COCHRANE CENTRAL EDITORIAL UNIT PRE-PUBLICATION SCREENING REPORT

1. REVIEW DETAILS

Title:	Vitamin C for preventing and treating the common cold (A066)
Authors:	Hemilä H, Chalker E.
CRG:	Acute Respiratory Infections Group
Review type:	Intervention (update)
Archie version no.:	16.78
DOI:	10.1002/14651858.CD000980.pub4
Editors:	Nuala Livingstone and Toby Lasserson, Cochrane Central Editorial Unit

Responses by Harri Hemilä 2017-3-2 are marked by "HH:"

There are 43 responses in all, by HH.

Over half of the responses (n=27) are related to reporting and do not challenge the scientific validity of our review. Reporting can easily be revised.

In my view, neither do any of the remaining comments challenge the scientific validity of our review, see the responses. Some of the CEU comments are clearly based on misunderstanding of our review. For many statements the CEU editors do not give any justifications.

2. SCREENING RESULT

Further actions

There are a number of serious issues with both the implementation of methods, and reporting of the findings in this review. Unit of analysis issues arising from counting episodes of cold as the denominator and not individual participants is a specific issue that must be rectified.

Authors' response

HH: We believe the reviewers have misunderstood our Methods section.

They write above: "Unit of analysis issues arising from counting episodes of cold as the denominator and not individual participants"

We write in the methods "In several studies participants had a few colds per person and such colds are correlated since they occur on the same person. However, Constantini 2011a reported that the duration of the third common cold episode was very weakly explained by the duration of the first and second common cold episodes ($R^2 = 0.05$). In most studies the average number of colds was less than 3 per person. Thus, the within-person correlation of cold durations probably has no relevant influence in our analysis."

Thus, although in some studies more than one common cold was observed per person for the analysis of common cold duration, the duration of separate colds within a single person has very weak if any correlation with the next cold.

3. SCREENING FINDINGS

Implementation of protocol methods (Search date, inclusion decisions, differences between protocol and review, additional considerations)

Updating methods

1. Several of the methods outlined in the Methods section of this review are insufficiently described or out of date. Specifically:

    a. Criteria for considering studies for this review

        i. **Types of studies** – Authors have not clearly described which study designs are eligible and which are not eligible for inclusion. Authors have not justified why "*we did not restrict our review to RCTs*".

Authors' response

HH: This is about reporting. We can update the review so this is clearer.

We are including alternative allocation trials. For example, if participants are divided into two groups on the basis of their birthday, it does not seem reasonable to assume that there is systematic bias by being born on even vs odd birthday. Thus, alternative allocation does not uniformly lead to systematic bias. We can extend this section.

ii. **Types of participants** – Authors have not clearly described the age range of eligible participants, specifically what age range they consider to be 'children', and what age they are 'adults'.

Authors' response

HH: This is about reporting. We do not believe this is possible for this review.

In the inclusion criteria, we did not report any limit for age range, which means that we include participants of all ages in our review.

For the protocol of a new RCT it is possible to set unambiguous inclusion criteria for participants that will be recruited to the trial. For example, if we carry out an RCT with "adults", we can set the inclusion criterion to >18.0 years so that participants aged 17.5 years are excluded from the trial.

However, for meta-analyses such unambiguous criteria are usually not possible since we are dependent on the inclusion criteria of the original RCTs that have already been conducted. For example, if we restrict a meta-analysis or a subgroup of a meta-analysis to adults >18 years, what should we do with an RCT that had included participants aged 15 to 25 years? If the majority of those participants are 18 to 25 years, it seems reasonable to ignore the few "children" and consider that the study is an "adult" study. On the other hand, if we exclude such a study entirely on the basis that there are also children among the participants, we would exclude much empirical data for >18 year old people. Thus, an exact cut off limit cannot always be used in meta-analysis; instead often judgments need to be made on a case by case basis.

In our review, this question is a problem with the Carr (1981) trial in which "there were 36 pairs of twins under 18 years, 34 pairs aged 18 to 30 years, and 25 pairs aged over 30 years."

Thus, for a meta-analysis we cannot define the inclusion criteria or subgroup criteria as unambiguously as for a novel RCT.

However, setting an age limit for children vs. adults cannot solve the problem of the Carr trial. If we set the limit to 18 years (or to 16 or 20 years etc), some of the data reported by Carr are for younger than the arbitrary limit and some are for older than the arbitrary limit.

We cannot divide the published data to one part that is for children (less than 16 or 18 or 20) and one part that is above the cut off limit.

iii. **Types of interventions** – Authors have not clearly described, or justified, what they consider to be an eligible intervention. Areas of ambiguity include exactly over what 'period' can vitamin C be administered? Can the vitamin C be administered with another intervention? If so, which co-interventions are eligible and which are not eligible? Setting a limit of 0.2 g/day requires a more reasonable justification than 'convenience'. Particularly because authors admit that studies with a limit < 0.2 g/day are not 'irrelevant' to the review question. It is particularly confusing because many of the analyses present subgroups of studies assessing < 0.1 g/day doses.

Authors' response

HH: This is about reporting. We don't believe that this is true.

The reviewers write: "Authors have not clearly described, or justified, what they consider to be an eligible intervention".

In the Abstract we wrote that the objective is "To find out whether vitamin C ..." indicating that the eligible intervention is vitamin C administration.

In the Methods section we write ("Types of interventions"): "We investigated orally administered vitamin C ≥ 0.2 g daily for a single day or over a period".

The reviewers write: "exactly over what 'period' can vitamin C be administered".

We wrote "The intervention considered was orally administered vitamin C of ≥ 0.2 g daily for a single day or for a period." Thus, we do not set a minimum or a maximum "period".

The reviewers write: "Can the vitamin C be administered with another intervention"

In our view stating "vitamin C" means pure vitamin C. However, we can add clarification if it is not clear that "vitamin C" means only vitamin C and not for example a multivitamin which happens to contain vitamin C.

