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

TITLE (PROVISIONAL)	Clinicopathologic Risk Factors for Gastric Cancer: A Retrospective Cohort Study in China
AUTHORS	Hu, Kongwang; Wang, Shuaili; Wang, Zikun; Li, Longlong; Huang, Zhiguo; Yu, Weiqiang; Chen, Zhongxue; Wu, Qing-Fa

VERSION 1 – REVIEW

REVIEWER	Wen-Chi Chou Chang Gung Memorial Hospital, Taiwan
REVIEW RETURNED	17-Apr-2019

GENERAL COMMENTS	This manuscript mainly conducted a retrospective analysis to evaluate the impact of potential clinicopathologic factors on survival outcome in patients with gastric cancer after cancer surgery. Total 716 gastric cancer patients treated at one medical institute between 2001 and 2012 in China were collected. They found that age, lymph node metastasis rate, tumor size, surgical type II, and clinical stage were the independent prognostic factors. Too much grammar errors need to correct. For example, first sentence of the introduction….. the thrid leading cause… The introduction is too long and does not provide the gap of current knowledge. The method and result section is mixed and difficult to interpretation. Mostly importantly, this study did not provide any novelty to the field.

REVIEWER	Amelia Tavares Centro Hospitalar de Vila Nova de Gaia Portugal
REVIEW RETURNED	03-May-2019

GENERAL COMMENTS	Pg 2: - The design and participants should be included in Methods - Line 38: type II and (remove the ,) - You have to improve the conclusion. This first sentence of the conclusion must be removed, or changed to introduction. Pg 4: - Line 20: "…the second highest..mortality rate…" but in line 9 you refer "third leading cause of cancer-related death…" - Line 22 "…in which alone accounts for 42% of all gastric cancer cases…"it´s incomprehensible - Line 25 "70% of the case translates into a high fatality…? This means? - Line 51: why to you use data from GLOBOCAN 2008 instead 2018 Pg 5:

| | - Line 12: "Of those…", rewrite this sentence, it´s incomprehensible
- Line 41: " Surgery…undergoing surgery remains very low" but it´s the unique curative treatment, rewrite the sentence
Pg 6:
- Line 41: " On the basis of sufficient…" what is sufficient?
Pg 7
- line 33: "…invasion degree tumor nodes number…" you have to explain this variable
Pg 9:
- Line 33: " Among the entire (total) patients, 673…radical resection, and 41…palliative resection" 673 + 41=714, what happened with the others two patients
- Line 38: "In addition, a total of 580 patients had received stomach surgery" ?? In Study cohort you refer seven hundred and sixteen patients with gastrectomy.
- Line 49 : you only refer the tumor size of 703 patients (299+275+129), if all your patients were submitted to gastrectomy?
- Line 56: you name 14 unfavorable prognostic factors but in line 17 "18 of them were found to have a significant influence on survival…"
Pg 10:
- Line 9: if 18 variables had influence in prognosis, 5 (23-18) hadn´t, but you named 7
- How to you explain that variables that were not significant in univariate analysis were significant in adjusted analysis? Usually the unadjusted analysis is used to sort the variables to adjusted analysis.

You have to improve conclusion.

Table 1:
- if all patients were submitted to surgery why you don´t present the data for all patients?
- what is the difference between age/ age2, explain the last 3 variables presented in table 1 |

VERSION 1 – AUTHOR RESPONSE

Responses to reviewer comments

Comments from Reviewer #1:

This manuscript mainly conducted a retrospective analysis to evaluate the impact of potential clinicopathologic factors on survival outcome in patients with gastric cancer after cancer surgery. Total 716 gastric cancer patients treated at one medical institute between 2001 and 2012 in China were collected. They found that age, lymph node metastasis rate, tumor size, surgical type II, and clinical stage were the independent prognostic factors.

