Peer Review File

Article Information: https://dx.doi.org/10.21037/atm-23-195

Round 1

1) First of all, my major concern for this study is the rationale and clinical needs for this network meta-analysis. The authors did not indicate the clinical needs for ranking and comparing the three treatments of interest and did not present the conflicting findings on the relative efficacy and safety of the head-to-head comparisons across the three treatments. The methodology is also problematic since no head-to-head trials of the three treatments were included and the authors only included place-controlled trials. The other methodology concern is the clinical heterogeneity in the three types of treatments; for example, there are different IL inhibitors and the efficacy and safety might be different across these inhibitors but the authors ignored such sources of variations in the treatment efficacy.

Reply 1): Thank you for pointing these out. We agree with what you say. In a study, the correct rationale is an important prerequisite for the results to be scientifically guided clinically. The clinical status of these three treatments has been briefly described in the introduction (lines 89-101), and the safety and efficacy of these drugs remains a research focus for the current rheumatology community. The focus of this manuscript is therefore to rank the strengths and weaknesses of these three treatments in terms of safety and efficacy, which are discussed extensively in the results section. In addition, by retrieving the database, we found that there are currently few head-to-head trials of these three treatments. Therefore, we use network meta-analysis for indirect exploration and comparison and strive for some reasonable results. Only one head-to-head trial was included in the study, and it is clear from the results of our analysis that the results of the head-to-head trial contradict those of our study. As you pointed out, we did not mention the possible cause of these contradictions, so we added that point to the manuscript. Changes in the text: Lines 400-404. There are indeed clinical differences in efficacy and safety among IL inhibitors, which may be related to differences in some of the molecular and pharmacokinetic properties of IL inhibitors. However, insufficient evidence has been obtained to suggest significant differences in the effectiveness between TNF-αand JAK inhibitors.

2) Second, the abstract is not adequate and needs further revisions. The background did not explain the clinical needs for comparing the three treatments, the clinical significance of this study, and the clinical controversy regarding the relative efficacy and safety of the three treatments. The methods did not define the inclusion of eligible studies by using the PICOS principles and did not describe the risk of bias assessment of included studies. The results need to report the sample size and number of included studies, as well as their risk of bias. Please report the head-to-head comparisons results between the three treatments in terms of efficacy and safety outcomes, not the treatment vs. placebo. Please also quantify these

findings by providing pooled effect size measures and accurate P values. The conclusion seems to be not supported by the current findings.

Reply 2): Thank you for your valuable advice. In response to the issues you mentioned above in the manuscript, we have made corresponding modifications and improvements to the abstract and background. (lines 34-37, 52-55, 62-67, 97-103) This manuscript is included in the study strictly under the conditions of the PICOS principle. E.g. P corresponds to '(I) The diagnosis of participants followed the 1984 modified New York Criteria for AS'; I, C and S corresponds to '(II) "randomized, double-blind, placebo-controlled" trials that compared IL inhibitors or JAK inhibitors or TNF-α inhibitors with a placebo for the treatment of AS were included'; O corresponds to '(III) the outcome measures were the number of patients satisfying the Assessment of SpondyloArthritis International Society (ASAS) 20 improvement criteria (ASAS 20), ASAS 40 improvement criteria (ASAS 40), at least a 20% improvement in at least five of six ASAS domains (ASAS 5/6), and at least a 50% improvement in the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI50), as well as the Bath Ankylosing Spondylitis Functional Index (BASFI) scores'. The risk assessment of bias included in the Study has been described in detail in the "Study characteristics and risk of bias" section of the results. (lines 215-228) The outcomes of the risk of bias assessment are presented in Figure 3. By retrieving the database, we found that there are currently few head-to-head trials of these three treatments. Therefore, we use network meta-analysis for indirect exploration and comparison and strive for some reasonable results. In statistics, binary outcomes were synthesized by calculating the relative risk (RR) with a 95% confidence interval (CI) using a random-effects model. A 95% CI crossing 1 indicates no significant differences between the groups and vice versa. Continuous variables were synthesized by calculating the mean difference (MD) with 95% CI using a fix-effects model. The analysis outcomes were converted into a likelihood or a rating for each condition measured by SUCRA expressed as a percentage.

3) Third, the introduction of the main text need to explain the clinical needs to compare the three treatments, clinical controversy regarding their relative efficacy, potential reasons for the controversy, clinical needs for ranking the three treatments, and why a network meta-analysis is suitable to answer the clinical question. The authors' comments on the limitations of available systematic reviews are difficult to understand, please further clarify "these reviews did not analyze "strictly randomized, double-blind, placebo-controlled" trials, or the interventions investigated by some of the included trials coincided with the drugs administered in stable doses in other included trials".

