Response to the reviewers' comments

Title: Estimating the Survival Benefit of Adjuvant Therapy Based on Bayesian Network in Curative Resected Advanced Gallbladder Adenocarcinoma

Ref. No.: 50116

Journal: World Journal of Gastroenterology

Dear reviewers and editor,

Thank you so much for your review and advice on our research work. Based on the associate editor and reviewers' comments and suggestions, we have revised the manuscript and provided point-to-point responses. We would like to submit a revised manuscript to *World Journal of Gastroenterology* for the consideration of potential publication.

The following paragraphs summarize our responses to the comments and suggestions. We hope you find the revisions acceptable.

Yours sincerely,

Shubin Si PhD
Professor
Department of Industrial Engineering
School of Mechanical Engineering
Northwestern Polytechnical University

General Responses to Associate Editor and Reviewers

We would like to thank the associate editor and three referees for their efforts in reviewing the early version of our manuscript. We are very appreciative to their insightful comments as well as the constructive suggestions which resulted in a much improved paper. For easy reference, reviewers' original questions/comments are reproduced below, followed by our responses highlighted in blue. If the title of the table (or figure) is highlighted, meaning the content of the table (or figure) has been updated as well.

Responses to Reviewer #1

1.This study supports the recent knowledge of the advantage of adjuvant radio chemotherapy in advanced stages of gall bladder carcinoma in a large amount of patients.

Response:

Thank you so much for your kind review.

Responses to Reviewer #2

1.The paper describes the factors associated with prognosis gall bladder cancer, which were identified by Bayesian network using database of Surveillance, Epidemiology, and End Results. The authors found that the survival time was associated with adjuvant chemoradiotherapy, radiotherapy and chemotherapy, rather than surgery alone. The model and statistical calculations appears reasonable. If the authors add the informations, the readers can understand the article easily. Figures needs more explanations. For example, it is unclear what "?" means in Figure 2 and what is the difference in black and blue arrows. There are many grammatical mistakes. Also, there are many strange sentences and wordings. The manuscript needs thorough revision.

Response:

Thanks for your suggestions. We have made correction according to the reviewer's comments. Figures had been added more explanations. The grammatical mistakes had been checked. The "?" in Figure 2 means that the variable has NA values, and we use

SEM (Structural Expectation-maximization) algorithm to imput the NA values in the dataset. Blue lines in Figure 2 represent the relationship between the attribute variable and the target variable. Meanwhile, black lines represent the relationship between the attribute variables.

Responses to Reviewer #3

General comments:

1. For the suture of this paper, I suggested that the authors performed the cox regression to determine the independent factors of survival, then evaluate the importance of these independent factors, and then using these independent factors to develop the prediction model and evaluate the efficacy of the model.

Response:

Thanks for your suggestions. It is true that using Cox regression to determine the independent factors of survival was a popular method in survival analysis. As we all know, the Cox regression is a kind of regression method. The attribute variables need to be independent to avoid collinearity for evaluating the importance. However, in this article, we applied the Bayesian network, which is a kind of probability graphic model. In fact, for multivariate problems, if we can find the multivariate joint probability distribution, the latter prediction problem is basically solved as solving the marginal probability problem. However, the complexity of the multivariate joint probability distribution is very high and difficult to solve. The Bayesian network can simplify the calculation and solve the prediction probability by establishing a Bayesian network graph model, using the structure graph to solve the joint probability distribution solution, and estimating the value of the joint probability distribution through the data. Therefore, we do not need to consider the independence between variables, even if there is a correlation between the variables, the Bayesian network can handle it. It could build the prediction model by estimating the conditional probability, and then using the structural of model to simplify the calculation of joint probability. It doesn't need to keep the independence of variables. Therefore, it is unnecessary to adopt the Cox regression method in this article.

2. I agree with the authors that using "survival time ≤ 9" and "survival time > 9" to split patients in table 2 and Figure 2. However, I am curious the results that using

“survival” and “death” to split patients. Can the authors performed the assays in comment 1 using “survival” and “death” to split patients?

