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

TITLE (PROVISIONAL)	White matter hyperintensities and their subtypes in patients with carotid artery stenosis: A Systematic Review and Meta-Analysis
AUTHORS	Ye, Huirong; Wang, Yujie; Qiu, Jianting; Wu, Qing; Xu, Mengmeng; Wang, Jian

VERSION 1 – REVIEW

REVIEWER	Peter Watson University of Cambridge UK
REVIEW RETURNED	30-Nov-2017

GENERAL COMMENTS	Carotid artery stenosis correlates with white-matter hyperintensities: a systematic review and meta-analysis bmjppen-2017-20830 The meta-analyses presented use the standard approaches with inverse variance, the addition of an additional random term (presumably using the (DerSimonian-Laird method) for handling heterogeneous effects and the testing of heterogeneity using I and Q statistics. On page 12 the authors conclude there is no publication bias so no need for an adjustment. I also like the explanation in the discussion of what factors have led to heterogeneity of variance amongst the studies. This is important as it suggests the additional influences which explain variation across the studies in the meta-analyses. You could reference the criteria used for I^2 statistic on line 23 of page 9 which are used, for example, in Higgins et al. (2003) giving heterogeneity thresholds for I^2 of 25% (low), 50% (heterogeneity) and 75% (high heterogeneity). I believe Eggers test (line 37 on page 9) is a test of funnel plot assymetry. I don't, however, see in the analysis section on page 9 or in the results section any reference or graphs of such funnel plots which are usually used to assess publication bias graphically. A visual display can be informative although I admit for just 8 articles (page 10 line 3) any test of publication bias might be underpowered which might be worth acknowledging in the discussion. (In passing I note that Peters et al. (2006) criticise the usual publication bias approach of Egger et al. (1997) (details here) who proposed a test for asymmetry of the funnel plot. In particular Peters et al. (2006) say the p-values of Egger et al.'s test are not as reliable as theirs because they have inflated type I errors). I also wonder about the sizes of the SMDs found in the results on

	page 12. The SMDs of 0.326 (page 12, line 12) and 0.412 (line 30) are small according to Cohen's (d) rules of thumb where anything below 0.50 is regarded as small. I wondered if the smallness of these effects could be acknowledged in the discussion rather than focusing on statistical significance since anything can be significant with a large enough sample size and the clinical relevance of the effects is more important. Is there a particular procedure in STATA 12 (page 9, line 13) that was used for the meta-analysis? References Higgins JP, Thompson SG, Deeks JJ and Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327 557-560. Peters, J. L., Sutton, A. J., Jones, D. R. and Abrams K. R. (2010). Assessing publication bias in meta-analyses in the presence of between-study heterogeneity. Journal of the Royal Statistical Society A 173(3) 575-591.

REVIEWER	Dr Laura Bonnett University of Liverpool, UK
REVIEW RETURNED	12-Dec-2017

GENERAL COMMENTS	This is a well-written systematic review and meta-analysis. However, I do have some concerns, particularly regarding the pooling of correlation coefficients in the meta-analysis and the deductions made based on the significance of the pooled correlation coefficients. This article would also benefit from improvements to the quality of the written English. Other comments are as follows: 1. Within the abstract, the outcome measure is listed as "CA stenosis is correlated with WMH and further with the subtypes, periventricular and deep." However, this is not an outcome measure, but a hypothesis instead. Either change the description of this section, or modify the text to truly be a description of the outcome measures. 2. Based on the abstract I believe that the numerical estimate which is being pooled is the correlation coefficient. As the authors will know, this ranges from 0 to 1 with 0 suggesting no correlation and 1 suggesting full correlation. A pooled correlation coefficient of 0.326 suggests weak correlation and the p-value is of limited value. Even within the subgroup analysis the pooled correlation coefficient only demonstrates moderate correlation. Therefore conclusions relating to the this pooled correlation coefficient should be appropriately described. 3. The abstract concludes that the results of this systematic review and meta-analysis will be of great help in elucidating risk factors etc. However, correlation and linear regression are quite different concepts - correlation relates to the strength of a relationship between variables while regression describes the relationship between the variables. Therefore, this conclusion should be appropriately reworded as the implication currently is that the results of this pooled correlation coefficient will inform linear regression results. 4. Within the Inclusion Criteria section on page 7, how are "definite