The reviewers further write: "Setting a limit of 0.2 g/day requires a more reasonable justification than 'convenience' "

The range of vitamin C administration in common cold studies has varied from 10 mg per day up to 8000 mg per day. Given that we are primarily interested in the effects of large doses, the studies with low doses are uninformative. In addition, many of the old studies were poorly conducted and reported. Setting the limit to 0.2 g/day has the added benefit of not having to get translations of some old non-English papers, since they used low doses.

However, this was not the main reason for the limit. We can add some more rationalization for excluding the low dose studies.

The reviewers write "many of the analyses present subgroups of studies assessing less than 0.1 g/day doses."

They do not give any examples of a subgroup that contained doses of less than 0.1 g/day, and since all included studies had doses of at least 0.2 g/day vitamin C there cannot be any subgroups that contain "less than 0.1 g/day doses".

Thus, ""many of the analyses present subgroups of studies assessing less than 0.1 g/day doses" is a false statement.

iv. **Types of outcomes measures** – authors have not provided a clear description of the diagnostic indicators of a 'cold'.

Authors' response

HH: This is about reporting. We believe the reviewers have misunderstood our description.

It seems that the reviewers have not understood our Background section in which we describe that there is no unambiguous definition for the common cold.

We wrote: "The term 'the common cold' does not denote any precisely defined disease, rather it is a cultural concept (Eccles 2013). Nevertheless, this illness is familiar to most people. Typically symptoms of the common cold ..."

We are limited to the operational definitions of the common cold that were used by the authors of the included studies. We could describe the definitions of the authors in the included studies section, but many papers did not provide an explicit definition.

b. Assessment of risk of bias in included studies

i. It is not clear whether non-randomised studies were excluded from the review as the authors ambiguously state that "*We did not restrict our review to RCTs'.* They have avoided describing any other study designs that may be relevant, nor provided details on how the Risk of Bias for non-randomised studies will be assessed.

Authors' response

HH: This is about reporting. We can revise the review to make this clearer.

In the Methods section we describe that "We undertook sensitivity analyses in Analysis 1.1 and Analysis 2.1 to test the robustness of our conclusions regarding the methodological

quality of the trials, in which we excluded all studies which were not randomised and double-blind."

Thus, we carried out sensitivity analyses to exclude alternative allocation trials. In the Results section we describe the findings of the sensitivity analyses.

We can add a more detailed description of the included trials to the Methods and Results sections.

ii. Authors have added several domains to the standard Cochrane Risk of Bias tool but have not justified these amendments.

Authors' response

HH: This is about reporting. The justifications were removed by the ARI Group Editorial Base.

In the Methods section of the July 2016 version we justified the additions.

In the July 2016 version of our update we wrote:

"indistinguishability of vitamin C and placebo"

"[Thomas] Chalmers 1975 proposed that the benefits of vitamin C supplementation on the common cold might be caused by "the result of the power of suggestion." His proposal was based on the Karlowski 1975a trial, in which placebo consisted of lactose which is sweet and differs by taste from ascorbic acid which was used in the vitamin C capsules. Therefore, we collected data on the reported indistinguishability of vitamin C and placebo preparations."

and

"baseline balance".

"In 2016 we added "baseline balance" as a new item to the Risk of Bias table, as encouraged by Corbett 2013. They commented that "if randomisation methods are unclear, then the risk of selection bias in the included studies will be unclear, with consequent reticence to draw firm conclusions from the review. But if baseline data demonstrate that all important prognostic factors were balanced across arms, then that reticence may be misplaced. Alternatively, randomisation methods may appear robust, but important group baseline imbalance not noticed, leading to unwarranted confidence in the findings of the study and hence in the broader findings of the review; selection bias is low as formally defined, but chance differences may need to be considered." (p. 80) ... "Use of suboptimal randomisation methods may be due to clinical practicalities or resource limitations. ... Suboptimal methods do not necessarily imply that the allocations were manipulated.

Examination of a study characteristics table may be able to clarify whether such bias is present. In some trials, adequate similarity across baseline will be achieved. The results of such studies could therefore be considered as being at a low risk of bias (p. 83)." Therefore we added descriptions of the available information on baseline balances of the included studies."

and

"In 2016 we also added "contamination" as a new item to the Risk of Bias table. In our review this means that the level of vitamin C in the placebo group is too high compared with the recommended levels of intake."

####################

However, the Cochrane Acute Respiratory Infections editorial office rewrote that part and deleted the sections above. On 10 October we received comments about the Methods section: Assessment of RoB: "This section has been re-written by the editors. Please retain." We assumed that statement meant that we should not touch that section.

Therefore, in the current version of the update, we describe those issues in the "Differences between protocol and review" section as items 4 and 5. We had added the "indistinguishability between vitamin C and placebo" item in a previous update and therefore it seems not to be in the "Differences between protocol and review" section, but we can add that.

Thus, in contrast to what the reviewers claim, we have justified our addition of the items, but they are available in "Differences between protocol and review" and not in the Methods section.

We are happy to put those sections back in the Methods section as the above versions or as shorter versions, if the Cochrane Acute Respiratory Infections editorial office allows us to do so.

c. Measures of treatment effect

    i. The detail provided in this section is unnecessary. Authors need only to clearly state and justify how dichotomous and continuous data will be assessed.

Authors' response

HH: This is about reporting. However, we need to justify our use of the percentage effect calculation.

The reviewers argue that the percentage calculation is not justified and that is the particular reason why we need to use space to justify it.

ii. Authors decision to standardise '*…the mean values and standard deviations (SD) in each group against the mean duration of the respective placebo group'* is not justified. The ratio of means method quoted relates to a different way of transforming the effect estimate to a ratio on the natural logarithmic scale rather than the difference in means that they have used.

Authors' response

HH: We do not agree with this comment.

The reviewers have not given any arguments to support their claim that "the decision to standardise … is not justified."

We do not refer to the Friedrich paper as a description for how we are calculating our analyses. The Friedrich method of calculation is not available in RevMan. The method which we are using is not identical, but it has the benefit of being transparent in the RevMan context.

