PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	What's the optimum systemic treatment for advanced/metastatic renal cell carcinoma of favorable-, intermediate- and poor-risk, respectively? A systematic review and network meta-analysis
AUTHORS	Cao, Guanghui; Wu, Xiaoqiang; Wang, Zhiwei; Tian, Xiangyong; Zhang, Chan; Wu, Xuan; Zhang, Haotian; Jing, Gaopeng; Yan, Tianzhong

VERSION 1 – REVIEW

REVIEWER	Jesse Y Hsu University of Pennsylvania, USA
REVIEW RETURNED	17-Dec-2019

GENERAL COMMENTS	Note: line numbering is based on ruler applied in Editorial Manager-generated PDF. 1. The statistical analyses were well conducted and the manuscript was clearly written. I hesitated and had concerns about the generalizability of the findings given the highly selected studies. 2. Page 3, lines 35–37: Because there was no start date in literature-search strategy, it would be inter-esting to know when the earliest possible article was published. Please describe when the first study for drugs targeting angiogenesis (among those listed in the inclusion and exclusion criteria #2) was available. 3. Page 4, line 4: Given the inclusion and exclusion criteria #3, the selection of study population was quite restricted. Please comment on how findings could be generalized as comparative effectiveness in the Discussion section. 4. Page 4, line 56: Should the availability of person-level be a criterion as well? 5. Page 5, lines 13–16: Please confirm this referred to the exclusion of "cytokines". 6. Page 5, line 33: Please elaborate what "as far as possible" meant. 7. Page 5, line 37: Was the disagreement about the quality of the included RCTs? 8. Page 5, line 52: What was the rationale of analyzing reported estimates given that person-level data were extracted?

	9. Page 5, lines 54–56: I suggest to briefly describe the pragmatic approach. 10. Page 5, lines 56–58: (a) Did this refer to the odds ratio of each drug-related adverse event or the odds ratio of each study having drug-related adverse event? (b) Because the drug-related adverse events likely occurred repeatedly during follow-up, how were the recurrent events treated in this calculation? How were different follow-up times from differ-ent studies handled? (c) What events were included in the drug-related adverse events? Were these events the same across studies? 11. Page 5, lines 58–60: This sentence was vague. Please provide more specifics. 12. Page 6, lines 17–21: Given the knowledge of potential violation of the transitivity assumption, how did you resolve the violation? 13. Page 6, line 35: The approved and not approved systemic therapies had not been defined. 14. Page 6, lines 41–59: It would be interesting to know what rationales led to the different iterations between PFS and adverse events. 15. Page 7, line 41: If the results were based on fixed-effects models, the findings would not be able to be generalized to other studies. Please comment on this in the Discussion section. 16. Page 8, line 4: What was SUCRA? 17. Page 12, line 4: What was OS? 18. Page 12, line 51: Please comment on the generalizability of the findings given a very restricted se-lected studies (Figure 1). 19. Figures 2–4: (a) In A, could you make the widths more distinguishable? They all looked the same width to me. (b) In C, please include x- and y-axis labels.

REVIEWER	Mehmet Bilen Winship Cancer Institute of Emory University
REVIEW RETURNED	06-Jan-2020

GENERAL COMMENTS	In this manuscript, authors reported Bayesian network analysis for optimum treatment for metastatic RCC. Comments: - They did not include recent options, pembrolizumab+ axitinib or

