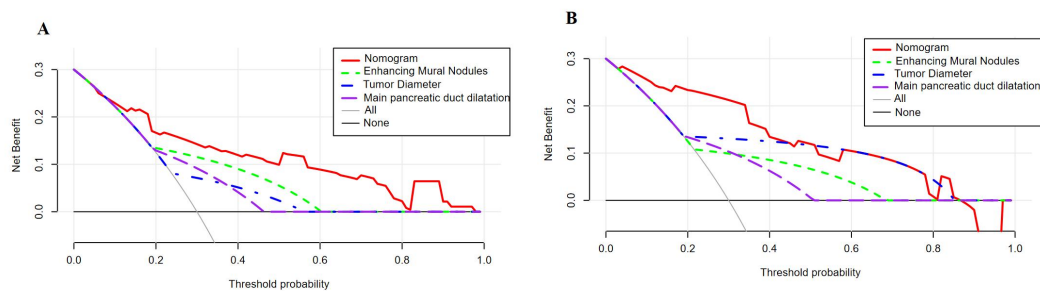
validation is the action of testing the original prediction model in a set of new patients to determine whether the model works to a satisfactory degree. External validation is necessary to assess a model's reproducibility and generalizability[4]. Secondly, the concordance index (C-index) was used to assess discrimination of the model. When computing the C-index, all possible pairs between a patient with and without the outcome are analyzed. A pair is concordant if the patient with the outcome has a higher predicted risk than the patient without the outcome. A C-index of 1 is perfect and 0.5 is equivalent to chance. A C-index of 0.60 means 60% of all possible pairs were concordant, and this is generally considered rather poor discrimination, while a C-index of 0.8 is usually considered good and ≥ 0.9 is excellent[5]. In conclusion, C-index assesses whether patients who experience the outcome have a higher predicted risk than patients who do not. For discrimination, it does not matter if the absolute value of C-index is 0.824 or 0.893, as long as the C-index maintain a high level in both the training and validation cohort. Therefore, it is not necessary to compare the C-index between the training and the validation cohort[4]. Meanwhile, the high level of C-index in the validation cohort proves that the model possesses good reproducibility and generalizability. We sincerely hope that our explanation will satisfy you, once again, our heartfelt thanks.

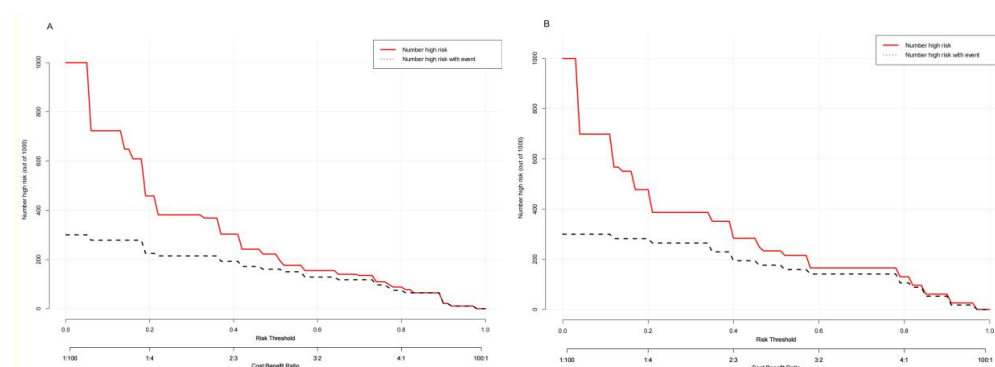7.Results, DCA and CICA is poorly interpreted. Normal readers should have more details explanations.

Response: Thank you very much for your useful comments. Under the guidance of the reviewers, new DCA curves have been drawn to compare the net benefit between the nomogram and the other three factors. Meanwhile, detailed interpretation of the DCA and CICA has been added in the methods and results. We have added the relevant content to the results. (The part "***Statistical analysis***" and "***Comparison of the performance of the nomogram and the risk factors identified in the relevant guidelines***" of the manuscript, rear position, marked in boldface letters)

DCA: In contrast to traditional performance measures, decision curve analysis (DCA) can assess the utility of models for decision making. DCA plots net benefit (NB) at a

range of clinically reasonable risk thresholds. As shown in the following figures, in the training cohort (A), DCA graphically shows that the use of the nomogram to predict the risk of malignancy when the threshold probability ranged from 0.2 to 1.0 added more net benefit than the other three factors. In the validation cohort (B), DCA graphically shows that the nomogram added more net benefit when the threshold probability ranged from 0.0 to 0.4.

CIC: The clinical impact curves of the nomogram indicated that the models had remarkable predictive power. At different threshold probabilities within a given population, the number of high-risk patients (solid red line) and the number of high-risk patients with the outcome (black dotted line) are shown. As shown in the following figures, it is obviously that in both training and validation cohort, the solid red line and black dotted line show a great fit, which indicated that the model had remarkable predictive ability.

8.Discussion: more comparisons with advanced machine learning models should be extensively discussed.

