PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Vitamin C may alleviate exercise-induced bronchoconstriction: a meta-analysis
AUTHORS	Hemilä, Harri

VERSION 1 - REVIEW

REVIEWER	Timothy Mickleborough, PhD Professor and Director of Graduate Studies School of Public Health Dept. Kinesiology Indiana University Bloomington Indiana USA
REVIEW RETURNED	18-Dec-2012

GENERAL COMMENTS	This is well-written succint meta-analysis of the efficacy of vitamin C on EIB. Unfortunately only 3 studies were included in the review, but none the less the data are convincing in terms of proof-of-concept. Since this review concerns the influence of vitamin C on EIB why is there a discussion on the influence of viatmin V on cold-like symptoms (p.13-14). My suggestion is to either delete this entire section or include these studies in the meta-analsysis and change the focus of the study.

REVIEWER	Wood, Lisa University of Newcastle, Respiratory Medicine
REVIEW RETURNED	23-Jan-2013

THE STUDY	I am not sure about the suitability of the data imputation.
RESULTS & CONCLUSIONS	Again, I am unsure about the imputation of the data.
GENERAL COMMENTS	1. Key messages – The first point should be reworded. There is no justification for this paper to encourage people to 'test vitamin C'. A more appropriate message would be 'Vitamin C may alleviate respiratory symptoms with exercise.' 2. Introduction – Some discussion of vitamin C as an antioxidant would be expected in the section on mechanisms. 3. Introduction – The author refers to 'severe errors' in the extraction of data of a previous Cochrane review. Can the author demonstrate that this is the case. This needs further explanation. 4. Methods – The following statement needs to be explained: 'The dose limit was set as a pragmatic choice.' 5. Methods – Why weren't 'vitamin C' and exercise-induced bronchoconstriction' included as search terms? 6. Statistical Methods – This reviewer must defer to a statistian to

	advise on the appropriateness of the imputation of data. Indeed, it does not appear to be very robust to impute ~25% of the data on which the analyses are based. 7. Discussion - The author claims that the imputation of data does not influence the conclusions, but the reality is that there are only 30 'real' data points, so it does reduce confidence in the data.

REVIEWER	Tim Pickles, Statistician, South East Wales Trials Unit, Cardiff University
REVIEW RETURNED	01-Feb-2013

GENERAL COMMENTS	Quick summary: This paper provides a good meta-analysis, and further analyses, on the topic of vitamin C in exercise induced bronchoconstriction. On the whole, this paper is very well written and the analysis is relevant and sound. The matter of missing data in the Cohen paper is dealt with sensibly and appropriate sensitivity analyses are presented. I do though have a few comments, corrections and questions I would like to highlight here, which I will do section by section. Abstract: • Last sentence: "Further research on the effects of vitamin C on EIB is warranted" Article Summary: • Key messages; point 2: I'm not sure you can say "simply". Also, and I will repeat this point later in the results section, I don't think you have any evidence to back this statement up. Introduction: • No comments Methods and Table 1: • Change "percentange" to "percentage" • Table 1: Please add a horizontal line between Intervention and Outcome for Cohen et al. 1997. Results: • Quote statistical heterogeneity figures here rather than in text describing Figure 1 • In the paragraph giving a summary of the effect of each trial, please provide some talk of significance and/or confidence intervals so we can tell if the effects are meaningful • You argue that a "… single fixed percentage point estimate of vitamin C effect may be simplistic." I don't know that the estimate from a linear regression is anything more than a single estimate. Also, I am unsure about the relevance of the word "fixed" here. Both the percentage point estimate, and that from linear regression, have a level of variation related to them • P = 0.054 is not "marginally significant". It is very much in that grey area of uncertainty that you get with P-values. I think "marginally not significant" would be more suitable here • So how is it "better"? Is it simply because it is now

