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)	The effect of nut consumption on markers of inflammation and endothelial function: a systematic review and meta-analysis of randomised controlled trials
AUTHORS	Neale, Elizabeth; Tapsell, Linda; Guan, Vivienne; Batterham, Marijka

VERSION 1 – REVIEW

REVIEWER	Giuseppe Grosso NNEdPro Global Centre for Nutrition and Health, St John's Innovation Centre, Cambridge, UK
REVIEW RETURNED	03-Apr-2017

GENERAL COMMENTS	The present article is a meta-analysis of RCTs on nut consumption and markers of inflammation. The study is well conducted, with proper methodology supporting the conclusions. One suggestion before approval, i can see from funnel plot that there is evidence of publication bias (and is also stated after test). I would like the authors to provide a sensitivity analysis by removing the paper(s) responsible for the publication bias, specify which are and giving potential reasons.

REVIEWER	Margaret Allman-Farinelli School of Life and Environmental Science Charles Perkins Centre University of Sydney AUSTRALIA 2006
REVIEW RETURNED	09-Apr-2017

GENERAL COMMENTS	This is a novel systematic review with several meta-analyses for determining the effects of nuts on biomarkers implicated in vascular disease (and inflammation). Overall the review is well conducted with extensive statistical analyses examining sensitivity and sub-groups. I have a few comments for consideration. Abstract Results: I think it should be clear that only 9 studies (not 32) contribute to the FMD meta-analyses and the sample size stated. Methods The last literature search was more than 12 months ago (i.e. January 2016). Obviously it takes a long time to conduct all the statistical analyses but I wonder if another search is indicated.

	It is stated that the GRADE system was used but this is not entirely clear here. Was Grade only applied with respect to quality of evidence and consistency because Grade encompasses generalisability of research results to the wider patient base - a variety of patient types were included in this review but number of participants is very small overall, in particular when individual biomarkers are considered. The limitations should address this as the number would be considered inadequate to reach a conclusion for FMD. While there is a statistically significant effect for FMD what does this equate to for the biological and clinical effects expected. The subgroup analyses show no significant effect for the studies that lasted more than three months nor for those with Type 2 diabetes. Some more discussion of FMD is needed here to guide practitioners. While RCTs and meta-analyses of this type provide highest level of evidence for treatments for dietary studies they only examine intermediates on the disease pathway and pooling of prospective cohorts yield better information on the relationship between nuts and cardiovascular disease and all cause mortality. Some consideration of this should be included in the limitations and discussion

REVIEWER	Malcolm Riley CSIRO Health and Biosecurity, Australia
REVIEW RETURNED	11-Apr-2017

GENERAL COMMENTS	This well-conducted systematic review and meta-analysis provides useful insight and guidance regarding its topic despite recently published reviews which partially cover the topic and which are cited by the authors. For inclusion in the review, the authors sought reports of randomised controlled trials, and prospective cohort studies relating to their chosen exposures and outcomes. No prospective cohort studies were found that met the search criteria (which should, of course, be reported) so the review material is entirely RCTs – therefore it is informative to note this in the title (eg '….:a systematic review and meta-analysis of RCTs'). The authors should justify their choice of outcomes of interest, and why an intervention duration of 3 weeks or more was chosen as an inclusion criterion. The authors have sub-grouped the Forest plots for each outcome into those studies for which an end of period value was provided for experimental and control groups, and those studies for which a change in value over the control or experimental period was provided. This would appear to provide a weighted mean difference for the final mean value between experimental and control, and a weighted mean difference for the change in value between experimental and control. These are then combined into an overall effect estimate – however aren't these different measures (i.e. not appropriate to be combined)?

	The caption for figure 2 states 'Change in FMD (%) between nut consumption and control (presented as sub-groups based on mean final or change values for readability) …' when it should be "Mean difference in final FMD (%) between nut consumption and control (group1) and mean difference in change in FMD (%) (group2)". Given effective randomisation, it might be assumed that these different effect measures would be close to the same thing – is this the assumption? The issue of potentially influential studies in the meta-analysis of CRP as an outcome is interesting. I checked the apparently aberrant study of Burns-Whitmore 2014 to find that mean final CRP for each dietary period is stated in ng/mL (2.36 for experimental, 1.95 for control) making the values used in the review correct. However, on checking Chiang YL 2012, their CRP values at the end of the dietary periods are also given in ng/mL (2.22 for experimental, 2.32 for control) – this is different to the result apparently used in the meta-analysis. This should be checked, of course. While it is possible to extract information from the large table of studies (Table 1) for each outcome, it is time intensive and would be much easier if this were summarised when the results for each outcome is reported. For example, while it is reported that only studies using walnuts found significant improvements in FMD (2 studies, I think), it is not directly stated that 5 of the 9 dietary comparisons were walnuts – all at a similar dose. The first sentence of the discussion appears to be incorrect – as quickly stated, the EFSA report was on walnuts, this review includes 9 dietary comparisons with FMD as an endpoint, most of which differed in walnut intake. I accept that this review is consistent with the conclusion of the EFSA report, and may have provided strong support had it been confined to walnuts. The issue of combining studies that used different nuts at different doses (stated very briefly as a limitation of this review in the discussion) might be expanded upon. There is presumably reason to group the interventions together, although compositional differences are noted for walnuts. The authors mention the need for appropriate dietary controls in their conclusions – is a recommendation able to be made about what an appropriate dietary control would be? The studies reviewed are striking in their diversity, with 'habitual diet' likely to be heavily influenced by the study population (although perhaps it is not a case of 'could be anything' as stated on p9). The authors warn to avoid study designs that increase total energy intake because they could 'skew results' – is there any information on which direction the outcome factors are changed by an increase in total energy intake? Minor issues: Many of the supplementary tables are incorrectly referenced in the text – there may have been a late change of table order – these should be carefully checked and corrected. The second last sentence of p22 mistakenly refers to CRP twice, the last sentence has an unnecessary text reference to Mazidi et al. Please include the units for 'effect estimate' in the sub-group analysis tables in the supplementary material.