Response: Thank you very much for your useful comments. By reading and learning a large number of advanced machine learning models, we have added relevant content

according to your guidance to the article. We are honored to work with you to raise the article to a higher level. Our deepest respect for you. (The part "**DISCUSSION**" of the manuscript, paragraph 6, marked in boldface letters)

Response: Dear Reviewer, many thanks for taking your precious time to review our work. Thank you very much for your useful comments. We are sorry that our format of the abstract does not meet the requirements of the WJG. We have revised the abstract as you indicated. I sincerely hope that the correction can meet your expectations. (The part "**Abstract**" of the manuscript, marked in boldface letters)

Response: Thank you very much for your comment. We are sorry that our expression caused you a misunderstanding. Actually, preoperative imaging is not necessary for pathological diagnosis, it is mainly used for clinical diagnosis. The pathological diagnosis is finally confirmed by the postoperative specimen examination. And we have revised it in the original manuscript. (The part "**MATERIALS AND METHODS**" of the manuscript, paragraph 1, marked in strikethrough)

3. How to interpret the statement that the appearance of high-risk disease was characterized as a study endpoint?

Response: Thank you very much for your comment. Construction of a nomogram requires precise definition of the primary outcome. The outcome is typically an event, such as diagnosis of a malignancy or time to event, such as time to recurrence or death[9]. Actually, in this study, we define our outcome as diagnosis of high-risk disease. In clinical trials, study endpoint was defined as an event or outcome that can be measured objectively to determine whether the intervention being studied is beneficial. However, in this retrospective study, when we developed the nomogram-based model, we have to classify patients into high-risk and low-risk groups based on the results of their pathology results. So that the pathology results were then divided into binary variables. And high-risk disease or low-risk disease can be understood as the outcome of our study. Meanwhile, it is easy to understand that the disease in our study was in either high-risk or low-risk groups, so when the high-risk group was identified, the low-risk group was identified accordingly. So we define it as study endpoint. We sincerely hope that our explanation will satisfy you, once again, our heartfelt thanks.

4. The sections of surgical procedure and postoperative management can be briefly described, which is not the focus of this article.

Response: Thank you for your recommendations. After discussion and reflection. We think that your suggestion is very reasonable and acceptable. In order to highlight the focus of the article and make the article more organized. We have streamlined the contents of these two parts and the unnecessary or controversial expressions were deleted. With the advices and guidance of the reviewers, in order to make the focus of the article more prominent, we have integrated preoperative, surgical and postoperative details into perioperative management. We sincerely hope that the revision may gain your understanding and support. (The part "***Perioperative Management***" of the manuscript, marked in strikethrough)

5. How to ensure that the sample size as training cohort and validation cohort is sufficient?

Response: Thank you very much for your comment. Pancreatic cystic neoplasms are very rare disease that present in only 2.5% of the population. And we have tried our best to collect the date from three high-volume pancreatic centers since January 2015 and December 2021. On the basis of Harrell's guidelines[10], when the outcome is binary, the minimum value of the frequencies to identify a significant effect estimate should be greater than 10 times the number of predictors. The number of predictors in our study is five, so the sample size of training cohort is sufficient. Meanwhile, a study suggests that externally validating a prognostic model requires a minimum of 100 events[11]. So the sample size of validation cohort is sufficient enough. In addition, the model performs pretty well in terms of predictive efficiency which reflects the adequacy of the sample size.

6. Which R packages were used for statistical analysis in statistical analysis?

Response: Thank you very much for your comment. In this study, packages named rms, nomogramFormula , DynNom and rmda were invoked for statistical analysis and figures drawing. Meanwhile, we have added the relevant content to the method. (The part "***Statistical analysis***" of the manuscript, rear position, marked in boldface letters)
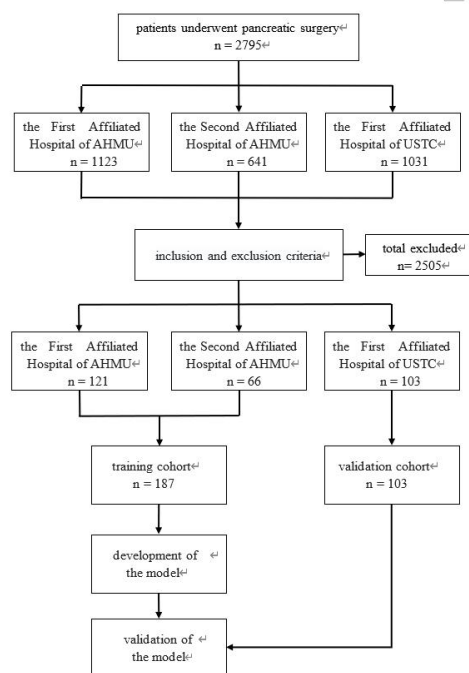
7. Recommend that the manuscript consistently use multivariable analysis. Multivariable cannot be used interchangeably with multivariate as these are different.

Response: Thank you for your careful review. We have now revised "multivariate" to "multivariable" throughout the manuscript. (Manuscript "multivariable", marked in boldface letters)

8. A flowchart of patient recruitment and diagnosis should be added.

Response: Thank you very much for your useful suggestion. A flowchart of patient recruitment and diagnosis has been added in the manuscript as Figure 1. We also

paste it here for your convenience.