	significant? Don't forget that you only have data on 12 individuals here. This hardly a large enough sample to be making distinct statistical statements about. I don't disagree that it is different, and valid, way of analysing these data but there needs to be more here to show me how it is better • For discussion of linear regression of the data from Schachter & Schlesinger and Cohen et al., could you quote values for R^2 and P in the text rather than in the description of figures (or not at all) • Report decimal points for upper bound of 95%CI in the first sensitivity analysis. Should be 12.3 • I miss any reference to this second sensitivity analyses in the tables. Could this analysis, and the first secondary analysis, be presented in the same format as figure 1 please. Discussion: • Again, I don't believe that you can make any statement about the linear regression being better • Low numbers are always a concern as it is highly unlikely that they are representative of any sample. The only advantage here is that the data is paired (from cross-over studies) and hence you have more than you would for non-cross-over studies. References: • No comments Figures: • Legends to Figures: remove all number here and put them in the results section • Figures 2 and 3: I cannot tell what the thin black line represents. Please explain better or remove the line. Supplementary Files: • Due to the excel to pdf transfer, some start and ends of lines are lost. Also, in the final table where you show the calculation of SE(z), the cells don't allow the reader to see all the text • Can you provide the data for all 20 of the Cohen patients of their Vitamin C day? • The Schachter & Schlesinger is presented poorly. Could you show it as before and after exercise, rather than making the reader decipher variable names to comprehend what is in the columns • For the sensitivity analysis using "no vitamin C effect" from the 9 missing Cohen patients, surely the SE is 3.24, not 3.03 (See: "The "no vitamin C effect" imputation leads to Mean = 11.2 and SE = 3.03 for the whole set of 20 participants") • PRISMA Checklist: • Can't say item 22 is covered on pages 9-10 when you state no page for items 12 or 15.

VERSION 1 – AUTHOR RESPONSE

Reviewer: Timothy Mickleborough, PhD
Professor and Director of Graduate Studies
School of Public Health
Dept. Kinesiology
Indiana University
Bloomington
Indiana USA

This is well-written succint meta-analysis of the efficacy of vitamin C on EIB. Unfortunately only 3 studies were included in the review, but none the less the data are convincing in terms of proof-of-concept.

Since this review concerns the influence of vitamin C on EIB why is there a discussion on the influence of viatmin V on cold-like symptoms (p.13-14). My suggestion is to either delete this entire section or include these studies in the meta-analsysis and change the focus of the study.

HH: I do not agree with the latter paragraph.
In the Discussion section, authors should discuss the context of the research findings and not just the findings of the current paper. Most vitamin C and common cold studies, and in particular all vitamin C-common cold studies that examined people under heavy short-term physical stress, have used a clinical definition of the cold such as the participant had cough (and/or other respiratory symptoms). However, cough is not specific to a virus infection, but for example EIB also causes cough. Therefore, in the trials of vitamin C for marathon runners, some of the coughs may have been caused by EIB. We do not need to trust the original authors' interpretation of “the common cold” when their diagnosis is based only on symptoms.
Furthermore, even if all the marathon runners of the vitamin C studies did have a virus infection, those studies are still relevant in parallel with the EIB studies. All 3 studies on EIB and all the mentioned 6 studies on the so called “common cold” indicate that vitamin C may be beneficial against respiratory symptoms of people under heavy acute physical stress - irrespective of the etiology of respiratory symptoms. I deleted one paragraph which substantially shortens the discussion of the common cold studies.

****

Reviewer: Lisa Wood
University of Newcastle, Respiratory Medicine

I am not sure about the suitability of the data imputation.

HH: First, if I would not do the imputations, I could not include the Cohen study to the meta-analysis. Cohen has explicit experimental FEV1 data for 11 out of the 20 participants and without imputations all Cohen's that experimental data should be discarded.
In systematic reviews, we are concerned with the possibility of publication bias, which means that some studies that have been carried out are not published. Therefore, we cannot include their experimental data in the systematic review. If we exclude the Cohen study, we generate a problem identical with publication bias. On the other hand, if we include the 11 published data points alone, we would generate bias, because the reported data were selected on the basis of positive result.
Thus, if we want to avoid the bias caused by the exclusion of the Cohen study, and the bias caused by including only the reported 11 positive data values, we must conservatively impute the missing values and thereby we can include the Cohen study to the meta-analysis. The reviewer does not suggest any other solution to avoid the two biases.

Second, I added a calculation of the effect of vitamin C on the proportion of participants suffering from EIB in the Cohen study. This approach does not need any imputations. However, the findings of this second approach are consistent with the imputed-data approach.

1. Key messages – The first point should be reworded. There is no justification for this paper to encourage people to 'test vitamin C'. A more appropriate message would be 'Vitamin C may alleviate respiratory symptoms with exercise.'

HH: Done

2. Introduction – Some discussion of vitamin C as an antioxidant would be expected in the section on mechanisms.

HH: One sentence and two references were added. There is much literature on oxidative stress caused by exercise and on the antioxidant role of vitamin C but it is not relevant to cover that literature in more detail here.