	results" on line 16 specifically defined? 5. Please justify the choice of including only English language studies. 6. Please justify the inclusion criterion of studies with at least 25 subjects. 7. I am concerned that requiring a minimum score of 6 for studies to be included in the analysis introduces bias to the results. Please justify this choice. 8. Within the Data Extraction section the authors discuss pooling mean values and standard deviations of WHM scores. This does not mention correlation described in the abstract. 9. Please be explicit with the model that be used when the random-effects model is preferred - the DerSimonian & Laird model I assume? 10. As in point 8, please clarify the scores pooled within the Meta-Analysis section. This reads as correlation scores to me. If this is the case, see my previous comment (point 2) regarding interpreting pooled correlation coefficients. 11. Correlation coefficients can only take values between 0 and 1. However, the pooled result in the Subgroup Analysis section has a value of 1.100. Please check this. 12. The first sentence of "Sensitivity analyses and publication bias" is unclear. Please re-write this. 13. Please explain why the I-squared value is 0% overall but 77% for subgroup analysis. This is a surprisingly large difference. 14. Figure 1 is incorrect. If there are only 219 records after duplicates have been removed, then 979 titles and abstracts cannot have been screened. Also, what do "summarize" and "no results" mean, and "summary of the meeting"? Additionally, why was "dichotomous data" excluded? 15. Please ensure that I-squared is appropriately described in the figure legends - I2 is incorrect; I-squared is better. 16. Please complete the PRISMA checklist as there is a blank next to number 5.

VERSION 1 – AUTHOR RESPONSE

Responses to the comments and requests from the editorial team.

We thank the editorial team for their positive and constructive comments and suggestions. Here are our responses

1. Question:Please revise your title. We ask authors to refrain from using declarative titles (i.e those that state the study's main findings). Please revise your title so that it frames the research question and includes the research design and setting.

Response: Revised title: “White matter hyperintensities and its subtypes in patients with carotid artery stenosis: A Systematic Review and Meta-Analysis”.

2. Question:The Abstract needs re-writing. For example, under Outcome measures you’ve written your findings. Under Conclusion you need to stick to the findings of this review, and not go beyond it (the review for example is not about risk factors).

Response: You are right. We have re-written our abstract (Page 36-37).

3. Question:Whilst you have cited and discussed previous systematic reviews and meta-analyses on the topic, please explain in more detail how this study is adding to the findings of those published studies. In the introduction you should go beyond saying that primary studies have been “conflicting”. Please explain how studies have been conflicting and how this systematic review and meta-analysis will help to resolve these inconsistencies.

Response: We have explained how this study is adding to the findings of those published studies, how present studies have been conflicting and how this systematic review and meta-analysis will help to resolve these inconsistencies in the text (Page 38, Line 34-36; Page 39, Line 13-29).

4. Question:The quality of English is not at the requisite standard for publication in places. Please thoroughly copy-edit the paper. We recommend consulting a native English speaker (if possible).

Response: We have tried our best to copy-edit the paper thoroughly and have consulted an international student who have translation certificate.

5. Question:Please revise the 'strengths and limitations' section on page e.g. the final point is a study finding not a strength or limitation. As a reminder, this section should contain up to 5 short bullet points, no longer than a single sentence each, that relate to the design or methods of the study reported.

Response: We have revised the 'strengths and limitations' section related to the design and methods of the study reported (Page 37, Line 49-52; Page 38, Line 3-8).

6. Question: Reporting of the methodological quality of the studies needs to be in the main body of the paper (not in the supplement).

Response: We have added the methodological quality of the studies (e.g. table 1) to the main body of the paper (Page 44).

Responses to the reviewers' comments
Responses to Reviewer: 1

1. Question:The meta-analyses presented use the standard approaches with inverse variance, the addition of an additional random term (presumably using the (DerSimonian-Laird method) for handling heterogeneous effects and the testing of heterogeneity using I and Q statistics. On page 12 the authors conclude there is no publication bias so no need for an adjustment. I also like the explanation in the discussion of what factors have led to heterogeneity of variance amongst the studies. This is important as it suggests the additional influences which explain variation across the studies in the meta-analyses.

