PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (Error! Hyperlink reference not valid.) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	A competing-risks nomogram for predicting cancer-specific death in upper-tract urothelial carcinoma: a population-based analysis
AUTHORS	Li, Chengzhuo; Han, Didi; Huang, Qiao; Xu, Fengshuo; Zheng, Shuai; Li, Xiang; Zhao, Fanfan; Feng, Xiaojie; Lyu, Jun

VERSION 1 – REVIEW

REVIEWER	Macek, Pawel Holycross Cancer Center, Department of Epidemiology and Cancer Control
REVIEW RETURNED	24-Jan-2021

GENERAL COMMENTS	Thank you for the opportunity to read this manuscript (Manuscript ID bmjopen-2020-048243). I have some comments and questions which I will present below. Main comments. I see some ambiguity in the text of the work between its methodological aspects (competing risks) and substantive (prognostic factors). These two aspects should be discussed separately in the DISCUSSION section. Specifically, you should explain how your methods improve the traditional Cox regression approach. I really do not understand why the analysis results obtained from the Fine and Gray models and Cox models were compared? In this study, the first method considers competing risk and the second method does not. First, it is obvious that competing risk estimates are more valuable than those that do not distinguish between them. Second, even if Cox models were used to analyze competing risks (such a possibility exists), the interpretation of the results on the basis of Fine and Gray models and Cox models is different. Moreover, the recommendations regarding the use of these two methods differ. In general, the Cox method is recommended for etiological studies which aim to establish a causal link between the risk factors and the respective outcomes. The Fine and Gray method is used in predictive studies, with a view to assessing the probability of a specific outcome, at a specific time, for a specific case. When interpreting the results obtained with the aid of the Cox method (for competing risks), it should be borne in mind that the competing events are treated as the censored observations. Consequently, the number of at-risk-cases decreases throughout the observation period. The

	estimated hazard ratio is interpreted among those cases which neither have experienced an event of interest, nor a competing event as yet. Regarding the Fine and Gray method, the cases which have experienced a competing event are not censored, and remain within the data set as the ones not actually at risk of the event of interest. It follows that sub-distribution hazard ratio should not be interpreted in the same way as the hazard ratio. Additional comments. Please explain all abbreviations in the ABSTRACT section. What are the endpoints of this study? Based on the ABSTRACT section, it can be seen that these are the models used. What is this study about? Competing risks, nomograms, or UTUC. Based on the INTRO and DISCUSSION sections, I think UTUC. Based on the RESULTS section, I think about competing risks. Based on the last paragraph of the DISCUSSION section, I think about nomograms. This has to be balanced somehow. Why were three tools used for numerical analyzes (SPSS, SAS, and R)? Each of them has sufficient potential to carry out these analyzes. METHODS section, lines: 28-36. Several studies, and only one footnote. How were predictors qualified for multivariate analyzes? What method was used? Please explain this in the METHODS section. In the text RESULTS section The authors repeat the results contained in the tables. I advise against. Tables 1, 2, and 3 require an explanation of the abbreviations used (under the tables). Did the training cohort and validation cohort differ significantly from each other due to the variables presented in Table 1?

REVIEWER	Lauseker, Michael Department of Medical Information Systems, Biometry, and Epidemiology (IBE), Ludwig Maximilians University, Munich
REVIEW RETURNED	01-Feb-2021