3. Introduction – The author refers to 'severe errors' in the extraction of data of a previous Cochrane review. Can the author demonstrate that this is the case. This needs further explanation.

HH: I have demonstrated the problems as a Feedback published within the Cochrane review (new ref. 18). My feedback (new ref. 19) is available at my home page:
http://www.mv.helsinki.fi/home/hemila/H35P.pdf
I cannot reiterate the details of the problems in the Introduction, since the Introduction would become too long for its purpose. If a reader wants to take a look at my demonstration of the severe errors, he or she should read ref. 19.

4. Methods – The following statement needs to be explained: 'The dose limit was set as a pragmatic choice.'

HH: I write: "If a trial with a low dose finds a negative result, the negative findings can be attributed to the low dose. Thus, trials with large doses are more critical for testing whether vitamin C is effective." Thus, low dose trials can give a negative result simply because the dose is too low. Therefore, they can be uninformative. For example, the average vitamin C intake in the Western countries is about 0.1 g/day. If a study administered 0.05 g/day of vitamin C, we would label it a "vitamin C trial" but it would be uninformative. For a study to give relevant information about vitamin C, the dose should be substantially higher than the ordinary variation around the average dietary intake.
Nevertheless, no studies were excluded with this pre-planned requirement.

5. Methods – Why weren't 'vitamin C' and exercise-induced bronchoconstriction' included as search terms?

HH: In Medline, "vitamin C" is classified in the MESH term system to "ascorbic acid". In the MESH terms, "exercise-induced asthma" covers EIB. Supplementary file 1 describes the search of Scopus: ("vitamin C" or ascorb*) AND ("exercise-induced asthma" or "exercise-induced bronch*").

6. Statistical Methods – This reviewer must defer to a statistian to advise on the appropriateness of the imputation of data. Indeed, it does not appear to be very robust to impute ~25% of the data on which the analyses are based.

HH: See below.

7. Discussion - The author claims that the imputation of data does not influence the conclusions, but the reality is that there are only 30 'real' data points, so it does reduce confidence in the data.

HH: First, I report sensitivity analysis to test the robustness of the conclusions, and the conclusions are not affected by the exclusion of the Cohen study. Thereby the conclusions do not depend on the Cohen study and its imputations.

Second, "there are only 30 'real' data points" is not correct.
I added a new analysis of Cohen's EIB cases as binary data and that does not require any imputations. However, this is a different outcome than the continuous variable of FEV1 decline and I cannot include this binary outcome in the meta-analysis of Fig. 1. Nevertheless, this binary outcome gives the same conclusions as the imputed-data analysis. The 2x2 table analysis of Cohen study has all 20 data points.

Third, the number of observations that are needed to show an effect depends on the magnitude of the effect. If the effect is small, tens of thousands of participants are needed. On the other hand, if the effect is large, then a small number of participants is sufficient.

Linear modelling of the Schachter data gives P=0.0003 for the vitamin C and placebo difference as the test of the slope. Thus, 12 participants is sufficient in this case to show that there is a highly significant difference between vitamin C and placebo.

Cohen reported the number of EIB cases (FEV1 decline 15% or over) after the exercise session on placebo and vitamin C days (20-0 vs 10-10, respectively).
There are many ways to calculate P-values for 2x2 tables, see e.g.:
http://www.ncbi.nlm.nih.gov/pubmed/19170020
The Fisher exact test has been widely used, but it is conservative (too large P-values), and the above paper strongly discourages its use. Instead the above paper encourages the use of the mid-P modification of the Fisher test. For the Cohen 2x2 table (20-0 vs. 10-10) the mid-P(1-Tail) = 0.00011. Although this P-value is very small, it is conservative (too large), since it ignores that all participants were selected as EIB cases. Still, in the Cohen study, the 20 participants is sufficient to show, as the 2x2 table (no imputations needed), that there is highly significant difference between vitamin C and placebo.

****

Reviewer: Tim Pickles, Statistician, South East Wales Trials Unit, Cardiff University

Quick summary:

This paper provides a good meta-analysis, and further analyses, on the topic of vitamin C in exercise induced bronchoconstriction. On the whole, this paper is very well written and the analysis is relevant and sound. The matter of missing data in the Cohen paper is dealt with sensibly and appropriate sensitivity analyses are presented.