Response:

Thank for your suggestions. For one thing, the doctor is concerned how long the patients can live after surgery and does not care about survival or death. Because, as time goes, each patient will die, and predictive models for establishing a state of survival will not help clinical treatment. For another thing, when we count the living status of patients at the cut-off of follow-up, we found that there were 194 patients alive and 621 patients dead. If we use the dataset to establish model, this will be an imbalanced classifier problem. Therefore, we give up to use this model in the article.

3. Disease specific survival is also a stable parameter to evaluate the survival of patients, and several publications which obtain patients from SEER database have included this endpoint. Thus, it will be very nice that the author can evaluate the prediction model in disease specific survival, or develop another model for predicting in disease specific survival. Thus, can the authors performed the assays in comment 1 using disease specific survival?

Response:

Thanks for your suggestion, it is really true as reviewer suggested that establishing the prediction model about disease specific survival has great significance. However, the problem is that we cannot obtain the variable of disease specific survival in SEER database. We will obtain the disease specific survival time of the patients in our hospital and establish the prediction model according to the reviewer's suggestion in the future.

2.Please use “chemotherapy”, “radiotherapy”, and “chemoradiotherapy” instead of “CTx”, “XRT”, and “cXRT” in the abstract and the whole manuscript.

Response:

Thanks for your suggestions. We have made correction according to the reviewer’s comments.

3.Please defined what is “Scope Reg LN Sur”, “Number Ln”, “Surg Oth Reg”, “Surg Prim Site”, “positive Ln” in the abstract and the whole manuscript?

Response:

We have made correction according to the reviewer's comments. We added the definition of these variables below table 1.

"Scope Reg LN Sur", Scope of Regional Lymph Node Surgery describes the procedure of removal, biopsy, or aspiration of regional lymph nodes performed during the initial work-up or first course of therapy at all facilities.

"Number Ln", Records the total number of regional lymph nodes that were removed and examined by the pathologist.

"Surg Oth Reg", Surgical procedure of Other Site describes the surgical removal of distant lymph node(s) or other tissue(s) or organ(s) beyond the primary site.

"Surg Prim Site" , Surgery of Primary Site describes a surgical procedure that removes and/or destroys tissue of the primary site performed as part of the initial work-up or first course of therapy.

"Positive Ln", Records the exact number of regional lymph nodes examined by the pathologist that were found to contain metastases.

4.Usually the median was followed by interquartile range. Thus, please provide the interquartile range in the whole manuscript.

Response:

We have added the interquartile range in the article according to the reviewer's comments. The median survival time was 9 months (interquartile, 4-19 months).

5. The authors statement that: "In order to evaluate the model performance more accurately, 70% (574) of patients formed the training dataset to establish Bayesian model and the remaining 30% (244) patients were considered as the testing dataset to test the model."

The data were spited randomly? If it is, please describe how they performed the randomization.

Response:

Thanks for your suggestions. We are very sorry for our negligence that did not write the method clearly in the article. We have made correction according to the reviewer's comments. "*In order to evaluate the model performance more accurately, the strategy*

of stratification sampling was adopted to split the dataset to training dataset and testing dataset. 70% (574) of patients formed the training dataset to establish Bayesian model and the remaining 30% (244) patients were considered as the testing dataset to test the model."

For strategy of stratification sampling, we first split the data randomly into two stratifications according to variable of survival time which has been discretized to categorical variables. Then, in every stratification dataset, we extracted 70% data randomly to compose training dataset, and the rest dataset were used for testing.

6. It would be very nice that the authors can reported their finding according to TRIPOD and cited the paper (Moons K G M, Altman D G, Reitsma J B, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Annals of internal medicine, 2015, 162(1): W1-W73.)

Response: Thanks for your suggestions. We have cited this paper in the Introduction.

7. I am do not understand which parameter were adopted to develop the prediction model. The parameters presented in Fig. 2 or parameters presented in Table 3. The author MUST defined clearly in the text. In addition, the cox regression is MUST be performed before the development of the model. This will allow that the authors only include independent parameters in this model.

