Author’s response to reviews

Title: Predicting Prognosis in Hepatocellular Carcinoma after Curative Surgery with Common Clinicopathologic Parameters

Authors:

Ke Hao (ke_hao@merck.com)
John M Luk (jmluk@hku.hk)
Nikki Lee (nikkilee@hkucc.hku.hk)
Mao Mao (mao_mao@merck.com)
Chunsheng Zhang (Chunsheng_Zhang@merck.com)
Mark Ferguson (Mark_Ferguson@merck.com)
John Lamb (John_Lamb3@merck.com)
Hongyue Dai (hongyue_dai@merck.com)
Irene Ng (iolng@hkusua.hku.hk)
Pak Sham (pcsham@hkucc.hku.hk)
Ronnie Poon (poontp@hkucc.hku.hk)

Version: 3 Date: 31 July 2009

Author's response to reviews: see over
RE: MS: 3181177452396835 - Predicting Prognostics in Hepatocellular Carcinoma after Curative Surgery with Common Clinicopathologic Parameters

Reviewer#1

I. General comments
The authors first develop a predictive model for post-resectional survival of hepatocellular carcinoma (HCC) by incorporating many clinicopathologic parameters, based on a large cohort of HCC. Then the model is confirmed to be highly predictive in another cohort. Similar survival predictive system using such comprehensive clinicopathologic parameters has not been found in previously published literatures, to my best knowledge. Therefore, this paper seems to be of originality. Besides, the data are well organized to get a reasonable conclusion, and the manuscript is well written.

II. Specific comments
These issues should be addressed.
1. Title: The word “Prognostics” can be changed to “prognosis”, according to the usual statement. Revised.

2. Study Subjects
The initial training dataset includes patients between 1993 and 2007, whereas the independent cohort for validation enrolls patients between 1990 and 2004. If and how many patients in these two cohorts overlap each other should be clearly described. If not, the selection criteria for patients who entered different cohorts should be stated.

This section has been rewritten to enhance the clarity. See the first section of the Materials & Methods. Essentially, this is a retrospective study on analyzing common clinicopathologic parameters from the same cohort of 600 primary HCC patients who underwent hepatectomy as the primary treatment option in our medical center in the period of 1990 to 2007. We further divided the patients into training and validation sets, with n=300 in each group. In the training set, we also conducted molecular profiling analysis for biomarkers exploration study, and thus excluded 28 cases who did not have sufficient follow-up time, lack of clinical data and poor tissue quality. Therefore, there is 272 cases in the training set. There was no overlap of patients
between the training and validation sets, and no significant difference on the demographic and clinicopathologic parameters.

3. Discussion
Because some important parameters, such as pTNM and venous infiltration status, can only be obtained in postoperative histological examination, this model might be of limitation for its utility in patients who accept potentially curative therapies (for example, locoregional ablative therapies) without resected samples. This limitation might be stated and discussed in Discussion.

The reviewer's comment is well taken. In the Discussion section, we added, "The purpose of our study was to develop a systematic model according to universally recognized clinicopathologic parameters for improved accuracy of prognostic outcome prediction in HCC patients after curative surgery. Therefore, all acknowledged factors associated with outcomes were evaluated. Notably, certain important parameters, such as tumor staging and venous infiltration status, would require postoperative histological examination of the resected tissues or biopsy samples of the patients, unless future radiological examination using dynamic MRI could provide such definitive diagnosis. Nevertheless our primary contribution is the modeling framework which incorporates multiple parameters in prediction prognosis. Its flexible nature allows us to easily remove parameters (e.g. VENINV when either biopsy samples are not available or diagnostic radiography data is not affirmative) or to add new biomarkers (e.g. newly identified gene signatures)."
Reviewer#2

This manuscript by Hao et al describes a new scoring system to predict survival in patients with HCC after survival. The study has a number of major flaws:

1. Page 2, line 1: Surgical resection is not the only standard curative treatment option for patients with HCC as suggested by the authors. RFTA is also considered as a curative therapy.
   The reviewer's comment is well taken. We rewrote the sentence as, "Surgical resection is one important curative treatment for hepatocellular carcinoma (HCC), but the prognosis following surgery differs substantially and such large variation is mainly unexplained."