I do though have a few comments, corrections and questions I would like to highlight here, which I will do section by section.

Abstract:

- Last sentence: “Further research on the effects of vitamin C on EIB is warranted”

HH: I do not know whether changes should be done. I suggest that because of the low cost and safety of vitamin C, and the strong evidence from the 3 EIB studies, people may test vitamin C if they have respiratory symptoms associated with exercise. Still, the small P-values do not mean that we are at the end of a road, instead more research should definitely be done.

Article Summary:

- Key messages; point 2: I’m not sure you can say “simply”. Also, and I will repeat this point later in the results section, I don’t think you have any evidence to back this statement up.

HH: I removed “simply”, yet I disagree with the reviewer's comment, see below

Introduction:

- No comments

Methods and Table 1:

- Change “percentange” to “percentage”
- Table 1: Please add a horizontal line between Intervention and Outcome for Cohen et al. 1997.

HH: Typo corrected.
There was (is) a horizontal line in Table 1, but apparently it was lost in the transformation to PDF.

Results:

- Quote statistical heterogeneity figures here rather than in text describing Figure 1

HH: Done

- In the paragraph giving a summary of the effect of each trial, please provide some talk of significance and/or confidence intervals so we can tell if the effects are meaningful

HH: Done. I had considered that the CI:s shown in Fig 1 would be sufficient.

- You argue that a “… single fixed percentage point estimate of vitamin C effect may be simplistic.” I don’t know that the estimate from a linear regression is anything more than a single estimate. Also, I am unsure about the relevance of the word “fixed” here. Both the percentage point estimate, and that from linear regression, have a level of variation related to them

HH: I calculate an estimate of 8 percentage points (pp) smaller decrease in FEV1 for persons classified with “EIB”, and many readers may assume that the 8 pp estimate is valid for all people who are labelled with “EIB.” However, there is much variation within the category of “EIB”. Some EIB cases have post-exercise FEV1 decline of 11% whereas some others have FEV1 decline of over 50% (e.g. participant #11 of Schachter had 52% FEV1 decline and participant #2 of Cohen had 50% FEV1

decline) yet both 11% and 50% FEV1 declines are labelled with the same term “EIB”.
By “fixed [8] percentage point estimate” I meant that the reader may assume from the 8 pp figure, that vitamin C causes a change in the FEV1 decline from 11% >> 3% and from 50% >> 42%. However, if the 50% relative effect better describes the influence of vitamin C, then the same people would have the FEV1 decline changed from 11% >> 5.5% and from 50% >> 25%.
Both the percentage point estimate and the slope of the linear regression do have variation. However, the variation is much smaller in the slope, as indicated by the ratio mean/SE which gives the P-value. Smaller variation around the mean indicates a better explanation, see below.

TP: “I don’t know that the estimate from a linear regression is anything more than a single estimate.”

HH: Linear regression does not give a constant percentage point estimate, instead its percentage point estimate depends on the placebo-day FEV1 decline as defined by the slope. Thus, the slope is a constant, but the percentage point estimate is not constant but varies depending on the placebo-day FEV1 decline. That is very different description of the vitamin C effect compared with a constant 8 pp effect for all people with a label “EIB” (with FEV1 decline varying from 11% to over 52%). Therefore this issue is important.

- P = 0.054 is not “marginally significant”. It is very much in that grey area of uncertainty that you get with P-values. I think “marginally not significant” would be more suitable here

HH: Done

- So how is it “better”? Is it simply because it is now significant? Don’t forget that you only have data on 12 individuals here. This hardly a large enough sample to be making distinct statistical statements about. I don’t disagree that it is different, and valid, way of analysing these data but there needs to be more here to show me how it is better

HH: I do not agree with the reviewer.
In linear modelling, the decision about which of two nested models is better is usually based on the comparison of their log-likelihood values (LL). The difference in the -2*LL between two nested models follows the chi-square distribution with degrees of freedom given by the number of variables added to the more complex model. See a short summary of the likelihood ratio testing:
http://en.wikipedia.org/wiki/Log_likelihood
http://en.wikipedia.org/wiki/Likelihood-ratio_test
Excerpt: “The model with more parameters will always fit at least as well (have a greater log-likelihood). Whether it fits significantly better and should thus be preferred is determined by deriving the probability or p-value” of the LL difference.
I revised the manuscript and used the likelihood ratio test to prove that the complex model containing the placebo-day FEV1 decline values is better than the simple model with only the intercept (corresponding to the t-test).
When I add the placebo-day FEV1 decline to the simple model containing only the intercept, -2*LL increases by 16.5 [=chi-square (1 df)], corresponding to P = 0.00005 as an answer to the question whether the difference between the simple model and the complex model might be explained purely by chance. Thus, the complex model fits significantly better than the simple model containing just the intercept. Therefore, the model including the placebo-day FEV1 decline should be preferred. The complex model is better because the slope effectively captures variation in the vitamin C effect.

