Peer Review File

Article information: https://dx.doi.org/10.21037/jgo-24-464

Reviewer A

The manuscript is well-structured and provides significant insights into the use of regorafenib in elderly mCRC patients. The findings support the use of higher doses in patients under 70 and careful dose management in older patients. However, the retrospective design and regional focus limit its generalizability and causal conclusions. Further prospective and randomized studies are needed to validate these findings and recommendations. A more thorough discussion of the limitations would provide a more balanced view of the study's findings.

With some additional methodological details and a more comprehensive discussion of limitations, the paper could provide an even stronger foundation for clinical decision-making in this patient population.

- Please consider revising the manuscript to use "association" instead of "correlation" when discussing the relationship between dosing regimens and patient outcomes. This change will align the terminology with standard practice in survival analysis and enhance clarity.
Reply: Thank you for your suggestion. We have replaced "correlation" in the text with "association".

- The study fails to compare baseline patient characteristics across dose groups before doing univariate and multivariate analyses. This can introduce bias and lead to incorrect conclusions. It's important to check and report characteristics and adjust for them in the analysis. Also, methods like PSM and IPTW need balanced groups to be effective, which requires a proper baseline assessment.
Reply:

Thank you for your suggestions. We further refined the differential analysis based on the basic characteristics within various dosage groups (Supplementary Tables 2, 3) and defined factors with $P<0.05$ as significant. In the analysis of the initial daily dose, we found no significant differences between the groups. However, in the analysis of the final daily dose, we found a significant sex difference ($P=0.049$).

We highly appreciate your suggestion to include factors that exhibit differences among groups in the multivariable analysis, which contributes to obtaining more accurate results. Consequently, we have developed a second multivariable model (Model 2) incorporating factors with $P<0.20$ from univariable analysis and those with $P<0.05$ from differential analysis,

and the results obtained align closely with our original model (Model 1). Due to the relatively small sample size, we have utilized our original model for the subsequent analysis.

Given the significant difference in sex among different final daily dose groups, it has been included in the PSM and IPTW analysis. In the PSM analysis, OS was insignificantly prolonged in the high-dose group compared to the low-dose group. After IPTW, the difference in OS between the high and low-dose groups becomes significant.

Changes in the text:

We have modified our text as advised. (Page 6, Lines 145, 151-153; Page 8, Lines 193-196; Page 9, Lines 229-240)

- The threshold for including variables in the multivariate model (P-value ≤ 0.20) should be justified. Including too many variables can lead to overfitting, while too few can omit important confounders.

Reply: Thank you for your suggestions. We acknowledge your concerns. For univariate analysis, a P-value <0.1 is commonly used as a screening criterion. In our preliminary research, we attempted to analyze only factors with P<0.1 from univariable analysis but did not yield significant results. Upon reviewing baseline characteristics, we have observed a potential association between RAS gene mutation status and the mechanism of small molecule TKIs. Besides, factors such as liver and lung metastasis, and tumor location directly impact prognosis. These factors may be relevant to the clinical outcomes of elderly patients treated with regorafenib monotherapy, most of whose P-values are distributed between 0.1 and 0.2. Therefore, we have slightly relaxed the P-value threshold to 0.20, which allows for the discovery of more potential prognostic factors and simultaneously ensures the statistical power of the multivariable analysis (The number of positive outcome events has reached at least ten times the number of variables in the multivariable model.).

Changes in the text:

We have modified our text as advised. (Page 6, Lines 148-150; Page 8, Lines 208-210)

- The use of PSM and IPTW, along with the presentation of adjusted HRs with 95% CIs, suggests that the authors followed a standard procedure for PSM analysis, which typically includes checking the balance of covariates between the matched groups. It is not clear whether the authors conducted balance checks post-PSM to ensure that the covariates were adequately balanced between the matched groups.

i) Provide detailed balance checks post-PSM to demonstrate that the matched groups are balanced and to validate the PSM methodology.

