

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	ABO Blood Groups and Hepatitis B Virus Infection: A Systematic Review and Meta-Analysis
<b>AUTHORS</b>	Jing, wenzhan; Zhao, Siyu; Liu, Jue; Liu, Min

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Xingshun Qi General Hospital of Northern Theater Command, China
<b>REVIEW RETURNED</b>	14-Sep-2019

<b>GENERAL COMMENTS</b>	<p>This paper is well written and organized. I recommend a revision before its publication in this journal. Some comments are listed.</p> <ol style="list-style-type: none"><li>1. In the Introduction section, the association of ABO with liver diseases should be reviewed. A recent meta-analysis found that HCC patients might have a lower proportion of blood type O than healthy subjects (PMID: 30004289). However, the meta-analysis found that the blood group O was associated with a higher risk of HBV infection. Please discuss it.</li><li>2. The search strategy was done on 2017. It must be updated. Some new papers may be obtained.</li><li>3. The authors restricted to inclusion of cross-sectional and cohort studies. Are there case-control studies included in the meta-analysis?</li><li>4. Did the authors put both cross-sectional and cohort studies into the statistical analysis together?</li><li>5. When both cross-sectional and cohort studies are included, is the calculation of RR right?</li><li>6. In the Table 1, there are a large heterogeneity in the sample size among studies. Some papers included more than 10000 patients, but others included less than 1000 patients. Is the combination of these data reasonable?</li><li>7. Are there any other confounding factors for HBV included in these evaluated studies?</li><li>8. As discussed, the authors said "the association between ABO blood group and HBV infection was only found in higher endemic areas but not in lower endemic areas." This association might be attributed to the regional factors.</li></ol>
-------------------------	--

<b>REVIEWER</b>	Terence Lao Department of Obstetrics & Gynaecology, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, Hong Kong SAR, PRC.
<b>REVIEW RETURNED</b>	29-Oct-2019

<b>GENERAL COMMENTS</b>	This report is well written and the study was conducted appropriately. The result would help to explain in part the different prevalence of HBV infection in different population groups, and possibly shed light on the factors affecting consequences of HBV infection, the response to HBV immunization, and the post-vaccination immunological memory.
-------------------------	--

<b>REVIEWER</b>	Nicola Armstrong Murdoch University Australia
<b>REVIEW RETURNED</b>	25-Nov-2019

<b>GENERAL COMMENTS</b>	<p>The article performs a review and meta-analysis of the association between blood groups and Hep B infection. While the article is well written, a brief review by a native English speaker would be beneficial. There are several places where the wrong tense is used, and some spelling errors exist (e.g Word Bank instead of World Bank on p5). Additionally, I would advise against the use of "Black" as a race, and instead advocate "African" or "African-American". It is a standard review article, my one concern is that the literature review was only up to Dec 2017, and the methods could be explained in a little more detail so as to be truly reproducible.</p> <p>Specific comments are as follows:</p> <p>In the abstract, the results section refers to subgroup analysis, but at this point it is unclear which were conducted. Please state explicitly that after adjusting for ** (subgroups), the effect remained. The conclusion should be that your analysis suggests that... not your data.</p> <p>In the introduction, I found the sentence on VWF to be unclear - could you please reword? Check the position of brackets and full stops in this sentence also.</p> <p>In the Methods, the statistical analysis lacks detail. Were the Egger's tests two-sided? What was deemed significant? What sort of random effects models were used (REML)? Why not conduct a full sensitivity analysis by removing all studies one by one? On p10, the sensitivity analysis is done for two studies, in contrast to the methods details. Please use the references of the studies where appropriate in the text (p10) and the tables. Were both studies removed at the same time to see sensitivity? What were the numbers of studies, and numbers of individuals involved, in the</p>
-------------------------	---