REVIEWER	Alma Adler London School of Hygiene & Tropical Medicine, UK
REVIEW RETURNED	09-Jun-2017

GENERAL COMMENTS	Dear Authors, 1. You need to make it clear what your outcomes of interest are at the beginning. in the introduction you state markers of inflammation and enothelial function. you need to state exactly what markers you are interested in. Were these listed in your registered protocol? 2. Pubmed is an engine for searching medline (it also searches some other smaller databases) which makes it redundant to search both Pubmed and medline. (See https://www.nlm.nih.gov/pubs/factsheets/dif_med_pub.html for more information) 3. it is considered bad practice in systematic reviews in cases of multiple reports of the same study to discard additional studies as they may include important information. You may want to look at section 7.6.4 of the Cochrane handbook for more information 4. On page 7 you state that where studies reported median rather than mean std deviation was imputed from interquartial range. I assume this means that you combined medians and means. Did you test to see if the data were symmetrical or skewed? 5. I am unsure what you mean when you say that cross-over studies were treated in the same way as parallel studies. Can you please explain this further. 6. Did you consider differing dosages of nuts? You state that dosages ranged from 18 to 85 grams a day, which seems to be a very large range. 7. When describing studies, you mention that duration ranged from 4 weeks to 5 years. Obviously this is a huge range. I would like to see a bit more information on this, for example, what was the median duration or some sort of information on what percentage of studies were over a certain length? 8. You mention that you did a risk of bias assessment using the Cochrane tool, but do not show support for your assessments. Generally when conducting risk of bias, you should give your assessment and provide support, particularly when stating there was a low risk of bias.

REVIEWER	Darren Greenwood University of Leeds, UK.
REVIEW RETURNED	21-Jun-2017

GENERAL COMMENTS	1) There are many good things about how this review was conducted: a) A good search strategy. The authors have not restricted search terms to a specific study designs, but this does not affect sensitivity. b) Statistical methods following the Cochrane handbook recommendations seem appropriate. c) Study authors were contacted for extra information where this was missing. This will have been a lot of work and enhances the quality of the results. d) Data screening and extraction were completed by two authors. 2) There were one or two sentences in the statistical methods that I thought could be clearer: a) Page 7, line 43. “Chi-squared tests were used to explore the consistency of the weighted mean differences for each outcome.” These are not Chi-squared tests, these are some other tests that use the chi-squared distribution. Please correct. This also needs to state what characteristics were explored as sources of heterogeneity. b) Page 7, line 45. “I2 was calculated based on the formula: I2 = 100% × (Q - df)/Q (where Q refers to the chi-squared statistic, and df refers to the degrees of freedom)”. There is no need to quote the formula here, just cite the reference. But at the start of this sentence should be the description of what this is estimating, e.g. “The proportion of total variation attributable to between-study heterogeneity was estimated using the I-squared test statistic [reference].” And it is also important to present the *absolute* heterogeneity, i.e. the range of estimates, not just the proportion of variation in the outcome attributed to between-study heterogeneity. c) Page 7, line 52. Funnel plot asymmetry is not necessarily publication bias, and is better thought of in the broader sense of small-study effects, of which publication bias is the most likely example. Here it’s possible that the smaller studies had better measures of exposure or outcomes, so could potentially be better than the larger studies in this context, but leading to funnel plot asymmetry that way. 3) Page 9, line 9. My main potential concern is whether there were any cohorts mixed in with the RCTs. It is inappropriate for these two study types to be in the same meta-analysis, because observational studies are more prone to bias. My reading of the results is that no cohorts were found, but I would feel more comfortable if the authors would confirm this please, and maybe comment on why no cohorts were found. 4) I am surprised that RCT type (parallel vs crossover) was not one of the pre-defined subgroup analyses. Given the need for adequate washout with the crossover trials, I would have thought that this would be an important exploration of possible between-study heterogeneity. 5) Page 2, line 44, and elsewhere throughout the results. My main problem with the presentation of the results is that they are almost devoid of any units. It is impossible to interpret the results without the units.

	How do we know if these are big and important effects or small and unimportant, or small and more easily attributable to potential biases? We need the units. This starts with the main significant result in the abstract, but continues throughout the text. 6) Page 2, line 44. The authors present the estimate only for the statistically significant result. Instead, the authors should present the results for the primary exposures they stated in their protocol, regardless of statistical significance. Focus should be on clinical importance, i.e. estimates, not statistical significance.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Comment: The present article is a meta-analysis of RCTs on nut consumption and markers of inflammation. The study is well conducted, with proper methodology supporting the conclusions. One suggestion before approval, i can see from funnel plot that there is evidence of publication bias (and is also stated after test). I would like the authors to provide a sensitivity analysis by removing the paper(s) responsible for the publication bias, specify which are and giving potential reasons.

