**Peer Review File**

**Article Information:** https://dx.doi.org/10.21037/jgo-21-82

**Reviewer A**

This manuscript pieced together by Xu et al. would make a good addition to current literature. The authors' study on nomograms for the prediction of pathological response to chemotherapy in patients with colorectal liver metastases, discusses an important and novel topic within the field. Chemotherapy response is an established predictor of long-term survival. They report important findings that have the potential to be easily integrated into clinical workflows and improve care.

The methods are appropriate and rigorously conducted. The statistical analysis of the data is sound. Their claims are fully supported by their experimental data and limitations appropriately discussed in the context of published literature.

**Comment**:

I recommend publication acceptance after clarifying few minor points:

Could the authors please make clear in methods how the variables were identified and inserted into the multivariate analysis. I presume after univariate analysis and scanning of literature?

**Reply to Reviewer A:**

Thanks very much for your considering of our manuscript. Just as you said, variables that were statistically significant in the univariate analysis were included in the multivariable logistic regression analysis to determine independent predictors of major pathologic response. We had described it in the text according to your request.

**Changes in the text:** Page 10, line 9-11; Page 12, line 11-14

**Reviewer B**

With great interest I read this manuscript by Xu et al. The authors described a single-center, retrospective study of patients who received neoadjuvant chemotherapy to treat colorectal liver metastases (CRLM) and who underwent resection subsequently. They aimed to identify factors predictive of a major pathological response (<50% viable tumor cells). Disease free interval (DFI), number and size of the metastases, and RAS status were independent factors influencing the chance of a major pathological response. The authors produced a nomogram, which was internally validated and tested in subgroups.

**Major points:**

**Comment 1**:

As acknowledged by the authors there was no external validation. Therefore publication of the nomogram is premature in my view. I would suggest to omit this.

**Reply 1:**

Thank you for your suggestion. Lack of external validation is one of the limitations of our article. Because of this, the reliability of the model was suboptimal. To reduce the bias and develop the reliability of the model, we did an internal validation randomly divided into the training cohort (n = 241) and validation cohort (n = 241). The 2 cohorts included totally different

patients, indicating that the model could predict pathologic response to some extent. Besides, in order to evaluate the applicability of the nomogram in patients with different characteristics, subgroup analyses were performed based on some important factors most related with prognosis and chemotherapy response. Patients with different characteristics also show high consistency with the results of the training and testing cohort. We believe the model would be helpful for clinicians to fully consider chemotherapy response before surgery and guiding the treatment strategy of performing local treatment for patients with CRLM.
**Changes in the text:** No changes were made in the text.

**Comment 2**:
In the methods section, it must become very clear which patient was scheduled for neoadjuvant chemo and which patient wasn't.
**Reply 2:**
Thank you for your recommendation. The detail of patients who was scheduled for neoadjuvant chemo in our daily practice were not described in the paper. We had added this part in MATERIALS AND METHODS section.
**Changes in the text:** Page 8, line 12-16; Page 9, line 1-7

**Comment 3**:
Different from the explanation of the authors on the role of DFI, it might be possible that patients with late metachronous CRLM and limited disease did not receive chemotherapy, and therefore the selection of patients with DFI>12 months is a selection of patients with a higher risk. And selection rather than administration of adjuvant chemo accounted for the lower chance of a major pathological response. Previous adjuvant therapy for the primary tumor should be included in the variables. The authors should address the effect of time (i.e. early versus late in the study) on outcome. Did they observe interaction between administration of targeted therapy and triplet therapy with time later in the study?
**Reply 3:**
You raised a very important question. We agree on your explanation that patients with DFI>12 months who receive neoadjuvant chemotherapy might have heavier tumor burden, leading to the probably of selection bias. This explanation is in line with our conclusion that patients with limited liver metastases are easier to show major pathologic response. In addition, we also believe that lots of patients with DFI>12 months might need to change to 2nd-line chemotherapy, since majority of them might have received adjuvant chemotherapy (1st-line) before. The efficiency of 2nd-line chemotherapy for CRLM was relatively low, and it might be more difficult to develop major pathologic response for these patients.

In this study, more than half patients with DFI < 12 months are those with synchronous liver metastases, which means that these patients are not suitable for adjuvant chemotherapy. That's why we didn't include adjuvant therapy as a factor in analysis.

We did analyze the effect of time on outcome before, our previous results showed although more patients received targeted drugs and aggressive chemotherapy, the response rate was not significantly improved (1). The reason might be explained by more CRLM patients with heavier tumor burden were treated in these years. The prognosis of these patients is inherently poor. A more severe liver metastases burden balances the improvement of chemotherapy on

pathologic response.
**Changes in the text:** Page 17, line 8-16; Page 18, line 1-3

**Comment 4**:
The authors should specify which patients were amenable to triplet and targeted therapy throughout the study.
**Reply 4:**
We had added the content of which patients were amenable to triplet and targeted therapy in MATERIALS AND METHODS part.
**Changes in the text:** Page 9, line 3-7

**Comment 5**:
In the statistics paragraph the author should state how they dealt with interaction and multiple testing. They should specify subgroup analyses. I personally don't like these subgroup analyses in such a retrospective study.
**Reply 5:**
Thanks for your suggestions. The variables that were statistically significant in the univariate analysis were included in the multivariable logistic regression analysis to determine independent predictors of major pathologic response. To evaluate the applicability of the nomogram in patients with different characteristics, subgroup analyses were performed based on 3 factors which was commonly regarded as main factors influencing survival and pathologic response. In subgroup analysis, receiver operating characteristic (ROC) curve analysis was conducted, and the area under the ROC curve (AUC) was calculated to further evaluate the predictive performance of the nomogram. We believe the data would be useful for clinicians to predict pathologic response in different situation.
**Changes in the text:** Page 10, line 9-11; Page 12, line 11-14