Reply 3): Thank you for your valuable advice. In response to the issues you mentioned above in the manuscript, we have made corresponding modifications and improvements to the abstract and background. (lines 34-37, 52-55, 62-67, 97-103) In response to your queries, we have found errors in the manuscript. The correct understanding of 'these reviews did not

analyze “strictly randomized, double-blind, placebo-controlled’ is that some of the studies included in these meta-analyses were not randomized, double-blind, placebo-controlled. In addition, in Cao 2022 (6), DMARDs were analyzed as part of the intervention along with other biologics or JAK inhibitors, but in other trials included in the study, steady doses of DMARDs were used together with biologic inhibitors or JAK inhibitors. So we say that 'the interventions investigated by some of the included trials coincided with the drugs administered in stable doses in other included trials'. Changes in the text: lines 103-106.

4) Fourth, in the methodology of the main text, the literature search has language bias since no non-English language databases were searched. The inclusion criterion on the placebo-controlled trials only is problematic since head-to-head comparison studies are more important in network meta-analysis. The authors need to use a separated paragraph to describe the risk of bias assessment details including the criteria for the three levels of risk of bias. In statistics, please describe the statistical test for publication bias, the test of heterogeneity and its sources, the test of the influence of level of risk of bias on the pooled results, and the P value for statistical significance.

Reply 4: We understand your concerns about linguistic bias arising from the inclusion of only English-language literature in the study of this manuscript. We quite agree with you. But this is undeniably a common problem for database-based analysis. Due to time and labor cost constraints, we still chose the mainstream language literature as the scope of our study. However, in future studies, we will try to expand the language range to make the results more rigorous. The assessment of literature risk bias has been described in detail in the manuscript. (lines 219-228) Detailed Statistical description has been given in the section "Statistical method".

Round 2

Reviewer A

First, the title is problematic such as “a “randomized, double-blind, placebo-controlled” trials”. Further, in the authors’ replies “Only one head-to-head trial was included in the study”, so the current title is not accurate.

Reply: Thank you for your careful review. We agree with you very much, so we have revised the title.

Changes in the text: Page one, Line 3-4

Second, in the abstract, the risk of bias assessment instrument and the assessment results on risk of bias of included studies are essential, but the authors still did not report these. The results must be quantified by reporting the pooled effect sizes and accurate P values but the authors still did not provide any figures.

Reply: Thanks, we have included the methodology and results of the bias risk assessment in the abstract as suggested by you. In addition, we have added the summary results of the analysis, where intervention differences in network meta-analyses are often judged by confidence intervals and therefore do not have a P-value.

Changes in the text: Page 2, Line 41-42，45-55.

Third, in the introduction the authors described “no direct head-to-head comparisons” but if this is the rationale the authors should do head-to-head RCTs, not this network meta-analysis, so the authors still did not explain why a network meta-analysis is suitable to address this clinical question.

Reply: Thank you for your careful review. We have deeply reflected on your questions and made appropriate modifications and additions in the manuscript.

Changes in the text: Page 4, Line 91-96, 101-109.

Fourth, in the methodology of the main text, the authors are still not willing to address the language bias by searching other language databases such as Chinese, which is very easy for the authors. In statistics, funnel plot alone is not adequate for assessing the publication bias. The authors also did not analyze the influence of level of risk of bias on the pooled results. An important methodology problem of this meta-analysis is that all placebos in different included studies are not equivalent but the authors hypothesized so.

Reply: Thanks, we have made a deep reflection on the questions you raised. So we retrieved and expanded the database. We retrieved literature from various databases including Web of Science, Cochrane, Embase, PubMed, China National Knowledge Infrastructure, Weipu Journal Database, SinoMed, and WanFang Data up to February 1, 2023. So we've updated the full manuscript. To more directly reflect publication bias results, we perform Egger's test on the data and describe and interpret publication bias results in the manuscript. In addition, placebos are generally considered to be made from substances that have no effect or toxic side effects on the subjects, so all placebos included in the study can be considered the same. If there are any other modifications we could make, we would like very much to modify them and we really appreciate your help. Thank you very much for your help.

Changes in the text: Page 7, Line 196. Page 11, Line 324-328. Page 14, Line 429-430.

