

On the evaluation of Cochrane review “vitamin C and the common cold” by Mark Jones on 15 December 2016

Mark Jones' review and responses by Harri Hemilä, summary on 2017-3-3

Mark Jones' comments 2016-12-15 were included within an email by Chris Del Mar in 2016-12-16. Chris Del Mar's email is on this page and the Mark Jones' reviewer comments are on the next page.

On 2016-12-16, ARI group co-ordinating editor Chris Del Mar wrote an email to Harri Hemilä:

From: Chris Del Mar <cdelmar@bond.edu.au>

Sent: Friday, December 16, 2016 01:38

To: Hemilä, Harri O

Cc: Louhiala, Pekka J; Liz Chalker; Ann Jones; Liz Dooley; Justin Clark; Mark Jones (m.jones@sph.uq.edu.au)

Subject: RE: Vitamin C reviews (A066, A105)

Dear Harri

I am writing to set out our concerns about these two reviews. Accordingly I am copying in your co-authors.

You have had extensive feedback from our editorial group, but you have largely chosen to decline the requests and advice we have offered. This can only lead to conflict.

For example the airing of possible benefits of vitamin C for CVD is something we cannot tolerate. It is not relevant, and distracts from the focus of the review. Readers do not want a treatise on vitamin C so much as a careful appraisal of the empirical evidence for its efficacy and safety for these indications. (Your point about needing to highlight the possible bias against ‘supplements’ in general is reasonable, and a sentence to that effect might be reasonable, although your citing the Commentary by Goodwin 1998 in Arch Int Med is unfortunate – it cites, as an exemplar of bias against efficacy, vitamin E for intermittent claudication, for which a more recent Cochrane review <https://www.ncbi.nlm.nih.gov/pubmed/10796571> suggests the evidence is too weak to conclude it is beneficial – and so some better evidence of such bias might be entertained).

You also undertake unorthodox forms of analysis that also raise concerns. Worrying about this, we have asked the Cochrane Central Editorial Unit (CEU) to look independently at the review and comment back to us. Enclosed is their assessment, which I trust you find helpful. **Some additional comments are set out below from the ARI Group's Deputy Coordinating Editor, Mark Jones.**

This is not the first time you have come into conflict with Cochrane editorial groups. In the end it can only lead to unhappiness on both sides: for you and your co-authors because it leads at the very least to delay, and possibly rejection of your work; and for us editorially because we hate delay, and finding another set of authors is time-consuming. The CEU are very worried about this, and wonder whether the review update should be rejected and the review itself also withdrawn.

If you feel so strongly you are fundamentally at odds with Cochrane methods and processes, then it might be best to terminate your collaboration with us early, and seek publication elsewhere. (If you exercise that option, please could you acknowledge the support we have provided to date). If you do that, then perhaps we should also re-visit the other review we invited you to undertake on zinc and the common cold.

Looking forward to hearing back from you soon.

Chris Del Mar

From: Ann Jones
Sent: Thursday, 15 December 2016 1:35 PM
To: Chris Del Mar <cdelmar@bond.edu.au>
Cc: Liz Dooley <lidooley@bond.edu.au>
Subject: Vitamin C reviews (A066, A105)

Mark commented after seeing the CEU report:

"The method they have used for analysis of cold duration looks dodgy to me as they have standardised the group means and standard deviations to reduce heterogeneity and used a non-standard method of analysis. If there is heterogeneity of treatment effect (which presumably would be important to know) then this method may obscure it. It would be interesting to see results from a "standard" analysis. I looked at the unit of analysis issue and after a bit of investigation it looks like they may have done as they said, i.e. "Incidence of colds during regular supplementation was assessed as the **proportion of participants** experiencing one or more colds during the study period". But I found some strange extractions for the placebo groups of the Anderson 1974 study. The numbers do not match the journal article numbers??? Furthermore it says in the characteristics of included studies that: "Problems with the placebo group #6; see p 40 (Table 36) in Hemilä 2006a. Therefore comparison in this review is restricted to the placebo group #4 which had close baseline values for "usual days indoors" and "usual days off work" and "contact with children" consistent with the baseline values in the 6 vitamin C groups" for studies Anderson 1974 a, b, c, and d which implies the same placebo group has been used 4 times??? However the numbers in the forest plot for each control group of these studies are different and none of them match the published data???"

Harri Hemilä's responses to the Mark Jones' review are on the following pages.

Our Cochrane review “vitamin C and the common cold”

<https://www.ncbi.nlm.nih.gov/pubmed/23440782>

has had a web page for over a decade:

<http://www.mv.helsinki.fi/home/hemila/CC>

The web page is mentioned in the 2016 review update:

At the end of the Background section:

“Links to cited publications are available from www.mv.helsinki.fi/home/hemila/CC.”

And in Published notes:

“Full-text versions of references which are available either free or at the publishers' databases can be accessed via the web page of the review: www.mv.helsinki.fi/home/hemila/CC.”