I modified the text and the linear modelling so that the likelihood ratio test is described in the revised

version. I also added the likelihood ratio test as prints of the R program to a new Supplementary File 3 so that a more interested reader can take a look at the actual figures.

- For discussion of linear regression of the data from Schachter & Schlesinger and Cohen et al., could you quote values for R2 and P in the text rather than in the description of figures (or not at all)

HH: Done

- Report decimal points for upper bound of 95%CI in the first sensitivity analysis. Should be 12.3

HH: Decimal point added. However, I revised the imputation so that I present only the "no vitamin C effect" imputation because it is simple and conservative. Therefore, only the sensitivity analysis in which I remove the Cohen study remains.

- I miss any reference to this second sensitivity analyses in the tables. Could this analysis, and the first secondary analysis, be presented in the same format as figure 1 please.

HH: Transparency and brevity are both important in scientific reporting, but sometimes they are conflicting. Adding figures and tables that are marginally relevant takes space and makes it more difficult for the ordinary reader to focus on the main issues, although transparency is increased. I added the sensitivity analysis as prints of the R program to a new Supplementary File 3 so that an interested reader can take a look at them. This increases transparency, but does not extend the length of the report.

Discussion:

- Again, I don't believe that you can make any statement about the linear regression being better

HH: I disagree. See above. The complex model fits significantly better and should thus be preferred.

- Low numbers are always a concern as it is highly unlikely that they are representative of any sample. The only advantage here is that the data is paired (from cross-over studies) and hence you have more than you would for non-cross-over studies.

HH: It is not the size of the trial that defines the possibility to extrapolate findings. Large trials are non-representative when the inclusion criteria are narrow, and in many large trials the inclusion criteria have been very narrow.
In this meta-analysis, there are 3 trials that were done in 3 different decades and on 2 different continents. In addition, the mean age was 14 yr in the Cohen study, but 25 and 26 yr in the two other studies and the criteria for EIB differed. Still, all of them found a 50% reduction in the FEV1 decline caused by exercise (with I-square = 0%). We do not know how far this 50% estimate can be generalized, but the concerns would be much greater if all 3 studies had been carried out by one research group in one university. In contrast, these three studies are far in time and geography, with different kinds of participants, etc., yet all of them found the 50% relative effect estimate.

References:

- No comments

Figures:

- Legends to Figures: remove all number here and put them in the results section

HH: Done

- Figures 2 and 3: I cannot tell what the thin black line represents. Please explain better or remove the line.

HH: The Figure legends stated: “the thin line indicates the identity of vitamin C and placebo treatments”. The purpose was to show the distance of the data points from the “no effect” level. Since linear modelling was modified to use the difference in FEV1 declines between vitamin C and placebo as the measure of the effect (see above), the Figures are new. In the new figures, I added a dash line to indicate the identity of vitamin C and placebo.

Supplementary Files:

- Due to the excel to pdf transfer, some start and ends of lines are lost. Also, in the final table where you show the calculation of SE(z), the cells don’t allow the reader to see all the text

HH: I submitted the supplementary table as an excel file and I did not know that it was being transformed to PDF. I will take a look at the layout. I hope I can include the supplementary tables in the excel form since then a critical reader can take a look also at the formulas that are used and not just the numbers.

- Can you provide the data for all 20 of the Cohen patients of their Vitamin C day?

HH: I revised the imputation so that I am presenting only the “no effect” imputation because it is much more simple and it is more conservative, yet the findings are not substantially different. There is explicit data for 11 participants. In the revised supplementary file 2 I show the imputed vitamin C and placebo day FEV1 decline values.