Response: Thank you for your favorable assessment of our manuscript.

2.Question:You could reference the criteria used for I^2 statistic on line 23 of page 9 which are used, for example, in Higgins et al. (2003) giving heterogeneity thresholds for I^2 of 25% (low), 50% (heterogeneity) and 75% (high heterogeneity).

Response: We have read the papers you recommended (Higgins et al. [2003]) and revised the descriptions: “Heterogeneity is assessed by χ2 and I2 values. If the P-value of χ2 is <0.1, homogeneity is rejected. For the I2 statistic, 25%, 50%, and 75% are the thresholds for low, moderate, and high heterogeneity.” (Page 42, line 11-16) Accordingly, the study of Higgins et al was cited.

3. Question:I believe Eggers test (line 37 on page 9) is a test of funnel plot assymetry. I don't, however, see in the analysis section on page 9 or in the results section any reference or graphs of such funnel plots which are usually used to assess publication bias graphically. A visual display can be informative although I admit for just 8 articles (page 10 line 3) any test of publication bias might be underpowered which might be worth acknowledging in the discussion.
(In passing I note that Peters et al. (2006) criticise the usual publication bias approach of Egger et al. (1997) (details here) who proposed a test for asymmetry of the funnel plot. In particular Peters et al. (2006) say the p-values of Egger et al.'s test are not as reliable as theirs because they have inflated type I errors).

Response: We agree that a visual display is informative. We have added a funnel plot in the part “Sensitivity analyses and publication bias” (Page 30) and acknowledged the limitations: “Fifthly, publication biases were known to be underpowered when there were only eight studies in our meta-analysis.” (page 48, line 31-36) Accordingly, the study of Peters et al was cited.

4. Question:I also wonder about the sizes of the SMDs found in the results on page 12. The SMDs of 0.326 (page 12, line 12) and 0.412 (line 30) are small according to Cohen's (d) rules of thumb where anything below 0.50 is regarded as small. I wondered if the smallness of these effects could be acknowledged in the discussion rather than focusing on statistical significance since anything can be significant with a large enough sample size and the clinical relevance of the effects is more important.

Response: We have acknowledged and discussed the smallness of the SMD of 0.326 in the text. The smallness of the SMD of 0.412 had similar explanation (page 47, line 6-21).

5.Question: Is there a particular procedure in STATA 12 (page 9, line 13) that was used for the meta-analysis?

Response: No, there is no particular procedure in STATA 12 that was used for our meta-analysis.

Responses to Reviewer: 2

Thank you for your positive and constructive comments and suggestions. We have tried our best to copy-edit the paper thoroughly and have consulted two international students who have translation certificate. Your concerns about “the pooling of correlation coefficients in the meta-analysis and the deductions made based on the significance of the pooled correlation coefficients” have been explained in detail below. Here are our responses.

1. Question:Within the abstract, the outcome measure is listed as "CA stenosis is correlated with WMH and further with the subtypes, periventricular and deep." However, this is not an outcome

measure, but a hypothesis instead. Either change the description of this section, or modify the text to truly be a description of the outcome measures.

Response: We have changed the description of this section on outcome measure (Page 37, Line 3-13).

2. Question: Based on the abstract I believe that the numerical estimate which is being pooled is the correlation coefficient. As the authors will know, this ranges from 0 to 1 with 0 suggesting no correlation and 1 suggesting full correlation. A pooled correlation coefficient of 0.326 suggests weak correlation and the p-value is of limited value. Even within the subgroup analysis the pooled correlation coefficient only demonstrates moderate correlation. Therefore conclusions relating to the this pooled correlation coefficient should be appropriately described.

