Peer Review File

Article Information: http://dx.doi.org/10.21037/jgo-20-329

Reviewer A

This retrospective study focuses on patients with colorectal cancer and synchronous liver metastases, who underwent simultaneous operation of both primary and metastatic tumors. Authors propose a preoperative nomogram to predict overall survival after resection. Compared to traditional Fong score, authors describe a slight improvement in predictive accuracy at time of 3-year OS. The text is well written and the proposed nomogram is well constructed and described.
However, there are several issues that severely limit the application of the score.

Comment1: Aim of the study is vaguely given. Do authors attempt to "improve" Fong score with additional preoperative factors, adjust the score for Chinese population or are authors trying to convince that simultaneous resection of both primary and metastasis may not be suitable for certain patients?

Reply1: The NCCN Guideline (Version 6.2020) has pointed out that curative hepatic resection provides survival superior to other treatments. And simultaneous resection has come into consideration among an increasing number of centers on account of safety and cost-effectiveness (See Abelson JS et.al, DOI 10.1007/s11605-017-3422-1). However, few studies so far focus on outcome of this subgroup of patients (who receive simultaneous resection for CRLM). *Fong's* score was initially formulated for metachronous CRLM resection and GAME score is not specific for simultaneous resection patients. Our study has proved substantial heterogeneity among such patients and more precise categorization is needed preoperatively to identify those who are at high risk of worse survival. Therefore, we provided this internally validated scoring system as an option. On the issue of ambiguous point of this article, we have modified Introduction and Discussion part of the text to stress the key point. The subjects are indeed from a single centered Chinese cohort, meaning application of our finding may be more suitable for Asian patients, which we also mentioned in the Discussion part. We hope our model can help identify patients with high risk of short-term mortality and assist physicians to decide on therapy schema.

Changes in the text: We have modified our text as advised (See Page 3, line 66)

Comment 2: The setting does not evaluate "survival benefit" from resection, since there is no control group, but is descriptive in nature, and the title and conclusions are thus misleading.

Reply 2: The guideline has acknowledged the survival benefit of hepatic resection for patients with CRLM. Also, some studies have confirmed the survival homogeneity of synchronous and metachronous resection (See ref #5-8 in the article). Therefore, we did not employ a control group to re-evaluate advantage of hepatic resection. In terms of misleading description in the title and main text, modification has been made for accuracy.

Comment 3: Surgical techniques are not described, 30-day mortality is not described. Have the techniques changed during the 10-year recruitment period?

Reply3: It is highly appreciated for pointing out our omission. We combed the clinical data and added surgical procedure information in our paper (See Page 5, line 110 and Table 1). For convenient classification, hepatic resection was divided into minor hepatectomy (<3 segments resected) and major

hepatectomy (≥3 segments resected).

Mortality within 30 days after surgery was added in a flowchart. (See Figure 1).

Stable surgical system has been established in affiliated hepatobiliary hospital of our college. We consulted cooperating specialists of hepatobiliary surgery to ensure that hepatic resection was conducted following the guidelines (See Page 5, line 107). We also added a Kaplan-Meier plot to prove stable outcome of hepatic resection through the whole course of observation (See Supplementary Figure1).

Comment 4: Authors also misinterpret the concept of validation: randomly choosing a third of patients for validation dataset may be viable in large datasets (thousands) or studies using machine learning. Validation dataset should be acquired from an independent source. Recommend merging all 234 patients in single cohort. Pg9 line 196, this study is NOT externally validated.

Reply 4: Cases in the training cohort were employed to construct the model and the result was internally validated with validation cohort. We admit lack of external validation and multi-center data as a limitation of our study which needs further improvement (Page 11, line 227). Although the value of simultaneous resection is increasingly accepted, unlike staged resection, few centers are capable of long-term conducting well quality-controlled simultaneous colorectal and hepatic resection. Designating a certain ratio to group subjects for training and validating has been applied in some studies with midsize datasets (e.g. Wang et al DOI: 10.2147/CMAR.S272797)

Changes in the text: We have modified our text as advised (See Page11, line 227).

Comment 5: Why molecular pathology is not used preoperatively like in GAME score?

Reply 5: Molecular pathology information was not preoperatively available for a large part of the patients. Most patients choose to receive surgery first and apply for gene status test with postoperative specimen in China, while our model was designed for preoperative assessment. We do hope with the popularization of more accessible genetic testing method, molecular pathology will serve a prior part in future scoring systems.

Changes in the text: We have modified our text as advised (See Page 11, line 231).

Comment 6: Number of excluded patients is not reported. Please also report the total pool of patients from the region to assess selection bias.

Reply 6: A flow-chart was added in the article to illustrate the process of case inclusion and exclusion.

Changes in the text: We have modified our text as advised (See Figure 1).