Response: Thank you for this feedback. We have conducted a sensitivity analysis removing three studies which appeared to be responsible for the asymmetry noted in the funnel plot for CRP (Parham et al., 2014, Sweazea et al., 2014; Kurlandsky and Stote, 2006) and Il-6 (Casas-Agustench et al., 2011). However after removal of these studies, there was still evidence of asymmetry (Egger's test: $p<0.05$). There thus appears to be evidence of asymmetry throughout the body of studies included. We have described this in the methods and results sections:
Methods: "Where funnel plot asymmetry was detected, sensitivity analyses were conducted to determine if removing studies eliminated the asymmetry."
Results: "Sensitivity analyses attempting to eliminate studies which appeared to be responsible for the small study effects did not alleviate the asymmetry found (data not shown)."

We have also commented on this in the limitations section: "Analysis of funnel plots suggested the results for CRP and IL-6 may have been influenced by small study effects (which could indicate publication bias), which resulted in downgrading the quality of the evidence for these outcomes. Funnel plot asymmetry remained after sensitivity analyses were conducted to remove the studies which appeared to be responsible for these effects."

Reviewer: 2

Comment: This is a novel systematic review with several meta-analyses for determining the effects of nuts on biomarkers implicated in vascular disease (and inflammation). Overall the review is well conducted with extensive statistical analyses examining sensitivity and sub-groups.

Response: Thank you

I have a few comments for consideration.
Abstract Results: I think it should be clear that only 9 studies (not 32) contribute to the FMD meta-analyses and the sample size stated.

Response: Thank you for your feedback. We have now amended the results presented in the abstract to read: “The effect of nut consumption on FMD was explored in n=9 strata from n=8 studies (involving n=652 participants), with consumption of nuts resulting in significant improvements in FMD (WMD: 0.79% [0.35, 1.23]).”

Methods
Comment: The last literature search was more than 12 months ago (i.e. January 2016). Obviously it takes a long time to conduct all the statistical analyses but I wonder if another search is indicated.

Response: We apologise again for this typographical error in the submitted manuscript. The final search was conducted in January 2017, however was incorrectly written as ‘2016’ in the submitted manuscript. We have now amended this.

Comment: It is stated that the GRADE system was used but this is not entirely clear here. Was Grade only applied with respect to quality of evidence and consistency because Grade encompasses generalisability of research results to the wider patient base - a variety of patient types were included in this review but number of participants is very small overall, in particular when individual biomarkers are considered. The limitations should address this as the number would be considered inadequate to reach a conclusion for FMD.

Response: Thank you for this comment regarding the use of GRADE. We applied all aspects of the GRADE system required to rate the quality of the evidence for each of our selected outcomes, namely study design, risk of bias, inconsistency, indirectness, imprecision, and other considerations such as publication bias. As part of the assessment of indirectness, we did consider the applicability of the research found to our pre-specified PICO components, and found no serious indirectness for all outcomes. As this manuscript involved a systematic review rather than guideline development, we did not determine the grade strength, in accordance with the guidance provided by Guyatt et al. (2011, Journal of Clinical Epidemiology). To clarify, however, we have amended in the methods section, to read:
“The quality of the body of evidence was then determined using GRADE32, which considers study design, risk of bias, inconsistency, indirectness, imprecision, and other considerations such as publication bias”.

To the point on generalizability, we acknowledge that the sample size for individual biomarkers was quite small and we have amended the limitations section to read:
“The size of the evidence base, including the small number of participants available for analyses of individual biomarkers, is a limitation, particularly with respect to generalisability and strength of the evidence.”
We have also addressed this issue specifically for FMD, adding the following statement to the discussion:
“This may have resulted from the small number of studies available for assessing FMD. Having few studies may have also played a role in the lack of significant effects observed in other FMD sub-group analyses. These include studies in participants with type 2 diabetes, or studies lasting longer than three months. Further research is therefore required in this area”

Comment: While there is a statistically significant effect for FMD what does this equate to for the biological and clinical effects expected. The subgroup analyses show no significant effect for the studies that lasted more than three months nor for those with Type 2 diabetes. Some more discussion of FMD is needed here to guide practitioners. While RCTs and meta-analyses of this type provide highest level of evidence for treatments for dietary studies they only examine intermediates on the disease pathway and pooling of prospective cohorts yield better information on the relationship between nuts and cardiovascular disease and all cause mortality. Some consideration of this should be included in the limitations and discussion

Response: Thank you for this feedback. Please see the comment above for our additional statements on FMD in the discussion section. We have further added:
"Despite the small sample size, the findings of this review relating to FMD are of value due to the associations between FMD and future cardiovascular events. A meta-analysis of cohort studies found a significant reduction in risk of cardiovascular events per 1% increase in FMD (RR: 0.872 [95% CI: 0.832 – 0.914])8. In comparison, the present study found an effect estimate of 0.79% for nut consumption compared to controls, suggesting these results are likely to be of clinical relevance to future cardiovascular risk."