We could calculate the relative effect by using the Friedrich approach on a spreadsheet and import the results into RevMan, but that would be less transparent compared to our approach.

Thus, we do not refer to the Friedrich paper to describe our method of calculation, instead we refer to that paper to justify our preference for the relative scale. We write "A recent comparison showed that the relative scale leads to less heterogeneity in meta-analyses compared with the absolute scale, which in this case means the calculation in days".

iii. All detail regarding how multiple study arms studies have been addressed needs to be incorporated in the section 'Unit of Analysis Issues'.

Authors' response

HH: This is about reporting. We can revise the text.

d. Unit of analysis issues

i. Authors' decision to analyse the number of colds, rather than the number of participants, is not appropriate and introduces a unit of analysis error. As an example - Anderson 1975; most of the sample sizes exceed the number of

participants randomised, indicating that the number of events is used as the denominator and not the number of participants randomised.

Authors' response

HH: We do not agree with this comment.

We write in the Methods ("Unit of analysis issues"):

"In several studies participants had a few colds per person and such colds are correlated since they occur on the same person. However, Constantini 2011a reported that the duration of the third common cold episode was very weakly explained by the duration of the first and second common cold episodes ($R^2 = 0.05$). In most studies the average number of colds was less than 3 per person. Thus, the within-person correlation of cold durations probably has no relevant influence in our analysis. "

Thus, although in some studies more than one common cold was observed per person, the duration of the first colds within a single person has very weak correlation with a following cold.

We do not have data for the duration of only the first cold per person. Thus, we can include the data with reasoning that the error is minimal, or we can exclude the date but exclusion of a whole study leads to much greater error in the meta-analysis.

Description of studies

2. The detail provided in the section 'Results > Description of studies' is insufficient. Specifically;
    a. Design - Authors state all studies were "*placebo-controlled parallel group comparison trials*". This does not clearly describe how many were RCTs, how many were not RCTs, the specific other designs that were included, etc.

Authors' response

HH: This is about reporting. We can add more detail.

There are 5 different main analyses and many of them have sub-analyses and subgroups. All analyses and all subgroups contain a different set of trials.

We had not included this detail, because although a description of the characteristics of trials may be useful for each specific analysis. The range of overall characteristics does not help the reader to consider the specific analyses that are restricted to a limited group of the trials.

For example, if we record how many of the overall included studies used alternative allocation (non-RCT), that does not help the reader to understand what methods were used in Analysis 1.1.3 which shows vitamin C to be beneficial.

Thus, to be informative of all analyses and their subgroups, the data should be reported for each analysis and subgroup, but that would be confusing.

In any case, we can add overall descriptive data.

b. Sample sizes – This is an insufficient description, particularly considering for many of the analyses, the number of 'colds' is counted, instead of the number of 'participants'.

Authors' response

HH: This is about reporting. We can add more detail.

Again, there are 5 different main analyses and many of them have sub-analyses and subgroups. All analyses and all subgroups contain a different set of trials.

Description of the characteristics of trials may be useful for each specific analysis. In contrast, the range of overall characteristics does not help the reader to consider the specific analyses that are restricted to a limited group of the trials.

Thus, the range of sample sizes do not give any information of the specific analyses that are restricted to a limited group of the trials. Furthermore, the number of participants and the number of colds in each analysis can be seen in the respective analysis table.

c. Participants – authors have not described the exact age range of participants, or what proportion were male and female. Authors have not clearly stated how many studies selected participants who were involved in strenuous activities (i.e. stating "*some studies*" is insufficient).

Authors' response

HH: This is about reporting. We can add more detail.

Again, there are 5 different main analyses and many of them have sub-analyses and subgroups. All analyses and all subgroups contain a different set of trials.

Description of the characteristics of trials may be useful for each specific analysis. In contrast, the range of overall characteristics does not help the reader to consider the specific analyses that are restricted to a limited group of the trials.

For example, if the reader considers how the results of Analysis 1.1.3 can be generalized, the age range that is relevant is the age range of the participants in Analysis 1.1.3. It does not help the reader to know the age range of trials that are not included in Analysis 1.1.3. The proportion of males and females, and their age range, etc are different for all different analyses and their subgroups. Therefore those distributions should be shown specifically

for each analysis if the purpose is to help reader to consider how far the findings can be generalized.
Analysis 1.1.3 shows "how many studies selected participants who were involved in strenuous activities". Evidently, the number of those studies can be listed in the text section.

d. Outcomes – authors have not described the specific range of 'follow up periods'. Authors have also not clarified if the outcome 'adverse effects' was reported by any studies.

Authors' response

HH: This is about reporting. We can add more detail.

Again, there are 5 different main analyses and many of them have sub-analyses and subgroups. All analyses and all subgroups contain a different set of trials.

Description of the characteristics of trials may be useful for each specific analysis. In contrast, the range of overall characteristics does not help the reader to consider the specific analyses that are restricted to a limited group of the trials.

For example, if the reader considers how the results of Analysis 1.1.3 can be generalized, the follow-up period range that is relevant is the range of the studies in Analysis 1.1.3. It does not help the reader to know the follow-up period range of trials that are not included in Analysis 1.1.3.

We discuss Adverse effects in a separate section, but we can mention that also in this section.

e. Time of publication of the studies – this detail is more relevant in the section 'Discussion > Completeness and Applicability of Evidence'. Authors have not factored this into their downgrading decisions on the basis of indirectness.

Authors' response

HH: This is about reporting. We do not agree with this comment.

The time of publication is a similarly interesting descriptive feature as the size of a study, the proportion of males and females, etc. Therefore the logical place for describing the time of publication is close to those other descriptive items.

The reviewers wrote that "authors have not factored this into their downgrading decisions on the basis of indirectness".

We do not really understand what this means and we would like to request clarification. Does this suggest that older trials should be downgraded purely on the base of age?

First, old trials may be methodologically less satisfactory than new trials. However, we should look directly at the methods of a specific trial and not on the year of its publication. Year of publication is a surrogate for quality and we should not focus on surrogates when we have data about the directly relevant items such as blinding and allocation.