Reply: We validated the balance before and after propensity score matching through qualitative and quantitative methods. Both histograms and jitter plots indicated that the post-matching propensity score distributions of the high-dose and low-dose groups were closer. Subsequently, we utilized the Standard Mean Difference (SMD) to further enhance the balance check,

revealing that after matching, the SMD for all variables did not exceed 0.1. It suggested an acceptable quality of matching for these variables and indicated a good balance post-matching.

Changes in the text:

We have modified our text as advised. (Page 7, Lines 162-164; Page 9, Lines 231-236)

ii) Conduct sensitivity analyses to assess the robustness of the findings.

Reply:

Thank you for your suggestion. We have added a section "Sensitivity Analysis" in the article to further validate the reliability of the conclusion regarding the association between final daily dose and overall survival. Initially, PSW and IPTW served as sensitivity analyses of the above results. We have incorporated the contents into the "Sensitivity Analysis" section. Given the significant difference in sex among different final daily dose groups, it has been also included in the analysis, revealing that in the PSM analysis, OS was insignificantly prolonged in the high-dose group compared to the low-dose group [HR (95% CI): 0.58 (0.30–1.12), P = 0.09]. After IPTW, the difference in OS between the high and low-dose groups becomes significant [HR (95% CI): 0.44 (0.23–0.83), P=0.01].

Additionally, an increased lines of therapy may indicate that patients are less likely to benefit from and tolerate the treatment. The majority of patients were treated with regorafenib monotherapy as third-line therapy. We proceeded with further analysis within this specified group. The results of the multivariate Cox analysis did not show any significant changes.

Finally, considering the close association between sex and cancer incidence as well as mortality, we excluded the confounding effects of sex through subgroup analysis. The conclusions drawn for male patients align with the original findings. However, no association was observed between the final daily dose and OS in female patients. Given the low proportion of female patients in our study, larger clinical trials are necessary to clarify the impact of sex.

Changes in the text:

We have modified our text as advised. (Page 7, Lines 156-172; Page 9, Lines 228-240; Page 10, Lines 241-253)

- The study is based on a real-world multicenter study in China. As a result, the findings may not be generalizable to other populations or healthcare settings. The authors should clearly state the limitations regarding generalizability in the discussion section of the paper. Future studies should aim to validate these findings in diverse populations to ensure broader applicability.

Reply: Thank you for your suggestions. We have further refined the "limitation" and emphasized the constraints of our study on generalizability.

Changes in the text:

We have modified our text as advised. (Page 13, Lines 351-356)

Reviewer B

You define progression-free survival (PFS) in this study; however, considering it utilizes real-world data without treatment efficacy data, I think it should be defined as "time to treatment failure" instead. This also applies to PFS2 and PFS3. If treatment efficacy across each treatment line cannot be confirmed, all endpoints should be described as "time to treatment failure" rather than "progression-free survival." If treatment efficacy is confirmed for all patients across all treatment lines, it should be explicitly stated in the manuscript.

Reply: Thank you for your suggestions. Time to treatment failure (TTF) is a comprehensive and readily accessible metric frequently used in real-world studies. It offers an intuitive approach for analyzing drug efficacy, especially when efficacy data is incomplete. In our study design phase, we intended to use progression-free survival (PFS) to represent long-term efficacy. More attention was paid to the time point of progressive disease (PD) through imaging and physicians' records of each patient. Most patients' times of medication discontinuation were also documented. Upon reevaluating the dataset, we found that the PFS data were comprehensive, whereas the TTF data were relatively insufficient (due to lacking records of medication discontinuation caused by adverse events or patient subjective preferences). This was one of the limitations of our study, which we have discussed in the "Discussion" section. To prevent the bias in statistical analysis due to missing TTF data, we utilized the relatively more adequate PFS data for subsequent analysis.

Changes in the text:

We have modified our text as advised. (Page 13, Lines 346-351)