	avelumab+ axitinib. Need to add those trial, and reanalyze since this is important for clinical usage of this data. Otherwise, results are not useful. - Recent studies used IMDC for risk stratification, older studies MSKCC. Although, those 2 are similar, there is difference. Need to include this to the manuscript, especially page 6 line 23, page 7 line 12. - The biology of RCC is heterogeneous, especially more sarcomatoid pt is either intermediate or poor risk, more so then good risk. This will effect the overall trial population. Please make sure to expend the discussion and include limitation. - Front line cabo trial is smaller, phase 2 trial. Making big conclusion might not be ideal. Please expend limitation. - Page 11, line 11 does not make sense. Pesudoprogression should not be different since RECIST criteria is the same regardless of risk group. But underline biology of good risk might be different, such as more angiogenic driven, etc. Please revise and use more uptodate reference that help this. (gene signature from pembro/axi and avelumab/axi trials). - They used PFS, not OS. Not clear why? They stated "In addition, PFS has been recognized as a surrogate endpoint of OS in medical oncology as well as in advanced/metastatic RCC in the TKI era" But we know it is not the case for immunotheraphy. Need further explanation and or revision. - There is recent paper: EBioMedicine. 2019 Sep;47:78-88. doi: 10.1016/j.ebiom.2019.08.006. Epub 2019 Aug 19. Role of immune checkpoint inhibitor-based therapies for metastatic renal cell carcinoma in the first-line setting: A Bayesian network analysis. what additional information is added? This paper is not cited although from same group? - They might revise conclusion. Not totally make sense intermediate risk patient benefit more from cabo. How they address particular trial and sample size issue? Need further explanation and possible revision. - Good risk patient will benefit from sunitinib might change when adding more uptodate trial, such as pembro/axi, avelumab/axi.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Comment 1- The statistical analyses were well conducted and the manuscript was clearly written. I hesitated and had concerns about the generalizability of the findings given the highly selected studies.

Response: We thank for your valuable comment. To revise, in Discussion, we added “Since the analysis was based on highly selected RCTs and the results were based on fixed-effects models, findings in this analysis may not be entirely generalized to real-world practice.” on page 15.

Comment 2- Page 3, lines 35–37: Because there was no start date in literature-search strategy, it would be interesting to know when the earliest possible article was published. Please describe when the first study for drugs targeting angiogenesis (among those listed in the inclusion and exclusion

criteria #2) was available.

Response: We are sorry for your confusion. On January 11, 2007, the New England journal published two phase 3 RCTs demonstrating the progression-free survival improvement of sorafenib and sunitinib for mRCC, respectively. They are the first RCTs to assess drugs targeting angiogenesis for mRCC. To be clear, in Introduction, we added “In 2007, results from two RCTs have been published reporting progression-free survival improvement of two newer targeted agents (sunitinib and sorafenib).” on page 5.

Comment 3- Page 4, line 4: Given the inclusion and exclusion criteria #3, the selection of study population was quite restricted. Please Comment on how findings could be generalized as comparative effectiveness in the Discussion section.
Response: We thank for your valuable Comment. In Discussion section, we added “Finally, findings in this meta-analysis were mainly based on patients with clear-cell advanced RCC, thus these results may not be generalizable to patients with non-clear-cell subtypes.” on page 16.

Comment 4- Page 4, line 56: Should the availability of person-level be a criterion as well?

Response: We thank for your valuable comment. We couldn’t get person-level data for most RCTs, so we didn’t set the availability of person-level data as an inclusion criterion. In addition, the use of reported estimators for meta-analysis is widely accepted.

Comment 5- Page 5, lines 13–16: Please confirm this referred to the exclusion of “cytokines”.

Response: Thanks for your kind remind. In Methods section, we replaced “Relevant interventions included, but were not restricted to: sorafenib, sunitinib, pazopanib, cabozantinib, nivolumab, ipilimumab, axitinib, tivozanib, everolimus, temsirolimus, bevacizumab, IFN-α or IL-2.” with “Relevant interventions included, but were not restricted to: sorafenib, sunitinib, pazopanib, cabozantinib, nivolumab, ipilimumab, axitinib, tivozanib, everolimus, temsirolimus, or bevacizumab plus IFN-α.” on page 6.

Comment 6- Page 5, line 33: Please elaborate what “as far as possible” meant.

Response: We are sorry for the confusion. To reduce misunderstanding, in Methods section, we replaced “Data were extracted from intention-to-treat analyses as far as possible.”with “We gave priority to extracting data from intention-to-treat analyses.” on page 7.

Comment 7- Page 5, line 37: Was the disagreement about the quality of the included RCTs?

Response: The methodological quality of included RCTs was assessed by two researchers (GH. Cao and XQ. Wu) independently. There are disagreements about the quality of the included RCTs, and they were resolved by consensus in consultation with a third reviewer (TZ. Yan).