9. The symbol font of "≥" should be corrected in Tables and Figures.

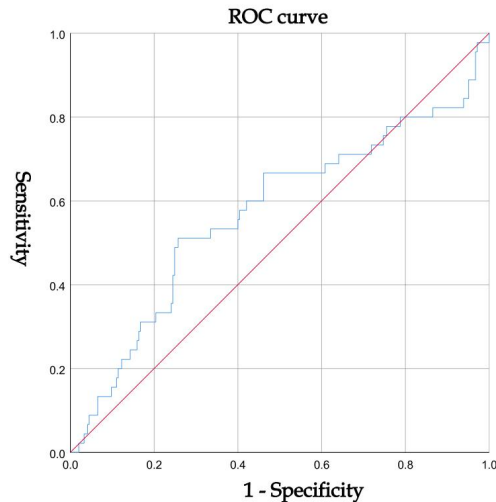**Response:** Thank you very much for pointing this out. We have now revised "≥" into the right format in Tables and Figures.

10. It would be preferred if you described the OR and the 95% CI of the factors, rather than just the p-value. Please do so throughout the text.

Response: Special thanks to you for your good comment. We have added the OR and the 95% CI of the factors throughout the text. Here we express our deepest respect.

11. Please plot the ROC curve to determine the optimal cutoff value of NLR.

Response: Thank you very much for your comment. The assigned cutoff values for NLR in our study was derived from the Youden index[12]. The ROC curve is shown below. According to Youden index = sensitivity + specificity – 1, the value of NLR was assigned corresponding to the maximum Youden index, which was 2.288. And the sensitivity and specificity were 0.532 and 0.744 respectively.

ROC curve

12. The AUC value of the prediction model and factors should provide confidence intervals.

Response: Thank you very much for your comment. We have added confidence intervals to the AUC value of the model and factors in both manuscript and figures. (The part "***Comparison of the performance of the nomogram and the risk factors identified in the relevant guidelines***" of the manuscript, front position, marked in boldface letters; Figure 4)

13. The C-index is not mentioned in the Methods section. Why use C-index in the third part and AUC in the fourth part? The C-index was 0.824 (95% CI, 0.735-0.914) and 0.893 (95% CI, 0.823-0.963) for the training cohort and the validation cohort, respectively. Was there a statistical difference in C-index between the two cohorts?

Response: Thank you very much for your comment. Especially thanks you for the pertinent question. We are sorry that we just briefly introduced the C-index, which didn't catch your attention. C-index is also known as concordance index, which is defined as the proportion of concordant pairs divided by the total number of possible evaluation pairs. It is often used to evaluate the accuracy of prediction models. So we used C-index in the third part to assess the discrimination of the model. AUC stands for "area under the ROC curve." That is, AUC measures the entire two-dimensional

area underneath the entire ROC curve. It is defined as the area under the curve of sensitivity and 1-specificity; this statistic is equal to the concordance of predicted and observed classes. So in the fourth part, though drawing the ROC curve, we used AUC to assess the diagnostic performance of the model with other risk factors.

The concordance index (C-index) was used to assess discrimination of the model. When computing the C-index, all possible pairs between a patient with and without the outcome are analyzed. A pair is concordant if the patient with the outcome has a higher predicted risk than the patient without the outcome. A C-index of 1 is perfect and 0.5 is equivalent to chance. A C-index of 0.60 means 60% of all possible pairs were concordant, and this is generally considered rather poor discrimination, while a C-index of 0.8 is usually considered good and ≥0.9 is excellent[5]. In conclusion, C-index assesses whether patients who experience the outcome have a higher predicted risk than patients who do not. For discrimination, it does not matter if the absolute value of C-index is 0.824 or 0.893, as long as the C-index maintain a high level in both the training and validation cohort[4]. Therefore, it is not necessary to compare the C-index between the training and the validation cohort. Meanwhile, the high level of C-index in the validation cohort proves that the model possesses good reproducibility and generalizability. We sincerely hope that our explanation will satisfy you, once again, our heartfelt thanks.

14. In the Result section, the values of AUC and C-index are the same for the training and validation cohorts. What is the difference between C-index and AUC?

Response: Thank you very much for your comment. The C-index was first proposed by Frank E Harrell Jr., professor of biostatistics at Vanderbilt University, in 1996. It is mainly used to calculate the distinction between predicted value and true value of COX model in survival analysis, and is often used to evaluate the prediction accuracy of patient prognosis model. In binary logistic regression model, the calculation process of C-index is to randomly form pairs of all the research objects in the research data and calculate the probability that the predicted results are consistent with the actual results. AUC is defined as the area under the curve of sensitivity and

1-specificity; this statistic is equal to the concordance of predicted and observed classes. In the case of a binary outcome like our study, the C-index is equivalent to the area under the curve (AUC)[4, 13]. The AUC of the ROC or C-index is simply a function of how well the ordered values of the continuous predictor correlate to the corresponding event status. In other areas of research, the C-index may be used for more types of studies during statistical analysis. Thanks again for your efforts and comments.