GENERAL COMMENTS	In their paper, Li et al. develop a model for cancer-specific death in upper-tract urothelial cancer (UTUC). They present a nomogram, which could help identifying patients that die of cancer and those that die of other reasons. An internal validation shows good discrimination and calibration. The language is clear and the methods seem to be appropriate in general. The strength of the analysis is the large sample size. The authors discuss some limitations, yet I have the concern that there could be more. I have some questions and comments with regard to the manuscript: - Data were from a period of 41 years- did the authors check whether there was an effect of the calendar year? And did the composition of the sample with regard to covariates change over the years? - Why were continuous variables categorised? This means a huge loss of information and seems to me not necessary (especially with this sample size). - In general, the cumulative incidence of death due to other causes seems a little bit high to me. Besides, cancer specific factors like the tumor grade and chemotherapy seem to influence

	the cumulative incidence of death due to other causes. That brings me to the question: How well can you differentiate between these causes, resp. what was the definition of “cancer-specific death”. This is a common problem, as causes of death may not unambiguously be cancer-related or not. - Table 3 – what does “Coefficient” refer to? This is obviously not the regression coefficient (which is the logarithm of the HR). - In general, when considering different causes of death, a competing risk analysis is seen as standard (at least from a statistical point of view). The question is usually (See Wolbers[17] or Putter[21]), whether the assumption of proportional subdistribution hazards or the assumption of proportional cause-specific hazards are more appropriate. Have you thought about this and have you tested this assumption? - As the authors write, their nomogram could guide clinicians. However, I am not sure, what the consequences would be. Should the clinician change treatment? Perhaps the authors could elaborate. And besides that, I am not sure, if every clinician is used to this technique. Therefore, I would suggest describing the use of the nomogram (e.g. with an example). - The SEER database contains data from the US. Could the authors please comment, whether their model is applicable for other countries as well? - References 17 and 23 are the same

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Main comments:
1.I see some ambiguity in the text of the work between its methodological aspects (competing risks) and substantive (prognostic factors). These two aspects should be discussed separately in the DISCUSSION section. Specifically, you should explain how your methods improve the traditional Cox regression approach.
Reply 1: Thanks, this is a good suggestion. We have added the relevant explanations in the DISCUSSION section.

2.I really do not understand why the analysis results obtained from the Fine and Gray models and Cox models were compared? In this study, the first method considers competing risk and the second method does not. First, it is obvious that competing risk estimates are more valuable than those that do not distinguish between them. Second, even if Cox models were used to analyze competing risks (such a possibility exists), the interpretation of the results on the basis of Fine and Gray models and Cox models is different. Moreover, the recommendations regarding the use of these two methods differ. In general, the Cox method is recommended for etiological studies which aim to establish a causal link between the risk factors and the respective outcomes. The Fine and Gray method is used in predictive studies, with a view to assessing the probability of a specific outcome, at a specific time, for a specific case. When interpreting the results obtained with the aid of the Cox method (for competing risks), it should be borne in mind that the competing events are treated as the censored observations.Consequently, the number of at-risk-cases decreases throughout the observation period. The estimated hazard ratio is interpreted among those cases which neither have experienced an event of interest, nor a competing event as yet. Regarding the Fine and Gray method, the cases which have experienced a competing event are not censored, and remain within the data set as the ones not actually at risk of the event of interest. It follows that sub-distribution hazard ratio should not be interpreted in the same way as the hazard ratio.

Reply 2: Thank you for giving such a pertinent opinion. It can be seen that you are a very professional scholar in this field. According to your opinions, we have deeply thought and discussed the methods of our research, so that our paper has been modified accordingly. In the current version, we deleted the original Cox regression analysis and added the cause-specific hazard function (CS) which is another approach of the competing risk model. According to the relevant literature, the CS model answers the epidemiological etiology question, and the subdistribution proportional-hazards function (SD) model answers the clinical absolute incidence questions. The SD model is more suitable for the construction of clinical risk scores. In addition, we also gave a more comprehensive explanation of the different results of SD model and CS model analysis in the discussion section.

Additional comments.
1.Please explain all abbreviations in the ABSTRACT section.
Reply 1: Thanks, this is a good suggestion. We have added explanations of all abbreviations in the ABSTRACT section.

2.What are the endpoints of this study?
Reply 2: Thank you. That's a very good question. The end point of our study were upper-tract urothelial carcinoma carcinoma (UTUC) cancer-specific death (CSD) (event of interest) and death due to other causes (DOC) (non-UTUC death, competing event).