In the 2013 version of our vitamin C and the common cold review, the web page was mentioned also at the end of the “Included studies” section in the “Results” section:

“Links to the trial reports and translations can be found at

<http://www.mv.helsinki.fi/home/hemila/CC>”

However, the latter citation seems to be removed by the ARI editors from the 2016 update.

Thus, there has been a web page for the review.

For the 2016 update we added a spreadsheet to the web page, which shows the extracted data and our calculations.

The spreadsheet makes our review transparent.

The spreadsheet is available at the top of the web page with a note:

“The spreadsheet describing the calculations from the reported data to the versions analyzed in our review is available here:

<http://www.mv.helsinki.fi/home/hemila/CC/CochraneVitcColds2016.ods>

[http://www.mv.helsinki.fi/home/hemila/CC/CochraneVitcColds2016.xls”](http://www.mv.helsinki.fi/home/hemila/CC/CochraneVitcColds2016.xls)

The spreadsheet is mentioned clearly in our 2016 update:

in section:

Data extraction and management

The authors constructed a spreadsheet to record all original data for the 2016 update. One author (HH) entered data and both authors (HH, EC) independently checked consistency of recorded data with original reports. There were no disagreements.

and in section:

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

6. We have not used data collection forms for collecting the data. Instead we have compared the entered data against the published data. In 2016 we constructed a spreadsheet in which we collected the original published data and carried out the calculations to transform the data to a version suitable for our meta-analyses. This makes the process transparent and helps us to check the collected data and the transformation processes. The spreadsheet is available on the web page of this review.

Our review on vitamin C and the common cold is long. The 2016 update is 182 pages in the final PEF format and includes a total of 46 studies, with a total of 77 separate reports.

Therefore it does take much time to go through our review and consider whether the methods are valid and whether all the calculations are appropriate.

Had Mark Jones wanted to consider the validity of our methods, he could have asked us for the details of our calculations and we could have described them, and we could also have forwarded him to the above mentioned spreadsheet.

In addition, by carefully reading our review he would have seen the existence of the spreadsheet since the spreadsheet was clearly mentioned, see above. Instead, Mark Jones speculated that the inconsistency between the published figures and our tables implied that we were sloppy.

Had Mark Jones read the Methods section of our review, he would have seen the answer to his comment "... the same placebo group has been used 4 times???" (Mark J: p. 2 of this document). Had Mark Jones read our review, he would have seen that Analysis 2.3 in our review showed the absolute scale ("It would be interesting to see results from a "standard" analysis").

Mark Jones' comment on the "Anderson 1974 study. ... the numbers in the forest plot for each control group of these studies are different and none of them match the published data???" had its explanation on the top (cells S6 to S9) of the sheet "Analysis 1" in the spreadsheet: <http://www.mv.helsinki.fi/home/hemila/CC/CochraneVitcColds2016.xls>

Harri Hemilä's responses to Mark Jones' comments 2016-12-15 are on the following pages. Those responses were sent to Chris Del Mar on 2016-12-16.

Neither Chris Del Mar, nor Mark Jones pointed out any problems in Harri Hemilä's 2016-12-16 responses to Mark Jones' reviewer comments.

From: Hemilä, Harri O
Sent: Friday, December 16, 2016 20:38
To: Chris Del Mar
Cc: Elizabeth Chalker
Subject: Re: Vitamin C reviews (A066, A105)

Dear Chris Del Mar,

I am quite puzzled with your letter.

... [other parts deleted; this text below is Harri Hemilä's responses to Mark Jones' comments]

... you sent also Mark Jones' comment:

“The method they have used for analysis of cold duration looks dodgy to me as they have standardised the group means and standard deviations to reduce heterogeneity and used a non-standard method of analysis. If there is heterogeneity of treatment effect (which presumably would be important to know) then this method may obscure it. It would be interesting to see results from a “standard” analysis.”

First, the two major scales that are usually used are absolute effect scale and relative effect scale. If there is a constant effect on one of them, there is obviously a heterogeneous effect on the other scale. The scale that produces the most uniform effect estimate is the most useful, since the most uniform effect estimate is most widely applicable.

For example smoking increases lung cancer by 10-fold is a relative effect that is quite widely useful. We can calculate what that 10-fold increase means in the absolute effect scale for various baseline risks depending on gender, age etc. People do not argue against the 10-fold relative effect on the basis that it indeed does obscure heterogeneity on the absolute scale, since that is the actual goal of using the relative scale.

We do not teach the effect of smoking to medical students on the absolute scale since there is no relevant estimate on the absolute scale that is widely useful.