Comment 8- Page 5, line 52: What was the rationale of analyzing reported estimates given that person-level data were extracted?

Response: We are sorry for the confusion. Performing meta-analysis using patient-level data can establish more robust and accurate conclusions in specific subgroups of patients. At the begging, we tried to extract person-level data for the analysis, but we can’t get them for most RCTs. Therefore, we performed the meta-analysis using reported estimates.

Comment 9-Page 5, lines 54–56: I suggest to briefly describe the pragmatic approach.

Response: We thank for your valuable suggestion. The methods were described in detail in the paper, and we calculated HRs from Kaplan-Meier curve and information on follow-up [25]. To be clear, in Methods section, we added “For studies not reporting HRs, we calculated them from Kaplan-Meier curve and information on follow-up with the pragmatic approach reported by Tierney et al [25].” on page 7.

Comment 10. Page 5, lines 56–58:
(a) Did this refer to the odds ratio of each drug-related adverse event or the odds ratio of each study having drug-related adverse event?
(b) Because the drug-related adverse events likely occurred repeatedly during follow-up, how were the recurrent events treated in this calculation? How were different follow-up times from different studies handled?
(c) What events were included in the drug-related adverse events? Were these events the same across studies?

Respond: (a) We are sorry for the confusion. To be clear, in Methods section, we replaced “For drug-related adverse events, we calculated odds ratios (ORs) of every trial for meta-analysis.” with “For drug-related adverse events, we calculated odds ratios (ORs) using the available raw data abstracted from the trials.” on page 7.

(b) In consistent with the included trials, we used the number of patients who had drug-related adverse events to calculate ORs, rather than the number of adverse events. We agree with the reviewer that different follow-up times may affect drug-related adverse events. We tried to perform sensitivity analyses adjusted for follow-up times. Only eight trials reported median follow-up, therefore it was impossible to perform sensitivity analyses adjusted for this factor. We stated “Moreover, the length of follow-up varied across studies, resulting in potential variations in survival benefits and adverse events of systemic treatments. Due to only eight trials reporting median follow-up, sensitivity analyses adjusted for this factor were impossible.” in Discussion section on page 16.

(c) Drug-related adverse events were evaluated by investigators in the included trials. The most common adverse events included diarrhea, hypertension, fatigue, and decreased appetite; and they were similar across studies. To be clear, we added “The most common adverse events included diarrhea, hypertension, fatigue, and decreased appetite.” in Results section on page 11.

Comment 11- Page 5, lines 58–60: This sentence was vague. Please provide more specifics.

Respond：We are sorry for the confusion. To be clear, we revised it with “For drug-related adverse events, we calculated odds ratios (ORs) using the available raw data abstracted from the trials.” in Methods section on page 7.

Comment 12. Page 6, lines 17–21: Given the knowledge of potential violation of the transitivity assumption, how did you resolve the violation?

Response: We are sorry for the inaccuracy of expression. It is believed that there is a treatment-by-risk group interaction. Taking no account of this interaction in the analysis, the transitivity assumption across all included trials may be violated. Therefore, we performed the network analyses separately by risk groups to assure transitivity assumption.
To be clear, we replaced “Taking no account of this possible interaction in the analysis, we suppose the transitivity assumption across all included trials would be violated. Consequently, we performed all

network analyses separately by risk groups (favorable-, intermediate-, and poor-risk groups) according to the MSKCC risk model." with "Taking no account of this possible interaction in the analysis, transitivity assumption across all included trials would be violated. Therefore, we performed all network analyses separately by risk groups (favorable-, intermediate-, and poor-risk groups) according to the MSKCC or IMDC risk model to assure transitivity assumption." in Methods section on page 8.

Comment 13- Page 6, line 35: The approved and not approved systemic therapies had not been defined.

Response: We are sorry for the confusion. In our updated manuscript, we replaced "Sensitivity analyses were performed restricted to trials that assessed approved systemic therapies"with "Sensitivity analyses were performed restricted to trials that assessed approved systemic therapies (sunitinib, sorafenib, pazopanib, cabozantinib, axitinib, everolimus, temsirolimus, bevacizumab plus IFN-α, and nivolumab plus ipilimumab)" in Methods section on page 8.