Response:

Thanks for your suggestions. We have re-written this part according to the suggestion. We first established the Bayesian network prediction model based all 17 variables in Figure 2. Then, we do not change the model, and just select four variables shown in Table 3 as observation variables to obtain the prediction table. This means that we just changed the status of the four variables based on established BN model. In other words, we use the established Bayesian network to solve marginal distribution for the four variables. It was one of useful method and the advantages of Bayesian network to calculate the posterior probability.

8. The authors state that: "Moreover, although this study was performed using the SEER database, the study population was only 818 patients after screening. Large

volume, prospective, randomized controlled clinical trials are therefore needed to validate the prediction model in the future."

To solve the limitation of small sample size, I suggested the author performed 10-fold cross-validation or bootstrap (Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection//Ijcai. 1995, 14(2): 1137-1145.)

Response:

Thanks for your suggestions. Hold-out, cross-validation and bootstrap are three commonly methods used for data split in machine learning. In our research, we split data into training dataset and testing dataset using the ratio with 7:3. According to the reviewers' suggestions, we also tried the 10-fold cross-validation method. The results showed that average accuracy of model was 67.97% and average AUC of ROC was 72.002%, which means that 10-fold cross-validation and hold-out method are basically similar in this dataset. However, considering that if 10-fold cross-validation was adopted in this paper, the overall framework, figures and tables of the article need to be revised and re-written. Therefore, this study still uses the hold-out method. But in the future research, we could use 10-fold cross-validation method.

9. For patient selection, as far as I know that some patients lost the information of lymph node, T-stage, radiation, chemotherapy, etc. (https://seer.cancer.gov/manuals/2018/SPCSM_2018_maindoc.pdf). Did the author also excluded these patients?

Response:

Thanks for your suggestions. Based on rigorous criteria of inclusion, we obtain 818 patients cases and 18 attribute variables. For the 818 patients, the variables of T stage, Number Ln, radiation and chemotherapy has not NA values. Statistics description can be found in Table1. While, the variables of positive Ln have 47.68% NA values. If we delete all patients with NA record directly, there will be only a few patient cases for study. So we use SEM (Structural Expectation-maximization) algorithm to impute the NA values in the dataset.

10. The author should declare the limitation of the radiation therapy and chemotherapy information of SEER database (https://seer.cancer.gov/data/ChemotherapyRadiation-SEER-DUA.pdf)

Response:

Thanks for your suggestions. We have made correction according to the reviewer's comments. Two main limitations affect recommended analyses using the SEER RT and chemotherapy data: 1) the completeness of the variables; and 2) the biases associated with unmeasured reasons for receiving or not receiving RT/chemotherapy.

11. The author should declare who access the SEER data (https://seer.cancer.gov/data/access.html) and Treatment Data (https://seer.cancer.gov/data/treatment.html)

Response:

Thanks for your suggestions. We have made correction according to the reviewer's comments. Dr. Feng Xue accessed the SEER data and Treatment Data and we declared it in the author contributions.

Other comments

1. Title

The main results and conclusions of this manuscript is that the authors developed a robust prediction model for curative resected advanced gallbladder adenocarcinoma. Thus, it would be very nice that the author can put "prediction model" in the title.

Response:

Thanks for your suggestions. We have made correction according to the reviewer's comments, as "Estimating the Survival Benefit of Adjuvant Therapy Based on Bayesian Network Prediction Model in Curative Resected Advanced Gallbladder Adenocarcinoma".

2. Abstract

"*The probability of a survival time of > 9.0 months was associated with, in order of the highest to lowest, adjuvant chemoradiotherapy (cXRT) > XRT >CTx> surgery alone, for patients with node-positive disease, the model predicted a larger benefit from xCRT*"

I do not understand what the authors want to tell the reader. Please re-write this sentence.