15. Was there any assessment of multicollinearity or effect modification with the multivariable model?

Response: Thank you very much for your comment. As we know, Relationships among the predictors, also known as multicollinearity, can influence the beta coefficients in the model, resulting in spurious associations and possibly unreliable effect estimates. Actually, we thought carefully about the problem when we first included the predictors. And we have tried our best to avoid the effect of multicollinearity. Firstly, we have considered both clinical and statistical significance when selecting variables for inclusion. And there is no underlying relationship between these factors in terms of clinical significance. Secondly, univariate and multivariable analysis were recognized as an important way to identify multicollinearity[9, 14]. Last but not least, in our study, no multicollinearity was identified though univariate and multivariable analysis. Once again, our heartfelt thanks.

16. How to develop the online calculator according to the nomogram? How were the factors assigned to points during the creation of the nomogram?

Response: Thank you very much for your comment. Especially thank you for the pertinent questions. The steps of creating the online calculator are as follows. Firstly, the package named DynNom was invoked in the RStudio to build the regression equation. Then the link to the associated terminal code was then generated for public use though the website shinyapps.io.

Nomogram points were assigned based on the weights for the relative importance of each variable in the final model. The total score (scaled to range from 0 to 100) for each patient was calculated as a weighted sum based on the contribution from the individual risk factors. The point system works by ranking the effect estimates, regardless of statistical significance, and it is influenced by the presence of other factors[15, 16]. Taking our model as an example, enhanced mural nodules have the highest effect, thus it is converted into100 points. A patient with enhanced mural nodules is assigned 100 points, whereas a patient without enhanced mural nodules gets 0 points. Regardless of statistical significance, the effect with the highest absolute value will be assigned 100 points on the scale, and the remaining variables are assigned a smaller number of points proportional to their effect size. This represents the relative importance of the least significant variable compared with the most significant variable. Finally, the total points accumulated by the various factors correspond to the predicted probability for a patient.

17. Is there a statistical difference in the AUC between the nomogram and the three factors (tumor diameter ≥ 40mm, enhancing mural nodules, and main pancreatic duct dilatation) in the training and validation cohorts?
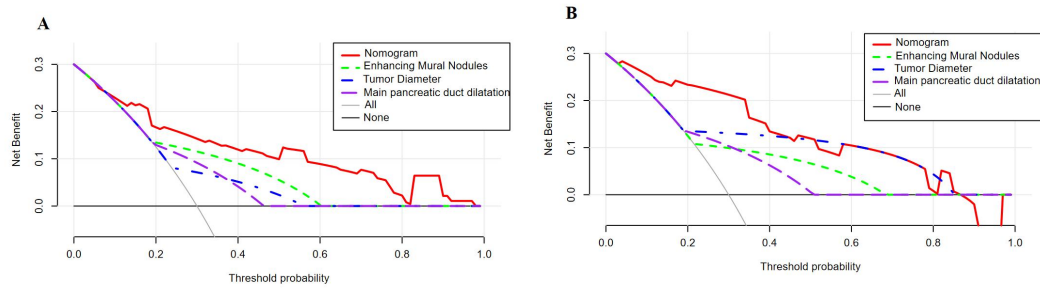
Response: Thank you very much for your comment. The answer is yes. First of all, The AUC of the nomogram is greater than that of other three factors in the training and validation cohorts. In the training cohort, the AUC of the nomogram, tumor diameter ≥ 40mm, enhancing mural nodules, and main pancreatic duct dilatation were 0.824, 0.619,0.692 and 0.653 respectively. In the validation cohort, the AUC of the nomogram, tumor diameter ≥ 40mm, enhancing mural nodules, and main pancreatic duct dilatation were 0.893, 0.718, 0.665 and 0.672 respectively. By comparing the AUC of the nomogram and the AUC of the other three factors, statistical differences were found. In the training cohort, the P-value between AUC of the nomogram and tumor diameter ≥ 40mm, enhancing mural nodules, and main pancreatic duct dilatation were $P < 0.001$, $P = 0.002$, $P = 0.006$ respectively. In the validation cohort, the P-value between AUC of the nomogram and tumor diameter ≥