3.Based on the ABSTRACT section, it can be seen that these are the models used. What is this study about? Competing risks, nomograms, or UTUC. Based on the INTRO and DISCUSSION sections, I think UTUC. Based on the RESULTS section, I think about competing risks. Based on the last paragraph of the DISCUSSION section, I think about nomograms. This has to be balanced somehow.
Reply 3: Thank you for your interest in our research. In summary, our study was to construct a nomogram of prognosis analysis for UTUC patients based on a competing risk model approach. According to your suggestion, we have made corresponding changes in the article to make the three parts of the whole article more balanced.

4.Why were three tools used for numerical analyzes (SPSS, SAS, and R)? Each of them has sufficient potential to carry out these analyzes.
Reply 4: Thank you for your patient guidance. Your opinions are useful to us, so we made corresponding changes and used R statistical software to complete all the statistical analysis.
5.METHODS section, lines: 28-36. Several studies, and only one footnote.
Reply 5: Thank you for pointing out it. As your advice, we have made corresponding modifications.

6.How were predictors qualified for multivariate analyzes? What method was used? Please explain this in the METHODS section.
Reply 6 :Thank you for your pertinent comments,we have added relevant definitions in the method section.

7.In the text RESULTS section The authors repeat the results contained in the tables. I advise against.
Reply 7: Thank you very much for your careful suggestion. We have deleted the repeated expressions in the RESULTS section.

8.Tables 1, 2, and 3 require an explanation of the abbreviations used (under the tables).
Reply 8: Thank you. We added an explanation of the abbreviations under the Table1/2/3.

9.Did the training cohort and validation cohort differ significantly from each other due to the variables

presented in Table 1?
Reply 9: Thank you for this very insightful comment, following your suggestion, we had added relevant contents in studies characteristic Table 1.

Reviewer: 2
1. Data were from a period of 41 years- did the authors check whether there was an effect of the calendar year? And did the composition of the sample with regard to covariates change over the years?
Reply 1: Thanks, that is a good question, and in accordance with your opinion, We added the variable of calendar year, and found no statistical significance of calendar year in both univariate and multivariate analyses. Therefore, it can be considered that there is no significant change in the covariate between years.

2.Why were continuous variables categorised? This means a huge loss of information and seems to me not necessary (especially with this sample size).
Reply 2:Thank you for pointing this out. According with your opinion, We reanalyzed age, LNE, LNP, LNR and other variables as continuous variables.

3. In general, the cumulative incidence of death due to other causes seems a little bit high to me. Besides, cancer specific factors like the tumor grade and chemotherapy seem to influence the cumulative incidence of death due to other causes. That brings me to the question: How well can you differentiate between these causes, resp. what was the definition of ‘cancer-specific death’. This is a common problem, as causes of death may not unambiguously be cancer-related or not.
Reply 3: Thanks. That is a good question. It is worth mentioning that in the SEER database, the variable 'SEER cause-specific death classification' is a clear definition of "cancer-specific death".
We used the SEER cause-of death item to determine whether a failure event occurred or not. Cause of death in SEER is based on death certificate reporting. Although accuracy of death certificates is imperfect, studies have shown that causes of death from death certificates are comparable to those obtained from autopsy in patients with malignancies.

4.Table 3-what does ‘Coefficient’ refer to? This is obviously not the regression coefficient (which is the logarithm of the HR).
Reply 4: Thanks for your friendly reminder. We have found our mistake, and we have modified the value of the coefficient in Table 3.

