Peer Review File

Article information: http://dx.doi.org/10.21037/tcr-20-3354

Reviewer A

Comment 1: This study compared outcomes of 2nd line combined versus monotherapy with novel agents in advanced UC patients. Since the 1st line treatments consisted of heterogeneous modalities including systemic chemo and ICIs, the reviewer wonders if this meta-analysis makes sense; in addition, the 2nd line treatments included ICIs, targeted agents, and their combinations. The treatment sequences would have impact on outcomes of subsequent treatments. However, this study did not concern this issue. In table 2, the superiority of survival outcomes were different between 12M and 24M between mono- vs. combined arms; this discrepancy would be attributed to the heterogeneity of the 1st line treatments, sequences, and 2nd line treatments. It would make sense if the authors separately analyze outcomes for 1st line chemo and 1st line ICIs to purely compare outcomes of 2nd-line treatments.

Reply 1: We really appreciate the thorough viewpoint of the reviewer on the heterogeneity of the 1st line treatments, sequences, and 2nd line treatments. Surely, the outcomes could be affected by the difference in application of chemo or ICIs in the 1st line treatment for advanced UC patients. However, the overwhelming majority of studies we included only enrolled the patients who progressed after platinum-based chemotherapy. Only a few articles contained the participants failed from 1st line ICIs, and additionally, the proportion of patients after 1st line ICIs was much smaller

compared with 1st line chemo patients in these articles. Furthermore, only the study by *Petrylak et al.* designed the sub-analysis researching the association of outcomes and CNA in 17 patients after 1st line ICIs, so that we could not separately compare the outcomes between CNA and MNA groups based on 1st line chemo and 1st line ICIs. However, we have added these information in the result of our study. Overall, the purpose of our meta-analysis was to explore the clinical advantages of CNA as a salvage treatment for advanced UC patients who failed from 1st line chemo.

Changes in the text: we had added the basic information about 1st line ICIs and chemo in the result section of our study, and explained these limitations in the discussion section (see Page 8 line 3-6, Page 14 line21-Page 15 line 1 and Page 15 line 3-4 with a different color of text).

Comment 2: There are many typographic errors. Please correct them.

Reply 2: Thanks for this proposal. We have modified our text as advised.

Changes in the text: We have corrected the typographic errors throughout our manuscript using a different color of text.

Comment 3: The term “TCC” should be replaced with “UC” throughout the manuscript.

Reply 3: Thanks for this proposal. We have modified our text as advised.

Changes in the text: We have replaced “TCC” with “UC” throughout the manuscript using a different color of text.

Comment 4: Please make a brief explanation for Bellumunt risk factor.

Reply 4: Thanks for this proposal. We have added an explanation for Bellumunt risk factor in our manuscript using a different color of text.

Changes in the text: Please see Page 7 line 10 with a different color of text.

Comment 5: p7, lines 16&22: Please make brief explanations why the authors excluded those studies from sensitivity analysis.

Reply 5: Thanks for this question. Take the disease control rate (DCR) of CNA. Initially, we got the pooled DCR with high heterogeneity (I^2=82.6%). Then, sensitivity analysis was conducted for exploring the reason of heterogeneity. Next, we excluded two studies which contributed to a high heterogeneity, to obtain the modified pooled DCR with lower heterogeneity (I^2=44.9%), and moreover, to test the stability of results and enhance the credibility of outcomes.

Changes in the text: We have highlighted the condition of omitting studies for a sensitivity analysis in methods section (see Page 7 line 3-5 with a different color of text).

Comment 6: PD-L1 profiling would be based on different assay conditions among studies with different antibodies. The reviewer does not consider a comparison according to PD-L1 status makes sense.

Reply 6: Thanks for this question. The positive or negative status of PD-L1 was

obtained from the original articles, and we could not make the assay conditions and antibodies entirely uniformed. However, the acceptable heterogeneities with less than 50% were found among studies according to the final results of these subgroup analyses. Though 28-8 and 22C3 PD-L1 assays were conducted in the 4 included studies, some current articles had demonstrated the analytical concordance between the two assays. Moreover, these subgroup analyses were principally to compare the outcomes between PD-L1 (+) and PD-L1 (-) based on reported population and to provide more meaningful clinical outcomes as much as possible. These results could make our study more comprehensive.

Changes in the text: None.

Reviewer B

Comment 1: The title is not accurate. Please indicate effectiveness and safety in the title.

Reply 1: Thanks for this proposal. We have modified our title as advised.

Changes in the text: We have added “effectiveness and safety” in the title of our manuscript (see Page 1 line 1 with a different color of text).