40mm, enhancing mural nodules, and main pancreatic duct dilatation were $P = 0.009$, $P < 0.001$, $P < 0.001$ respectively. There is a significant statistical difference in the AUC between the nomogram and the three factors. The relevant conclusion has been added to the manuscript and figures. (The part "***Comparison of the performance of the nomogram and the risk factors identified in the relevant guidelines***" of the manuscript, middle position, marked in boldface letters; Figure 4)
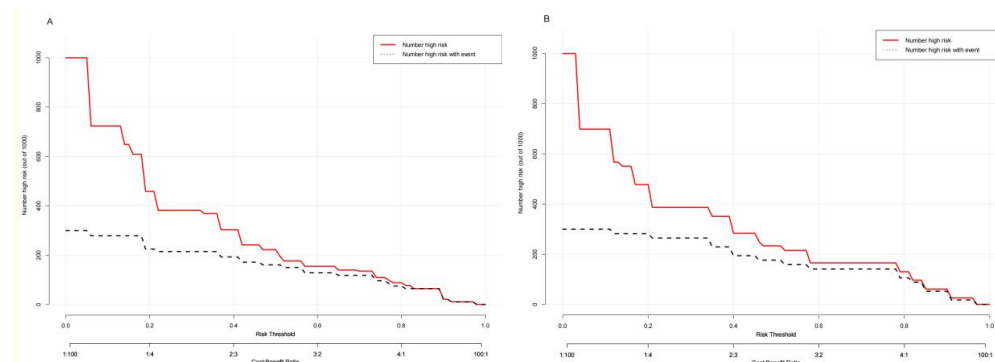
18. How to interpret decision curve analysis and clinical impact curves? More detailed information is needed to help the reader understand the information in Figure 4 and Figure 5.

Response: Thank you very much for your useful comments. Under the guidance of the reviewers, new DCA curves have been drawn to compare the net benefit between the nomogram and the other three factors. Meanwhile, detailed interpretation of the DCA and CICA has been added in the methods and results. We have added the relevant content to the results. (The part "***Statistical analysis***" and "***Comparison of the performance of the nomogram and the risk factors identified in the relevant guidelines***" of the manuscript, rear position, marked in boldface letters)

DCA: In contrast to traditional performance measures, decision curve analysis (DCA) can assess the utility of models for decision making. DCA plots net benefit (NB) at a range of clinically reasonable risk thresholds. As shown in the following figures, in the training cohort (A), DCA graphically shows that the use of the nomogram to predict the risk of malignancy when the threshold probability ranged from 0.2 to 1.0 added more net benefit than the other three factors. In the validation cohort (B), DCA graphically shows that the nomogram added more net benefit when the threshold probability ranged from 0.0 to 0.4.
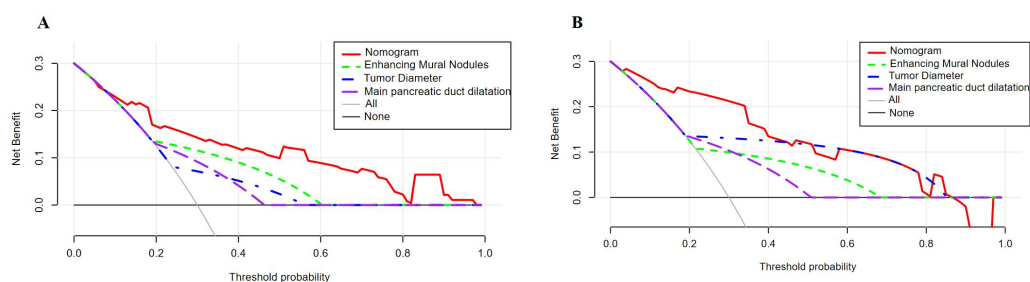
CIC: The clinical impact curves of the nomogram indicated that the models had remarkable predictive power. At different threshold probabilities within a given population, the number of high-risk patients (solid red line) and the number of high-risk patients with the outcome (black dotted line) are shown. As shown in the following figures, it is obviously that in both training and validation cohort, the solid red line and black dotted line show a great fit, which indicated that the model had remarkable predictive ability.

19. Whether nomogram provides more net benefits than the other three factors (tumor diameter ≥ 40mm, enhancing mural nodules, and main pancreatic duct dilatation)?

Response: Thank you very much for your comment. The answer is yes. After careful consideration of your comment and discussion, we draw a new DCA curve which compares the net benefits between the nomogram and the other three factors in both training and validation cohort. As shown in the following figures, in the training cohort (A), DCA graphically shows that the use of the nomogram to predict the risk of malignancy when the threshold probability ranged from 0.2 to 1.0 added more net benefit than the other three factors. In the validation cohort (B), DCA graphically

shows that the use of the nomogram to predict the risk of malignancy when the threshold probability ranged from 0.0 to 0.4 added more net benefit than the other three factors. And the new DCA curves and relevant contents has been added to the manuscript. (The part "***Comparison of the performance of the nomogram and the risk factors identified in the relevant guidelines***" of the manuscript, rear position, marked in boldface letters; Figure 5)

20. Rather than reporting only the AUC for comparison between the training and validation cohorts, it would be preferred for you to describe the sensitivity, specificity, accuracy, positive predictive value, and negative predictive value with the associated 95%CI of the model performance in the training and validation cohorts, respectively.

Response: Thank you very much for your useful suggestions. After research and discussion, we all think that your proposals are pretty valuable and useful. We have supplemented the relevant statistical indices to evaluate the performance of the model in both training and validation cohort. In the training cohort, the AUC and 95%CI is 0.824(0.733-0.915), the accuracy is 0.829, the sensitivity and specificity are 0.643 and 0.862 respectively and the positive predictive value and negative predictive value are 0.451 and 0.932 respectively. In the validation cohort, the AUC and 95%CI is 0.893(0.822-0.964), the accuracy is 0.925, the sensitivity and specificity are 0.882 and 0.826 respectively and the positive predictive value and negative predictive value are 0.501 and 0.973 respectively. These indices undoubtedly enrich the data of the article. Meanwhile, we have added the relevant

content to the method. Once again, our heartful thanks. (The part "*Comparison of the performance of the nomogram and the risk factors identified in the relevant guidelines*" of the manuscript, middle position, marked in boldface letters)

21. All patients underwent at least two preoperative imaging examinations among ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI), and 2-18F-fluoro-2-deoxy-D-glucose positron emission tomography (PET)-CT. However, how to reduce the heterogeneity between different imaging examinations?