Comment 7: Also, the nomogram is rather inconvenient for clinical use without computer assistance, as compared to Fong and GAME scores. How should the score be interpreted? Author state on pg 8 that cut-off 170 is based on two-thirds of scores. What is the rationale for this? Should patient with score 160 be operated simultaneously for primary and if score 170 not?

Reply 7: We adopted *maxstat*, an algorithm in R to decide on a cut point with the most significant difference between groups. After modification, patients were stratified into two groups, namely high-risk and low risk. Nomogram is a proven method of quantify and visualize the weight of every influence factor and provide a statistical description of total risk. The assessment of nomogram should be taken as an assistance and physicians and patients still hold the final decision.

Changes in the text: We have modified our text as advised (See Page 8, line 167).

Minor comments

Comment 1: References in introduction (#2-5) do not adequately cover survival prognosis in mCRC. They should focus on references from China/Asia. Also, different preoperative scores such as Fong and GAME are not referred.

Reply 1: We updated references in the text and added results of study concerning Chinese population. Fong's and GAME score were also discussed in Introduction and Discussion parts.

Comment 2: Use term "node-positive primary" in analogy to Fong score to differentiate from LN metastases in abdominal cavity

Reply 2: We have modified our text as advised.

Comment 3: "histological type" should read "tumor grade"

Reply 3: We have modified our text as advised.

Comment 4: Term "multivariate" should be "multivariable"

Reply 4: We have modified our text as advised.

Comment 5: "EMVI" is radiological term, "LVI" is pathology

Reply 5: We have modified our text as advised.

Comment 6: Confidence intervals should be in Mat&med, not in results section

Reply 6: We have modified our text as advised.

Comment 7: For CEA and tumor size, please see Fong et al for cut-off values (CEA should be >200 not >= 200, which is actually correct in Fig 1)

Reply 7: We have modified our text as advised.

Comment 8: Table 3 & Fig 4, number of total patients is missing

Reply 8: We have redefined risk groups (described above) and added number of cases as required.

Comment 9: Fig 3 it is unclear which is the experimental nomogram and which is Fong score

Reply 9: Necessary annotation was added to the figure to make it clear.

Comment 10: Fig 4. Test p-values between low-med and med-high

Reply 10: After regrouping, a binary classification was used, and K-M curve was renewed correspondingly.

Changes in the text: We have modified our text as advised (See Figure 5).

Reviewer B

This study is to investigate the prognostic factors for survival in patients who underwent synchronous resection colorectal primary and hepatic metastases.

My comments and suggestions are below.

Comment 1: Please describe the clear meaning of 'tumor deposit' in the variables of the method section.

Reply 1: The concept of tumor deposit (TD) varies with update of AJCC guideline. Through the course of follow-up in our work, our center adopted AJCC 7th to assess the presence of TD. Pathologists were consulted to ensure consistency of criteria in tumor deposit.

Changes in the text: We have modified our text as advised (See Page 5, line 102).

Comment 2: Regarding the gene mutations, authors have roughly mentioned just as 'status of KRAS, NRAS, and BRAF' in the method description and 'wildtype, mutant type, and absent of KRAS, NRAS, and BRAF' in table 3. The mutations of genes they identified should be clearly described in detail.

Reply 2: The previous description in the text was vague as "absent of KRAS" meaning unknown gene status (already corrected). For economic concern, routinely tested gene loci for colorectal cancer in our center are exon 2, 3 and 4 of KRAS and NRAS and codon 600 of BRAF which are common and significant in colorectal cancer. We have clarified the expression and made further explanation in Table 3 and the text.

Changes in the text: We have modified our text as advised (See Page 5, line 105 and Table 3).

Comment 3: In the statistical analysis, it is necessary to describe why authors selected the variables with $p<0.15$ in univariate analysis for the analysis of multivariate Cox-regression model. Because the selection of variables for multivariate analysis are different from the criteria commonly used in other studies, so detail explanation is needed.

Reply 3: For the sake of comprehensive screening of relevant factors, we adopted relatively loose p-value cutoff in univariable analysis and restrained with $p<0.05$ in multivariable Cox. A relatively loose standard for inclusion in multivariable analysis has been adopted by other published articles (e.g. Bachet JB et al. DOI: 10.1002/bjs.11180).

Comment 4: The authors need to describe in detail the criteria for classifying the risk of 234 cases into low, medium, and high risk using a nomogram in the result section.

Reply 4: We re-classified all the cases with well-accepted maxstat algorithm and identified 135 as the cut point to make our result scientifically rigorous. This new classification showed high degree of distinguishing in Kaplan-Meier analysis.

Changes in the text: We have modified our text as advised (See Page 8, line 176 and Figure 5).

Comment 5: Please present consistently the decimal point of p-value on page 8.

Reply 5: Corresponding adjustment has been made in the text and figure.