Second, in the GRADE assessment, "indirectness" is defined as follows (Cochrane Handbook 12.2.2):

"Two types of indirectness are relevant. First, a review comparing the effectiveness of alternative interventions (say A and B) may find that randomized trials are available, but they have compared A with placebo and B with placebo. Thus, the evidence is restricted to indirect comparisons between A and B. Second, a review may find randomized trials that meet eligibility criteria but which address a restricted version of the main review question in terms of population, intervention, comparator or outcomes."

Thus, the "time of publication of the studies" has nothing to do with the GRADE definition of "indirectness" and we are unsure what the reviewers meant by their comment.

Exclusion decisions

3. Authors have removed studies due to "*administration of vitamin C with other substances"* and *"not a parallel comparison*". Neither of these factors was clearly listed as specific reasons for exclusion in the section 'Criteria for considering studies for this review'.

Authors' response

HH: This is about reporting. We can make this clearer.

In our review wee wrote: "We investigated orally administered vitamin C ≥ 0.2 g daily for a single day or over a period". This statement means that we are studying vitamin C alone. Obviously we can add a further statement that we are analyzing the effects of vitamin C alone, if the reviewers consider that our statement is not clear enough.
We also write "We included placebo-controlled trials." This implies a parallel comparison, but we can explicitly add "parallel-comparison" if the reviewers are concerned that some readers may consider that there are placebo-controlled trials which have administered placebo for one period and vitamin C for a different period.

Assessment of risk of bias

4. Many of the judgements made during the 'Risk of Bias' assessment were inappropriate, or insufficiently justified (See table 8.5.d in the Cochrane Handbook for more information), specifically:

    a. Review authors judged many studies in the domain 'Sequence Generation (Selection Bias)' as 'Low Risk of Bias', because they were described as '*randomised'*, or '*randomly allocated'*. This minimal level of detail would not justify a 'low' risk of bias.

Authors' response

HH: We do not believe this is a scientifically valid comment.

The RoB table has a column "Author's judgment" which indicates that we should consider ourselves, as Authors, whether it is likely that the particular item actually might cause risk of bias in the observed findings.

There are several reports which have shown that trials usually are methodologically better than their reports indicate:

Hill et al (2002) reported:

"in contrast to previous reports, inadequate random-sequence generation and allocation concealment, per se, may not be a major problem in RCTs. Characterizing RCTs as "good" or "poor" quality based on the published report is likely to be inappropriate."

https://www.ncbi.nlm.nih.gov/pubmed/12384192

Soares et al. (2004) reported:

"The reporting of methodological aspects of randomised controlled trials does not necessarily reflect the conduct of the trial. Reviewing research protocols and contacting trialists for more information may improve quality assessment."

https://www.ncbi.nlm.nih.gov/pubmed/14703540

Devereaux et al. (2004) reported:

"Readers should not assume that bias-reducing procedures not reported in an RCT did not occur."

https://www.ncbi.nlm.nih.gov/pubmed/15617948

Thus, when authors report that they randomized participants to two groups, there is no justification to assume that such a description is fabricated. In contrast, there is evidence that, on average, the published trials are methodologically better than reported as shown in the papers mentioned above. One evident reason for brief reporting of methods is that journals set word limitations for manuscripts to 2000-3000 and therefore all methodological detail cannot be reported. Space is also needed for the other sections such as introduction, results and discussion.

Another important issue is that potential bias in this context means that we consider that observed finding are not correct, but that the true effect is either lower or higher compared with the observed effect. The possibility of bias depends on what is actually found.

For example, in our Analysis 1.1.1 and 1.1.2 we calculated that vitamin C does not influence the average number of colds in the general community. Some of the included studies used alternative allocation and were thus not randomized. If we consider that the methodological shortcomings may lead to bias in the calculated estimate, that means we should assume that the true average effect of vitamin C in the general community is not null.

In our view this kind of reasoning is not sound. Instead it seems much more reasonable to trust the calculated null effect and consider that the alternative allocation studies – which are consistent with the null effect of the randomized trials – are not biased.

b. Review authors judged many studies in the domain 'Allocation Concealment (Selection Bias)' as 'Low Risk of Bias', as t it was a "*double-blind comparison*". The domain 'Allocation Concealment' refers not to the blinding of group status, but to the extent to which the Random Sequences generated are protected by adequate concealment of the allocation sequence from those involved in the enrolment and assignment of participants. It is therefore a major assumption that any 'double blinded' study had adequate 'allocation concealment'.

Authors' response

HH: We do not agree with this comment.

The time scale of double-blinding is long and covers the time range from allocation until the trial is stopped.

Allocation concealment refers to the time point when people are being divided into the trial groups.

Thus, allocation concealment is possible even though blinding might not be possible at the later stages of the trial, eg if surgery is compared with drug treatment that cannot be blinded. However, if a study is double-blinded, its allocation must have been blinded, otherwise it could not be blinded at any later stage of the study.

Thus allocation concealment can occur without blinding at the later stages of the study, ie treatment and observing the outcomes.

In contrast, double-blinding cannot occur without participants being blinded also at the stage of allocation to groups. If participants or researchers knew to which group participants were allocated (ie no allocation concealment), they cannot be made blinded afterwards (ie double-blinding cannot be started after no allocation concealment).

Our reasoning is basis logic, it is called "syllogism", and it is similar to the reasoning:

All men are mortal, Socrates is a man, therefore Socrates is mortal.

Our argument is:

Double-blinding means that over all the time points within a controlled trial, from the very beginning to the very end, participants and researchers are unaware about the treatment of a particular patient in a controlled trial.

Allocation is a time point within a controlled trial,

therefore the existence of double-blinding means that participants and researchers are unaware of the treatment at the time point of allocation.

c. In the section 'Results > Effects of interventions', authors state "*we removed four studies which were not randomised and double-blind (Charleston 1972; Clegg 1975; Coulehan 1974a; Coulehan 1974b).*" If these four studies were not randomised, they should all be high risk of selection bias.

Authors' response

HH: We do not agree with this comment.