Response: Thank you very much for your comments. Actually, the preoperative imaging data of all patients included in this study were reviewed by a specialized imaging expert. For the imaging features consistent across different examinations, imaging features were extracted routinely. For the imaging features inconsistent across different examinations, imaging features were extracted based on the most reliable imaging examination. In this way, the heterogeneity between different imaging examinations can be minimized. However, after discussion and research, the heterogeneity is still unable to be avoided completely. We decide to add it to the limitation. We sincerely hope that the answer may satisfy you. Once again, our heartful thanks. (The part "**DISCUSSION**" of the manuscript, paragraph7, marked in boldface letters)

22. The English writing of this paper needs to be greatly improved. It is strongly recommended to seek the help/service from professional English editor or company to improve this paper.

Response: Dear Reviewer, many thanks for taking your precious time to review our work. We are very sorry for the mistakes in this manuscript and inconvenience they caused in your reading. The manuscript has been thoroughly edited by the editor and a professional institution (Elsevier) that your esteemed journal recommended, and the certificate is attached. We hope it can meet the journal's standard. Thanks so much for your useful comments.

Response: Dear Reviewer, many thanks for taking your precious time to review our work. Special thanks to you for your good comment. An extra study limitation has been added at the end of the discussion.

(The number of patients included in the analyses is relatively small and possible heterogeneity in pathological diagnosis determining the grade of dysplasia or malignancy.) (The part "**DISCUSSION**" of the manuscript, paragraph 7, marked in boldface letters).

Response: Thank you so much for your careful review. We are sorry for the inconvenience this mistake caused for your review. And we have revised it.

Response: Thank you so much for your careful review. We are sorry for the inconvenience this mistake caused for your review. And we have revised it. (Figure 4)

Specific Comments to Authors:

1.Abstract: statistical analysis methods should be indicated in the part of method.

Response: Dear Reviewer, many thanks for taking your precious time to review our work. Thank you very much for your suggestion. We have revised the part "method" as you suggested. Statistical analysis methods have been added to the method of the abstract. I sincerely hope that the correction can meet your expectations. (The part "**Abstract**" of the manuscript, Methods, marked in boldface letters)

2.Study population: what's the "similar therapeutic approaches for PCNs"？It should be explained.

Response: Thank you very much for your comment. The treatment plan for each patient in each center strictly follows the guidelines and consensus and adopts similar preoperative examination and operation methods, so there is no great difference in the treatment plan for PCN patients in the three centers. For patients who met the indications for surgery, based on the intention of patients, surgery was performed according to the surgical scheme recommended in the guidelines. For patients who were unable or unwilling to undergo surgery, regular follow up was executed. We are honored to work with you to raise the article to a higher level. Our deepest respect for you.

3.Method: "In this study, patients were categorized as showing low-risk (low- or intermediate-grade dysplasia) or high-risk (high-grade dysplasia or invasive carcinoma) disease on the basis of the pathological diagnosis.", is there any references for this category?

Response: Thank you for your helpful comments. After referring to the relevant literatures[17-19], we find that in most relevant studies[17-19], researchers tend to category

the people like this. Therefore, we think such category is reasonable. Once again, our heartful thanks.

4.Preoperative evaluation: "In accordance with the same preoperative evaluation protocol at all centers", what's the evidence about the same preoperative evaluation protocol?

Response: Thank you very much for your comment. The three centers are all in the same province, they belong to the same pancreatic collaborative group and frequent academic exchanges are held regularly. All patients included in our study underwent laboratory inspection, imagological examination and other relevant examinations before surgery. In addition, patients who did not perfect these examinations were excluded when screening patients. However, we are sorry that our expression caused you a misunderstanding. "The same preoperative evaluation protocol" may be an over absolute expression. What we initially meant was that all centers specify their preoperative evaluation strategies according to guidelines and consensus. We think it might be more appropriate to change the word "same" to "similar". We sincerely hope that our explanation will satisfy you. We are honored to work with you to raise the article to a higher level.
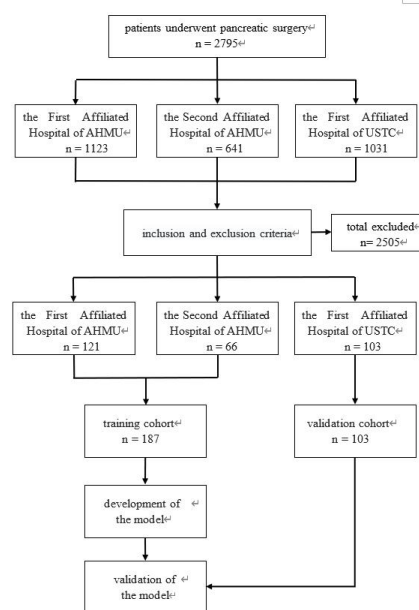
5.Postoperative management: "A digestive secretion inhibitor and broad-spectrum antibiotics were administered immediately after surgery.", is there any references or guidelines about the timing of antibiotic administration in this study?

