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)	Effect of levothyroxine on the progression of carotid intima-media thickness in subclinical hypothyroidism patients: A meta-analysis
AUTHORS	Zhao, Tong; Chen, Baomin; Zhou, Yingying; Wang, Xinyi; Zhang, Yuanyuan; Wang, Haoyu; Shan, Zhongyan

VERSION 1 – REVIEW

REVIEWER	Marijana Tadic Charity Virchow Klinikum, Cardiology department, Berlin, Germany
REVIEW RETURNED	22-Feb-2017

GENERAL COMMENTS	The article is interesting, but several topics deserve to be commented. 1. You put strength and limitations at the very beginning of the article. I believe that this is not the best place for these bullets. 2. The Introduction section is too long. Please shorten it on 1 page and focus only on your topic. 3. You are discussion to many studies that are not related with your results. You have included 52 references, which is too many for one meta-analysis. 4. Please add the limitations at the end of the Discussion section. 5. Conclusion should be extended and written as separate section.

REVIEWER	Christian Selmer Copenhagen University Hospital Denmark
REVIEW RETURNED	28-Feb-2017

GENERAL COMMENTS	In this study by Zhao Tong et al. the effect of L-T4 replacement therapy on C-IMT in individuals with subclinical hypothyroidism (SCH) is evaluated by meta-analysis of existing published RCTs and self-controlled trials. The main finding is a significant reduction in C-IMT in SCH individuals < 65 yrs with a treatment-period longer than 6 months. The study subject is important as SCH is a common condition and C-IMT leading to arteriosclerotic diseases is potentially associated with high morbidity and mortality. The meta-analysis seems to be well-performed using the Cochrane principles and Strobe Guidelines. The introduction is good, the presentation of methods and results to the point and the discussion is through and reads well. The authors have pointed out relevant strengths and limitations in the study.

I only have minor comments:

1. Page 19 Line 28: The term "mixed gender" could be confusing to some readers.

2. Table 2: P-values = 0 are given for SBP, DBP, LDL, FMP and Lp(s). When reading the text I would expect it to be at least $p < 0.05$

3. Page 19 Line 35: "L–T4 therapy of SCH participants also resulted in a significant reduction in TC, TG, LDL, SBP, DBP, LP(a), and FMD." This sentence does not correspond to the results presented in Table 2 (see minor comment no. 2)

4. Fig 1: Has a number of typos: "articals", "relevent"

5. Fig 3: Legend for x-axis is missing (and should maybe also include an explanation e.g. "Negative values equals improvement in IMT")

6. Fig 2 and fig 3: The studies should be presents as in Fig. 4 with year of publication.

7. Fig 4: Legend for x-axis missing

8. Fig 5: "mix" should be "mixed genders" or similar

9. Fig 5: Legend for x-axis missing

10. Fig 6: "long" and "short" should be "long study duration" and "short study duration" in the figure

11. Fig S2: This figure is quite dark. Could it be presented in standard Stata graphics as Fig S3?

12. Table 1A: The symbol for micro should be corrected.

13. Could you include a discussion of whether the mean change in CIMT of -0.05 mm is expected to be clinical relevant i.e. if a large-scale RCT was performed on L-T4 replacement in SCH with hard endpoints (all-cause mortality, MACE etc.)

REVIEWER	John Ohrvik Karolinska Institutet, Sweden
REVIEW RETURNED	27-Mar-2017

| GENERAL COMMENTS | This is an interesting meta-analysis evaluating the effect of levothyroxine on the progression of carotid intima media thickness in patients with subclinical hypo-thyroidism. From a statistical point of view, I have the following comments:

1. There is no unique definition of weighted mean difference, the weighting could be done in different ways. Thus the authors have to explain how the weighting was performed.

2. The chi-square approximation of Cochran's Q statistics under the null hypothesis – no heterogeneity among the studies – is not accurate for small and moderate study sizes. |

	Kulinskaya et al (Biometrics 67, 203-212, March 2011) suggest a method based on fractional degrees of freedom (df) which substantially improve the original chi-square distribution with s-1 df, where s is the number of studies included in the meta analysis. 3. The random effect model used in case of significant heterogeneity has to be better explained. In general random effect models are not a cure for the difficulty to generalize the results of a meta analysis in case of heterogeneity. 4. In Table 1(A) the standard deviation (=0.99 mm) for CIMT at baseline in the Dursun study looks completely out of bounds to me, please check! You can see the effect in the forest plots (Figure 4-6) and Funnel plot (Figure S4). Further in the same Table I assume that the doses in the Monica and Ilknur study also are in ug/d. 5. In Table 2 the scales for all the metabolic and anthropometric variables as well as FMD are missing. Further a p-value can't be exactly 0, please change to e.g. < 0.001.