5.In general, when considering different causes of death, a competing risk analysis is seen as standard (at least from a statistical point of view). The question is usually (See Wolbers[17] or Putter[21]), whether the assumption of proportional subdistribution hazards or the assumption of proportional cause-specific hazards are more appropriate. Have you thought about this and have you tested this assumption?
Reply 5: Thanks, that is a good question. According to your suggestion, we have added the comparison of proportional subdistribution hazards and proportional cause-specific hazards in the article.
According to the previous study, proportional cancer-specific hazards are appropriate to answer the upstream epidemic etiology questions, while proportional subdistribution hazards answer the downstream clinical absolute incidence questions. Proportional subdistribution hazards models are increasingly being used for clinical prediction models and risk scoring.Since the purpose of our study is to analyze the prognostic factors of UTUC cancer-specific death (CSD), we believe that assumption of proportional subdistribution hazards are more appropriate.
In addition, we also gave a more comprehensive explanation of the different results of SD model and CS model analysis in the discussion section.

6.As the authors write, their nomogram could guide clinicians. However, I am not sure, what the consequences would be. Should the clinician change treatment? Perhaps the authors could elaborate. And besides that, I am not sure, if every clinician is used to this technique. Therefore, I would suggest describing the use of the nomogram (e.g. with an example).
Reply 6: Thank you. That's good advice.Following your instructions, we have added an example of nomogram application in the DISSCUSION section. Finally, the predicted value calculated by the nomogram does not represent an absolute prognosis and is only suitable for clinician reference, rather than providing absolutely accurate prognoses. Future research could use the current findings to develop a widely accepted UTUC risk prediction tool.

7. The SEER database contains data from the US. Could the authors please comment, whether their model is applicable for other countries as well?
Reply 7:Thank you for pointing out it. Because the SEER database is based on a research cohort of the US population, it may not well reflect the prognosis of UTUC patients in other countries, and it still needs to be further verified in other populations. But at present, it can already guide clinicians well, and at the same time, it can also provide further research directions for researchers in other countries.

8.References 17 and 23 are the same.
Reply 8:Thanks for your friendly reminder. We had done the correction.

VERSION 2 – REVIEW

REVIEWER	Macek, Pawel Holycross Cancer Center, Department of Epidemiology and Cancer Control
REVIEW RETURNED	15-Mar-2021

GENERAL COMMENTS	I can see that the Authors have done a great job. Thank you. I have no more comments. Good luck

REVIEWER	Lauseker, Michael Department of Medical Information Systems, Biometry, and Epidemiology (IBE), Ludwig Maximilians University, Munich
REVIEW RETURNED	26-Mar-2021

GENERAL COMMENTS	Thank you for you extensive answers! I think, the manuscript has been improved in some points. However, in my opinion, some points still remain unclear. Major points: - The problems regarding the definitions of cancer-related death should be discussed in the limitation section (see reviewer 2's comment). Besides, to be honest, I do not completely understand the autors' reply „It is worth mentioning that in the SEER database, the variable 'SEER cause-specific death classification' is a clear definition of "cancer-specific death".'

	- It might be true that „the gold standard treatment for UTUC is still surgery.“. However, I do not think that an analysis on this data is able to contribute to this question as these are purely observational data with multiple potential biases present (e.g. time-dependent bias, possible correlation with of treatment with severity...). - Regarding reviewer 1's comment on the aims of the paper: I am still not sure what the aim of this manuscript is - There are tests to assess the proportional hazards assumptions. Did the authors test for proportional cause-specific or proportional subdistribution hazards? - The paper and especially the new parts are a little bit difficult to read and understand. Being a non-native speaker myself, I would therefore strongly recommend involving a native speaker in writing. Besides, the model is usally called „proportional subdistribution hazards model“ not „subdistribution proportional hazards model“. Minor points: - Is there really an 8-year estimate of patients diagnosed from 2013 to 2015 (when there is none for the patients diagnosed earlier)? - Figure 2: What do „1“ and „2“ refer to? Please label correctly. - I had the impression that „cause-specific“ and „cancer-specific“ are sometimes used interchangeably, which is confusing.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Comments to the Author:
Thank you for you extensive answers! I think, the manuscript has been improved in some points. However, in my opinion, some points still remain unclear.
Major points:
1.The problems regarding the definitions of cancer-related death should be discussed in the limitation section (see reviewer 2's comment). Besides, to be honest, I do not completely understand the authors' reply. It is worth mentioning that in the SEER database, the variable 'SEER cause-specific death classification' is a clear definition of "cancer-specific death".'