Response: Thank you very much for your comment. In fact, there are indeed some literatures supporting this practice[20-22]. However, no guidelines about the timing of antibiotic administration have been reached. There is still much controversy about this. We believe that postoperative infections and complications may be controlled by prophylactic antibiotics. Actually, the study aimed to develop and verify a nomogram model, the part "Postoperative management" is not the focus, in order to highlight the focus of the article and make the article more organized. And we believe that it is improper to publish the controversial opinions in this magazine. After discussion and

reflection, the controversial expressions were deleted. We sincerely hope that the revision may gain your understanding and support. Our deepest respect for you.

Response: Thank you very much for your useful suggestion. A flowchart of patient recruitment and diagnosis has been added in the article as Figure 1. We also paste it here for your convenience.

(1) Science editor:

(2) Company editor-in-chief:

manuscript is too long and must be shortened to meet the requirement of the journal (Title: The title should be no more than 18 words). Before final acceptance, uniform presentation should be used for figures showing the same or similar contents; for example, "Figure 1Pathological changes of atrophic gastritis after treatment. A: ...; B: ...; C: ...; D: ...; E: ...; F: ...; G: ...". Please provide decomposable Figures (in which all components are movable and editable), organize them into a single PowerPoint file. Please authors are required to provide standard three-line tables, that is, only the top line, bottom line, and column line are displayed, while other table lines are hidden. The contents of each cell in the table should conform to the editing specifications, and the lines of each row or column of the table should be aligned. Do not use carriage returns or spaces to replace lines or vertical lines and do not segment cell content. Please check and confirm whether the figures are original (i.e. generated de novo by the author(s) for this paper). If the picture is 'original', the author needs to add the following copyright information to the bottom right-hand side of the picture in PowerPoint (PPT): Copyright ©The Author(s) 2022. Before its final acceptance, the author(s) must provide the Signed Informed Consent Form(s) or Document(s). For example, authors from China should upload the Chinese version of the document, authors from Italy should upload the Italian version of the document, authors from Germany should upload the Deutsch version of the document, and authors from the United States and the United Kingdom should upload the English version of the document, etc.

Response: Thank you for your positive feedback and helpful suggestions. We have re-submitted a manuscript titled "**Online Calculator for Predicting the Risk of Malignancy in Patients with Pancreatic Cystic Neoplasms: A Multicenter, Retrospective Study**", which meets the requirements of the journal. For figures showing the same or similar contents, uniform presentation has been used. We have adjusted the figures and tables according to the requirements and provided the figures as decomposable images in a single PowerPoint file. Meanwhile, the relevant informed consents and documents required have been correctly uploaded. It is an great honor to work with you to raise the article to a higher level.

Thank you for your consideration. I look forward to hearing from you.

Sincerely,
Fu-bao Liu

Department of General Surgery, the First Affiliated Hospital of Anhui Medical University (AHMU)

No. 218, Jixi Road, Hefei 230022, China

Email: lancetlfb@126.com

REFERENCES

1 **Hämmerle M** and Bergmann F. [Rare pancreatic tumors]. *Der Pathologe* 2021;**42**: 484-490 [PMID: 34402979 10.1007/s00292-021-00967-0]

2 **Burk KS**, Knipp D and Sahani DV. Cystic Pancreatic Tumors. *Magnetic resonance imaging clinics of North America* 2018;**26**: 405-420 [PMID: 30376978 10.1016/j.mric.2018.03.006]

3 **Yoon JG**, Smith D, Ojili V, Paspulati RM, Ramaiya NH and Tirumani SH. Pancreatic cystic neoplasms: a review of current recommendations for surveillance and management. *Abdominal radiology (New York)* 2021;**46**: 3946-3962 [PMID: 33742217 10.1007/s00261-021-03030-x]

4 **Ramspek CL**, Jager KJ, Dekker FW, Zoccali C and van Diepen M. External validation of prognostic models: what, why, how, when and where? *Clinical kidney journal* 2021;**14**: 49-58 [PMID: 33564405 10.1093/ckj/sfaa188]

5 **Moons KG**, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG and Grobbee DE. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart (British Cardiac Society)* 2012;**98**: 683-690 [PMID: 22397945 10.1136/heartjnl-2011-301246]

6 **Harish KB**, Price WN and Aphinyanaphongs Y. Open-Source Clinical Machine Learning Models: Critical Appraisal of Feasibility, Advantages, and Challenges. *JMIR formative research* 2022;**6**: e33970 [PMID: 35404258 10.2196/33970]