We also acknowledge that while biomarkers can provide information on disease progression, they do not displace exploration of the effects of food consumption of disease endpoints, which is required for guiding practice and policy. As such, we have added the following information to the limitations section of the discussion:

"These biomarkers were selected to reflect changes in disease progression and amelioration, in order to explore mechanisms responsible for the favourable effects of nut consumption on cardiovascular disease10 and other chronic conditions11 12. However we fully acknowledge that the measures explored here are not interchangeable with disease endpoints such as mortality and morbidity."

Reviewer: 3
Comment: This well-conducted systematic review and meta-analysis provides useful insight and guidance regarding its topic despite recently published reviews which partially cover the topic and which are cited by the authors.

Response: Thank you

Comment: For inclusion in the review, the authors sought reports of randomised controlled trials, and prospective cohort studies relating to their chosen exposures and outcomes. No prospective cohort studies were found that met the search criteria (which should, of course, be reported) so the review material is entirely RCTs – therefore it is informative to note this in the title (eg '…:a systematic review and meta-analysis of RCTs').

Response: Thank you for your feedback. We have amended the title to read: "The effect of nut consumption on markers of inflammation and endothelial function: a systematic review and meta-analysis of randomised controlled trials".

In line with the issue raised here and also by Reviewer 5, we have clarified in the methods that no cohort studies met the eligibility criteria for inclusion in the review:
"Although prospective cohort study designs were also considered, no cohort studies met the overall inclusion criteria for the review. The most common reason was that the cohort studies did not report on the association between nut consumption and an outcome of interest".

Comment: The authors should justify their choice of outcomes of interest, and why an intervention duration of 3 weeks or more was chosen as an inclusion criterion.

Response: To clarify, we have added a justification of our choice of biomarkers and intervention duration in the methods:
"The outcomes of interest were selected to cover a suite of biomarkers regularly used in the literature to indicate changes to inflammation and endothelial dysfunction, including in previous meta-analyses exploring the effects of foods and dietary patterns27 28"
"This minimum duration was selected to ensure included studies reflected sustained changes to inflammation and endothelial function, and to align with similar cut-offs used in other meta-analyses exploring the impact of dietary components on inflammation27 or the effect of nut consumption on other physiological measures29 30"

Comment: The authors have sub-grouped the Forest plots for each outcome into those studies for which an end of period value was provided for experimental and control groups, and those studies for which a change in value over the control or experimental period was provided. This would appear to provide a weighted mean difference for the final mean value between experimental and control, and a weighted mean difference for the change in value between experimental and control. These are then combined into an overall effect estimate – however aren't these different measures (i.e. not appropriate to be combined)? The caption for figure 2 states 'Change in FMD (%) between nut consumption and control (presented as sub-groups based on mean final or change values for readability) …' when it should be "Mean difference in final FMD (%) between nut consumption and control (group1) and mean difference in change in FMD (%) (group2)". Given effective randomisation, it might be assumed that these different effect measures would be close to the same thing – is this the assumption?

Response: We have included both final and mean change values in the same meta-analysis as it is accepted by the Cochrane Handbook of for Systematic Reviews of Interventions (Higgins and Green, 2011, Section 9.4.5.2). This is due to the effect of randomisation, as the reviewer has highlighted. We have presented the data as recommended by the Handbook (with the change and final values in separate sub-groups in Figures 2 and 3, to avoid any confusion for readers), and have referred to the reasoning for including both final and change values in the methods as: "Mean changes in relevant outcomes were extracted where possible, and in the case that this data was not available, mean final values were retrieved as recommended by the Cochrane Handbook for Systematic Reviews of Interventions31".
To avoid confusion, we have amended the figure captions throughout to refer to 'difference' rather than 'change', for example: "Figure 2: Difference in FMD (%) between nut consumption and control (presented as sub-groups based on mean final or change values for readability). Diamond indicates weighted mean difference with 95% confidence intervals"

Comment: The issue of potentially influential studies in the meta-analysis of CRP as an outcome is interesting. I checked the apparently aberrant study of Burns-Whitmore 2014 to find that mean final CRP for each dietary period is stated in ng/mL (2.36 for experimental, 1.95 for control) making the values used in the review correct. However, on checking Chiang YL 2012, their CRP values at the end of the dietary periods are also given in ng/mL (2.22 for experimental, 2.32 for control) – this is different to the result apparently used in the meta-analysis. This should be checked, of course.

Response: Thank you for this comment. The reviewer is correct in that Chiang et al. (2012) report their CRP data in ng/mL. However, during the process of the review we contacted the study authors to confirm the units reported in the publication. They confirmed that the units were reported incorrectly in the published article, and should have been reported in mg/L.

We have used the correct values (mg/L) in the meta-analysis, however the reviewer's comment highlights that this process not clear in our submitted meta-analysis. As similar issues were also found with other articles in the review, we have now presented information on units in Table 1, and have indicated where units were corrected following correspondence with authors.

Comment: While it is possible to extract information from the large table of studies (Table 1) for each outcome, it is time intensive and would be much easier if this were summarised when the results for each outcome is reported. For example, while it is reported that only studies using walnuts found significant improvements in FMD (2 studies, I think), it is not directly stated that 5 of the 9 dietary comparisons were walnuts – all at a similar dose.