Potential bias in alternative allocation trials depends on how the alternative allocation was done, what was studied, and what was found.

If participants are divided into two groups on the basis of their birthday, it is not reasonable to assume that there is systematic bias between those born on even vs odd birthday. Thus, alternative allocation does not uniformly lead to bias and a universal suggestion that non-random allocation automatically means "high risk of selection bias" is not sound.

On the other hand, in treatment trials on surgery, if the researcher alternatively divides patients into two groups from a queue: surgery vs drug treatment, the researcher might switch the order of patients on the basis of the disease severity of a particular patient and, therefore, there is risk of bias in such a version of alternative allocation. Severity of disease is an important predictor in hospital treatments and that is a characteristic available to the physician who sees the patient.

However, when a school class is divided into two groups alternatively, there is no basis to assume that the resulting groups are systematically different over the (unknown) characteristics that are correlated with common cold incidence and severity.

There are no useful predictors for the incidence or duration of colds. Thus even if the researcher might be able to change the order in the alternative allocation in a school class, there is no variable that he or she could use to generate imbalance in the same way as a physician who can see the severity of the disease. Therefore, in the context of prevention, labelling an alternative allocation common cold trial mechanistically as "high risk of bias" is not reasonable.

Furthermore, given that Analysis 1.1.1 found no effect of vitamin C, assuming "bias" means we would expect that the true effect is not null. That does not seem reasonable.

Finally, we reported a sensitivity analysis in which we showed that exclusion of those alternative allocation trials did not influence the conclusions.

Subgroup analysis

5. The subgroup analysis presented in Analysis 1.1 is inappropriate, as there is no consistent reason to divide this analysis into these three groups. Subgroups 1.1 and 1.1.2 are variations in the type of intervention, and Subgroup 1.1.3 is a variation in the type of participants.

Authors' response

HH: We do not agree with this comment.

In the previous version of our vitamin C and common cold review (2013), we divided Analysis 1.1 into "general community" and "physical stress" subgroups.

That subgroup division was based on the increase in oxidative stress during heavy exercise, and on an earlier meta-analysis which found that vitamin C prevented colds in people

under heavy acute physical stress, published in 1996:
https://www.ncbi.nlm.nih.gov/pubmed/8858411
and on a 1997 meta-analysis which showed that large doses of vitamin C in large trials did not prevent colds, indicating that potential preventive benefits are restricted to special conditions:
https://www.ncbi.nlm.nih.gov/pubmed/9059230.
Both of the previous meta-analyses were published well before the first version of the Cochrane review (2000)
https://www.ncbi.nlm.nih.gov/pubmed/10796569
Thus, there was and still is good justification to separate the heavy acute physical stress studies as a separate subgroup from the general community studies.
Furthermore, the current Analysis 1.1.3 has four additional studies (ie after the 1996 paper) with physically stressed people and all the new studies are consistent with the findings of the 1996 paper and there is no heterogeneity over the current seven trials, with $I^2 = 0\%$ and $P = 0.84$. Thus this subgroup 1.1.3 appears to capture conditions when vitamin C has a genuine biological effect.
The existence of subgroup 1.1.3 has no logical relation to dividing or not dividing the general community trials into subgroups by vitamin C dosage.
In the Background section we describe that the great interest in vitamin C and the common cold originated from Linus Pauling's 1970 book and he proposed large doses of vitamin C for preventing colds. Therefore it is justified to test Pauling's proposal by restricting to the large dose studies in current subgroup 1.1.1. That subgroup is much more informative for the question whether vitamin C in high doses (as proposed by Pauling) might or might not influence the number of colds.
The studies that used low doses of vitamin C, <1 g/day, are not relevant for testing Pauling's hypothesis.
In any case, the division of the general community trials by dosage does not imply that the physical activity trials should be divided by the same dosage.
We might divide subgroup 1.1.3 by dosage if there were many more studies with varying vitamin C doses. Currently, the confidence intervals of the seven small trials are wide and the division to low and high doses would not have statistical power for any meaningful comparison. Instead, such a division would make the presentation confusing.
In science we may be interested in different hypotheses and different subgroups in a meta-analysis may have different justifications.
Reviewers do not formulate any valid arguments against our presentation of subgroups in Analysis 1.

6. The planned subgroups have been inconsistently performed and presented throughout the section 'Results > Effects of interventions', and is therefore difficult to follow. Specifically:

i. 'Vitamin C dosage' was explored in Analysis 1.1. Yet, this analysis is not appropriately constructed, as all the studies in analysis 1.1.3 should be assigned to either 1.1.1 (> 1 g/day) or 1.1.2 (< 1 g/day).

Authors' response

HH: We do not agree with this comment.

See our previous response.

Furthermore, the purpose of dividing vitamin C dosage into high and low in Analysis 1.1.1 and 1.1.2. was not to "explore vitamin C dosage", but to formulate a stronger refutation of common cold prevention by high vitamin C doses in the general community. If there is no effect by >=1 g/day doses it is quite evident that smaller doses do not have effects either, except possibly under circumstances that are different from the "general community".

ii. Authors have not provided a clear description and justification for why some subgroups were not performed/were not possible for every comparison and outcome (e.g. no formal 'age' subgroup analysis was presented for the 'incidence' outcome, and no 'dosage' subgroup analysis was presented for the 'duration' or 'severity' outcome).

Authors' response

HH: We do not agree with this comment.

In science we may be interested in different hypotheses and different subgroups in a meta-analysis may have different justifications.

In medicine, we expect that different variables are relevant under different conditions.

There is no basis to assume that the same subgroup divisions are most informative for all outcomes.

As described above, the purpose of dividing vitamin C dosage into high and low in Analysis 1.1.1 and 1.1.2.was not to "explore vitamin C dosage", but to formulate an stronger evidence against any average benefits of high doses of vitamin C, which is the justification for 1.1.1.

There is no benefit in dividing Analysis 1.1.1 further to child and adult studies since all studies (both child and adult studies) consistently found no average preventive effect. Such presentation would simply be confusing to the reader.