7 **Sendak M**, Gao M, Nichols M, Lin A and Balu S. Machine Learning in Health Care: A Critical Appraisal of Challenges and Opportunities. *EGEMS (Washington, DC)* 2019;**7**: 1 [PMID: 30705919 10.5334/egems.287]

8 **Decherchi S**, Pedrini E, Mordenti M, Cavalli A and Sangiorgi L. Opportunities and Challenges for Machine Learning in Rare Diseases. *Frontiers in medicine* 2021;**8**: 747612

[PMID: 34676229 10.3389/fmed.2021.747612]

9 **Iasonos A**, Schrag D, Raj GV and Panageas KS. How to build and interpret a nomogram for cancer prognosis. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2008;**26**: 1364-1370 [PMID: 18323559 10.1200/jco.2007.12.9791]

10 **Steyerberg EW**, Schemper M and Harrell FE. Logistic regression modeling and the number of events per variable: selection bias dominates. *Journal of clinical epidemiology* 2011;**64**: 1464-1465; author reply 1463-1464 [PMID: 22032755 10.1016/j.jclinepi.2011.06.016]

11 **Vergouwe Y**, Steyerberg EW, Eijkemans MJ and Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *Journal of clinical epidemiology* 2005;**58**: 475-483 [PMID: 15845334 10.1016/j.jclinepi.2004.06.017]

12 **Youden WJ**. Index for rating diagnostic tests. *Cancer* 1950;**3**: 32-35 [PMID: 15405679 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3]

13 **Su W**, He B, Zhang YD and Yin G. C-index regression for recurrent event data. *Contemporary clinical trials* 2022;**118**: 106787 [PMID: 35568377 10.1016/j.cct.2022.106787]

14 **Vatcheva KP**, Lee M, McCormick JB and Rahbar MH. Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. *Epidemiology (Sunnyvale, Calif.)* 2016;**6**: [PMID: 27274911 10.4172/2161-1165.1000227]

15 **Stephenson AJ**, Scardino PT, Eastham JA, Bianco FJ, Jr., Dotan ZA, DiBlasio CJ, Reuther A, Klein EA and Kattan MW. Postoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2005;**23**: 7005-7012 [PMID: 16192588 10.1200/jco.2005.01.867]

16 **Diblasio CJ** and Kattan MW. Use of nomograms to predict the risk of disease recurrence after definitive local therapy for prostate cancer. *Urology* 2003;**62 Suppl 1**: 9-18 [PMID: 14747038 10.1016/j.urology.2003.09.029]

17 **Attiyeh MA**, Fernández-Del Castillo C, Al Efishat M, Eaton AA, Gönen M, Batts R, Pergolini I, Rezaee N, Lillemoe KD, Ferrone CR, Mino-Kenudson M, Weiss MJ, Cameron JL, Hruban RH, D'Angelica MI, DeMatteo RP, Kingham TP, Jarnagin WR, Wolfgang CL and

Allen PJ. Development and Validation of a Multi-institutional Preoperative Nomogram for Predicting Grade of Dysplasia in Intraductal Papillary Mucinous Neoplasms (IPMNs) of the Pancreas: A Report from The Pancreatic Surgery Consortium. *Annals of surgery* 2018;**267**: 157-163 [PMID: 28079542 10.1097/sla.0000000000002015]

18 **Shimizu Y**, Hijioka S, Hirono S, Kin T, Ohtsuka T, Kanno A, Koshita S, Hanada K, Kitano M, Inoue H, Itoi T, Ueki T, Matsuo K, Yanagisawa A, Yamaue H, Sugiyama M and Okazaki K. New Model for Predicting Malignancy in Patients With Intraductal Papillary Mucinous Neoplasm. *Annals of surgery* 2020;**272**: 155-162 [PMID: 30499803 10.1097/sla.0000000000003108]

19 **Akhtar MZ**, Hughes DL and Silva M. Comment on "New Model for Predicting Malignancy in Patients With Intraductal Papillary Mucinous Neoplasm by Shimuzi et al". *Annals of surgery* 2021;**274**: e872-e873 [PMID: 33443898 10.1097/sla.0000000000004740]

20 **Swan CD**, Nahm C, Samra JS, Mittal A and Figtree M. Microbiology of pancreatoduodenectomy and recommendations for antimicrobial prophylaxis. *ANZ journal of surgery* 2020;**90**: 283-289 [PMID: 31743952 10.1111/ans.15560]

21 **Pham H**, Chen A, Nahm CB, Lam V, Pang T and Richardson AJ. The Role of Targeted Versus Standard Antibiotic Prophylaxis in Pancreatoduodenectomy in Reducing Postoperative Infectious Complications: A Systematic Review and Meta-analysis. *Annals of surgery* 2022;**275**: 315-323 [PMID: 33630442 10.1097/sla.0000000000004816]

22 **Cengiz TB**, Jarrar A, Power C, Joyce D, Anzlovar N and Morris-Stiff G. Antimicrobial Stewardship Reduces Surgical Site Infection Rate, as well as Number and Severity of Pancreatic Fistulae after Pancreatoduodenectomy. *Surgical infections* 2020;**21**: 212-217 [PMID: 31697194 10.1089/sur.2019.108]