2. How were the patients chosen for this study? Are these all patients treated at the authors institution?
   All the patients were treated at the authors' institution, Queen Mary Hospital, Pokfulam, Hong Kong SAR, China. Please see the first section of the Materials & Methods. "Essentially, this is a retrospective study on analyzing common clinicopathologic parameters from the same cohort of 600 primary HCC patients who underwent hepatectomy as the primary treatment option in our medical center in the period of 1990 to 2007."

3. Page 3, line 4: The reference is wrong. The referenced paper discusses patients after ablative therapy!
   We have added 2 new references:

4. The authors missed a number of very relevant studies in the field:
   i. Llovet et al. 1999 describes portal hypertension and bilirubins as predictors for survival after surgical resection
   ii. Wayne et al, 2002 describethe Child-Pugh score and the tumor differentiation as predictive factors for survival after surgical resection
   iii. Vauthey et al. 2002, describe macroscopic and microscopic vascular invasion, number of tumors and fibrosis as potential factors for survival after surgical resection
How do these studies differ from the one presented here?

The authors thank the reviewer for pointing us the important studies in the field. We agree that the above-mentioned are of prognostic values. However, these data may be somewhat subjective and highly depends on the quality of the liver pathologists, and often times, are not available in smaller clinics. Nevertheless, we have referenced and discussed these papers in the revised manuscript. Our results are consistent with these papers in terms of hazard ratio and significance level, although, there are variables we did not measure and could not evaluate (e.g. portal hypertension and bilirubin). The aim of this study is to develop an objective and flexible framework, and use the derived hazard score to predict HCC prognosis. The framework must be flexible to adapt new predictors (e.g. molecular biomarkers) for future research. We were not aiming to compare the hazard score to existing HCC staging systems. In the discussion section, we added, "The prediction power of the predicted hazard (h) was demonstrated in this report, which is literally the linear combination of many predictors. A number of studies were conducted quantifying the predictive power of clinicopathological parameters, and our results are consistent with previous reports on vascular invasion [14, 15], AFP level [11, 15] and tumor size [15] in term of hazard ratios. However, this study did not measure some variable, e.g., portal hypertension and bilirubin [16], therefore, we could not directly assess their predictive value. 97.3% of our patients were classified as Child-Pugh grade A, and we found Child-Pugh grade not to be a significant predictor (possibly due to lack of statistical power). Nevertheless, our primary goal is to develop the objective and flexible framework, which can easily accommodate additional biomarkers when becoming available."

5. There has been outstanding study by Hoshida et al in 2008 in the NEJM, which should be discussed in this context.

Actually, we do have some reservation on Hoshida paper, which brought a “misleading” message that prognostic gene signature could only be found in the adjacent non-tumor tissues, but not in the tumor per se. We believe that both the tumor and adjacent tissues should contribute to the prognostic outcome of HCC patients. Anyhow, we have discussed that incorporating gene expression to our model as an important future. In the revision, we further described the work of Hoshida et al and how to extent our model to molecular level biomarkers, "Recent studies [19, 20] revealed gene expression in tumor and adjacent normal liver tissues (suggesting a so called "field-
effect") were predictive for HCC prognosis. It has been suggested that mechanistically these signatures capture tumor status, damage to liver tissue and the state of inflammation which relates to the likelihood of subsequent tumors arising [18, 20]. However, all such studies failed to incorporate clinicopathological and expressional predictors together. In such cases, the identified expressional predictors first capture the same information as clinicopathological parameters (e.g. cancer stage), and the performance for the expressional biomarkers would not necessarily outperform clinicopathological predictors (unpublished results). Herein, we argue that in order to capture information from gene expression that was not redundant to clinicopathologic data, the clinicopathologic parameters would be included during the search for expression signatures, and our hazard score model will serve for this purpose."

6. How did the authors define moderate smoking/drinking?

   moderate drinking: <= 2 drinks/per day
   heavy drinking: > 2 drinks/per day
   moderate smoking: <= 1 pack/day
   heavy smoking: > 1 pack/day

7. Patients with more than 6 nodules are not considered to be eligible for resection. These papers should be treated with TACE and should therefore not be included in this study.