Analysis 2 was divided into children and adult studies for the reasons we describe in our review.

The reviewers wrote that " no 'dosage' subgroup analysis was presented for the 'duration' or 'severity' outcome"

That is not quite correct.

We do not present the dosage as subgroups in Analysis 2.1. However, we show the results of ≥1 g/day studies in Analysis 2.2 and we write in our Results "Analysis 2.2 was restricted to studies which administered ≥ 1 g/day of vitamin C since they are most informative about the possible effects of high-dose supplementation."

Thus, we did explore the potential role of dosage, but there are not so many <1 g/day studies that a direct comparison of high and low dosage would be informative.
In analysis 2.1.2 (Child studies) the low dose studies are dominated by the Wilson (1973) studies which we critically comment in our Included studies section (Notes) as follows: "Complicated classification system makes comparison with other trials difficult. Kinlen and Peto pointed out that Wilson calculated 48 different P-values in the report without considering the multiple-comparison problem."
Thus, a direct comparison of <1 g/day children studies against ≥1 g/day children studies is not informative for that reason. Nevertheless, we can restrict to high dose studies and calculate the estimated specific effect of the higher doses. In such a way we can take dosage into consideration without directly comparing low and high doses.
Reviewers do not formulate any valid arguments against our presentation of subgroups in Analysis 2.

7. Some studies may not be assigned to the correct subgroups in the subgroup analyses of 'adults versus children'. For example, the study 'Carr 1981' is currently presented in Analysis 2.1.1 (Adults). Yet in the text, authors state that this study examined 'twin children'. Finally, in the characteristics of included studies table, the age range for this study was '14-64 years'. This also relates back to **point 1.a.ii** in this report, and emphasises the need for a clear description of included participants and a clear definition of the age range of a 'child'.

Authors' response

HH: We do not agree with this comment.

When HH first worked on the review with Bob Douglas for the 2004 update, HH included the Carr (1981) study with studies in children as there were many children in the study. Bob pointed out then that "there were 36 pairs of twins under 18 years, 34 pairs aged 18 to 30 years, and 25 pairs aged over 30 years".
Thus, 38% of the participants were <18 years. and 62% were above 18 years. So we moved that trial to the adult studies group because the majority were adults.
The above comment states: "This also relates back to point 1.a.ii in this report, and emphasises the need for a clear description of included participants and a clear definition of the age range of a 'child'"
As described above, for a single new RCT it is possible to set unambiguous inclusion criteria. However, if the age range is from 14 to 64 years, with close to half below 18 years and half above 18 years, deciding an exact cut off age of 18 (or 17 or 19 or any other age) for the division between children and adults does not help.
We cannot divide the Carr study participants into subgroups by 18 years or any other age. We only have pooled data for all of the participants. The data are partly from children and partly from adults, irrespective of the exact limit, unless we set the limit at 14 years which is not otherwise reasonable.

In Analysis 2.1, the weight of the Carr (1981) study is 2.5% + 1.4% = 3.9% of all studies. In the Adult group, that is just 5% of the total weight (3.9%/76% weight of the total adult subgroup 2.1.1). If the Carr study is removed from the adult studies, the pooled estimate in 2.1.1. becomes -7.32 [-11.22 to -3.42](P = 0.0002) compared with -8.09 [-11.89 to -4.29](P < 0.0001) when Carr is included. That difference is not practically important. We can add this kind of sensitivity analysis to check the impact of that study that has partly children and partly adults.
The total weight of the child subgroup is just 23.6% of the whole Analysis 2.1. Thus, if we moved the Carr study to the child group, it would have a substantially greater weight on the estimate with a weight of 14% in the children subgroup (ie 3.9% /(23.6%+3.9%))
Thus, the Carr study does not have a great weight in the Adult studies (5%), and the majority were adults (ie 62%), and thus it seems reasonable to include it to the Adult group.
When a study has both children AND adults (over and above 16 and 18 and 20 years), it cannot be shown in a child OR adult subgroup, without generating some errors. Inclusion of the Carr study to the adult group, however, leads to minor errors.

Evidently, we could set a new subgroup of "children and adults" and include just the Carr study in that subgroup, but that would become confusing for the reader.

We can correct "twin children" to "twins".

Interpretation (GRADE, SoF tables, full text discussion and conclusions)

Interpretation of results

8. The presentation of results in the section 'Results > Effects of interventions' is confusing. Ideally, the wording and presentation of this section should follow the below example. If data are not available for an outcome within each comparison, the subheading should still be presented to maintain consistency, along with a statement such as "no data were available for this outcome".
    - 1.1. Comparison 1: Regular supplementation trials
        - 1.1.1. Outcome 1: Incidence of colds
        - 1.1.2. Outcome 2: Duration of cold
        - 1.1.3. Outcome 3: Severity of cold
        - 1.1.4. Outcome 4: Adverse effects
    - 1.2. Comparison 2: Therapeutic studies
        - 1.2.1. Outcome 1: Incidence of colds
        - 1.2.2. Outcome 2: Duration of cold
        - 1.2.3. Outcome 3: Severity of cold
        - 1.2.4. Outcome 4: Adverse effects
    - 1.3. Comparison 3: Studies of naturally-occurring colds
        - 1.3.1. Outcome 1: Incidence of colds
        - 1.3.2. Outcome 2: Duration of cold
        - 1.3.3. Outcome 3: Severity of cold
        - 1.3.4. Outcome 4: Adverse effects

1.4. Comparison 3: Laboratory trials with artificially infected volunteers
- 1.4.1. Outcome 1: Incidence of colds
- 1.4.2. Outcome 2: Duration of cold
- 1.4.3. Outcome 3: Severity of cold
- 1.4.4. Outcome 4: Adverse effects

Authors' response

HH: We do not agree with this comment.

In our review, we divide naturally-occurring common cold trials to two categories: (a) regular supplementation and (b) therapeutic supplementation trials. Thus, "naturally occurring colds" cannot be presented as a separate group listed "1.3" above, since all the "naturally occurring colds" were already included in categories "1.1" and "1.2" according to the reviewers' list above.