Response: There are two summary statistics used for meta-analysis of continuous data, the mean difference (MD) and the standardised mean difference (SMD) (effect measures for continuous outcomes). The numerical estimate which is being pooled is not the correlation coefficient. Selection of summary statistics for continuous data is principally determined by whether the outcome uses the same scale (when the MD is used) or uses different scales (when the SMD is used).
SMD=difference in mean outcome between groups/standard deviation of outcome among participants.
SMDs are small according to Cohen's (d) rules of thumb where anything below 0.50 is regarded as small. [1] We have acknowledged and discussed the smallness of the SMD of 0.326 in the text. The smallness of the SMD of 0.412 had similar explanation.

[1]. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions [Internet]. 2009. http://www.cochrane-handbook.org/

3. Question:The abstract concludes that the results of this systematic review and meta-analysis will be of great help in elucidating risk factors etc. However, correlation and linear regression are quite different concepts - correlation relates to the strength of a relationship between variables while regression describes the relationship between the variables. Therefore, this conclusion should be appropriately reworded as the implication currently is that the results of this pooled correlation coefficient will inform linear regression results.

Response: We have reworded our conclusion (Page 37, Line 36-39).

4. Question:Within the Inclusion Criteria section on page 7, how are "definite results" on line 16 specifically defined?

Response: "Definite results" means there was a quantitative or semiquantitative assessment of WMH imaging in patients with CA stenosis. We have reworded this to make it clear in the text (Page 40, Line 13-18).

5. Question:Please justify the choice of including only English language studies.

Response: It is found that the exclusion of trials reported in a language other than English does not significantly affect the results of the meta-analyses. [2] Thus, we did not included studies in a language other than English. We have described this in the limitations (page 48, line 6).

[2]. Moher D, Pham B, Lawson ML, Klassen TP. The inclusion of reports of randomised trials published in languages other than English in systematic reviews. Health Technology Assessment 2003; 7: 1–90.

6. Question:Please justify the inclusion criterion of studies with at least 25 subjects.

Response: Studies of small sample size are insufficient in statistical power. However we did not find out any relevant standard regarding how many samples are required for meta-analysis. We cancelled this criterion for inclusion. The sample size of our included studies was at least 29 subjects. Thus the results were not affected.

7. Question:I am concerned that requiring a minimum score of 6 for studies to be included in the analysis introduces bias to the results. Please justify this choice.

Response: A score over 6 of the Newcastle-Ottawa Scale (NOS) for studies is of high quality.[3] NOS of our included studies ranged from 6 to 8 (Page 43, Line 6). So it is not necessary to set “a minimum score of 6 for studies”.

[3] Lin Q, Li Z, Wei R, et al. Increased Risk of Post-Thrombolysis Intracranial Hemorrhage in Acute Ischemic Stroke Patients with Leukoaraiosis: A Meta-Analysis. PLoS One 2016; 11:e0153486.

8. Question:Within the Data Extraction section the authors discuss pooling mean values and standard deviations of WHM scores. This does not mention correlation described in the abstract.

Response: There are two summary statistics used for meta-analysis of continuous data, the mean difference (MD) and the standardised mean difference (SMD) (effect measures for continuous outcomes). The numerical estimate which is being pooled is not the correlation coefficient.

9. Question:Please be explicit with the model that be used when the random-effects model is preferred - the DerSimonian & Laird model I assume?

Response: We are explicit DerSimonian & Laird model was used with the random-effects model.

10. Question: As in point 8, please clarify the scores pooled within the Meta-Analysis section. This reads as correlation scores to me. If this is the case, see my previous comment (point 2) regarding interpreting pooled correlation coefficients.

Response: Please refer to the response to point 8.

11. Question:Correlation coefficients can only take values between 0 and 1. However, the pooled result in the Subgroup Analysis section has a value of 1.100. Please check this.

Response: The value of 1.100 is the upper limit of the confidence interval not correlation coefficient.

12. Question:The first sentence of "Sensitivity analyses and publication bias" is unclear. Please re-write this.

Response: We reworded the descriptions: “In the sensitivity analysis, we subsequently omitted each individual study to recalculate the SMDs. The re-evaluated SMDs had no obvious fluctuation (Figure 5) (Page 46, Line 41-44).”

13. Question:Please explain why the I-squared value is 0% overall but 77% for subgroup analysis. This is a surprisingly large difference.

Response: The primary reasons for the heterogeneity were the difference in the pathogenic mechanisms of periventricular and deep WMH, and smallness of the sample size.