**Comment 6**:
I'm in doubt whether most oncologists would agree with the distinction of just two categories. I would like to also see the data of complete response, near complete response (<10% vital tumorcells). I would suggest adding survival curves for the used groups and suggested groups above instead of the nomorgram.
**Reply 6:**
The most commonly used tumor regression grade (TRG) criteria in CRLM were MD Anderson(2) and Rubbia-Brant criteria(3). MD Anderson criteria divided patients into 2 groups, major response (1-49% residual cancer cells) and minor response (≥50% residual cancer cells). Rubbia- TRG scoring system includes 5 categories based on the amount of residual cancer cells, fibrosis and necrotic area. The MD Anderson criteria evaluates the pathologic response based simply on the residual tumor amount, which is more objective and convenient to use. Although Rubbia-Brant criteria could stratify patients more detailed but rely more on the experience of pathologists and has a certain degree of subjectivity in judgment. Besides, since the aim of the study is to establish a predictive model, patients need to be clearly divided into two groups. That's why we choose MD Anderson criteria as distinction of two categories. In addition, since many previous studies have proved that the MD Anderson

criteria can well differentiate the survival of patients, we didn't add the survival analysis.

Just as you said, complete response or near complete response (<10% vital tumor cells) might indicates a more favorable prognosis of patients. However, the proportion of pathological complete response (pCR) and nearly complete patients is very low. Our previous studies have shown that this ratio was about 6%(1). Other previous research also confirmed this result(4). Therefore, taking this into consideration, focusing on this small number of patients as the prediction end point will result in a large prediction error. That's why we didn't used the suggested groups as study point.

**Changes in the text**: No change was made

**Comment 7**:

The authors state that BRAF status is determined but do not mention this in the tables.

**Reply 7:**

About 5% patients with metastatic colorectal cancer would have BRAF mutation. However, patients with BRAF mutation often had a more aggressive biological behavior and heavier liver burden or more extensive extrahepatic metastases. Thus, the number of patients with BRAF mutation who are suitable for hepatectomy was relatively low. In the whole cohort of this study, only 6 patients were BRAF mutation. The patients' number were too small to include in analysis. That's why we didn't included in the tables.

**Changes in the text**: No change was made

**Comment 8**:

The authors should describe heterogeneity between de CRLM in the results section

**Reply 8:**

**Thanks for your suggestion.** For patients with multiple liver metastases, the TRG value was calculated using the average TRG value of the largest 3 lesions in order to reduce the heterogeneity between each CRLM tumors.

**Changes in the text:** Page 8, line 1-3

**Minor**

**Comment 9**:

This paper would benefit by correction of the text by a native English speaker.

**Reply 9:** The paper has been edited by native speaker. The editing certificate has been attached in supplementary file.

**Changes in the text**: No change was made

**Comment 10**:

In table 1 and 2, give age, tumor size, tumor number and number of cycles as a continuous variable with median and range. For RECIST assessment use the categories complete response, partial response, stable disease and progressive disease. Add doublet or triplet systemic chemotherapy. Add previous adjuvant therapy for the primary tumor. Add type after RAS (i.e. Wild-type)

**Reply 10:**

Thank you for your suggestion. In order to do the multivariate analysis to explore the predictive

factors influencing pathologic response, all the factors included were classified as categorical variables. The RECIST assessment, targeted drugs, RAS types has been added in the table.

Since more than half patients with DFI < 12 months are those with synchronous liver metastases, which means that these patients are not suitable for adjuvant chemotherapy. That's why we didn't include adjuvant therapy as a factor in analysis.

**Changes in the text:** Table 1

**Comment 11**:

Why did the authors choose a cut off for CEA of 50? Fong et al. provided the most practical risk classification and chose 200.

**Reply 11:** Just as you said, in Fong et al. CRS criteria, the cut-off value of CEA is 200ng/ml. However, as the development of systematic chemotherapy, the CEA of most patients would decrease, and few patients would still show CEA over 200 ng/ml after chemotherapy. In the whole cohort, no more than 20 patients had CEA over 200 ng/ml preoperatively. If we choose CEA 200ng/ml as cut-off value, it will be inaccurate in multivariate analysis, thereby affecting the accuracy of establishment model.

**Changes in the text**: No change was made

**References:**

1. Xu D, Yan XL, Liu JM, Li J, Xing BC. The characteristics and long-term survival of patients with colorectal liver metastases with pathological complete response after chemotherapy. J Cancer. 2020;11(21):6256-63.

2. Blazer DG, 3rd, Kishi Y, Maru DM, Kopetz S, Chun YS, Overman MJ, et al. Pathologic response to preoperative chemotherapy: a new outcome end point after resection of hepatic colorectal metastases. J Clin Oncol. 2008;26(33):5344-51.

3. Rubbia-Brandt L, Giostra E, Brezault C, Roth AD, Andres A, Audard V, et al. Importance of histological tumor response assessment in predicting the outcome in patients with colorectal liver metastases treated with neo-adjuvant chemotherapy followed by liver surgery. Ann Oncol. 2007;18(2):299-304.

4. Adam R, Wicherts DA, de Haas RJ, Aloia T, Levi F, Paule B, et al. Complete pathologic response after preoperative chemotherapy for colorectal liver metastases: myth or reality? J Clin Oncol. 2008;26(10):1635-41.