Response: In line with this feedback, we have added further detail to the results section, for example:
"A total of nine strata from eight studies14 17 18 22 39 53 58 60 explored the effect of nut consumption on FMD. Of the nine strata, five explored the effect of walnut consumption on FMD17 18 22 60, and six had a duration of less than three months14 17 18 22 53 58", and "A total of 26 strata from 25 studies13-16 18 19 21 34 35 39-41 43 46-51 53-57 59 explored the effect of nut consumption on CRP. Almonds were the most common nut type used in these analyses (seven strata21 40 47 53 55 57), followed by walnuts18 49 51 54 59 and mixtures of more than one nut type15 16 34 35 43 (each used in five strata). A total of 17 strata from 16 studies had a duration of less than three months14 15 18 21 34 35 40 41 46 48 51 53-55 57 59."
Further, we have clarified the description of the sub-group results for FMD to read:
"No significant differences were found for sub-group analyses (Supplementary material 4) although it was noted that when sub-group comparisons were made according to nut type, only the walnut sub-group found significant improvements in FMD."

Comment: The first sentence of the discussion appears to be incorrect – as quickly stated, the EFSA report was on walnuts, this review includes 9 dietary comparisons with FMD as an endpoint, most of which differed in walnut intake. I accept that this review is consistent with the conclusion of the EFSA report, and may have provided strong support had it been confined to walnuts.

Response: Thank you and we acknowledge that the original wording may be unclear. We have amended the first paragraph of the discussion to read:
"The results of this systematic review and meta-analysis suggested favourable effects of nut consumption on FMD, a measure of endothelial function. These findings align with a review conducted in 2011 by the European Food Safety Authority (EFSA), which explored the effects of walnut consumption on endothelium-dependent vasodilation 62".
To further clarify, we have also amended the discussion to read:
"Sub-group analyses found significant improvements in FMD only in those studies using walnuts, consistent with the EFSA report which only examined walnut consumption, although the test for sub-group differences in the present study did not reach statistical significance."

Comment: The issue of combining studies that used different nuts at different doses (stated very briefly as a limitation of this review in the discussion) might be expanded upon. There is presumably reason to group the interventions together, although compositional differences are noted for walnuts. The authors mention the need for appropriate dietary controls in their conclusions – is a recommendation able to be made about what an appropriate dietary control would be? The studies reviewed are striking in their diversity, with 'habitual diet' likely to be heavily influenced by the study population (although perhaps it is not a case of 'could be anything' as stated on p9).

Response: We did previously consider differences in nut type, as well as a range of other components, in our sub-group analyses (supplementary material 4). Following feedback from reviewers we have added two more post-hoc sub-group analyses (based on study design and nut dose). Significant sub-group differences were only found for CRP for the following:
- Comparison of studies which included the energy value of nuts versus those that did not
- Comparison of studies which incorporated >50 grams of nuts per day vs <50 grams/day
Borderline significant (p=0.05) sub-groups differences in CRP were also found based on study design (parallel versus cross-over).
We do acknowledge however that this may have been influenced by the small number of studies available – if there was a larger sample size it is possible additional significant sub-group differences may have been detected. We have added this information to the limitations section:

"Statistically significant sub-group differences were found only for CRP when studies were grouped according to whether they incorporated the energy value of nuts into the diet, and based on nut dose (<50 grams/day versus >50 grams/day). However due to the small number of studies, it is possible that other sub-group differences may have been found if the sample size was larger. For example, borderline significant differences (p=0.05) were found between the study designs, with larger reductions in CRP found for cross-over design studies".

We have also added further information to the conclusion to expand on our recommendations regarding appropriate control diets, and highlight the importance of dietary modelling: "These could include healthy dietary patterns (not including nuts), with a greater emphasis on dietary modelling required to ensure nutrient intakes are matched between control and intervention groups, minimising the risk of confounding."

In accordance with the above comment, we have also amended the statement in the methods regarding habitual diets, to read:
"Background diets consisted of either participant's habitual diet, or a prescribed diet aligned with healthy lifestyles such as the NCEP Step I or II diet, a Mediterranean-style diet, the Therapeutic Lifestyle Changes diet or another prudent style diet in line with dietary guidelines".

Comment: The authors warn to avoid study designs that increase total energy intake because they could 'skew results' – is there any information on which direction the outcome factors are changed by an increase in total energy intake?

Response: In line with this comment we have added the following information to the discussion:
"There is also evidence to suggest markers of inflammation such as CRP may be reduced following periods of energy restriction72, highlighting the importance of considering total energy intake when exploring the effects of individual foods."

Minor issues:
Many of the supplementary tables are incorrectly referenced in the text – there may have been a late change of table order – these should be carefully checked and corrected. The second last sentence of p22 mistakenly refers to CRP twice, the last sentence has an unnecessary text reference to Mazidi et al.

Response: Our apologies for these errors. We have updated the supplementary material to match the order presented in text.

We have amended the wording on page 22, and have removed the unnecessary reference to Mazidi et al. This text in the discussion now reads:
“Although including fewer studies than in our review, a recently published review by Mazidi et al.23 also found non-significant differences in inflammatory biomarkers (CRP, IL-6, adiponectin, ICAM-1, and VCAM-1), although in contrast to our review they observed a small increase in CRP levels. The review by Mazidi et al.23 appeared to have broader eligibility criteria which also included post-prandial studies and those exploring the effects of soy consumption.”