Too much grammar errors need to correct. For example, first sentence of the introduction….. the thrid leading cause…

Response: We tried to avoid any grammar error, but unfortunately, a few still occurred in the previous version. However, we respectfully disagree with the comment "Too **much** grammar **errors**". We have found and corrected some typos in the revision. In addition, we have asked a native English speaker to proofread this revised manuscript.

The introduction is too long and does not provide the gap of current knowledge.

Response: In the introduction section of the original submission we tried to provide more relevant and important information from the literature. But we agree that people may have different opinions about the optimal length of this section. We have shortened this section in the revision. However, we respectfully disagree with the comment "…does not provide the gap of current knowledge". In fact, for instance, we did mention some gaps in the first sentence "The epidemiological and clinicopathological characteristics of gastric cancer still largely remain uncertain, although some risk factors have been identified in the previous literature" (now the revised first sentence of the second paragraph), and the second sentence "It is not easy to identify the most significant factors concerning prognosis because of the high correlations among those variables" (now the revised second last sentence of the last paragraph).

The method and result section is mixed and difficult to interpretation.

Response: We respectfully disagree with the comment "The method and result **section is** mixed", rather we believe that the two section**s are** not mixed. In addition, we are not sure the exact meaning of the comment "difficult to **interpretation**". We believe that our results are not difficult to interpret. However, we have revised the two sections based on comments from the other reviewer.

Mostly importantly, this study did not provide **any** novelty to the field.

Response: We respectfully disagree with this comment. First, it is unclear to us what "**mostly importantly**" means by the reviewer. Second, our study does provide additional useful knowledge and information in the area based on the new data from a specific population. In addition, we don't understand why "**novelty**" is "most important" for a manuscript published in BMJ Open.

Comments from Reviewer #2:

We are very grateful to the helpful and constructive comments offered by this reviewer. We have revised our manuscript accordingly.
Pg 2:

- The design and participants should be included in Methods

Response: They have been moved to the method section in the revision accordingly.

- Line 38: type II and (remove the ,)

Response: It has been changed in the revision accordingly.

- You have to improve the conclusion. This first sentence of the conclusion must be removed, or changed to introduction.

Response: It has been removed.

Pg 4:

- Line 20: "…the second highest..mortality rate…" but in line 9 you refer "third leading cause of cancer-related death…"

Response: We have revised the information in the revision.

- Line 22 "…in which alone accounts for 42% of all gastric cancer cases…"it´s incomprehensible

Response: According to reviewer 1, we have shortened the introduction and removed this sentence.

- Line 25 "70% of the case translates into a high fatality…? This means?

Response: We have rewritten the sentence in the revision "… 70% of the gastric cancer cases had high fatality… ".

- Line 51: why to you use data from GLOBOCAN 2008 instead 2018 **Response**: The date from the updated reference has been used in the revision.

Pg 5:

- Line 12: "Of those…", rewrite this sentence, it´s incomprehensible

Response: According to reviewer 1, we have shortened the introduction and removed this sentence.

- Line 41: " Surgery…undergoing surgery remains very low" but it´s the unique curative treatment, rewrite the sentence

Response: In the revision, it has been rewritten as "Surgery is the most preferred treatment for gastric carcinoma, but the survival rate of patients undergoing surgery remains very low".

Pg 6:

- Line 41: " On the basis of sufficient…" what is sufficient?

Response: According to reviewer 1, we have shortened the introduction and removed this sentence.

Pg 7

- line 33: "…invasion degree tumor nodes number…" you have to explain this variable

  Response: It should be "invasion degree". We have corrected this typo.

Pg 9:

- Line 33: " Among the entire (total) patients, 673…radical resection, and 41…palliative resection" 673 + 41=714, what happened with the others two patients

Response: This variable has two missing values in this data set. We have rewritten this sentence and reported the number of missing values for this variable in Table 1.

- Line 38: "In addition, a total of 580 patients had received stomach surgery" ?? In Study cohort you refer seven hundred and sixteen patients with gastrectomy.