REVIEWER	Vasilios Pergialiotis National and Kapodistrian University of Athens, Greece
REVIEW RETURNED	10-May-2017

GENERAL COMMENTS	The statistical analysis is adequately presented and well described. A question i would ask is whether the authors believe that the level of carotid intima measurement actually may result in significant heterogeneity between studies (and; thus affect measurements) as well as the gender of patients. If so, the I2 test should be disregarded and the random effects model should be applied. Moreover, the authors should adequately describe the limitations of their study (small number of studies and patients, heterogeneity, difficult assessment of the potential publication bias mainly due to the previous two factors).

VERSION 1 – AUTHOR RESPONSE

Reviewer 1:

Reviewer Name: Marijana Tadic

1. Response to comment: We put strength and limitations at the very beginning of the article.
Response: We really want to change the place of strength and limitations, but according to the contribution regulations of BMJ Open, we have put these there.

2. Response to comment: The Introduction section is too long
Response: It is really true as Reviewer suggested, so we have deleted a lot of sentences that have nothing to do with the topic in the introduction section.

3. Response to comment: You are discussion to many studies that are not related with your results.
Response: We are sorry for the prolix writing. We have deleted a lot of sentences that have nothing to do with the topic in the discussion section. Through deletion, we have 32 references in this meta-analysis.

4. Response to comment: Please add the limitations at the end of the Discussion section.
Response: As Reviewer suggested, we have added the limitations at the end of the Discussion section.

5. Response to comment: Conclusion should be extended and written as separate section.
Response: We are very sorry for our negligence to the format of the conclusion. We have re-written this part according to the Reviewer's suggestion.

Reviewer: 2
Reviewer Name: Christian Selmer
1. Response to comment: The term "mixed gender" could be confusing to some readers.
Response: We are very sorry for using the confusing term, because we can not find the better term to describe the population of the sub-group. We have made a note to the term.

2. Response to comment: P-values = 0 are given for SBP, DBP, LDL, FMP and Lp(s).
Response: We are very sorry for using the incorrect writing of P-values. We have changed the P-values as p <0.001.

3. Response to comment: L–T4 therapy of SCH participants also resulted in a significant reduction in TC, TG, LDL, SBP, DBP, LP(a), and FMD." This sentence does not correspond to the results presented in Table 2 (see minor comment no. 2)
Response: We have changed the expression in the paper.

4. Response to comment: Fig 1: Has a number of typos: "articals", "relevent"
Response: We have changed the words in Fig 1.

5. Response to comment: Fig 3: Legend for x-axis is missing.
Response: We have added the legend for x-axis in Fig 3.

6. Response to comment: Fig 2 and fig 3: The studies should be presents as in Fig. 4 with year of publication.
Response: We have added the year of publication in Fig 2 and fig 3.

7. Response to comment: Fig 4: Legend for x-axis missing.
Response: We have added the legend for x-axis in Fig 4.

8. Response to comment: Fig 5: "mix" should be "mixed genders" or similar
Response: We have changed the expression in Fig 5.

9. Response to comment: Fig 5: Legend for x-axis missing
Response: We have added the legend for x-axis in Fig 5.

10. Response to comment: Fig 6: "long" and "short" should be "long study duration" and "short study duration" in the figure.
Response: We have changed "long study duration" and "short study duration" instead of "long" and "short" in Fig 6.

11. Response to comment: Fig S2: This figure is quite dark.
Response: We are very sorry for our negligence of that. We have adjusted the brightness of the picture.

12. Response to comment: Table 1A: The symbol for micro should be corrected.
Response: We have adjusted the symbol for micro in Table1A.

13. Response to comment: Could you include a discussion of whether the mean change in CIMT of -0.05 mm is expected to be clinical relevant.
Response: We are sorry that we have not found the exact change in CIMT of -0.05 mm to clinical relevant outcomes. But For a 0.1 mm difference in C-IMT, the prospective risk of myocardial infarction increased from 10% to 15%, and the stroke risk was increased from 13% to 18%. We have added that in the discussion section.

Reviewer: 3
Reviewer Name: John Ohrvik
1. Response to comment: the authors have to explain how the weighting was performed.
Response：As Reviewer suggested, we have explained the way of weighting to weighted mean difference(WMD) in the method section.