In the review draft we write “The duration of colds is a type of outcome which should be analysed on the relative scale (Hemilä 2016b)” and that reference illustrates the benefit of the relative scale for continuous outcomes. I write for example: “Our Cochrane review on vitamin C and the common cold showed the shortest average duration of colds in a placebo group of children to be 2.6 days, whereas the longest was 14 days (Analysis 2.1.2 in [2]). Such a great variation in cold duration in the placebo groups is explained by variations in viruses and in operational outcome definitions between different trials, and so forth. An ideal treatment might shorten the long colds by 10 days, whereas such an effect on short colds is impossible. Thus, the absolute effect scale is not feasible for analyzing effects on the duration of colds.”

In our review manuscript we also refer to one of Friedrich's papers “A recent comparison showed that the relative scale leads to less heterogeneity in meta-analyses compared with the absolute scale (calculation in days) (Friedrich 2011).” Thus, there is empirical evidence for using the relative scale for some or many continuous outcomes.

Continuous outcomes is a heterogeneous category and we should not consider that all continuous outcomes should be analyzed on the relative scale. For example, blood pressure changes may be much more reasonable on the absolute scale, taking into account the initial level. However, analysis of common cold duration is more reasonable on the relative scale than on the absolute scale.

Second, we do calculate the “standard” analysis (Mark: “It would be interesting to see results from a “standard” analysis”).

We show the absolute scale for the duration of colds with ≥ 1 g/day vitamin C (Analysis 2.3 in our review).

We don't show the absolute effect for other outcomes and groups, otherwise the review would become complex and confusing.

In our review, we compare the two scales (Analysis 2.3): “The reason we used the relative scale in our primary analysis (i.e. percentages) is because it adjusts for variations in untreated (placebo group) colds. For children the calculation of effect on days led to greater heterogeneity (46% versus 33%) and a larger P value (10×10^{-5}) versus 3×10^{-5}) compared with the calculation on the relative scale. These differences illustrate the benefit of the percentage scale, though the difference is not dramatic.”

Mark: “non-standard method of analysis”

I do not quite understand what this means. In the RevMan context, the Friedrich approach to calculate relative effects is non-standard: it is not available in RevMan.

Our linear transformation (division by placebo group mean is a linear transformation in mathematics) causes that the t-test of the transformed figures corresponds to t-test identical with the original values (the exactly same t-value) and therefore the analysis is “standard” as the mean difference analysis of RevMan. Weighing is different and the measure is percent benefit, but I would not call those non-standard.

Mark Jones also writes

“which implies the same placebo group has been used 4 times???”

In our Unit of analysis section we write:

“In four trials (Anderson 1974a; Anderson 1975a; Audera 2001a; Karlowski 1975a), more than one vitamin C group was compared with a single placebo group. For these studies, we divided the respective placebo group for all vitamin C arms weighted with the number of participants or common cold episodes of the vitamin C arms. This avoided double counting placebo arm participants in the analysis.”

In addition, in the “Differences between protocol and review” section we describe: Item 11 “A few studies had several vitamin C arms which were compared against a single placebo arm. In 2004 we pooled the vitamin C arms to a single vitamin C arm which we compared with the single placebo arm. However, such pooling decreases transparency. In addition, such pooling is based on an assumption that the true effect in the pooled arms is equal, so that the differences between the arms are caused just by random variation. This assumption need not be correct. Therefore in 2016 we show all vitamin C arms separately and we divide the placebo arms evenly between the vitamin C arms.”

Thus, although the previous calculation in earlier versions was not unsound, it was non-transparent.

Mark also wrote about the Anderson 1974 trial:

“However the numbers in the forest plot for each control group of these studies are different and none of them match the published data???”

We cannot describe all calculations in the review. We have previously opened a web page where we give links to the studies.

<http://www.mv.helsinki.fi/home/hemila/CC/>

We added a link to that web page to our calculations and we planned to mention that the

calculations are available through the above web page. The calculations are here:

<http://www.mv.helsinki.fi/home/hemila/CC/CochraneVitcColds2016.xls>

As to Mark's question above, it seems that he did not see that for the incidence of colds (Analysis 1.1), in addition to placebo group #4, also groups #7 and #8 were placebo groups. They were given vitamin C therapeutically, but that means they were administered placebo as the regular treatment until they caught a cold. Thus they received placebo until the outcome occurred and thus they were placebo arms for Incidence. Addition of those two arms as placebo groups for the incidence analysis increases the size of the placebo group and accuracy of the comparisons.

We divided the control group participants (#4, #7, #8; N= 285, 275, 308 and colds-N = 233, 209, 234) to control groups for arms #1, #2 #3 and #5 so that we weighed by the number of participants in the vitamin C groups. Thus, in analysis 1.1., the sum of placebo participants in the four Anderson 1994 arms should sum up to 676 colds and 868 participants.

In Analysis 1.1. (ie weighed by the vitamin C group size):

Anderson 1974a has 157 colds among 202 participants

Anderson 1974b has 156 colds among 200 participants

Anderson 1974c has 175 colds among 225 participants

Anderson 1974d has 188 colds among 241 participants

The above sums up to 676 colds among 868 participants as they should.