Comment 14- Page 6, lines 41–59: It would be interesting to know what rationales led to the different iterations between PFS and adverse events.

Response: We are sorry for the confusion. We performed the Bayesian network analysis using OpenBUGS version 3.2.2 for PFS, and Gemtc version 0.14.3 for adverse events. In the beginning, we tried to apply the same iterations for PFS and adverse events. However, compared with GeMTC, OpenBUGS requires too much time for data analysis. Therefore, we applied lesser iterations for PFS without loss of convergence and model fit.

Comment 15- Page 7, line 41: If the results were based on fixed-effects models, the findings would not be able to be generalized to other studies. Please Comment on this in the Discussion section.

Response: We thank for your valuable Comment. To revise, we added "Since the analysis was based on highly selected RCTs and the results were based on fixed-effects models, findings in this analysis may not be entirely generalized to real-world practice." in Discussion section on page 16.

Comment 16- Page 8, line 4: What was SUCRA?

Response: We are very sorry for the confusion. SUCRA is short for the surface under the cumulative ranking curve. To be clear, we added "The treatments were ranked in terms of PFS, OS, and high-grade AEs, respectively, using the surface under the cumulative ranking curve (SUCRA) and the distribution of the ranking probabilities [32]." in Methods section on page 8.

Comment 17- Page 12, line 4: What was OS?

Response: We are very sorry for the confusion. OS is short for overall survival. We have replaced "OS" with "overall survival (OS)" in Methods section on page 6.

Comment 18- Page 12, line 51: Please Comment on the generalizability of the findings given a very restricted selected studies (Figure 1).

Response: We thank for your valuable Comment. To revise, we added "Since the analysis was based on highly selected RCTs and the results were based on fixed-effects models, findings in this analysis may not be entirely generalized to real-world practice." in Discussion section on page 16.

Comment 19- Figures 2–4:
(a) In A, could you make the widths more distinguishable? They all looked the same width to me.
(b) In C, please include x- and y-axis labels.

Response: We are very sorry for the confusion. To make the figures clearly visible, we have re-done Figures 2–4 the following:

Figure 2- Analysis of progression-free survival for patients with favorable-risk disease

Figure 3-Analysis of progression-free survival for patients with intermediate-risk disease.

Figure 4-Analysis of progression-free survival for patients with poor-risk disease.

Reviewer: 2

Comment 1- They did not include recent options, pembrolizumab+ axitinib or avelumab+ axitinib. Need to add those trial, and reanalyze since this is important for clinical usage of this data. Otherwise, results are not useful.

Response: Thanks for pointing this out. To include recent treatments for advanced/metastatic RCC, we performed an updated literature search in February 2020 and included two more valuable RCTs [JAVELIN Renal 101 and KEYNOTE-426]. We reanalyzed the data and updated the results. In the section of Results, we revised them to "3.2.1. For advanced/metastatic RCC of favorable-risk, 13 trials enrolling 2 514 total patients reported adequate information on progression-free survival and contributed to network meta-analysis (Fig. 2A)[10-13 15 16 34-37 39-42]. Fig. 2B summarizes the results of the network meta-analysis for PFS. Compared with sunitinib, IFN-α and nivolumab plus ipilimumab were associated with significantly worse PFS (HR 2.69, 95% CI 1.54-4.67; and HR 2.18, 95% CI 1.47-3.25, respectively). Network meta-analysis showed that only avelumab plus axitinib was associated with a significantly higher improvement in PFS than sunitinib (HR 0.57, 95% CI 0.34-0.96). Based on the results of ranking, there was a 45% chance that avelumab plus axitinib provided the greatest PFS benefit for patients with favorable-risk disease (SUCRA = 92.3%)(Fig. 2C).