How these patients being treated rely on the salient procedures and aggressiveness of the surgeons as well as how the multiple nodules are distributed – cluster in one lobe or spread out to multiple lobes. In our medical center, we do conduct HCC patients with >6 nodules if they are not near any major vessels and grouped into 1 or 2 clusters as well as good liver function. Of course, this kind of HCC patients receiving surgery is very rare.
The paper by Hao et al. describes prognostic factors in HCC after curative surgery. Even though the authors state that their system is superior to the classification systems published so far (CLIP, Barcelona, GRETCH, Okuda, etc.) the factors found in their study are in most cases already well known and evaluated. Therefore the authors of this study should definitely demonstrate the differences of their findings in comparison to the known classification and predicting models published so far. A table in which the factors are compared should be incorporated to better understand the superiority of their prediction system. Furthermore the evaluation of the system should be done in a prospective study. Otherwise the study is well written and the methods used are standard.

We appreciate the reviewer's comments. We stated that our hazard score model was superior to other classification systems for its flexible and objective nature, rather than its better performance in prediction accuracy. 97.3% of our patients were classified as Child-Pugh grade A, and we found Child-Pugh grade not to be a significant predictor (possibly due to lack of statistical power). Other classifications were not used in our hospital practice, and we could not directly compare them. Instead, we theoretically discussed our model along with other classification rules, "In addition, the existing prediction rules appear to be ad hoc, involve multiple arbitrary cutoffs and lack statistical rigor. Lastly, the performance of such rules has not been formally assessed by such analyses, e.g., ROC." Further, one of our primary goals of our study is to lay out the framework which can readily adapt additional predictors (e.g. gene expression biomarkers). In the discussion section, we wrote, "There are well known scoring systems to classify HCC, including Child-Pugh [14], the Oku [9] and Barcelona Clinic Liver Cancer Group [16, 17]. As the drawback, these scores have been devised by a series of ad hoc rules. In the other hand, predicted hazard is objective and can readily incorporate a new patient and biomarkers information. This can be simply performed by updating the linear model and the set of coefficients and will allow us to continuously update the model when new data becomes available. This is particularly important as new biomarkers based on molecular studies of HCC could be incorporated into the model. HCC has a heterogeneous etiology and many factors (e.g. patients' ethnicity and genetic background) may affect prognosis. Therefore, this model may not be directly applicable to a different HCC cohort. This approach should serve as
a general framework, where the Cox linear classifier can be trained on a particular cohort and applied to future patients." Moreover, we added comments on how to accommodate molecular level predictors into our model, "Recent studies [19, 20] revealed gene expression in tumor and adjacent normal liver tissues (suggesting a so called "field-effect") were predictive for HCC prognosis. It has been suggested that mechanistically these signatures capture tumor status, damage to liver tissue and the state of inflammation which relates to the likelihood of subsequent tumors arising [18, 20]. However, all such studies failed to incorporate clinicopathological and expressional predictors together. In such cases, the identified expressional predictors first capture the same information as clinicopathological parameters (e.g. cancer stage), and the performance for the expressional biomarkers would not necessarily outperform clinicopathological predictors (unpublished results). Herein, we argue that in order to capture information from gene expression that was not redundant to clinicopathologic data, the clinicopathologic parameters would be included during the search for expression signatures, and our hazard score model will serve for this purpose."
Although the authors studied an interesting question, I have several concerns. First, a clear distinction between methods and results is not present. I do not know whether this kind of structure of the manuscript is acceptable for BMC Cancer. At least for me, the structure complicates the reading of the manuscript. In the following, I present a list of my concerns as they came up by reading the manuscript:

1. Important details of the methods are missing. The cohort of the training set was recruited from 1993 through 2007 from the same hospital as the cohort of the training set recruited from 1990-2004. The recruitment periods overlap. It is unclear to me how the authors decided which patient would belong to which subcohort. Did refusals occur? If yes, were refusers different compared to participants? What about an IRB? An ethical review?