	<p>subgroup analyses? Was there enough power to detect an association in all cases? It would be good to highlight the single cohort study in Table 1, and in the methods state that Liu et al dominated results of the meta analysis specifically due to its size in comparison to the other studies (at least I assume this was the case?). What were the "other sources" that 10 articles were found from? Is there any reason why the literature search stopped in Dec 2017, almost 2 years ago? On p6, the use of words vs numbers for the number of studies is inconsistent, please unify. What studies represent endemic areas? Were any studies removed due to I<sup>2</sup> values?</p> <p>In the discussion on p11, line 22 - "was few reported explicitly" does not make sense. The addition of references here, or expansion of the "different analysis methods" is needed. Also, on p12, line17 - what do the authors mean when they say that group O should be "given more attention"?</p> <p>With respect to the figures and tables: in figure 1, please clarify and make consistent the use of either records or articles. There doesn't appear to be a caption for figure 2? Again in the supplementary tables, highlight the cohort study. Supp table 1 repeats information from Table 1 - is it necessary?</p>
--	--

### VERSION 1 – AUTHOR RESPONSE

Response to Reviewer 1 Comments:

This paper is well written and organized. I recommend a revision before its publication in this journal. Some comments are listed.

1. In the Introduction section, the association of ABO with liver diseases should be reviewed. A recent meta-analysis found that HCC patients might have a lower proportion of blood type O than healthy subjects (PMID: 30004289). However, the meta-analysis found that the blood group O was associated with a higher risk of HBV infection. Please discuss it.

Response: Thank you. According to the Reviewer's suggestion, we have added that "Recently, a meta-analysis also found that hepatocellular carcinoma (HCC) patients might have a lower proportion of O subjects than healthy subjects" in the introduction section.

In the discussion, we have also supplemented that "Our meta-analysis was inconsistent with the recently meta-analysis, which found that HCC patients might have a lower proportion of O subjects than healthy subjects. The possible explanation for the inconsistency is the long-term and complicated process from HBV infection to the occurrence of HCC."

2. The search strategy was done on 2017. It must be updated. Some new papers may be obtained.

Response: Thanks for the Reviewer's suggestion. According to the Reviewer's suggestion, we have updated our search.

3. The authors restricted to inclusion of cross-sectional and cohort studies. Are there case-control studies included in the meta-analysis?

Response: Thanks for the Reviewer's suggestion. Case-control studies were not included in the meta-analysis. We need to calculate the HBV prevalence of the included studies, and conduct corresponding subgroup analysis. Because case-control studies can't be used to calculate this prevalence, we did not include them in this meta-analysis.

4. Did the authors put both cross-sectional and cohort studies into the statistical analysis together?

Response: Thanks for the Reviewer's suggestion. We put them into the statistical analysis together. And we have also done the subgroup analysis according to the different study design, which has been shown in table 2.

5. When both cross-sectional and cohort studies are included, is the calculation of RR right?

Response: Thanks for the Reviewer's suggestion. ABO blood group is an invariant factor, so we think the calculation of RR is appropriate. Additionally, RR is easier to interpret than OR.

6. In the Table 1, there are a large heterogeneity in the sample size among studies. Some papers included more than 10000 patients, but others included less than 1000 patients. Is the combination of these data reasonable?

Response: Thanks for the Reviewer's suggestion. Although there was a large heterogeneity in the sample size among included studies, dropping any small sample studies may also be unreasonable. Therefore, considering to the heterogeneity in the sample size, subgroup analysis have been done by sample size  $\geq 2000$  or  $< 2000$ .

7. Are there any other confounding factors for HBV included in these evaluated studies?

Response: Thanks for the Reviewer's suggestion. Yes, a few studies have reported the distribution of HBV infection among Rh-positive and Rh-negative blood group. However, the relationship is controversial. Additionally, age group is a confounding factor for HBV infection. However, HBV infection by both age group and ABO blood group was not reported in these evaluated studies. Therefore, we could not do further analysis. In the limitations, we have stated that "Third, few published studies on the association between HBV infection and blood group have controlled HBV infection related risk factors such as family history of HBV infection, age group, blood transfusion, and acupuncture, thus we were not able to conduct the corresponding subgroup analyses."