3.2.2. For advanced/metastatic RCC of intermediate-risk, 14 trials enrolling 5 473 total patients contributed to the analysis of PFS (Fig. 3A) [10-13 15 16 33-37 39 40 42]. Network meta-analysis demonstrated that cabozantinib, nivolumab plus ipilimumab, pembrolizumab plus axitinib, and avelumab plus axitinib were associated with significantly higher improvement in PFS than sunitinib (HR 0.63, 95% CI 0.44-0.97; HR 0.66, 95% CI 0.53-0.81; HR 0.58, 95% CI 0.44-0.80; HR 0.62, 95% CI 0.47-0.83, respectively). Everolimus, bevacizumab plus IFN-α, and temsirolimus plus bevacizumab were significantly less efficacious for PFS than sunitinib (HR 1.50, 95% CI 1.11-2.01; HR 1.69, 95% CI 1.18-2.41; HR 1.88, 95% CI 1.22-2.81, respectively) (Fig. 3B). Based on the analysis of ranking, pembrolizumab plus axitinib had the highest probability (49%) to be the best treatment for intermediate-risk patients (SUCRA = 90.7%). Avelumab plus axitinib and cabozantinib had a similar likelihood of being the second-best option for patients with intermediate-risk disease (Fig. 3C).

3.2.3. Based on data that was available for advanced/metastatic RCC of poor-risk, the network involved seven trials comparing nine different treatments (721 total patients; Fig. 4A) [15 16 33-35 37 39 40 42]. Network meta-analysis demonstrated that nivolumab plus ipilimumab and pembrolizumab

plus axitinib were associated with significantly higher improvement in PFS than sunitinib (HR 0.57, 95% CI 0.43-0.76; HR 0.48, 95% CI 0.30-0.82, respectively) (Fig. 4B). On the base of ranking analysis, there was a 60% probability that pembrolizumab plus axitinib had the greatest PFS for poor-risk patients (SUCRA = 91.3%) (Fig. 4C).

3.3. Overall survival
Five RCTs reported OS according to risk subgroup, and data from three of them contributed to the network meta-analysis (572, 1801, and 407 patients for favorable-, intermediate-, and poor-risk, respectively)[16 38 39]. For advanced/metastatic RCC of favorable-risk, there is no significant OS benefit between sunitinib and pazopanib (HR 0.88, 95% CI 0.64-1.21) or pembrolizumab plus axitinib (HR 0.64, 95% CI 0.24-1.70) (Fig. 5A). For intermediate-risk patients, pembrolizumab plus axitinib and nivolumab plus ipilimumab were associated with significantly higher improvement in OS than sunitinib (HR 0.53, 95% CI 0.34-0.81; HR 0.66, 95%CI 0.50-0.87, respectively)(Fig. 5B). For advanced/metastatic RCC of poor-risk, pembrolizumab plus axitinib and nivolumab plus ipilimumab were significantly more efficacious for OS than sunitinib (HR 0.43, 95% CI 0.23-0.80; HR 0.57, 95% CI 0.39-0.883, respectively) (Fig. 5C). Based on the results of ranking, there were 81% and 78% probabilities for pembrolizumab plus axitinib to be the best choice for intermediate- and poor-risk patients, respectively (SUCRA =93.1% ; SUCRA= 91.4%, respectively) (appendix Fig.2-3).” on pages 9-11

Comment 2- Recent studies used IMDC for risk stratification, older studies MSKCC. Although, those 2 are similar, there is difference. Need to include this to the manuscript, especially page 6 line 23, page 7 line 12.

Response: Thanks for your kind remind. As is shown in the Table, the CABOSUN and KEYNOTE-426 trials used the IMDC for risk stratification, the other trials used MSKCC criteria. To be clear, we replaced “Consequently, we performed all network analyses separately by risk groups (favorable-, intermediate-, and poor-risk groups) according to the MSKCC risk model .” with “Therefore, we performed all network analyses separately by risk groups (favorable-, intermediate-, and poor-risk groups) according to the MSKCC or IMDC risk model.” in Methods section on page 7, and replaced “According to the MSKCC criteria, there were 2 318, 4 413 and 517 participants had favorable-, intermediate-, and poor-risk disease, respectively. ” with “According to the MSKCC or IMDC criteria, there were 2 783, 5 474 and 721 participants had favorable-, intermediate-, and poor-risk disease, respectively.” in Results section on page 9.