Response:

Thanks for your suggestions. We have made correction according to the reviewer's comments. "The distribution of the survival time (> 9.0 months) was affected by different treatments with the order of adjuvant chemoradiotherapy (cXRT) > XRT >CTx> surgery alone. For patients with node-positive disease, the larger benefit predicted by the model is xCRT".

3. Introducrion

"The role of adjuvant therapy for GBC, however, is not well-known at this time."

Please add one reference here.

Responses:

Thanks for your suggestions. We have made correction and added three references according to the reviewer's comments.

4. Method

4.1 *"We removed patients who did not have the ICD-0-3 codes 8140, 8141, 8143, or 8147, which designate adenocarcinoma."*

Please add one reference which use the ICD-0-3 code to collect patients from SEER database.

Response:

Thanks for your suggestions. We have added reference [18] according to the reviewer's comments.

Lau CSM, Zywot A, Mahendraraj K, Chamberlain RS. Gallbladder Carcinoma in the United States: Population Based Clinical Outcomes Study Involving 22,343 Patients from the Surveillance, Epidemiology, and End Result Database (1973-2013). HPB Surg. 2017;2017:1532835. doi: 10.1155/2017/1532835.

4.2 *"The variable of age was divided into three intervals of 19 to 64 years, 65 to 75 years, and 76 to 97 years. Positive Ln was divided into three intervals of 0, 1 to 3,*

and >3. Number Ln was divided into four intervals of 0, 1 to 3, 4 to 6, and > 6. Tumor size was divided into four intervals of 0 to 10 mm, 11 to 30 mm, 31 to 50 mm, and > 50 mm based on medical definitions"

To determine the threshold (cut-off), the author should perform ROC curve and Youden Index (Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. Indian pediatrics, 2011, 48(4): 277-287.)

Response:

Thanks for your suggestions. We have discussed the method of ROC to determine the cut-off of variables. We found that method of ROC mainly was used for binary variable to determine the cut-off. In our research, the variables will be divided into three or four intervals. Therefore, we adopt the equal frequency to split the variable of age, and split medical variables according to medical definitions.

4.3 *"In some datasets, when the number of negative and positive cases varies, the accuracy may not be the appropriate criteria. Considering this condition, the receiver operating characteristic (ROC) curve and the area under the curve (AUC) were calculated to measure the overall performance of the classification model further."*

Please write clearly in which dataset, such as in Fig.x or in Table x.

Response:

Thanks for your suggestions. We have made correction according to the reviewer's comments, as "when the number of negative and positive cases are imbalance in dataset, the accuracy may not be the appropriate criteria. Considering this condition, the receiver operating characteristic (ROC) curve and the area under the curve (AUC), one of useful evaluating criteria, was calculated to measure the overall performance of the classification model further."

4.4 *"All continuous variables were transformed to discrete variables for BN analysis and expressed with frequency and percent. Categorical variables were presented with frequency and percent. Survival curve was estimated with the Kaplan–Meier method and the results were compared with the log-rank test. "*

See comments 4.2. To determine the threshold (cut-off) of continuous variables, the author should perform ROC curve and Youden Index (Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. Indian pediatrics, 2011, 48(4): 277-287.)

Responses:

Thanks for your suggestions. Please refer to the response for general comments 4.2.

5. Results

5.1 The authors state that: *"Out of the patients who received radiation, only 4 (0.49%) received it before surgery, while 222 (27.14%) received it after surgery. Regarding chemotherapy, 345 (42.18%) underwent it, while 473 (57.82%) did not."*

For patients receiving radiation, 0.49% received it before surgery and 27.14% received it after surgery, however, what the other 72.37% of patients received? For patients not receiving chemotherapy (473), my feeling is that some of these patients were indicated "Unclear" or "Not available" (https://seer.cancer.gov/data/treatment.html). If this is true, please defined these patients in the text, figures, and the tables. Please do the same with patients receiving radiation.