14. Question:Figure 1 is incorrect. If there are only 219 records after duplicates have been removed, then 979 titles and abstracts cannot have been screened. Also, what do "summarize" and "no results" mean, and "summary of the meeting"? Additionally, why was "dichotomous data" excluded?

Response: Our detailed search identified 1198 studies through PubMed, EMBASE and Cochrane Library. After excluding 219 duplicate records, 979 abstracts were identified. “Summarize” we means review, “no results” means there were no results relevant to our studies; “the summary of the meeting” means only conference abstract can be obtained. We have reworded our Figure 1(Page 25).
We planned to include studies in which WMH were assessed by MRI and evaluated quantitatively or semiquantitatively. The results were continuous data. So, dichotomous data was excluded.

15.Question: Please ensure that I-squared is appropriately described in the figure legends - I2 is incorrect; I-squared is better.

Response: What you recommended has been revised in the text (Page 56, Line 6、24、42).

16.Question: Please complete the PRISMA checklist as there is a blank next to number 5.

Response: A blank next to number 5 in the PRISMA checklist has been added (Page 32, Line21).

VERSION 2 – REVIEW

REVIEWER	Dr Laura Bonnett University of Liverpool, UK
REVIEW RETURNED	23-Jan-2018

GENERAL COMMENTS	This is a much improved manuscript performing a systematic review and meta-analysis of white matter hyperintensities and its subtypes in patients with carotid artery stenosis. The statistical aspects of the manuscript are now appropriate and well-described. The quality of written English could be improved further but I appreciate that it is much improved on the initial submission. Very minor comments are as follows: 1. Within the Methods section, the authors say "Search terms were combination of subjective and random words". Random words is not a phrase usually associated with search terms. What do the authors mean? Equivalent words perhaps? The same phrase appears in the next but one sentence too. 2. Within the Statistical Analysis section, please mention the model to be used to fit a random-effects model - i.e. the DerSimonian & Laird model. Also, be aware that the latest recommendations suggest that the model (fixed effects or random effects) should be decided upon based on the sampling frame rather than the I-squared statistic. 3. Table 1 - A better phrase for the filled in stars would be "Score 1 point" rather than "Get one score". 4. Statistical language usually prefers to say something is not rather than it is. For example in a hypothesis test, we would never conclude that the null hypothesis was true. Instead we would say that there was no evidence to reject the null hypothesis. The same applies for conclusions about I-squared statistics. Within the WMH

	and CA stenosis subsection I would recommend that instead of saying "the I-squared statistic showed homogeneity" that the authors instead said "there was no evidence of heterogeneity" and then quote the values as in the manuscript already. 5. Please add a date of access to reference 16.

VERSION 2 – AUTHOR RESPONSE

Responses to the comments and requests from the editorial team.

We thank the editorial team for their positive and constructive comments and suggestions. Please find are our responses below.

1. Question: Unfortunately, the quality of English still needs improving in places e.g. “We aimed to perform a systematic review and meta-analysis to elucidate the association between white matter hyperintensities (WMH) and carotid artery (CA) stenosis that is controversial.” You do not need to state that the association is controversial here. Please amend to: “We aimed to perform a systematic review and meta-analysis to elucidate the association between white matter hyperintensities (WMH) and carotid artery (CA) stenosis.”
The following also needs revising to improve the quality of English (NB: this is not an exhaustive list):
“But the etiology and mechanism are not all clear” (introduction section)
“Fifthly, publication biases were known to be underpowered when there were only eight studies in our meta-analysis” (page 15)
We strongly recommend that you consult a native English speaker/ professional copy-editing service.

Response:
We have consulted a native English speaker and revised our manuscript to improve the quality of English. The changes to our manuscript are highlighted in green.

2. Question: Regarding your abstract: Please ensure that you are reporting all information recommended in the PRISMA extension for abstracts (see: http://www.prisma-statement.org/documents/PRISMA%20Abstracts%20Checklist.pdf). For example, the abstract currently does not provide the full search dates and it does not provide the methods of assessing the risk of bias/ quality of included studies.
Is it possible to briefly elaborate on the conclusion in the abstract? For example, are there any implications of these findings for other researchers or physicians?