Response: Although all participants received gastrectomy, there were 7 patients in the data set had missing values for this variable. We have rewritten this sentence with updated information (please also see Table 1).

- Line 49 : you only refer the tumor size of 703 patients (299+275+129), if all your patients were submitted to gastrectomy?

Response: 13 patients had missing values for this variable (please also see Table 1).

- Line 56: you name 14 unfavorable prognostic factors but in line 17 "18 of them were found to have a significant influence on survival…"

Response: we have corrected this information in the revision.

Pg 10:

- Line 9: if 18 variables had influence in prognosis, 5 (23-18) hadn´t, but you named 7

  Response: In the revision, we updated this information. We actually included 24 independent variables (one is "age square", derived from age).

- How to you explain that variables that were not significant in univariate analysis were significant in adjusted analysis? Usually the unadjusted analysis is used to sort the variables to adjusted analysis.

Response: This is a typical phenomenon in statistical data analysis. In a univariate analysis, if some of the confounding variables were significant but not adjusted we may have false negatives. For a data set with large variables, the way you mentioned for variable selection can be used. But standard statistical methods for variable selection (e.g., backward elimination method used in this study) are more sensible for our study.

You have to improve conclusion.

Table 1:

4

- if all patients were submitted to surgery why you don´t present the data for all patients?
  Response: Some variables have missing values, we have updated this information in the revision (see Table 1).
- what is the difference between age/ age2, explain the last 3 variables presented in table 1
  Response: age2 is the variable "age square", we have changed it in Table 1 in the revision. The meanings of the other variables were also explained in the subsection of "study cohort" in the revision.

VERSION 2 – REVIEW

REVIEWER	Wen-Chi Chou Chang Gung Memorial Hospital at Linkou,Taiwan
REVIEW RETURNED	01-Jul-2019

GENERAL COMMENTS	The authors have conducted a study to analyze the risk factor for patient with gastric cancer receive surgical treatment based on single medical institute from China. In brief, they identified 6 important clinicopathologic factors relevant to gastric cancer in this manuscript. I do have several concerns about the methology and statistical analysis. 1.Table 1 present the unadjust analysis for each variables. Please add the event number, median survival time, and hazard ratio for each variables. 2.Please clearly define and elaborate who did author select variables in the multivariable model ? 3.The patient numbers is not consistent in Table 1, for example N1 stage (n=170)+ N2 stage (n=176) should equal to positive LN number 1-6, but its n=338 in Table 1. Patient with M1 stage =28, not equal to stage IV (n=29). 4.Why no patients allocated into T3 classification, it's a serious selection bias in such large patient cohort. The majority of gastric cancer originated from antrum and pylorus, but proximal cancer dominate in this study. Why ? 5.Why age, age square, and LN metastatic rate did not reach statistical difference in unadjust analysis, but were independent prognosticators in the adjusted analysis. Please discuss the reasons. 6.Who did author avoid the interaction of variables in statistically analysis, for example, surgical type with palliative resection is associated with large tumor size and advanced clinical stage, for which these three factors were identified as the independent variables in your table 2. Minor suggestion: 1.please use the same term in the manuscript, for example, stomach cancer vs. gastric cancer 2. page 5, 2nd paragraph: "Previous studies have revealed that the average survival time of patients with advanced gastric cancer is less than 12 months". Please mention which instance for patients had an average survival time less than 12 months after surgical resection. 3. page 5, 3rd paragraph: "However, most of the previous cohort studies in this area had small sample sizes and each focused on the effect of a single pathological factor18.". The citation (ref 18) did not support author's claim as the authors included 1614patients numbers which is bigger than yours.

Responses to reviewer comments

Reviewer: 1
The authors have conducted a study to analyze the risk factor for patient with gastric cancer receive surgical treatment based on single medical institute from China. In brief, they identified 6 important clinicopathologic factors relevant to gastric cancer in this manuscript. I do have several concerns about the methology and statistical analysis.