Comment 3- The biology of RCC is heterogeneous, especially more sarcomatoid pt is either intermediate or poor risk, more so than good risk. This will effect the overall trial population. Please make sure to expend the discussion and include limitation.

Response: We agree with the reviewer. As we showed in the Results section, Thirteen trials selected for clear-cell carcinoma subtypes [10-12 15 16 33-40], and two trials also included small subsets of non-clear-cell histotypes, each comprising 11% and 14% of the study population, respectively [41 42]. In our opinion, the small subsets of non-clear-cell histotypes might somewhat damage the results of our analysis. Accordingly, in Discussion section, we added “two trials included small subsets of non-clear-cell histotypes (11% and 14% of the study population), which might somewhat damage the results of our analysis.” on page 16.

Comment 4- Front line cabo trial is smaller, phase 2 trial. Making big conclusion might not be ideal. Please expend limitation.

Response: We agree with the reviewer. In our updated analysis, pembrolizumab plus axitinib had the highest probability to be the best treatment for intermediate-risk patients. Accordingly, we revised our

conclusion to “pembrolizumab plus axitinib was most likely to be the best option for intermediate- and poor-risk patients.” on page 15, and added “In addition, three included trials (CABOSUN, ROSORC, and RECORD-3) are phase 2 RCTs with smaller sample size, and they may be less authoritative compared with phase 3 RCTs.” in Discussion section on page 16.

Comment 5- Page 11, line 11 does not make sense. Pesudoprogression should not be different since RECIST criteria is the same regardless of risk group. But underline biology of good risk might be different, such as more angiogenic driven, etc. Please revise and use more uptodate reference that help this. (gene signature from pembro/axi and avelumab/axi trials).

Response: We agree with the reviewer. Some patients treated with immunotherapies were observed to experience initial increased size of tumor, confirmed as inflammatory cell infiltrates or necrosis by biopsy [1,2]. This phenomenon is defined as “Pesudoprogression”. In our updated analysis, avelumab plus axitinib was associated with a significantly higher improvement in PFS than sunitinib (HR 0.57, 95% CI 0.34-0.96) for patients with favorable-risk. Accordingly, we deleted “It’s reported that immunotherapy might be associated with “pseudo-progression” phenomenon according to the (Response Evaluation Criteria in Solid Tumors) RECIST criteria.” on page 12.

Reference

[1] George S, Motzer RJ, Hammers HJ, Redman BG, Kuzel T, Tykodi SS, et al. Safety and efficacy of nivolumab in patients with metastatic renal cell carcinoma treated beyond progression a subgroup analysis of a randomized clinical trial. JAMA Oncol 2016;2(9):1179–86.
[2] Chiou VL, Burotto M. Pseudoprogression and immune-related response in solid tumors. J Clin Oncol 2015;33(31):3541.

Comment 6- They used PFS, not OS. Not clear why? They stated "In addition, PFS has been recognized as a surrogate endpoint of OS in medical oncology as well as in advanced/metastatic RCC in the TKI era" But we know it is not the case for immunotheraphy. Need further explanation and or revision.