Reply 1: Thank you for pointing out it. We apologize for our vague reply. The following are the definitions of all-cause and cancer-specific deaths in the SEER database. We used these two variables to generate the outcome variable of our study, and the variable values include: alive, cancer-specific death(CSD), and death due to other causes(DOC). The cause of death for SEER was that according to the death certificate report, some deaths may have been misclassified, so we added relevant content in the limitation part of the article.

2. It might be true that the gold standard treatment for UTUC is still surgery. however, I do not think that an analysis on this data is able to contribute to this question as these are purely observational data with multiple potential biases present (e.g. time-dependent bias, possible correlation with of treatment with severity...).

Reply 2:Thank you. That is a good question. Surgery is generally considered the gold standard for UTUC treatment. In our study, surgery was an important prognostic factor in patients with UTUC, which further confirms this conclusion. However, as you said, there may be multiple biases in the study of purely observational data. we strongly agree with your opinion, so we have added relevant content in the limitation part of the discussion. Despite these limitations, the large sample size and detailed clinical follow-up along with relevant demographic and tumour information can still provide a guarantee for the accuracy of our study to a certain extent.

3. Regarding reviewer 1's comment on the aims of the paper: I am still not sure what the aim of this manuscript is.

   Reply 3: Thank you for your pertinent comments. The objective of this study was to use a competing risks model to assess the effect of potential prognostic factors on cancer-specific death in upper-tract urothelial carcinoma, and to construct a more comprehensive and accurate nomogram containing significant prognostic factors to guide clinical practice.

4. There are tests to assess the proportional hazards assumptions. Did the authors test for proportional cause-specific or proportional subdistribution hazards?

   Reply 4: Thank you, that's a good question. Through our test, both models of the multivariate analysis satisfy the proportional hazards assumptions.

5. The paper and especially the new parts are a little bit difficult to read and understand. Being a non-native speaker myself, I would therefore strongly recommend involving a native speaker in writing. Besides, the model is usally called proportional subdistribution hazards model" not subdistribution proportional hazards model.

   Reply 5: Thank you for your patient guidance. We have found a native English speaker to help us rewrite the new section of the article. In addition, we have also modified all the misrepresentations of the models in the paper.

   Minor points:
   1. Is there really an 8-year estimate of patients diagnosed from 2013 to 2015 (when there is none for the patients diagnosed earlier)?
   Reply 1: Thank you for your careful review. We examined the raw data and found that the survival time of 2 patients was miscalculated.After modification, we have re-analyzed and modified the relevant parts of the article.We are very sorry for our mistake. Thanks again for your help.

   2. Figure 2: What do '1'and '2' refer to? Please label correctly.
   Reply 2:Thank you for pointing out it. We have redrawn Figure 2 and annotated it.

   3. I had the impression that "cause-specific" and "cancer-specific" are sometimes used interchangeably, which is confusing.
   Reply 3:Thank you for your pertinent comments.We apologize for the confusion caused by our negligence.In present study, 'cause-specific' was used to define the 'cause-specific hazards model(CS)', which is a derivative model of Cox proportional hazards model, it was first introduced into the analysis of competing risk data by Prentice[10] and has been widely used. While 'cancer-specific' was used to define the 'cancer-specific death(CSD)',which is an outcome defined in our study. We apologize for the confusion caused by our negligence.

VERSION 3 – REVIEW

REVIEWER	Lauseker, Michael
REVIEW RETURNED	27-May-2021

GENERAL COMMENTS	Thanks for your patience, I have nothing to add!