The method section on Study subjects has been rewritten to enhance the clarity. “In this retrospective study, we analyzed common clinicopathologic data from 600 HCC patients at Queen Mary Hospital, Pokfulam, Hong Kong in the period of 1990 to 2007. These patients were diagnosed with primary HCC and received hepatic surgery as the primary treatment option. Patients with other malignancies and non-resectable HCC were excluded. Preoperative investigation of the patients included blood biochemistry, alpha-fetoprotein assay, chest x-ray, percutaneous ultrasonography, computed tomography (CT), and hepatic angiography in selected patients. Liver function was assessed by the Child's-Pugh grading. All patients were treated and received the same postoperative care by the same team of surgeons, and had postoperative follow-up every month for the first year, and every 3-6 months thereafter. The minimal duration of follow-up time of the surviving patients was 12 months. Those cases lacking sufficient clinical and follow-up data were not included. Disease-free survival time was calculated from the date of hepatectomy to the date when recurrence was diagnosed.”
First, we randomly selected in the training set, 300 patients with available frozen tissue samples (both tumor and adjacent non-tumor) for biomarkers exploration studies by genotyping and mRNA expression profiling. 28 cases that were found missing clinical data or poor sample quality were thus excluded. Table 1 summarizes the demographic and clinicopathologic features of the remaining 272 patients in the training set. The other 300 cases were included in the validation set. There were no significant differences (p-value>0.05, two-sided tests) in the demographic and clinicopathologic features of HCC patients between the training and validation dataset. The study protocol was approved by the Institutional Review Board of Queen Mary Hospital, and informed consent was obtained from patients regarding to the use of the liver specimens for research.

2. The authors stated “there is no significant difference in the demographic and clinicopathologic features of HCC patients between the training and validation dataset”. What do these authors mean with significant? Clinically relevant or statistically significant? If statistical significance was meant, which alpha did they use (0.05?), were the tests one- or two-sided?

We examined the demographic and clinicopathologic features of the training and validation data, and found there were no statistically significant differences (two-sided test at 0.05 alpha level). In the manuscript, we rewrote, "There were no significant differences (p-value>0.05, two-sided tests) in the demographic and clinicopathologic features of HCC patients between the training and validation dataset."

3. In the introduction section, the authors state that overall survival and disease-free survival are the most important endpoints. As an explicit study question is missing, I can only guess that the
authors were interested in developing a prediction model of the endpoints overall survival and disease-free survival. In their method section, they state that “tumor recurrence could not be observed at the time of surgery” and therefore “it was not included into the prediction model”. It is no surprise that recurrence of disease is not put into the prediction model as independent variable because recurrence is an outcome (disease-free survival)!

The authors appreciate the reviewer's comment. In the introduction section, we added, "Our primary goal is to develop and validate models that use clinicopathologic parameters and common biomarkers observed at the time of surgery to predict the HCC prognosis. Further, such model must be flexible in accommodating addition factors (e.g. gene expression biomarkers) when becoming available." As the results, we did not include recurrence into the model to predict survival, since recurrence was not observed at the time of surgery.

4. The selection of variables that qualify for the multiple regression model (proportional hazards regression) is only vaguely described. From which list of potential predictors did the authors start? How many null hypotheses tests (univariate analyses) to study the associations between single predictors and outcomes were undertaken? If alpha was set to 0.05, we can expect that one out of 20 tested variables will show up to be significantly associated with the outcome. What was their alpha? Any corrections for multiple testing? Did they test one- or two-sided?

The reviewer's comment is well taken. In the Methods section, we added, "After evaluating all variables in Table 2, we selected those of significant association with survival and DFS into the multiple regression model (proportional hazards regression). This Cox model employed AFP, ALBU, veninv, tumor size, new AJCC and NOTN to predict HCC prognosis (survival and DFS)."

In this variable selection, we set alpha as 0.05 and used two-sided tests (Table 2). We evaluated around 20 potential predictors, and did not perform multiple-testing adjustment for the following reasons. (1) Most of the selective predictors showed highly significant association (e.g.}
pvalue≤0.002) with HCC prognosis during variable selection. These variables would remain significant after Bonferroni correction. (2) Albumin level was significantly associated with both survival and DFS, therefore, we also included it. (3) The primary goal of our model was for prediction. Inclusion of marginally significant predictors may result in certain degree of over-fitting. But such potential over-fitting will be quantified in the downstream cross-validation (LOO) and independent validation, where we found the predictive model performed very well.