- The Schachter & Schlesinger is presented poorly. Could you show it as before and after exercise, rather than making the reader decipher variable names to comprehend what is in the columns

HH: Schachter does not report the FEV1 level before and after exercise, but they report the difference in FEV1 and the pre-exercise FEV1 value. The percentage change can be calculated from these data, but the original data cannot be presented in the same way as the Cohen data. I improved the table. However, I keep the letters for identification because they help the reader to follow the subtractions and divisions.

- For the sensitivity analysis using “no vitamin C effect” from the 9 missing Cohen patients, surely the SE is 3.24, not 3.03 (See: “The “no vitamin C effect” imputation leads to Mean = 11.2 and SE = 3.03 for the whole set of 20 participants”)

HH: On the EXCEL sheet “Cohen 1997” I calculate that assuming “no vitamin C effect” gives mean

effect = 11.20 and SE = 3.03. This SE(c) = 3.03 is the correct SE value for the 20 observations. However, meta-analysis programs assume that the SE originates from large studies (i.e. they assume infinite degrees of freedom: they use "z" instead of "t"). Therefore, on the EXCEL sheet "Fig.1" I need to transform the SE(c) so that the transformed SE(z) gives the correct confidence interval in the meta-analysis program. The transformed SE(z) is 3.24 for the Cohen study with the 9 "no effect" imputations.
The confidence intervals in text and in the forest plots must be correct, even if that requires modification of the SE values for the meta-analysis program.

•

PRISMA Checklist:

- Can't say item 22 is covered on pages 9-10 when you state no page for items 12 or 15.

HH: Prisma item 22 states: "Present results of any assessment of risk of bias across studies."
In Table 1 and at the beginning of the Results, I describe that all included trials were randomized and double-blind, in addition to the inclusion criterion of being placebo-controlled. This means that there is no great risk of bias in the included studies.
Items 12 and 15 are empty because I did not pre-plan assessment of risk of bias. I reasoned that if identified studies have relevant shortcomings, I will consider their influence on conclusions when I see the shortcomings. Still, describing that all included trials were randomized and double-blind means that item 22 was briefly covered.

VERSION 2 – REVIEW

REVIEWER	Tim Pickles, Statistician, South East Wales Trials Unit, Cardiff University
REVIEW RETURNED	21-Mar-2013

GENERAL COMMENTS	Quick summary: This paper has been improved wonderfully following a review. An important query surrounding the statement of linear regression against simple t-test has been answered and is now a very useful argument. Most of my original points have been covered though I still have a couple points to raise and amendments to suggest. <u>Important Points</u> Discussion: • Your response to my point regarding low numbers is relevant and also is available in the text. I would remove the sentence beginning 'However, a low number …' and would jump straight to the next paragraph, inserting a 'However, ' before it. That you have highly significant results from 40 is neither here nor there (it could be that you managed to collect 40 very unlikely outliers), but what you write about the differences in the studies in terms of time and geography, as well as getting very similar results has far more relevance. Original review and response … *TP: Low numbers are always a concern as it is highly unlikely that they are*

	representative of any sample. The only advantage here is that the data is paired (from cross-over studies) and hence you have more than you would for non-cross-over studies. *HH: It is not the size of the trial that defines the possibility to extrapolate findings. Large trials are non-representative when the inclusion criteria are narrow, and in many large trials the inclusion criteria have been very narrow.* *In this meta-analysis, there are 3 trials that were done in 3 different decades and on 2 different continents. In addition, the mean age was 14 yr in the Cohen study, but 25 and 26 yr in the two other studies and the criteria for EIB differed. Still, all of them found a 50% reduction in the FEV1 decline caused by exercise (with I-square = 0%). We do not know how far this 50% estimate can be generalized, but the concerns would be much greater if all 3 studies had been carried out by one research group in one university. In contrast, these three studies are far in time and geography, with different kinds of participants, etc., yet all of them found the 50% relative effect estimate.* • I don't think you can state that 'the risk of bias between trial periods is low'. It probably is but you have not assessed it (see below). You may wish to change this statement. PRISMA Checklist: • Item 22 – I do understand where you are coming from but the statement alone is not a 'result' and no analysis has been done. I refer you to Figures 2 and 3 on pages 21 and 22 of http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001865.pub3/pdf for the results of an assessment for bias. I think it would be best to remove the page numbers from item 22. Original review and response … *TP: Can't say item 22 is covered on pages 9-10 when you state no page for items 12 or 15.* *HH: Prisma item 22 states: "Present results of any assessment of risk of bias across studies."* *In Table 1 and at the beginning of the Results, I describe that all included trials were randomized and double-blind, in addition to the inclusion criterion of being placebo-controlled. This means that there is no great risk of bias in the included studies.* *Items 12 and 15 are empty because I did not pre-plan assessment of risk of bias. I reasoned that if identified studies have relevant shortcomings, I will consider their influence on conclusions when I see the shortcomings. Still, describing that all included trials were randomized and double-blind means that item 22 was briefly covered.* Minor amendments. Text in bold is either an addition or a change to the text. Abstract: Methods:

	• 3rd line: 2) relative **effect** in the Abstract: Methods: • 3rd line: 95% CI: 4.6 to 12**.2** • 4th line: 95% CI: 33**%** to 64% • 7th line: 50 percentage points (**95%CI:** 23 to 68) Abstract: Conclusions: • 4th line: vitamin C on EIB **is** warranted Article summary: Article focus: • 3rd bullet point The aim of this **research** Introduction: • Spaces between references 12 and 13, and 14 and 15 Methods: • 2) relative **effect** in the Results: • All p-values given to 3 decimal places. Also, any p-values like p = 0.0003 should just be given as p < 0.001 • In the 2nd paragraph. The upper bound of each 95% CI doesn't have a decimal place whilst the lower bound does. Please give 1 decimal place to the upper bound • EIB is not a dichotomous condition**;** instead • Adding the placebo-day post-exercise FEV_1 decline values to the **null** linear model**, which is equivalent to a t-test, improved the model fit** by • decline (95% CI: 19**%** to 64%) • Therefore, **as a secondary measure,** the pooled estimate of the relative • decline by 48% (96% CI: 33**%** to 64%; P < **0.001**) [NB: see bullet point 2 above as to why I would wish the p-value to be changed] Discussion • Space between references 14 and 15 Supplementary Files: • Due to the excel to pdf transfer, some start and ends of lines are lost • There are some overlapping cells in the 'Cohen 1997 2x2' sheet, rows 74 and 78

VERSION 2 – AUTHOR RESPONSE

Reviewer: Tim Pickles, Statistician, South East Wales Trials Unit, Cardiff University

Quick summary:
This paper has been improved wonderfully following a review. An important query surrounding the statement of linear regression against simple t-test has been answered and is now a very useful argument. Most of my original points have been covered though I still have a couple points to raise and amendments to suggest.

Important Points
Discussion:
• Your response to my point regarding low numbers is relevant and also is available in the text. I would remove the sentence beginning 'However, a low number …' and would jump straight to the next paragraph, inserting a 'However, ' before it. That you have highly significant results from 40 is neither here nor there (it could be that you managed to collect 40 very unlikely outliers), but what you write about the differences in the studies in terms of time and geography, as well as getting very similar results has far more relevance.

HH: Done
However, I do not agree with the reasoning "managed to collect 40 very unlikely outliers".
With that kind of reasoning we should not calculate P-values at all, because all small P-value can be caused by chance, from the philosophical point of view.
Small P-values are difficult to explain by chance and the P-values are important. Thus, the small P-values are essential, yet I agree that generalization is much more important issue in the above context.

Original review and response …
[lines deleted]

• I don't think you can state that 'the risk of bias between trial periods is low'. It probably is but you have not assessed it (see below). You may wish to change this statement.

HH: Deleted.
Still, I do not agree with the reviewer. Randomization and double-blinding does imply that there is low risk of bias, see below. Randomization and double-blinding give a positive mark to 4 of the 9 items listed below, cross-over design gives one more positive mark. The methodology of the 3 RCTs does imply low risk of bias, but I leave the note on methodology to the text but delete the implied conclusion.

PRISMA Checklist:
• Item 22 – I do understand where you are coming from but the statement alone is not a 'result' and no analysis has been done. I refer you to Figures 2 and 3 on pages 21 and 22 of http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001865.pub3/pdf for the results of an assessment for bias. I think it would be best to remove the page numbers from item 22.