Comment: Please include the units for ‘effect estimate’ in the sub-group analysis tables in the supplementary material.

Response: In line with this comment we have added units for effect estimates to the sub-group analysis tables (supplementary material 4).

Reviewer: 4

Comment 1. You need to make it clear what your outcomes of interest are at the beginning. In the introduction you state markers of inflammation and enothelial function. You need to state exactly what markers you are interested in. Were these listed in your registered protocol?\

Response: Thank you for this feedback. While the outcomes of interest are currently listed in the first paragraph of the introduction, we have now amended the statement of aim to ensure specific outcomes of interest are listed:
“The aim of the review reported here was to examine the effect of nut consumption on markers of inflammation and endothelial function (CRP, adiponectin, TNF-α, IL-6, ICAM-1, VCAM-1, FMD) in adults.”
We can confirm that individual outcomes of interested were listed in our registered protocol on PROSPERO.

Comment 2. Pubmed is an engine for searching medline (it also searches some other smaller databases) which makes it redundant to search both Pubmed and medline. (See https://www.nlm.nih.gov/pubs/factsheets/dif_med_pub.html for more information)

Response: Thank you for this comment. While PubMed does encompass Medline, a recent paper by Rosen and Suhami (2016, BMC Medical Research Methodology) highlighted that differences between databases in the time for articles to appear leads to some articles not being detected if only Medline is searched. As a result, their recommendation is for systematic review authors to search both PubMed and Medline to ensure they obtain recent articles. We have added the following information to the methods to clarify this:
“In line with recommendations by Rosen and Suhami26 both Medline and PubMed were searched to ensure recent studies were detected”.

Comment 3. it is considered bad practice in systematic reviews in cases of multiple reports of the same study to discard additional studies as they may include important information. You may want to look at section 7.6.4 of the Cochrane handbook for more information

Response: We apologise for any apparent lack of clarity in this part of the methods. During the process of study selection, in the case of multiple articles from a single study, we checked all articles to ensure all relevant information was collected. In the case that multiple articles reported different information from the same study (as in the case of Lopez-Uriarte et al. 2010, and Casas-Agustench et al. 2011), both articles were included in the review. However, where the same outcomes were reported for a single study across multiple articles, we selected the article with the longest follow-up time to include in the meta-analysis, to avoid duplication of populations in the meta-analysis. This aligns with Australian regulatory guidelines for conducting systematic reviews (Food Standards Australia New Zealand, 2016), as well as the strategy used by a meta-analysis on a similar topic (Blanco Mejia et al., 2014, BMJ Open)

To further clarify we have added this information in the methods:
"In the case that results from one study were reported in multiple articles, all articles were checked to avoid duplication of study populations in the analysis or overlooking new information on outcomes. Where different information on outcomes were reported across articles, all relevant articles were included in line with the guidelines of the Cochrane Handbook31. Where the same outcomes from a single study were reported across multiple articles, decisions relating to article inclusion were based first on the length of follow-up for the outcome, and then by sample size."

Comment 4. On page 7 you state that where studies reported median rather than mean std deviation was imputed from interquartial range. I assume this means that you combined medians and means. Did you test to see if the data were symmetrical or skewed?

Response: Thank you. As we used published data in the meta-analysis, we unable to explore the distribution of the data. We did however make the assumption that if a median was presented, the data was likely to be skewed. The reviewer is correct in that where medians were provided, these were combined with means. We have now clarified this in the methods: "Where studies reported median rather than mean, medians were used in the meta-analysis, and standard deviation was imputed from interquartile range."
We have also discussed this as a limitation of the review: "Furthermore, although we were unable to explore the distribution of the published data included in this meta-analysis, the fact that several studies reported median values rather than means suggests some of the data may have been skewed, which may have impacted upon our analyses."

Comment 5. I am unsure what you mean when you say that cross-over studies were treated in the same way as parallel studies. Can you please explain this further.

Response: We apologise for any apparent confusion. We have added further detail to this section to clarify:
"In initial analyses, cross-over studies were treated in the same way as parallel studies by comparing measurements from the intervention periods with the control periods via a paired analysis, as the most conservative approach to managing cross-over studies31"

Comment 6. Did you consider differing dosages of nuts? You state that dosages ranged from 18 to 85 grams a day, which seems to be a very large range.

Response: Thank you. We have now added nut dose as one of the post-hoc sub-group analyses in the methods. Nut dose was grouped as <50 grams/day and >50 grams/day. This cut-off was selected to reflect the mid-point of doses used, and also to align with cut-offs used in published literature (Blanco Mejia et al., 2014, BMJ Open). Results have now been added to the tables in Supplementary material 4.

Significant differences based on nut dose were found only for CRP ($Chi^2 = 5.74$, df = 1 (P = 0.02), $I^2 = 82.6\%$), with larger reductions in CRP following nut consumption found for studies which included 50 grams or more nuts per day. We have now commented on these results in the discussion:
"Sub-group analyses found significant reductions in CRP when studies incorporated 50 grams or more of nuts per day. This finding aligns with previous research suggesting a dose-response effect of nut intake on other outcomes such as cholesterol70. However, these findings should be interpreted with caution, as several studies14 18 19 21 34 49 57 58 incorporated nuts as a proportion of total energy, resulting in substantial variation between individuals in the dose consumed."