Response:
According to the PRISMA extension for abstracts, we have provided the full search dates and methods of assessing the quality of included studies. (Page 36, line 50; Page 37, line 12) Our findings suggested that WMH may be considered an individual risk stratification score when choosing a proper plan for therapy of CA stenosis. The changes to our manuscript have been highlighted in green. (Page 37, line 42-44)

3. Question: Strengths and limitations section (pages 4-5): it needs to be clearer why the second point is a strength of your study. Likewise, your final bullet point is not very clear. What sample size are you referring to here?

Response:
Strengths and limitations section (pages 4-5):
The second point is strength of our study. Our analysis only included studies in which all WMH were assessed using MRI (not CT). Previous meta-analysis included studies in which WMH were assessed

using MRI or CT, which increased the heterogeneity of the results. This point was revised as follows: Our analysis only included studies in which all WMH were assessed using MRI, and the including criteria for the severity of CA stenosis was set at ≥50%.(Page 38, line 9-12)
Our final bullet point was revised as follows: Only 3 of the 8 included studies reported data on the association between subtype of WMH and CA stenosis and the SMD is small, which made the conclusion less persuasive. (Page 38, line 14-19)

4. Question: The introduction section still does not explain why the results of studies looking at the relationship between carotid artery (CA) stenosis and WMH are conflicting and how a systematic review and meta-analysis may help resolve these inconsistencies. Likewise, you have not explained why findings from previous meta-analyses are inconsistent or why the current study will help resolve these differences.

Response:
The introduction section was revised to explain why the previous results of studies on the relationship between CA stenosis and WMH are conflicting and how our study may help resolve these inconsistencies. The changes to our manuscript are highlighted in green. (Page 38, line 47-55; Page 39, line 3-32)

5. Question: Please ensure that your discussion section covers the following areas recommended in our instructions for authors: a statement of the principal findings; strengths and weaknesses of the study; strengths and weaknesses in relation to other studies, discussing important differences in results; the meaning of the study: possible explanations and implications for clinicians and policymakers; and unanswered questions and future research.

Response: The discussion section has been revised to cover the bullet points you recommended. (Page 47, line 6-55; Page48, line 4-17; Page48, line 55; Page49, line 4-37)

Responses to the reviewers' comments
Responses to Reviewer: 2
Thank you for your positive and constructive comments and suggestions. Here are our responses.
1. Question: Within the Methods section, the authors say "Search terms were combination of subjective and random words". Random words is not a phrase usually associated with search terms. What do the authors mean? Equivalent words perhaps? The same phrase appears in the next but one sentence too.

Response:
By “Random words,” we mean free-text terms. We have revised this term in the Methods section, to make it clearer. (Page 40, line 4)

2. Question: Within the Statistical Analysis section, please mention the model to be used to fit a random-effects model - i.e. the DerSimonian & Laird model. Also, be aware that the latest recommendations suggest that the model (fixed effects or random effects) should be decided upon based on the sampling frame rather than the I-squared statistic.

Response:
We have added the sentence “Otherwise, when significant heterogeneity among the studies was detected, a random-effects model (DerSimonian & Laird) was used” to the Statistical Analysis section. (Page 42, line14-17)

3. Question: Table 1 - A better phrase for the filled in stars would be "Score 1 point" rather than "Get one score".
Response: The legend of Table 1 has been revised as recommended. (Page44, line35)

4. Question: Statistical language usually prefers to say something is not rather than it is. For example in a hypothesis test, we would never conclude that the null hypothesis was true. Instead we would say that there was no evidence to reject the null hypothesis. The same applies for conclusions about I-squared statistics. Within the WMH and CA stenosis subsection I would recommend that instead of saying "the I-squared statistic showed homogeneity" that the authors instead said "there was no evidence of heterogeneity" and then quote the values as in the manuscript already.

Response: The Statistical analysis section has been revised as recommended. (Page 46, line 9; Page 46, line 32; Page 46, line 37)

5. Question: Please add a date of access to reference 16.

Response: The access date for reference 16 has been added. (Page 53, line 50)