HH: I deleted pages of Prisma item 22.
However, I do not agree with the reviewer.
Figures 2 and 3 in the linked document list the following items:
****
1. Random sequence generation
2. Allocation concealment

3. Blinding of participants and personnel
4. Blinding of outcome assessment
5. Incomplete outcome data
6. Selective reporting
7. Funding for the screening test
8. Baseline comparability
9. Measure against contamination
****

As to the 3 trials included in the manuscript:
Because the 3 included trials were randomized, item 1 is satisfied
Because the 3 trials used double-blinding, item 2 is satisfied, otherwise the studies could not be double-blind
Because the 3 trials used double-blinding, items 3 and 4 are satisfied
Because the 3 trials used cross-over, item 8 is satisfied
Because the 3 trials were short, item 5 is not of concern and there is no mention that any participants would have dropped out. Cohen reported FEV1 decline values for only 11 of 20 participants, but they reported data on the occurrence for all 20 participants. In my manuscript I imputed data for the 9 participants with missing data.
Because the meta-analysis focuses on FEV1, item 6 is not of concern. Item 6 may be of concern for example in measurements of pain if many methods are used to measure pain and only the most positive outcome is reported. FEV1 is the standard measurement for EIB and other outcomes are not relevant when an EIB analysis focuses on FEV1.
On the basis of the papers, the studies were not funded by commercial sources, meaning that item 7 is not of concern. Funding can be a source of bias for patented pharmaceutical products but much less so for vitamins.
I do not easily see what the above paper specifically means with item 9. In cross-over studies such a term might mean that e.g. if the wash-out period is too short, then the vitamin C level does not fall to baseline level between a vitamin C first and placebo second -sequence. This issue could lead to a false negative finding but cannot generate a false positive finding. Besides, there were reasonable wash-out periods.

Original review and response …
[original lines deleted]

Minor amendments. Text in bold is either an addition or a change to the text.
Abstract: Methods:
- 3rd line: 2) relative effect in the

Abstract: Methods:
- 3rd line: 95% CI: 4.6 to 12.2
- 4th line: 95% CI: 33% to 64%
- 7th line: 50 percentage points (95%CI: 23 to 68)

Abstract: Conclusions:
- 4th line: vitamin C on EIB is warranted

Article summary: Article focus:
- 3rd bullet point The aim of this research

HH: Done all above

Introduction:
- Spaces between references 12 and 13, and 14 and 15

Methods:

- 2) relative effect in the

Results:

- All p-values given to 3 decimal places. Also, any p-values like p = 0.0003 should just be given as p < 0.001

HH: Done all above

- In the 2nd paragraph. The upper bound of each 95% CI doesn't have a decimal place whilst the lower bound does. Please give 1 decimal place to the upper bound

HH: Done
However, the exact value of the CI-limit closer to the null effect is much more relevant than the CI-limit further from the null effect. Therefore in my previous papers I have essentially always given the former more accurately. Rounding of the more distant CI level does not influence conclusions but makes text easier for a reader.

- EIB is not a dichotomous condition; instead
- Adding the placebo-day post-exercise FEV1 decline values to the null linear model, which is equivalent to a t-test, improved the model fit by
- decline (95% CI: 19% to 64%)

HH: Done

- Therefore, as a secondary measure, the pooled estimate of the relative

HH: This is not correct. The Methods section describes that both absolute effect and relative effect were used as primary outcomes: "The primary outcome … the relative FEV1 decline caused by exercise (as a percentage). The measures ... were: 1) the arithmetic difference ..., and 2) the relative effect .... A secondary outcome in this meta-analysis was the proportion of participants who suffered from EIB after the exercise test"
When a RCT is being planned, a single primary outcome is needed for the calculation of study power. However, in meta-analyses a few primary outcomes of equal importance can be used and here I use both absolute and relative differences as two measures for the primary outcome.
The proportion of participants is classified in Methods as a secondary outcome.
The Abstract states: "The primary measures of vitamin C effect used in this study were: 1) the arithmetic difference, and 2) the relative effect... " and this was accepted by the statistician (he suggested a minor change in the sentence so he carefully read it).
Thus, the suggested change was not done.

- decline by 48% (96% CI: 33% to 64%; P < 0.001) [NB: see bullet point 2 above as to why I would wish the p-value to be changed]

Discussion

- Space between references 14 and 15

HH: Done

Supplementary Files:

- Due to the excel to pdf transfer, some start and ends of lines are lost
- There are some overlapping cells in the 'Cohen 1997 2x2' sheet, rows 74 and 78

HH: I hope and assume that the EXCEL sheet can be added as an supplementary file (instead or in addition to a PDF). I had checked that the contents of the cells would be seen in the PDF version but apparently some escaped my eyes.
The ends of lines 74 and 78 were moved to new rows.