8. As discussed, the authors said "the association between ABO blood group and HBV infection was only found in higher endemic areas but not in lower endemic areas." This association might be attributed to the regional factors.

Response: Thanks for the Reviewer's suggestion. According to the Reviewer's suggestion, we have added this sentence in the discussion. As shown in P10 Line 27-28, we added that "Additionally, this

association might be partly attributed to the regional factors, due to the high relevance between HBV endemic and region.”.

Response to Reviewer 2 Comments:

This report is well written and the study was conducted appropriately. The result would help to explain in part the different prevalence of HBV infection in different population groups, and possibly shed light on the factors affecting consequences of HBV infection, the response to HBV immunization, and the post-vaccination immunological memory.

Response: Thanks for the Reviewer's recognition.

Response to Reviewer 3 Comments:

The article performs a review and meta-analysis of the association between blood groups and Hep B infection. While the article is well written, a brief review by a native English speaker would be beneficial. There are several places where the wrong tense is used, and some spelling errors exist (e.g Word Bank instead of World Bank on p5). Additionally, I would advise against the use of "Black" as a race, and instead advocate "African" or "African-American". It is a standard review article, my one concern is that the literature review was only up to Dec 2017, and the methods could be explained in a little more detail so as to be truly reproducible.

Response: Thanks for the Reviewer's suggestion. Firstly, according to the Reviewer's suggestion, we have asked my colleague to check the English of my manuscript, and we have revised the wrong tense and the spelling errors. According to the published articles, we have changed the "black" as "African". Additionally, we have updated our search strategy up to Dec 2019, and uploaded a full electronic search strategy for PubMed as an additional file. And we have also explained more detail about the methods.

Specific comments are as follows:

In the abstract, the results section refers to subgroup analysis, but at this point it is unclear which were conducted. Please state explicitly that after adjusting for \*\* (subgroups), the effect remained. The conclusion should be that your analysis suggests that... not your data.

Response: Thanks for the Reviewer's suggestion. According to the Reviewer's suggestion, we have revised the corresponding results as "In the subgroup analyses, the inverse relationship between blood group B and HBV infection remained stable in higher endemic areas (HBV prevalence  $\geq$  5%), Asian people, larger sample size studies ( $\geq$  2000), general population and blood donors, lower middle income group and studies published before 2010 years." We have revised the conclusion as "Our data suggested that the blood group B was associated with a lower risk of HBV infection. More researches are needed to clarify the precise role of ABO blood group in HBV infection to address the global question of HBV infection."

In the introduction, I found the sentence on VWF to be unclear - could you please reword? Check the position of brackets and full stops in this sentence also.

Response: Thanks for the Reviewer's suggestion. According to the Reviewer's suggestion, we have reworded this sentence as "For instance, by expressing on N-glycans of von Willebrand factor (VWF), ABH antigens (H antigen is the biosynthetic precursor to A and B antigens) impact the half-life of VWF, so VWF survival in O subjects is significantly shorter versus (vs.) in non-O subjects. Therefore, because of the lower VWF levels, O subjects have lower risk of venous thromboembolism."

In the Methods, the statistical analysis lacks detail.

Response: Thanks for the Reviewer's suggestion. According to the Reviewer's following specific suggestions, we have supplemented the detail of the statistical analysis.

Were the Egger's tests two-sided? What was deemed significant?

Response: Thanks for the Reviewer's suggestion. The Egger's tests were two-sided.  $P < 0.05$  was deemed significantly. We have revised the corresponding sentence as "Publication bias was evaluated by funnel plots and two-sided Egger's tests, and  $P < 0.05$  was deemed significantly."

What sort of random effects models were used (REML)?