Response:

Thanks for your suggestions. After calculating again, for variable of radiation sequence, 0.49% received it before surgery and 27.14% received it after surgery. The other 72.37% were patients that do not receive the radiation or unclear. This is due to the limitation of the SEER database. Regarding chemotherapy, 345 (42.18%) underwent it, while 473 (57.82%) did not or unclear. It is pity that there were not the variable described the order between surgery and chemotherapy. We have made correction according to the reviewer's comments in the paper.

5.2. *"When the follow-up was cut-off in December 2015"*

How the authors obtained this information?

Response:

Thanks for your suggestions. When we download the SEER dataset, the description PDF showed that the follow-up cutoff date was December 31, 2015.

SEER RESEARCH DATA RECORD DESCRIPTION CASES DIAGNOSED IN 1973-2015.

https://seer.cancer.gov/data-software/documentation/seerstat/nov2017/TextData.FileDescription.pdf

5.3 *"The results of confusion matrix are listed in Table 2. There were 114 patients whose survival time was less than 9 months and 130 patients who had a survival time of longer than 9 months. "*

In table 2 the authors only presented results of testing data set. I would be very nice that the authors can show the results of tanning dataset in the same table.

Response:

Thanks for your suggestions. The accuracy of training dataset was 70.21%, and AUC under ROC was 77.50%.The confusion matrix of training dataset was showed in the table below. Since the accuracy of the testing dataset is generally concerned, the results of the training set do not need to be placed in the text.

Confusion matrix of training dataset

	Survival time (n)	≤9m (n=269)	>9m (n=305)
Confusion matrix (n)	≤9m (298)	198	100
	>9m (276)	71	205
Reliability (%)	≤9m (298)	66.44%	33.56%
	> 9m (276)	25.72%	74.28%
Accuracy (%)	≤9m (298)	73.61%	32.79%
	> 9m (276)	26.39%	67.21%

5.4 "A total of 83 patients were correctly classified as having survival time ≤ 9 months and 86 patients were classified as having survival time of > 9 months, based the probability threshold of 0.5."

I agree with the authors that using "survival time ≤ 9" and "survival time > 9" to split

patients in table 2 and Figure 2. However, I am curious the results that using "survival" and "death" to split patients. Can the authors show me the results as they did in table 2 and Figure 2.

In addition, how the authors defined the threshold of 0.5? Please write clearly.

Response:

Thanks for your suggestions. For establishing model using "survival" and "death", we explained it in general comments 2. Threshold of 0.5 is very common used in machine learning problems. If we thought that the cost of recall and precision is similar, the threshold of 0.5 could be used for classifier problems. In fact, the accuracy would be different with the threshold varying, so we introduced the ROC curve to evaluate the accuracy of the model.

5.5 ***"Prognostic factors ranked by importance"***

I think the cox regression followed by the importance measure should be presented before the authors introduce the model. This means I suggested that the authors performed the cox regression to determine the independent factors of survival, then evaluate the importance of these independent factors, and then using these independent factors to develop the prediction model and evaluate the efficacy of the model.

Responses:

Thanks for your suggestions. Please refer to the response of general comments1.

5.6 *"The results indicated that the radiation was the most important prognosis factor influencing survival time after radical resection for advanced GBC patients."*

I still think the cox regression is necessary to support this conclusions. This means only adjusted other risk factors, the radiation is still an independent risk factor. Then the authors can concluded that radiation is a prognosis factor of patients. Only based on this point (radiation is a prognosis factor of patients), evaluating the importance is

valuable, and this is same to other parameters.

Responses:

Thanks for your suggestions. Please refer to the response of general comments 1.

5.7 "We combined the BN model and importance measures to select radiation, chemotherapy, T-stage and N-stage as the observation variables."

I do not understand that why combined the BN model and importance measures, the authors can select radiation, chemotherapy, T-stage and N-stage as the observation variables. Please write clearly.

Responses:

We have re-written this part according to the reviewer's suggestion. "We combined the BN model and importance measures to identify radiation, chemotherapy were important prognosis factor. Meanwhile, T-stage and N-stage are always used to determine the severity of the patient's illness. As a result, we select radiation, chemotherapy, T-stage and N-stage as the observation variables to obtain prediction table".