Comment 7. When describing studies, you mention that duration ranged from 4 weeks to 5 years. Obviously this is a huge range. I would like to see a bit more information on this, for example, what was the median duration or some sort of information on what percentage of studies were over a certain length?

Response: In line with this comment, we have added further detail to the results:
"Amongst all studies, duration ranged from four weeks to five years, although 2014 15 17 18 21 22 34 35 40 41 46 48 51-55 57-59 out of 32 studies (63%) had a duration of less than three months."

Comment 8. You mention that you did a risk of bias assessment using the Cochrane tool, but do not show support for your assessments. Generally when conducting risk of bias, you should give your assessment and provide support, particularly when stating there was a low risk of bias.

Response: We apologise for not including this information in the initial submission. We have now presented support for the assessments as Supplementary material 9.

Reviewer: 5
Comment 1) There are many good things about how this review was conducted:
a) A good search strategy. The authors have not restricted search terms to a specific study designs, but this does not affect sensitivity.
b) Statistical methods following the Cochrane handbook recommendations seem appropriate.
c) Study authors were contacted for extra information where this was missing. This will have been a lot of work and enhances the quality of the results.
d) Data screening and extraction were completed by two authors.

Response: Thank you for this feedback

Comment 2) There were one or two sentences in the statistical methods that I thought could be clearer:
a) Page 7, line 43. "Chi-squared tests were used to explore the consistency of the weighted mean differences for each outcome." These are not Chi-squared tests, these are some other tests that use the chi-squared distribution. Please correct. This also needs to state what characteristics were explored as sources of heterogeneity.

Response: Thank you, we have now removed the sentence "Chi-squared tests were used to explore the consistency of the weighted mean differences for each outcome" from the methods. We have also reworded the methods to clarify the exploration of heterogeneity: "I2 values were generated for each analysis, including sub-group analyses (outlined below)."

Comment b) Page 7, line 45. “I2 was calculated based on the formula: I2 = 100% × (Q - df)/Q (where Q refers to the chi-squared statistic, and df refers to the degrees of freedom)”. There is no need to quote the formula here, just cite the reference. But at the start of this sentence should be the description of what this is estimating, e.g. “The proportion of total variation attributable to between-study heterogeneity was estimated using the I-squared test statistic [reference].” And it is also important to present the *absolute* heterogeneity, i.e. the range of estimates, not just the proportion of variation in the outcome attributed to between-study heterogeneity.

Response: In line with this comment, we have amended the methods to read: “The proportion of total variation attributable to between-study heterogeneity was estimated using the I2 test statistic32”. We have also presented the range of estimates for each outcome in Table 2, and Supplementary material 3.

Comment c) Page 7, line 52. Funnel plot asymmetry is not necessarily publication bias, and is better thought of in the broader sense of small-study effects, of which publication bias is the most likely example. Here it’s possible that the smaller studies had better measures of exposure or outcomes, so could potentially be better than the larger studies in this context, but leading to funnel plot asymmetry that way.

Response: In line with this comment, we have amended the manuscript to refer to small study effects, rather than publication bias alone:
Methods: “For outcomes with ten or more strata, funnel plots were generated to explore small study effects, with Egger’s test used to determine the extent of funnel plot asymmetry33”
Results: “Egger’s test indicated asymmetry in funnel plots for CRP (bias = -0.69 [95% CI = -1.07 to -0.31], P = 0.001) and IL-6 (bias = -0.80 [95% CI = -1.45 to -0.16], P = 0.02), suggesting the presence of small study effects which may have been attributable to publication bias.” Discussion: “Analysis of funnel plots suggested the results for CRP and IL-6 may have been influenced by small study effects (which could indicate publication bias).”

Comment 3) Page 9, line 9. My main potential concern is whether there were any cohorts mixed in with the RCTs. It is inappropriate for these two study types to be in the same meta-analysis, because observational studies are more prone to bias. My reading of the results is that no cohorts were found, but I would feel more comfortable if the authors would confirm this please, and maybe comment on why no cohorts were found.

Response: Thank you for this comment, it is correct that no cohort studies were found to be eligible for inclusion in this review. We acknowledge that this may not have been clear in the submitted manuscript, and therefore have added this information to the results section: “Although prospective cohort study designs were also considered, no cohort studies met the overall inclusion criteria for the review. The most common reason was that the cohort studies did not report on the association between nut consumption and an outcome of interest”

Comment 4) I am surprised that RCT type (parallel vs crossover) was not one of the pre-defined subgroup analyses. Given the need for adequate washout with the crossover trials, I would have thought that this would be an important exploration of possible between-study heterogeneity.