Response: We appreciate the reviewer for careful reading of our manuscript and providing helpful comments which resulted in an improved presentation of our revised manuscript.

1. Table 1 present the unadjust analysis for each variables. Please add the event number, median survival time, and hazard ratio for each variables.

Response: For each categorical independent variable, we have added the event number, the median survival time and the estimated hazard ratio for each level from the unadjusted analyses. Please see the updated Table 1.

2. Please clearly define and elaborate who did author select variables in the multivariable model ?

Response: We guess the question was "which author did variable selection in the adjusted analysis". We used the SAS procedure Cox proportional hazard regression with the option "backward" variable selection. In the original model in the revised analysis, we entered all of the main effects and the possible two-way interaction terms. The program automatically selected the significant main and interaction effects which were used in the final model (please see Table 2).

3. The patient numbers is not consistent in Table 1, for example N1 stage (n=170)+ N2 stage (n=176) should equal to positive LN number 1-6, but its n=338 in Table 1. Patient with M1 stage =28, not equal to stage IV (n=29).

Response: Thank you for the comment. We have double checked the database and made changes/corrections accordingly. Now we present the updated results in Tables 1 and 2.

4. Why no patients allocated into T3 classification, it's a serious selection bias in such large patient cohort. The majority of gastric cancer originated from antrum and pylorus, but proximal cancer dominate in this study. Why ?

Response: Thank you for this insightful comment. In the original database, the level of T3 was recorded as T4. Now we have corrected those records (please see Table 1).

5. Why age, age square, and LN metastatic rate did not reach statistical difference in unadjust analysis, but were independent prognosticators in the adjusted analysis. Please discuss the reasons.

Response: This is a common phenomenon in statistical practice. It has been recognized long time ago that conditional association (e.g., none zero adjusted effect) does not imply marginal association (e.g., none zero unadjusted effect), and vice versa.

6. Who did author avoid the interaction of variables in statistically analysis, for example, surgical type with palliative resection is associated with large tumor size and advanced clinical stage, for which these three factors were identified as the independent variables in your table 2.

Response: Thanks for the comment. In the original analyses, we did not consider the interactions as we thought our sample size might not be large enough to detect interaction effects. In the revised adjusted analysis, as suggested we included all two-way interactions in the initial model and the model selection indeed identified a significant interaction term. Please see the new results as displayed in Table 2 of the revised manuscript.

Minor suggestion:

1. please use the same term in the manuscript, for example, stomach cancer vs. gastric cancer

Response: We have replaced "stomach cancer" with "gastric cancer" in the revision.

2. page 5, 2nd paragraph: "Previous studies have revealed that the average survival time of patients with advanced gastric cancer is less than 12 months". Please mention which instance for patients had an average survival time less than 12 months after surgical resection.

Response: Thank you for the comment. In the revision we have updated the right references which mentioned several examples for this case.

3. page 5, 3rd paragraph: "However, most of the previous cohort studies in this area had small sample sizes and each focused on the effect of a single pathological factor18.". The citation (ref 18) did not support author's claim as the authors included 1614patients numbers which is bigger than yours.

Response: Thank you for pointing this out. We did some literature review and found that for old studies, the statement was right. But for current studies, this statement was not all correct. Therefore, we decided to remove this sentence in the revision.

VERSION 3 – REVIEW

REVIEWER	Wen-Chi Chou Chang Gung Memorial Hospital at Linkou, Taiwan
REVIEW RETURNED	11-Aug-2019

GENERAL COMMENTS	Thanks the efforts make by authors to improve the quality of this manuscript. However, I still concern the data of the statistics and interpretation of the results from the statistical analysis.

VERSION 3 – AUTHOR RESPONSE

Responses to reviewer comments

My major concerns are about the statistics analysis of the manuscript entitled "Clinicopathologic Risk Factors for Gastric Cancer: A Retrospective Cohort Study in China".