5.8 *"For patients with node-negative disease, the model estimated the similar survival benefit from the addition of XRT and cXRT, regardless of T3 or T4 stage. For example, for a patient with T3N0 disease, his/her probability of a survival time of > 9 months with surgery alone, CTx, XRT, cXRT was 41.35%, 58.29%, 75.42% and 76.62%, respectively. For patients with node-positive disease, the model predicted a small survival benefit from CTx and XRT, and a larger benefit from xCRT. For example, for a patient with T4N1 disease, his/her probability of a survival time of > 9 months with surgery alone, CTx, XRT, cXRT was 14.85%, 37.03%, 43.14% and 57.97%, respectively."*

Please take out "similar", "large" from the text, because these subjective words do not give objective information to the readers. Please write down the true number, such as "his/her probability of a survival time of > 9 months with surgery alone, CTx, XRT,

cXRT was 41.35%, 58.29%, 75.42% and 76.62%, respectively"x

Response:

Thanks for your suggestions. We have made correction according to the reviewer's comments."For patients with node-positive disease, the model predicted survival benefit from CTx, XRT and xCRT. Meanwhile, patients acquired more benefit from xCRT than CTx and XRT".

5.9 *"The median OS for the advanced GBC patients……. (Figure 3)"*

In Figure 3, the authors performed this analysis using the entire patients, thus *advanced GBC patients*, is not correct, because some of the patients are early stage of GBC.

Responses:

We are very sorry for our writing that may cause misunderstandings. In our research, we only selected patients whose T stage were T3 and T4, which means that all 818 patients were advanced GBC patients.

5.10 *"There was a significant difference among the different adjuvant therapy groups (log rank, P=0.000) (Figure 3)."*

Firstly, *P=0.000*is not correct, because P-value never be zero. Please correct as *P*<0.001. Secondly, what is the meaning of the *P*-value, for comparing Surgery alone vs. CTx? Or for comparing Surgery alone vs. XRT. Please write the *P*-value of each comparing in the figure legends. Thirdly, the information of patients receiving chemoradiotherapy is missing in Table 1. Please add these information.

Response:

Thanks for your suggestions. For the first two suggestions, we have made correction according to the reviewer's comments. For the third suggestion, there weren't directly variable which described the information of patients receiving chemoradiotherapy.

However, for every patients, we know that whether he or she received radiation or chemotherapy. By combining the values of these two variables, we could get the information of chemoradiotherapy for patients.

For p values of each pair, we showed the results in table below.

Pairwise Comparisons									
		0(Surgery alone)		1(CTx)		2(XRT)		3(cXRT)	
	Method	Chi-Square	Sig.	Chi-Square	Sig.	Chi-Square	Sig.	Chi-Square	Sig.
LogRank (Mantel-Cox)	0			15.917	.000	6.032	.014	57.060	.000
	1	15.917	.000			.841	.359	13.660	.000
	2	6.032	.014	.841	.359			1.602	.206
	3	57.060	.000	13.660	.000	1.602	.206		

5.11 In the discussion section, please avoid using words such as “advanced” to describe GBC. In addition, avoid using “small” and “large” to describe survival benefit.

Response:

Thanks for your suggestions. Advanced GBC refers to gallbladder cancer with late TNM stage. In many literatures, advanced GBC is T2 or higher compared with early gallbladder cancer. In our research, the inclusion criteria are above T3, which accounts for the majority of patients with gallbladder cancer, and the prognosis is worse. The below are some references. About “small” and “large”, we have made correction.

1.Indications for major hepatectomy and combined procedures for advanced gallbladder cancer. Br J Surg 2017;104:257-266.

2. Surgical Treatment of Advanced Gallbladder Cancer. Am J Clin Oncol 2015; 38: 5-10

3.Aggressive Surgery for Locally Advanced Gallbladder Cancer with Obstructive Jaundice: Result of a Prospective Study. Dig Surg 2016;33:213–219