2. Response to comment: The chi-square approximation of Cochran's Q statistics is not accurate for small and moderate study sizes.
Response：We have read the paper carefully (Kulinskaya et al Biometrics 67, 203-212, March 2011) that the reviewer recommended. To meta-analysis of three RCTs, we have re-calculated according the method based on fractional degrees of freedom (df). The corrected P-value is also>0.05. We have selected a fixed-effect model. To meta-analysis of nine self-controlled experiments, because of the relatively high clinical heterogeneity, we have chosen a random-effect model. As a result of this change, the corresponding results have also changed. We have made the modify in the result.

3. Response to comment: Check the standard deviation (=0.99 mm) for CIMT at baseline in the Dursun study and the doses in the Monica and Ilknur study.
Response: We have checked the value of you mentioned. The standard deviation for CIMT at baseline in the Dursun study is actually 0.99 mm. That is obviously wrong. We are very sorry for making such a clear mistake. We have sand an email to Dursun for getting the correct value of CIMT. Fortunately, we have received the reply. The correct standard deviation for CIMT at baseline is 0.09 mm. As a result of this change, the corresponding results have also changed. We have made the modify in the result. We have also checked the doses in the Monica and Ilknur study. We are very sorry for our mistake. We have changed the units of L-T4 doses in the Monica and Ilknur study according to the original articles. We have also checked other values of variables in the original articles we included.

4. Response to comment: In Table 2 the scales for all the metabolic and anthropometric variables as well as FMD are missing. Further a p-value can't be exactly 0, please change to e.g. < 0.001.
Response: We are very sorry for our negligence of the scales for all the metabolic and anthropometric variables as well as FMD. We have added the corresponding scales in Table 2. We are very sorry for using the incorrect writing of P-values. We have changed the P-values as p <0.001.

Reviewer: 4
Reviewer Name: Vasilios Pergialiotis
1. Response to comment: whether the authors believe that the level of carotid intima measurement actually may result in significant heterogeneity between studies (and; thus affect measurements) as well as the gender of patients.

Response: To this comment, we really admit that the level of carotid intima measurement actually may result in clinical heterogeneity.

In this situation, we have disregarded the methodological heterogeneity and selected a random-effect model to the meta-analysis of eight self-controlled experiments. Whereas, to the meta-analysis of RCTs, we have also chosen a fixed-effect model. Because the levels of carotid intima measurement in three RCTs are same.

2. Response to comment: The authors should adequately describe the limitations of their study. Response: We are very sorry for not adequately describing the limitations of our study. According to your comments, we have added more details to describe the limitations. Especially, forces on small number of studies and patients, heterogeneity etc.

In all, we appreciate for Editors/Reviewers' warm work earnestly, and we found the reviewer's comments are quite helpful. We really hope that the correction will meet with approval. Once again, thank you very much for your comments and suggestions.

VERSION 2 – REVIEW

REVIEWER	Christian Selmer Rigshospitalet Department of Endocrinology Denmark
REVIEW RETURNED	12-Aug-2017

GENERAL COMMENTS	I have no further comments.

REVIEWER	John Ohrvik Karolinska Institutet, Sweden
REVIEW RETURNED	23-Aug-2017

GENERAL COMMENTS	The presentation has improved and statistical methods have been partly clarified in the revised manuscript. I have just two remaining comments: 1. The explanation of the weighting when applying a random effect model is still unclear. 2. The authors should add Kulinskaya et al to the reference list since they have applied their method to calculate the degrees of freedom of Cochran´s Q statistic. General comment: The written English needs improvement.

VERSION 2 – AUTHOR RESPONSE

To Reviewer: 3
Reviewer Name: John Ohrvik
Institution and Country: Karolinska Institutet, Sweden
1.The explanation of the weighting when applying a random effect model is still unclear.

Response: We are so sorry about the ambiguous expression. We have revised the explanation of the weighting when applying a random effect model in the text. We have highlighted it in the yellow color.

2.The authors should add Kulinskaya et al to the reference list since they have applied their method to calculate the degrees of freedom of Cochran´s Q statistic.

Response: We are sorry that we have omitted this paper to the reference list. We have added it to the reference list in the method section.
General comment: The written English needs improvement.

Response: We are so sorry that our written English is not good. We have tried our best to improve it. We have asked a native English speaking colleague to read our manuscript. She has modified some inappropriate expressions which are marked in yellow in the paper.
Thank you very much for your comments and suggestions again.

VERSION 3 – REVIEW

REVIEWER	John Ohrvik Uppsala University
REVIEW RETURNED	09-Sep-2017

GENERAL COMMENTS	The manuscript has improved and the comments and questions have been addressed in a satisfactory manner.