Response: We appreciate the reviewer for willing to spend time to write the comments. But we respectfully disagree the points made in the comments.

As I commended in my previous review (R1 version), the authors did not provide adequate reason how to select significant variables from the univariate analysis in the multivariate analysis.

Response: First of all, we **DID NOT** do multivariate analysis for this study as we only considered a single dependent variable (i.e., survival time). Second, in the previous revision, we already provided details on how to select significant independent variables (i.e., backward variable selection in the Cox PH model) in the final adjusted analysis (see lines 30-44 on page 7, and the response to comment #2 in the previous response letter as attached) .

Some of the important factors that had significant value in univariate analysis (for example, surgical margin , p = 0.020 and tumor stage, p<0.01) did not included for further analysis. **Response**: It is very common that an independent variable is significant in an unadjusted (adjusted) analysis but is not significant in the adjusted (unadjusted) analysis. We **DID** consider all variables in the adjusted analysis, but from the variable selection procedure, some were not significant in the adjusted analysis and were not included in the final model. We already explained the reason in the previous response letter (see the response to comment #5 in the previous response letter as attached).

In addition, the patients numbers in the table 1 still inconsistent (for example, page 16 line 37 of table , N0 stage patient number n=253; but page 17 line 42 positive lymph node number 0, n=257.

Response: We double checked and are very confident that the two numbers mentioned (line 37 on page 16, and line 42 on page 17) were the **SAME** (i.e., 257) in the previous revised manuscript. The mentioned inconsistence **DOES NOT** exist.

Finally, the interaction among variables of multivariate analysis in Table 2 are obvious, such as tumor size, lymph node metastatic rate and invasion depth, all are associated with tumor stage, but authors neglect this factor into multivariate analysis. The miss interpretation of statistics analysis ineviable provide wrong information of this manuscript.

Response: Fist, again we **DID NOT** do multivariate analysis. Second, the final model was fitted using the variables selected from the result of a variable selection procedure, which is a very common and standard data analysis strategy. Third, an independent variable is associated with another one **DOES NOT** imply that their interaction is significantly associated with the dependent variable. Forth, to our best knowledge, there is **NO** misinterpretation in our data analyses, **NOR** wrong information was provided by our manuscript.

Appendix: the following are the responses to the previous comments from reviewer 1.

2. Please clearly define and elaborate who did author select variables in the multivariable model ?
Response: We guess the question was "which author did variable selection in the adjusted analysis". We used the SAS procedure Cox proportional hazard regression with the option "backward" variable selection. In the original model in the revised analysis, we entered all of the main effects and the possible two-way interaction terms. The program automatically selected the significant main and interaction effects which were used in the final model (please see Table 2).

\# 5. Why age, age square, and LN metastatic rate did not reach statistical difference in unadjust analysis, but were independent prognosticators in the adjusted analysis. Please discuss the reasons.
Response: This is a common phenomenon in statistical practice. It has been recognized long time ago that conditional association (e.g., none zero adjusted effect) does not imply marginal association (e.g., none zero unadjusted effect), and vice versa.

VERSION 4 – REVIEW

REVIEWER	Wen-Chi Chou Chang Gung Memorial Hospital at Linkou, Taiwan
REVIEW RETURNED	30-Aug-2019

GENERAL COMMENTS	The authors didnot provide adequate response to my concerns. The only question is: Sixteen significant prognostic factors of gastric cancer including Borrmann's type, surgical margin, M stage, N stage, T stage, lymph node metastasis rate group, surgical type II group, clinical stage, number of cancer nodules group, tumor size group, invasion depth group, positive lymph nodes number group, tumor location, positive lymph nodes number, number of retrieved lymph nodes, and number of cancer nodules were identified (p<0.05) from the unadjusted analyses." dose all these 16 factors put into multivariate analysis? if yes, please provide the result of multivariate for all the 16 clinical variables. If no, please address the reasons for choose clinical factors into multivariate analysis.