Response: We have now added RCT type as one of the post-hoc sub-group analyses in the methods, and results have now been added to the tables in Supplementary material 4.
Borderline significant (p=0.0500) differences based on study design were found only for CRP (Chi^2 = 0.58, df = 1 (P = 0.45), I^2 = 0%), with larger reductions in CRP following nut consumption found for cross-over design studies. We have now commented on these results in the discussion: “For example, borderline significant differences (p=0.05) were found between the study designs, with larger reductions in CRP found for cross-over design studies. As the nature of cross-over studies eliminates between-subject variation75, they may provide superior insights when exploring the impact of dietary interventions on biomarkers such as CRP, however their results may also be impacted by carry-over effects31. Given the short or absent wash-out periods of some of the included studies18 35 50 54 57, the potential impact of carry-over effects cannot be ruled out”

Comment 5) Page 2, line 44, and elsewhere throughout the results. My main problem with the presentation of the results is that they are almost devoid of any units. It is impossible to interpret the results without the units. How do we know if these are big and important effects or small and unimportant, or small and more easily attributable to potential biases? We need the units. This starts with the main significant result in the abstract, but continues throughout the text.

Response: We apologise for this oversight, we have now added the units to all results reported in the manuscript and supplementary materials.

Comment 6) Page 2, line 44. The authors present the estimate only for the statistically significant result. Instead, the authors should present the results for the primary exposures they stated in their protocol, regardless of statistical significance. Focus should be on clinical importance, i.e. estimates, not statistical significance.

Response: Thank you for this feedback. We appreciate the importance of including all results regardless of significance, however as this was in the abstract, the limited word count restricts the ability to present all results. We have added further detail to the abstract: “Nut consumption resulted in small, non-significant differences in CRP (WMD: -0.01mg/L [95% CI: -0.06, 0.03]), although sensitivity analyses suggest results for CRP may have been influenced by two individual studies. Small, non-significant differences were also found for other biomarkers of inflammation.”

VERSION 2 – REVIEW

REVIEWER	Margaret Allman-Farinelli University of Sydney Australia
REVIEW RETURNED	22-Aug-2017

GENERAL COMMENTS	Detailed responses and revision made in manuscript.

REVIEWER	Darren Greenwood University of Leeds, UK
REVIEW RETURNED	01-Aug-2017

GENERAL COMMENTS	The authors have addressed my concerns, but in revising their manuscript, one other issues has arisen. I was unhappy with the statistical aspects of the authors' response to reviewer #1's comments. Leaving aside the issue of whether funnel plot asymmetry can be ascribed to publication bias (point 2c in my earlier review) any asymmetry due to publication bias would be because studies are *missing*. So first off, removing more studies is unlikely to be helpful. Second, the point is not whether this removes the asymmetry (that's a rather circular argument, whether removing studies that lead to asymmetry actually remove the asymmetry. Instead, the point is whether the overall pooled estimate is sensitive to removing those studies. Third, if the funnel plot asymmetry is because of publication bias, then any sensitivity analysis should instead be adding in additional hypothetical studies (e.g. trim and fill method) to see if this changes the overall estimate. I'm not a fan of this approach, but in my opinion that would be a more useful sensitivity analysis. All other points have been adequately addressed.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Detailed responses and revision made in manuscript.
Thank you for this feedback

Reviewer: 5
Comment: The authors have addressed my concerns, but in revising their manuscript, one other issues has arisen. I was unhappy with the statistical aspects of the authors' response to reviewer #1's comments. Leaving aside the issue of whether funnel plot asymmetry can be ascribed to publication bias (point 2c in my earlier review) any asymmetry due to publication bias would be because studies are *missing*. So first off, removing more studies is unlikely to be helpful. Second, the point is not whether this removes the asymmetry (that's a rather circular argument, whether removing studies that lead to asymmetry actually remove the asymmetry. Instead, the point is whether the overall pooled estimate is sensitive to removing those studies. Third, if the funnel plot asymmetry is because of publication bias, then any sensitivity analysis should instead be adding in additional hypothetical studies (e.g. trim and fill method) to see if this changes the overall estimate. I'm not a fan of this approach, but in my opinion that would be a more useful sensitivity analysis.

All other points have been adequately addressed.

Response: Thank you for this feedback. In light of the reviewer's comments, we have removed the added sensitivity analysis from the amended version of the paper. In accordance with the reviewer's suggestion, we conducted sensitivity analyses using the trim-and-fill method to explore potential publication bias, however this did not affect the overall estimate. We have amended the manuscript to reflect this, and trust that this addresses the reviewer's concerns.
Methods: "Where funnel plot asymmetry was detected, sensitivity analyses using the trim-and-fill method were conducted to explore potential publication bias34. Egger's test and the trim-and-fill method were conducted using Stata (Stata Statistical Software [Computer program]. Release 15. College Station, TX: StataCorp LLC, 2017)."
Results: "Funnel plots were generated for outcomes with ten or more strata (CRP, IL-6, ICAM-1, and VCAM-1) (Supplementary material 7). Egger's test indicated asymmetry in funnel plots for CRP (bias = -0.68 [95% CI = -1.06 to -0.31], P = 0.001) and IL-6 (bias = -0.81 [95% CI = -1.45 to -0.16], P = 0.02), suggesting the presence of small study effects which may have been attributable to publication bias. Use of the trim-and-fill method did not change these results (data not shown)."
Discussion: "Funnel plot asymmetry remained after sensitivity analyses were conducted."

VERSION 3 – REVIEW

REVIEWER	Darren Greenwood University of Leeds, UK. None that I'm aware of.
REVIEW RETURNED	18-Oct-2017

GENERAL COMMENTS	The authors have addressed all my points.