Response: We thank you for pointing this out. In our updated analysis, we used overall survival as co-primary endpoint. In the section of Result, we added “
3.3. Overall survival
Five RCTs reported OS according to risk subgroup, and data from three of them contributed to the network meta-analysis (572, 1801, and 407 patients for favorable-, intermediate-, and poor-risk, respectively)[16 38 39]. For advanced/metastatic RCC of favorable-risk, there is no significant OS benefit between sunitinib and pazopanib (HR 0.88, 95% CI 0.64-1.21) or pembrolizumab plus axitinib (HR 0.64, 95% CI 0.24-1.70) (Fig. 5A). For intermediate-risk patients, pembrolizumab plus axitinib and nivolumab plus ipilimumab were associated with significantly higher improvement in OS than sunitinib (HR 0.53, 95% CI 0.34-0.81; HR 0.66, 95%CI 0.50-0.87, respectively)(Fig. 5B). For advanced/metastatic RCC of poor-risk, pembrolizumab plus axitinib and nivolumab plus ipilimumab were significantly more efficacious for OS than sunitinib (HR 0.43, 95% CI 0.23-0.80; HR 0.57, 95% CI 0.39-0.883, respectively) (Fig. 5C). Based on the results of ranking, there were 81% and 78% probabilities for pembrolizumab plus axitinib to be the best choice for intermediate- and poor-risk patients, respectively (SUCRA =93.1% ; SUCRA= 91.4%, respectively) (appendix Fig.2-3).” on page 10.
In the section of Discussion, we deleted “There are several factors which make an analysis of OS challenging, therefore meta-analysis on OS was not performed. First, the OS data were not reported in five included studies. Second, for the available OS data, there were issues of confounders, such as crossover to more efficacious treatment and the influence of subsequent anticancer therapies [33, 36, 65]. These factors can substantially underestimate the difference between two treatments in terms of

OS. In addition, PFS has been recognized as a surrogate endpoint of OS in medical oncology as well as in advanced/metastatic RCC in the TKI era [66].", and added "Moreover, most of the trails did not perform the analysis of OS in risk subgroup, which made it impossible to assess the OS benefits of all the existing treatments for different risk patients." on page 16.

Comment 7- There is recent paper:
EBioMedicine. 2019 Sep;47:78-88. doi: 10.1016/j.ebiom.2019.08.006. Epub 2019 Aug 19. Role of immune checkpoint inhibitor-based therapies for metastatic renal cell carcinoma in the first-line setting: A Bayesian network analysis.
what additional information is added? This paper is not cited although from same group?

Response: We are sorry for the confusion. The meta-analysis above used aggregate data and did not perform subgroup analysis based on risk strata. In the present study, we performed all network analyses separately by risk groups (favorable-, intermediate-, and poor-risk groups), thus providing physicians with the optimal treatment for different risk patients. To be clear, in Discussion, we have added "Recently, several network meta-analyses were attempted to investigate the comparative effects of different systemic agents for treatment of advanced/metastatic RCC [60-63]. However, trials included in the meta-analyses enrolled patients with different risk groups. The analysis used aggregate data and did not perform subgroup analysis based on risk strata. In the present study, we performed a network meta-analysis to compare first-line systemic treatments for advanced/metastatic RCC of favorable-, intermediate-, and poor-risk, respectively, thus providing physicians with the optimal treatment for different risk groups." on page 15.

Comment 8- They might revise conclusion. Not totally make sense intermediate risk patient benefit more from cabo. How they address particular trial and sample size issue?
Need further explanation and possible revision.

Response: We agree with the reviewer. In our updated analysis, pembrolizumab plus axitinib had the highest probability to be the best treatment for intermediate-risk patients. Accordingly, we revised our conclusion to "pembrolizumab plus axitinib was most likely to be the best option for intermediate- and poor-risk patients." on page 16.

Comment 9- Good risk patient will benefit from sunitinib might change when adding more uptodate trial, such as pembro/axi, avelumab/axi.

Response: We agree with the reviewer. In our updated analysis, avelumab plus axitinib was associated with significantly higher improvement in PFS than sunitinib (HR 0.66, 95% CI 0.54-0.81). Based on the analysis of SUCRA, there was a 41.7% chance that avelumab plus axitinib provided the greatest PFS benefit for patients with favorable-risk disease. We have revised the conclusion accordingly. In the section of conclusion, we have replaced "sunitinib might be the optimum treatment for advanced/metastatic RCC of favorable-risk" with "avelumab plus axitinib might be the optimum treatment for advanced/metastatic RCC of favorable-risk". on page 16.

