Peer Review File

Article Information: https://dx.doi.org/10.21037/jgo-22-410

Reviewer A

Comment 1: Is there any duration different between training and validation cohort?
Reply 1: Thank you for your comments. The duration between the training and validation cohort is not different. We randomly divided all samples into a training cohort and validation cohort at a ratio of 7:3. The training cohort was used to build the model and the validation cohort was used to verify the accuracy of the model. The baseline data were similar between the training and validation cohort.
Changes in the text: N/A

Comment 2: Is there any duration different between recurrent and non recurrent group?
Reply 2: Thank you very much for your comments. The duration between recurrence group and nonrecurrence group is not different. Each patient completed a follow-up of 2 years to assess recurrence of adenomas by colonoscopy.
Changes in the text: We have modified the relevant contents in the Materials and Methods section--- page 6, line 126-127.

Comment 3: How about the CRC family history information for all study patient?
Reply 3: Thank you very much for your comments. We think your comment is very meaningful. We have improved the study by collecting the CRC family history information. It was found that the CRC family history was not an independent risk factor for adenoma recurrence in this study. We will continue to expand the sample size for further studies and improve the model. We've added text into the Discussion on this point and redrawn Table 1-3.
Changes in the text: We have collected the CRC family history as a parameter--- page 7, line 141-142/page 9, line 196. We have supplemented the relevant contents into the Discussion section--- page 12-13, line 274-280. We have redrawn Table 1-3, with the changes marked in red.

Reviewer B

Comment 1: In the result section, line 194, authors need emphasize the large adenoma(>10mm), multiple adenoma and histology finding..... were first colonoscopy result to avoid misunderstanding.
Reply 1: Thank you for the comments. We fully agree with your point of view. We've emphasized “first colonoscopy result” in the Results section.
Changes in text: We have supplemented the relevant contents into the Results section--- page 9, line 192-193.

Comment 2: The figure 5 was too blurred to identify, please re-drawing.
Reply 2: Thank you for your corrections. The figure in Word documents might be compressed so that they are not clear. We have redrawn and re-uploaded the Figure 5

as required by the journal.
Changes in text: We have re-uploaded Figure 5 as a separate file.

Comment 3: In several studies had demonstrated the high NLR was related to metabolic syndrome. Hyperglycemia or diabetes were also the risk factors of CRA. Do author had collected the biochemical data of participants to survey the possible confounding factors?
Reply 3: Thank you very much for your comments. We think this issue is worth exploring. High NLR was indeed associated with metabolic syndrome, which was also a risk factor for recurrence of adenoma, and our results were consistent with the above conclusions. However, NLR was not only related to metabolism, but also reflected the state of inflammation and immune function of the body. So it was not a conflict to include it in the study with diabetes. In this study, a collinearity test was used before data analysis to ensure that there was no highly correlated relationship between independent variables, so as to avoid affecting the prediction accuracy of the regression model. The collinearity of the independent variables is determined by using tolerance and variance inflation factor (VIF). The tolerance is defined as the reciprocal of the VIF. A tolerance value below 0.1 indicates that there is a serious collinearity problem. The following general rules are applied to interpret the VIF values: VIF<3, no collinearity; 3<VIF<10, moderate collinearity; VIF>10, high collinearity. In this study, the tolerance value among all independent variables was greater than 0.1 and the VIF value among all independent variables was lower than 3, so no collinearity was detected (as shown in the following table). In the absence of significant collinearity of independent variables, Logistic analysis showed that NLR was an independent risk factor for adenoma recurrence. As blood glucose was a dynamic and unstable indicator related to hypoglycemic drugs, diet and exercise, it was not included in this study.

Variable	Tolerance	VIF
Smoking	0.914	1.094
Diabetes	0.960	1.041
Adenoma location	0.671	1.491
No. of adenomas	0.667	1.499
Adenoma size	0.623	1.605
Differentiation	0.723	1.383
Villous component	0.770	1.299
NLR	0.553	1.808
PLR	0.810	1.234
LMR	0.606	1.649
PNI	0.760	1.316
FLR	0.626	1.598

Changes in text: N/A

Comment 4: In the method section authors were defined "301 patients were recruited

in the training cohort and 120 patients in the validation cohort". But the whole study design could not show why authors need "training cohort" and "validation cohort". The results and discussions sections were focus in " recurrence group" and "non-recurrence group" Authors need explain why they want defined "training cohort" and "validation cohort"? Or modified table 1 item.

Reply 4: Thank you very much for your comments. In order to build a new prediction model, it usually needs a training cohort to build the model and a validation cohort to verify the prediction ability of the model. We randomly divided all samples into a training cohort (n = 301) and validation cohort (n = 120) at a ratio of 7:3. Next, univariate analysis and multivariate analysis of factors associated with CRA recurrence were performed in the training cohort, and a nomogram for predicting recurrence of CRA was established. Then the accuracy and applicability were verified by the validation cohort. Figure 3B, 4B, 5B were all drawn by the validation cohort.

Changes in text: N/A

Comment 5: The case numbers might increase because the study participants and might not represent the whole recurrence CRA patients.

Reply 5: Thank you for your comments. In order to make a clear judgment, we need to develop strict inclusion and exclusion criteria to ensure the rigor of the study. With a follow-up of 2 years to assess recurrence of adenomas by colonoscopy, 421 patients were included in the analysis finally. When performing logistic analysis like this study, the sample size was often determined based on events per variable (EPV). Statistical studies showed that as per rule of thumb derived from the simulation study for logistic regression, at least 10 EPV were required for the robustness of the results. Six independent risk factors associated with adenoma recurrence were finally included in the study. A study has shown showed that 20 to 50% of patients with CRA developed recurrence within 2-5 years. If the recurrence rate was 20-50%, assuming EPV=10, the number of patients with adenoma recurrence was 10×6=60, and the total sample size (patients with adenoma recurrence and those without recurrence) was 60÷ 50% - 60 ÷20%= 120-300. In this study, 301 patients were enrolled in the training cohort, so this nomogram model can be established with some reliability.

Changes in text: N/A

Comment 6: Discussion should be based on comparing the current findings and prior reports. Limitations and future research directions should be improved.

Reply 6: Thank you very much for your comments. We have added text based on comparing the current findings and prior reports into the Discussion. In addition, Limitations and future research directions have been improved.

Changes in text: We have supplemented the relevant contents into the Discussion section--- page 12, line 269-272/page 13-14, line 287-298/page 14, line 305-308/page 14, line 310-316. We have improved the limitations and future research directions--- page 15, line 324-336.