In our review, we put "Laboratory trials with artificially infected volunteers" in a separate group, since it is not obvious whether the results can be directly extrapolated to people who contract natural colds.

In any case, the 3 trials on laboratory common colds with artificially infected volunteers are such small studies so that they cannot give any meaningful information about incidence. Thus we disagree with the reviewers' suggestion above for "1.4.1".

Furthermore, the 3 trials on laboratory infections are clinically so different, and the measured outcomes are so different, that it makes no sense to pool their results for "1.4.2" or "1.4.3".

Finally, the laboratory studies were short and the number of participants was so low that they are un-informative about adverse effects, so that the suggestion "1.4.4" is also not sound. There is much better evidence about adverse effects from larger and longer trials.

Had the reviewers read our Table 2 about the laboratory colds and the description of the 3 trials in the Characteristics of Included Studies table, they would have seen that the reported outcomes of the 3 laboratory trials cannot be reasonably pooled to a single estimate of effect on "duration" and "severity".

Above, the reviewers propose that the regular supplementation studies should be pooled to a single estimate of effect for incidence, duration, and severity.

This also reveals that they have not properly read/understood our review. We describe in our Results section that in Analysis 1.1 there is significant heterogeneity in the effect of vitamin C on common cold incidence with $P = 0.02$. When there is strong evidence of heterogeneity, no single estimate of effect is consistent with all the included trials and therefore giving a single estimate of effect as "1.1.1" proposed above would mislead the readers.

Given our reasons described in the review as to why we present analyses on common cold duration separately for children and adults, it would not be sound to present the child and adult results together as "1.1.2" as proposed above by the reviewers.

9. Some of the findings described in the section 'Results > Effects of the intervention' were not directly planned as outcomes of interest in the Methods section, and should therefore be removed. Specifically:

    a. The subsection 'Other effects of regular vitamin C in Analysis 1.1'. (e.g. the proportion of chest colds versus head colds, or 'other effects of regular vitamin C' were not specifically listed in the Methods section, and is therefore beyond the scope of this review).

Authors' response

HH: We can revise this section.

However, when substantial within-trial heterogeneity is observed, that should not be swept under a carpet. When, for example, Anderson (1972) found that one in eight and Coulehan (1974) found that one in six benefited of vitamin C, those are relevant observations. The substantial within-trial heterogeneity which we show in this section indicates that the findings of Analysis 1.1.1 should not be generalized to all people. It seems possible that there is a subgroup of the general population who benefit from vitamin C.

We do not suggest that people should take vitamin C because a subpopulation may benefit. However, this kind of review is not done purely for immediate practical suggestion, but this review is also intended to guide further research. If there are significant within-trial difference in vitamin C effect, that should be described.

    b. The subsection 'Possible differences in the effects of vitamin C between subgroups'. (e.g. findings related to contact with young children, and twins living apart or separately is beyond the planned scope of this review).

Authors' response

HH: We do not agree with this comment.

When substantial differences between subgroups are observed, they should be described to readers and they should be taken into account in planning further studies on vitamin C and the common cold. Therefore they are informative in our review.

10. Inappropriate instances of 'vote counting' occur in the section 'Results > Effects of interventions'. For example, statements such as "*although the null effect with the narrow 95% CI in Analysis 1.1.1 should discourage routine use of vitamin C to prevent colds in ordinary people, several studies in Subgroup 1 indicate that a proportion of participants did benefit from the regular vitamin C supplementation*" are misleading and unnecessary.

Authors' response

HH: We do not agree with this comment.

The Cochrane Handbook describes vote counting as follows (9.4.11): " 'vote counting' to compare the number of positive studies [ie P<0.05] with the number of negative studies [ie P>0.05)]".

In the above section referred to by the reviewers, we are not doing any "vote counting". Instead, we are analysing the within-trial heterogeneity in the effects of vitamin C on common cold duration.

If there is within-trial heterogeneity in treatment effect, that has nothing to do with vote counting.

11. Authors have not provided a clear justification for their decision to restrict consideration of adverse effects to the "*largest studies that administered ≥ 1 g/day of vitamin C and had a follow-up period of > 50 person-years in the vitamin C group, and to the two trials that administered the highest doses, 2 g/day of vitamin C, to children*".

Authors' response

HH: This is about reporting. We do not agree with this comment.

If a large dose of vitamin C for a long period does not cause adverse effects, it is highly unlikely that smaller doses or shorter periods will cause harm. Therefore trials with large doses and for long periods of time are much more crucial when considering the possible adverse effects of vitamin C.

Furthermore, all the trials on vitamin C and common cold are small compared with the Cook 2007 and Sesso 2008 trials, which we mention in the Background. In addition, all vitamin C doses used in the vitamin C and common cold studies are very low compared with the Padayatty 2004 which we mention in the Background section.

If we are interested in the safety of vitamin C, there is no justification to restrict ourselves to the common cold studies. Instead, we can use other studies that were much larger and longer, or that used much higher doses, when we consider the safety of vitamin C.

In particular small and short vitamin C and common cold studies are uninformative.

Discussion

12. The discussion section is too long and unnecessarily detailed. It is not a brief 'summary' of the results, and in many instances, just a replication of the section 'Results > Effects of interventions'.

Authors' response

HH: This is about reporting. We can shorten the discussion.

Some medical topics are very simple, whereas some other topics are very complex.

For example, we can ask whether 3 day or 5 day antibiotics for ear infection have similar efficacy or not, and we do not need long discussions for such a pragmatic issue, neither in the Background section, nor in the Discussion section.

In contrast, the non-scurvy effects of vitamin C is a complex topic, which is fundamentally inconsistent with the widely spread belief that vitamin C is just for scurvy. If there is effect by vitamin C, we should expect heterogeneity in vitamin C effects depending on the dietary intake level, on various forms of physiological stress since they lead to increased oxidative stress, etc. Expecting and observing heterogeneity in the effects of vitamin C means that there is a need for detailed discussion in Background and Discussion.