Reviewer B

1 The article follows the PRISMA-NMA checklist for reporting standards. Please revise your manuscript according to the checklist.

Reply: We have revised our manuscript according to the attached checklist.

2. Please add citation of references for the previous meta-analyses.

657 in these meta-analyses overlapped others. Some of our findings through this network
658 meta-analysis echo the conclusions of some previous meta-analyses. For example, IL-
659 6 inhibitors and IL-23 inhibitors have poor or no therapeutic effects, compared with the
660 placebo, and IL-17 inhibitors present significantly good treatment effects. Previous
661 meta-analyses have reported that TNF-α inhibitors were the best option for people with
662 AS who suffer rapid disease progression; however, these analyses found that TNF-α
663 inhibitors were the safest drug and IL-17 inhibitors were irreplaceable for treating AS,
664 which differs from the findings in our study. The distinct differences between our study
665 and previous meta-analyses may lie in the types of trials included for analysis, as

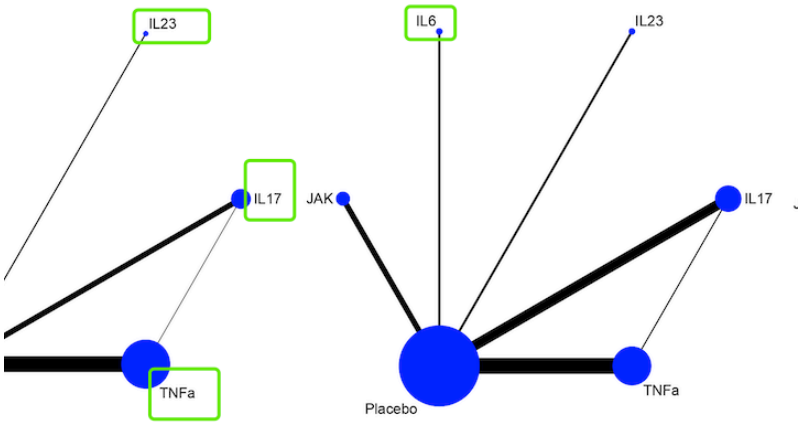
Reply: Thanks. We have added citation of references.

3. The below number is wrong. Should it be "47"?

378 BASFI was estimated in this network meta-analysis. Among 46 RCTs, 13 investigated
379 IL inhibitors, including nine RCTs on IL-17 inhibitors (Bimekizumab, Ixekizumab,
380 Secukinumab, and Netakimab), with 1,724 participants in the intervention group and

Reply: Thank you for your careful review. We have corrected it.

4. Figures 2, 4-5:
Please revise all word "IL6" to "IL-6", "IL17" to "IL-17", "IL23" to "IL-23", "TNFa" to "TNF-α" and "ASAS56" to "ASAS5/6" in Figures 2, 4-5.



ASAS56		
IL17		
0.69 (0.09, 3.22)	IL23	
1.53 (0.60, 3.80)	2.23 (0.39, 18.66)	IL6

Reply: Revised. Thanks.

5. Figure S1, S3-S5:
1) Please revise all word "IL6" to "IL-6", "IL17" to "IL-17", "IL23" to "IL-23", "TNFa" to "TNF-α" in Figure S1, S3-S5.
2) Please check whether the word below should be "(95% CI)" in your Figure S1, S4-S5.

Risk Ratio (95% CrI) Study P−value

TNFa vs IL17

Reply: Thanks, Revised. ‘Crl’ stands for ‘Credible Interval’. The term ‘Credible interval’ was used in Bayesian analysis, the meaning of credible interval was the most possible parameter.

5. Figure S2:

Please complete all number “.5” to “0.5”.

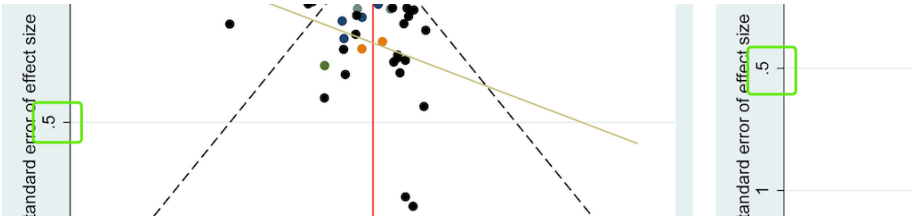


Figure: Revised. Thanks.

6. Please check all “ASAS56” in your whole manuscript. It should be “ASAS5/6”. Please revise.

Reply: Thank you for your careful review. We've updated the full manuscript.