Response: Thanks for the Reviewer's suggestion. The random effects model uses the method of DerSimonian & Laird, with the estimate of heterogeneity being taken from the Mantel-Haenszel model. We have revised the corresponding sentence as "RRs and 95% CIs (A vs. non-A, B vs. non-B, O vs. non-O, AB vs. non-AB) were pooled by using of random-effect models with the estimate of heterogeneity being taken from the Mantel-Haenszel model, and  $P < 0.05$  was deemed significantly."

Why not conduct a full sensitivity analysis by removing all studies one by one?

Response: Thanks for the Reviewer's suggestion. Eventually, 38 eligible articles were included in the meta-analysis. The results of the full sensitivity analysis by removing all studies one by one are more complex and difficult to display clearly than the sensitivity analysis by excluding the studies with the largest sample size which dominated the results of the meta-analysis.

On p10, the sensitivity analysis is done for two studies, in contrast to the methods details.

Response: Thanks for the Reviewer's suggestion. We have revised corresponding methods as "Sensitivity analyses were performed by excluding large sample size studies orderly or at the same time, which dominated the results of the meta-analysis."

Please use the references of the studies where appropriate in the text (p10) and the tables.

Response: Thanks for the Reviewer's suggestion. According to the Reviewer's suggestion, we have supplemented the references.

Were both studies removed at the same time to see sensitivity?

Response: Thanks for the Reviewer's suggestion. According to the Reviewer's suggestion, we have added the sensitive analysis, which removed both studies at the same time to see sensitivity. The results were still stable and were added in the table 2.

What were the numbers of studies, and numbers of individuals involved, in the subgroup analyses?

Response: Thanks for the Reviewer's suggestion. According to the Reviewer's suggestion, number of studies and numbers of individuals involved in the subgroup analyses were shown in the table 2.

Was there enough power to detect an association in all cases?

Response: Thanks for the Reviewer's suggestion. In our subgroup analysis, the smallest number of individuals involved was 2655 in the patient's population. And even the smallest sample size in our subgroup analysis was larger than the sample size of 14 included articles with < 2000 individuals. Additionally, the number of individuals involved in other subgroups were more than 10,000. So, we think there was enough power to detect the association.

It would be good to highlight the single cohort study in Table 1, and in the methods state that Liu et al dominated results of the meta analysis specifically due to its size in comparison to the other studies (at least I assume this was the case?).

Response: Thanks for the Reviewer's suggestion. According to the Reviewer's suggestion, we have highlighted the single cohort study in Table 1. Additionally, in the methods stated that "Sensitivity analyses were performed by excluding large sample size studies orderly or at the same time which dominated the results of the meta-analysis.". Due to the sample size of Liu et al was more than 3.8 million and the sample size of Mohammadali et al was more than 2.0 million, so sensitivity analyses were performed by excluding the two largest studies orderly or at the same time.

What were the "other sources" that 10 articles were found from?

Response: Thanks for the Reviewer's suggestion. A manual search was conducted by reviewing the list of references. We have added the sentence that "Meanwhile, highly relevant reference articles were also searched by reviewing the list of references." in the search strategy.

Is there any reason why the literature search stopped in Dec 2017, almost 2 years ago?

Response: Thanks for the Reviewer's suggestion. There is no special reason. We have updated the search strategy and updated the results and discussion.

On p6, the use of words vs numbers for the number of studies is inconsistent, please unify.

Response: Thanks for the Reviewer's suggestion. According to the Review's suggestion, we have changed the words as numbers for consistence.

What studies represent endemic areas?

Response: Thanks for the Reviewer's suggestion. There were 14 studies in higher endemic areas (HBV prevalence  $\geq 5\%$ ), which was shown specially in the figure 2.

Were any studies removed due to  $I^2$  values?

Response: Thanks for the Reviewer's suggestion. There were no studies removed due to  $I^2$  values.