Comment 2: Abstract. Methods part, the authors should briefly describe the inclusion of studies according to PICOS criteria, outcomes extracted from included studies and the assessment of risk of bias of included studies.

Reply 2: Thanks for this proposal. We have added relevant description of included patients, interventions, comparisons, and the assessment of risk of bias.

Changes in the text: See Page 2 line 7-8 and line 11-12 with a different color of text.

Comment 3: Introduction. Line 9, page 4, please provide more specific examples on the side effects of CNA. In this part, the authors reviewed several studies, however, I did not why the current meta-analysis is needed, because, in terms of efficacy, it seems no controversies. The authors may consider to explain more on this. I suggest the authors to have some comments on the limitations of previous studies, to support the necessities of the current research topic.

Reply 3: We really appreciate this valuable suggestion. And we have added specific examples of grade >3 adverse effects rates in CNA and comments on the limitations of published studies in Introduction to highlight the necessities of this meta-analysis.

Changes in the text: See Page 4 line 9-11 and line 12-14 with a different color of text.

Comment 4: Methodology. The literature search has language bias because only English-language databases were searched. The authors should strictly define the inclusion of related studies according to the PICOS criteria. It is strange that the authors did not specify the clinical research design of studies to be included. For the risk of bias assessment, Cochrane risk of bias tool is not feasible for studies that were not designed as RCTs. The authors need to consider the clinical heterogeneity in the treatment methods to be studies, in fact, they included a range of medications, the

clinical heterogeneity is significant.

Reply 4: Thanks a lot for the questions above.

1) Only English articles according with our inclusion criteria could be available in the databases of PubMed, Cochrane Library, Embase, and Web of Science. Though some abstracts in non-English language could be also found in these databases after screening, the full-texts were not useful to our study. Therefore, we decided not to include non-English articles. According to reviewer's suggestions, study type, patients, interventions and outcomes were further defined in the part of Selection criteria.

2) Surely, it is undoubted that Cochrane risk of bias tool is applied for the quality judgment of randomized controlled studies. And we only used it for 12 RCTs (Choueiri 2012, Fradet 2019, Gallagher 2010, Herbst 2019, Jones 2017, Parikh 2018, Petrylak 2016, Petrylak 2020, Powles 2018, Rosenberg 2018, Sharma 2019, and Wong 2012). For single-arm studies, we did not find credible tools to access their quality.

3) We really appreciate the thorough viewpoint of the reviewer on the clinical heterogeneity due to the usage of different medications. The aim of this meta-analysis is to compare the efficacy and safety of CNA and MNA, so we included a large number of relevant studies to expand the patients' number for the enhancement of result credibility. Furthermore, we have omitted several studies with large heterogeneity for sensitivity analysis to verify our results. What's more, subgroup analyses were also carried out to explore the differences between ICIs and TDs, PD-L1(+) and PD-L1(-), etc. In summary, we have done our best to decrease possible

heterogeneity derived from medication differences. More importantly, we have further emphasized this issue in the part of limitation.

Changes in the text: We have added study type in Selection criteria (see Page 5 line 16-17 with a different color of text). We have added relevant explanations about the bias risk judgment of RCTs and single-arm studies (see Page 6 line 15-16 with a different color of text).

Comment 5: Statistics. Please specify which modules of Stata were used for the current analyses. In general, Stata is more flexible than Revman, it remains unclear why both software were used. The authors need to explain why Begg's test was not used for testing publication bias. Please specify the sensitivity analysis based on "Certain studies which possibly contributed to a high heterogeneity".

Reply 5: Thanks a lot for the questions above.

1) Certainly, Stata has some advantages than Revman. But study quality assessment can not be conducted by State, so we used Cochrane risk of bias tool. Furthermore, Revman can not be applied for sigle-arm studies. Therefore, Stata was used as the main software in our study for meta-analysis, heterogeneity test, sensitive analysis, publication bias test and sub-group analysis.

2) We used Egger's test rather than Begg's test, because Egger's method had relatively stronger statistical power than Begg's method in terms of publication bias test, which was demonstrated in many previous studies (PMID: 16276033/31743750/29663281/25666576).

3) Sensitivity analysis was applied when heterogeneity among studies was significant ($P < 0.10$), and several studies which could led to a large heterogeneity were omitted. For example, we found a DCR with high heterogeneity (I^2=82.6%) in CNA group. Accordingly, sensitivity analysis was performed to explore why there was the heterogeneity. After then, two studies were omitted and we got a DCR with lower heterogeneity (I^2=44.9%), which contributed to testing the stability of results.

Changes in the text: None.