A brief Background and Discussion on a complex issue is not sufficient to describe to the reader the rationale for the review and how the findings should be interpreted. Thus, the effect of vitamin C for the common cold cannot be presented as a "yes-or-no" issue in a similar way as the results of a 3 vs 5 day antibiotics meta-analysis might be.

Nevertheless we can shorten the Discussion.

13. Authors have not activated or completed all the standard headings in the RevMan file, specifically:
    a. Summary of main results
    b. Overall completeness and applicability of evidence
    c. Quality of the evidence
    d. Agreements and disagreements with other reviews and studies

Note: Headings have been activated.

Authors' response

HH: This is about reporting. We can add the missing standard headings.

As far as we have read the instructions, the Discussion headings have not been obligatory and Cochrane ARI editors have not instructed us that they are obligatory.

Furthermore, for Methods and Results sections uniform headings are useful, but a Discussion section usually needs more freedom because there can be a variety of considerations that differ between topics. Our review was originally written in 2004, long before the current instructions.

We can add the above headings and reorganize the Discussion.

Authors' conclusions

14. The section 'Implications for practice' does not provide a clear, concise and accurate summary of the evidence only, as authors' 'opinions' are described.

Authors' response

HH: This is about reporting. When the findings are heterogeneous, then there is no "clear, concise" summary that properly covers the findings.

15. The section 'implications for practice' must only base conclusions on findings from the synthesis. Detail on asthma, for example, was not part of the formal review process and is inappropriate in this section.

Authors' response

HH: This is about reporting. We can remove the comment on asthma.

16. Authors have not provided a general interpretation of the evidence only, and instead have given inappropriate, direct recommendations, such as "*it may be worthwhile for common cold patients to test on an individual basis whether therapeutic vitamin C is beneficial for them, especially given the safety and low cost of the vitamin*".

Authors' response

HH: This is about reporting. We would be grateful for some clarity on the reviewers' specific concerns with the above statement.

Many suggestions for patients are soft. For example, when a patient has head ache, we say that you can try some of the NSAIDs. Usually we do not say that you need to take NSAID. Etc.

Summary of Findings

17. Authors have not presented one SoF table for each comparison. The current version of the table combines two different comparisons in one table (Regular supplementation trials, and therapeutic studies).

Authors' response

HH: This is about reporting. We can divide the SoF table into separate tables.

18. Authors have not included all relevant detail on the PICO in the second row of each SoF table. (i.e. '*see below*' is insufficient).

Authors' response

HH: This is about reporting. We are not sure how the reviewers consider this is best addressed.

We cannot describe the patients and conditions in the heading of the table, since the patients and the settings are different for many of the estimates. For example, the first row of our table describes the effect of vitamin C for the general community and the second row describes the effect of vitamin C for people under short-term exposure to severe physical stress and/or cold. They cannot be combined to a "single setting" in the heading.

Similarly, the third line describes the estimates calculated for adults, and the fourth describes the estimates calculated for children. They cannot be combined to a meaningful "patient or population" in the heading other than "see below".

We can revise the text if there are clear suggestions how to improve the presentation.

19. Authors have not clarified in the table when the 'number of participants (studies)' column actually presents 'number of colds'.

Authors' response

HH: This is about reporting. We can revise the table.

20. Authors have not presented absolute effects in each table.

Authors' response

HH: This is about reporting. We don't believe this is relevant for this review.

There is no relevant "absolute effect" on common cold duration. Mathematically we can calculate a pooled "absolute effect" as we are doing in Analysis 2.2. However, as Analysis 2.2 shows, some of the studies reported 2.6 day colds in their placebo groups, whereas some other studies reported up to 14 day colds in their placebo groups. Thus, it is evident that the average "absolute effect" has no practical relevance.

In analogy, smoking increases lung cancer by 10-fold and that is a widely useful relative effect. We can calculate what that 10-fold increase means on the absolute effect scale for numerous baseline risk combinations depending on gender, age etc. However, 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.

We can add absolute effect, but we do not believe that is informative for readers.

21. Authors have presented results from subgroups, regardless of whether they are necessary to explain the heterogeneity in the overall result.

Authors' response

HH: This is about reporting. We believe there are reasons other than heterogeneity for presenting results for subgroups.

There are clinically relevant groups of people for whom it is meaningful to report subgroup specific effects irrespective of whether there is or is not statistically significant heterogeneity.

For example, usually the effects of treatments are shown separately for males and females, irrespective of whether there is a statistically significant difference between the sexes. The estimates and 95%CIs are interesting even if the difference is nonsignificant.

In our case the question is about children and adults, and they are also interesting separately for the reasons we describe in the text.

22. The Summary of Findings table does not present a balanced overview of the evidence, as the outcome 'Adverse effects' is not included in the table.

Authors' response

HH: This is about reporting. We can include adverse effects in the table.

When there are no adverse effects reported in the large trials, it does not mislead the reader if we do not mention that in our SoF table. Furthermore, as pointed out above, the common cold studies are not very informative about the lack of adverse effects. There are much larger and longer studies, and there are studies with much higher doses. Those studies are much more informative about safety.

However, we can revise the text.

GRADE Considerations

23. Authors have not downgraded the evidence for study limitations. (i.e. if authors addressed **Point 4** in this report, many of the domains would be changed to 'unclear' and 'high' risk of bias, which would downgrade the quality of evidence for each outcome).

Authors' response

HH: As described in our response to point 4 of this report, we do not agree with this comment.

24. Authors have not downgraded for indirectness (see **Point 2e** in this report).

Authors' response

HH: As described in our response to point 2e of this report, we are confused by this comment.

"Time of publication", which is the issue in point 2e has nothing to do with GRADE.

Consistency (abstract, PLS, results in text and SoF tables)

Consistency of Summary Versions

25. In the section 'Results > Description of Studies > Results of the search', authors state "*Our searches identified a total of 127 records for this update (Figure 2). After removal of duplicates, we assessed 110 records for possible inclusion*". According to Figure 2, the search identified 129 records for this update, and all 129 were screened.

Authors' response

HH: This is about reporting. This is a typo. We can revise the text.