VERSION 2 – REVIEW

REVIEWER	Jesse Y Hsu University of Pennsylvania, USA
REVIEW RETURNED	15-Mar-2020

GENERAL COMMENTS	Note: Line numbering and page numbering (pages 2–50) are based on ruler applied in Editorial Managergenerated PDF. The responses to the previous comments were helpful and the revised manuscript

	had improved its clarity. Further clarifications and suggested revisions are listed as follows. 1. Page 7, lines 17–33: Following your responses to previous Comments #4 and #8,I suggest clarifying in the manuscript how the patient-level data were used For example, in Section 2.3, please confirm and specify that the patient characteristics, treatment strategies and interest outcome were extracted at the study-level for meta analyses even if the patient-level were available. In Section 2.4, the patient-level data were used to calculate the odds ratios for drug-related adverse event. Were there other places where the patient-level data were used? 2. Page 7, lines 54–57: Following your response to previous Comment #10a, I assumed the available raw data referred to the available patient-level data. Please consider replacing “raw” with ‘patient-level’ for consistency. 3. Page 7, lines 57–59: My apologies for not making previous Comment #11 clear. Specifically, if the sentence, “both random-effects and fixed-effects models were performed for the analyses”, referred to all analyses in this manuscript, please consider writing ‘both random-effects and fixed-effects models were performed for all Bayesian network meta-analyses’. I struggled to understand what “the analyses” referred to. 4. Page 8, lines 43–51: Following your response to previous Comment #14, please consider adding a sentence ‘we performed fewer iterations for PFS to reduce computational burden without loss of convergence and model fit’.

REVIEWER	Mehmet Asim Bilen Winship Cancer Institute of Emory University, ISA
REVIEW RETURNED	24-Mar-2020

GENERAL COMMENTS	Authors provided satisfactory revision for my comments.

VERSION 2 – AUTHOR RESPONSE

Comment 1- Page 7, lines 17–33: Following your responses to previous Comments #4 and #8,I suggest clarifying in the manuscript how the patient-level data were used For example, in Section 2.3, please confirm and specify that the patient characteristics, treatment strategies and interest outcome were extracted at the study-level for meta analyses even if the patient-level were available. In Section 2.4, the patient-level data were used to calculate the odds ratios for drug-related adverse event. Were there other places where the patient-level data were used?

Response: We thank for your valuable suggestion. To be clear, we added “The patient characteristics, treatment strategies, PFS and OS were extracted at the study-level for meta-analyses even if the patient-level were available. For drug-related adverse events, the patient-level data were extracted for meta-analyses. ” in lines 27-31 on page 7. In Section 2.4, we replaced “For drug-related adverse events, we calculated odds ratios (ORs) using the available raw data abstracted from the trials.” with “For drug-related adverse events, we calculated odds ratios (ORs) using the available patient-level data abstracted from the trials.” in line 60 on page 7 and line 4 on page 8. The patient-level data were

not used in other places in the manuscript. Thanks for your kind remind.

Comment 2- Page 7, lines 54–57: Following your response to previous Comment #10a, I assumed the available raw data referred to the available patient-level data. Please consider replacing “raw” with ‘patient-level’ for consistency.

Response: We thank for your kind remind. To be clear, we replaced “For drug-related adverse events, we calculated odds ratios (ORs) using the available raw data abstracted from the trials.” with “For drug-related adverse events, we calculated odds ratios (ORs) using the available patient-level data abstracted from the trials.” in line 60 on page 7 and line 4 on page 8.

Comment 3- Page 7, lines 57–59: My apologies for not making previous Comment #11 clear. Specifically, if the sentence, “both random-effects and fixed-effects models were performed for the analyses”, referred to all analyses in this manuscript, please consider writing ‘both random-effects and fixed-effects models were performed for all Bayesian network meta-analyses’. I struggled to understand what “the analyses” referred to.

Response: We are sorry for the confusion. To be clear, we replaced “Both random-effects and fixed-effects models were performed for the analyses” with “Both random-effects and fixed-effects models were performed for all Bayesian network meta-analyses.” in lines 4-6 on page 8.

Comment 4- Page 8, lines 43–51: Following your response to previous Comment #14, please consider adding a sentence ‘we performed fewer iterations for PFS to reduce computational burden without loss of convergence and model fit’.

Response: We thank for your valuable comment. In Section 2.4, we added “We performed fewer iterations for PFS to reduce computational burden without loss of convergence and model fit.” in lines 50-52 on page 8.