In the discussion on p11, line 22 - "was few reported explicitly" does not make sense. The addition of references here, or expansion of the "different analysis methods" is needed. Also, on p12, line17 - what do the authors mean when they say that group O should be "given more attention"?

Response: Thanks for the Reviewer's suggestion. According to the Reviewer's suggestion, we have added expansion of the "different analysis methods" in the discussion. As shown in P11 Line21, we have added that "..., such as the different reference of blood group in analysis". Additionally, we have deleted the sentence that "group O should be given more attention".

With respect to the figures and tables: in figure 1, please clarify and make consistent the use of either records or articles. There doesn't appear to be a caption for figure 2?

Response: Thanks for the Reviewer's suggestion. According to the Reviewer's suggestion, we have made consistent the use of either records or articles. And the caption for figure 2 is that "Figure 2. Forest plots by prevalence: (A) B vs. non-B; (B) O vs. non-O."

Again in the supplementary tables, highlight the cohort study. Supp table 1 repeats information from Table 1 - is it necessary?

Response: Thanks for the Reviewer's suggestion. We have highlighted the cohort study in the supplementary tables, and we have deleted the Supplementary table 1.

#### VERSION 2 – REVIEW

<b>REVIEWER</b>	Xingshun Qi General Hospital of Northern Theater Command, China
<b>REVIEW RETURNED</b>	20-Dec-2019

<b>GENERAL COMMENTS</b>	The reviewer completed the checklist but made no further comments.
-------------------------	--

<b>REVIEWER</b>	Nicola Armstrong Murdoch University Australia
<b>REVIEW RETURNED</b>	05-Jan-2020



<b>GENERAL COMMENTS</b>	<p>The vast majority of my previous comments have been satisfactorily answered by the authors. I have only two additional minor comments - RR and OR are used but undefined as being the relative risk and odds ratio in the methods. OR is also not mentioned in the methods section - this needs to be rectified. When is the RR used, and when is the OR used?</p> <p>The article should again be check for fluidity and English grammar. For instance, the addition of the Mantel-Haenszel sentence in the methods means that the following sentences referring to I<sup>2</sup> as heterogeneity need to be adjusted slightly. A p&lt;0.05 is (or was) deemed significant, not significantly. There are still several other places where the tense is wrong and needs correcting before publication.</p>
-------------------------	---

### VERSION 2 – AUTHOR RESPONSE

Response to Reviewer 3 Comments:

The vast majority of my previous comments have been satisfactorily answered by the authors. I have only two additional minor comments.

Response: Thanks for the Reviewer's suggestions, which are very helpful for revising and improving our paper.

RR and OR are used but undefined as being the relative risk and odds ratio in the methods. OR is also not mentioned in the methods section - this needs to be rectified. When is the RR used, and when is the OR used?

Response: Thanks greatly for the Reviewer's suggestion. First, only RR should be used in this manuscript. The OR was displayed because of our typographical errors. We are very sorry for this mistake, which has caused your confusion. We have revised this mistake, and have carefully checked this manuscript to avoid similar errors. Second, in the Statistical analysis section, we have mentioned that "The relationship between the ABO blood groups and HBV infection was quantified using RR values and the corresponding 95% confidence intervals (CIs)." Finally, thanks again for pointing out this mistake.

The article should again be check for fluidity and English grammar. For instance, the addition of the Mantel-Haenszel sentence in the methods means that the following sentences referring to I<sup>2</sup> as heterogeneity need to be adjusted slightly. A p<0.05 is (or was) deemed significant, not significantly. There are still several other places where the tense is wrong and needs correcting before publication.

Response: Thanks for the Reviewer's suggestion. First, the article has been checked again for fluidity and English grammar. Second, we have reversed the sentences referring to I<sup>2</sup> as "Between-study heterogeneity was evaluated with the I<sup>2</sup> statistic." Third, the relevant sentence was revised as "A p < 0.05 was deemed significant".