5. Typo: page 5: “the patients were into three equal sized groups…”

We rewrote the sentence as "To further demonstrate robustness of the data, the patients were divided into three equal sized groups according to h (Figures 1C and 1D)."

6. Their tables 1-4 look strange with respect to the effect estimates. As the authors used Cox proportional hazards regression, I would expect to see hazards rate ratios and their 95% confidence intervals. Instead, the authors name it coefficients. As the hazard rate ratio is calculated by exponentiation of the regression coefficient, I conclude that the presented effect estimates are NOT hazard rate ratios.

The reviewer's comment is well taken. Tables 2 and 4 present the variable selection using Cox regression model. The primary goal was to identify the potential predictors for HCC prognosis, rather than quantify the hazard ratios. In the Tables, we showed the coefficients in the Cox model, which was the logarithm of the hazard rate ratio. In the Table legend, we added, "The regression coefficient in the Cox proportional hazards model. The hazard rate ratio can be calculated as the exponentiation of the regression coefficient."
7. The title of table 1 states that 272 subjects are included in the training set. However, the title of table 2A states that 277 subjects are included. Why?

It was a typo and has been corrected. We appreciate this comment. Throughout the manuscript (Abstract, Results, Materials & Methods and Table 1), we stated N=272 patients were used in the training dataset. We examined the data again, and confirmed this number. The number “277” in Table 2A was a typo, and we have corrected it in the revised manuscript.

8. Some distributions are highly skewed, some of them by expectation. For example, the survival time is highly skewed. The mean of the survival time is about the same as the standard deviation. To describe survival times, median is always preferred as long as not every cohort member has reached the endpoint of interest (see also disease-free survival). Liver function parameters, tumor size, and other factors are also highly skewed so that the mean values do not tell the reader much. The reviewers comment is well taken. In the revised Tables 1 and 3, we present median values for all continuous variables.

9. The authors never defined what they mean by moderate or heavy alcohol drinking and smoking (see Table 1).

Definitions have been provided as stated in the revised Table 1.

- moderate drinking: <= 2 drinks/per day
- heavy drinking: > 2 drinks/per day
- moderate smoking: <= 1 pack/day
- heavy smoking: > 1 pack/day

10. It is striking that there were no missing data for all covariates presented in table 1. As the percentages do not allow to check whether the sums of subjects add up to the expected total (either 272 or 277? to be clarified), I cannot check whether missing data were simply excluded and not presented or were not present.
The reviewer's comment is well taken. We have rewritten the Study subjects In the Materials and Methods section to enhance the clarity.

11. Overall, 67.8% of all observations were right-censored. Given the poor prognosis of HCC, the amount of right-censoring appears to be high. Reasons for right-censoring are missing. For example, how many patients were lost to follow-up? How many patients died due to competing risks?

Patients lost to follow-up and with insufficient clinical data were excluded from this study. When we analyze the data, many patients were alive, therefore, become right-censored. At the next scheduled follow-up, we will update their prognosis and the proportion of right-censoring may decrease.

12. Table 2a has one wrong label for the fourth column.
I have corrected that label in Table 2A.

13. In table 2b, the authors present family history. How did they define a positive family history? First degree only?
Positive family history does not limit to first degree only.

14. The figures 1-2 are hard to interpret as all graphs have the same symbols and line sizes. The legends do not help to identify which graph corresponds to which group. The titles of the y-axes are not exact. It is just stated SURVIVAL. However, it is the “estimated cumulative survival probability”.
The reviewer's comment is well taken. In the revised Figure, we used different color to distinguish various groups, which would be very clear to readers. We also changed the Y-label as "estimated cumulative survival probability".

15. Figure S1A shows two Kaplan-Meier plots. Why does the lower graph end at about 105 months? Was month 105 the month with the last observation censored or is it an artefact due to the software package used?

Among the initial selected patients in the training set, there was one patient survived for 170 months and still alive. However, this patient did not have venous infiltration status and was thus removed from the modeling. In other words, this patient was not part of the N=272 training data.

In Fig S1A, we were aimed to compare the distribution of survival and DFS between training and testing set rather than other clinicopathological parameters. In this case, we looked all the 272 patients enrolled for the training (Fig S1A).