These below are Mark Jones' comments on our 2016 update for:

Cochrane Review “vitamin C for pneumonia”, which is available at:

http://www.mv.helsinki.fi/home/hemila/CC/CochranePneu_2013.pdf

Here below are responses in yellow by Harri Hemilä 2017-2-13

##

Statistical Review of Vitamin C for preventing and treating pneumonia (Jan 2017)

Dear ARI Team,

I have conducted a statistical review of this study. My findings are as follows:

Some methodology decisions appear inconsistent, for example:

1. Prevention trials included those with no treatment control groups but treatment trials excluded those with no treatment control groups.

HH: Mark Jones does not seem to have read our review carefully.

We write

> “*We included controlled trials to assess the preventive effects of vitamin C administration. The use of placebo was not required as it seems unlikely that being aware of, or not, taking vitamin C would affect the occurrence of a severe infection. A recent meta-analysis of trials comparing placebo groups with no treatment groups found no evidence of a placebo effect on binary outcomes (Hrobjartsson 2010), thus there is no empirical evidence that suggests the placebo effect might affect the occurrence of pneumonia*.”

I cannot see how our description can be transformed to Mark Jones' version “Prevention trials included those with no treatment control groups”. That comment does not make any sense when compared what we wrote ourselves, see above. It is obvious from our text above and from our review that we have not “included” prevention trials “with no treatment control groups”.

We also write

> “*We used placebo-controlled trials to assess treatment effects of vitamin C on the severity and duration of pneumonia since the outcome (for example, severity) may be affected by the awareness of the treatment by the patients. The recent meta-analysis of trials that compared placebo groups with no treatment groups found evidence of a placebo effect in trials focusing on pain (Hrobjartsson 2010). A placebo control may, therefore, be more useful for the validity of treatment observations.*”

I cannot see how this description can be transformed to "*treatment trials excluded those with no treatment control groups*".

When we restrict to "controlled trials" it is obvious that we exclude studies that did not have "treatment control groups". However, we clearly set an additional criterion that there needs to be a placebo control to include a study in our review. Thus, Mark Jones' sentence does not capture the meaning of our description also in this case.

Thus, it seems that Mark Jones did not read our review thoughtfully.

I cannot see what Mark Jones tries to imply with the above sentence. If there is some real argument behind the sentence, it should be written more clearly so that a reader such as me can understand it.

2. Trials where other substances were administered along with vitamin C were excluded. But trials with vitamin C administered with another co-intervention were included (as long as the control group included the co-intervention).

HH: This comment makes no sense. Given that the topic of our review is vitamin C specifically, we cannot include trials in which the difference between the trial groups is more than vitamin C alone. Pneumonia patients need to take standard medicines such as antibiotics, but when the other medicines are the same in both groups, then the only difference between the study groups is vitamin C, as is the purpose of our review.

Does Mark Jones suggest that we should include trials in which the difference is something else than "vitamin C alone" or does he suggest that we should exclude trials that used other medicines, such as antibiotics, for both groups? His above comment is obscure.

It is not possible to carry out treatment trials on pneumonia with a restriction that no other medicines are allowed. Essentially all pneumonia patients are given antibiotics and other medicines.

A critical reader may question whether these decisions were made after initially observing the trial data (prior to writing the study protocol). I suggest these decisions should be made consistent to reassure critical readers of the validity of the review.

HH: As pointed above, the two comments above make no sense. I do not understand what Mark Jones means by "*I suggest these decisions should be made consistent*". The protocol was written over 10 years ago and I cannot understand Mark Jones' rational to speculate on our old protocol when the question here should be the scientific validity of our update.

Baseline balance is included as an additional risk of bias measure. This is not standard methodology used in Cochrane reviews. It is not so much baseline balance of the measured variables that is the issue in non-randomised studies (as these can be adjusted for in the analysis) but rather the baseline balance of the non-measured

variables that can introduce a bias in a non-randomised study. A critical reader may see the addition of this non-standard measure as a way to justify or enhance the status of non-randomised studies included in the review. I suggest it should be removed.

HH: We give reference to the paper that described that baseline balance is an important issue, it is the goal of successful allocation. Mark Jones does not formulate any arguments against that paper or against presenting the baseline balance data.

In the Differences between protocol and review we justify the addition as follows:

> *"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)."*

Mark Jones does not challenge our justification above.

The second author of the Corbett (2013) paper is Julian Higgins, who is the editor of Cochrane Handbook. Therefore, it does not seem reasonable to propose that the consideration of baseline balance is inconsistent with the Cochrane approach. Given that the paper is authored by Higgins, it is even possible that the topic will be added to Cochrane Handbook in the future, because of the arguments described in the paper and summarized above.

Mark Jones uses term "non-randomised studies" without specifying what he means with that term in this context. In epidemiology, we carry out cohort studies which are non-randomized studies. In cohort studies we "can be adjusted for in the analysis" as Mark Jones describes.

However, allocation of trial participants to two groups by birthday is not random allocation in the Cochrane definitions. Eg. Handbook section 8.9.2.1 describes

> *"Randomization with no constraints to generate an allocation sequence is called simple randomization or unrestricted randomization"*
>
> and section 8.9.2.2 describes
>
> *"Inadequate methods of sequence generation: Systematic methods, such as alternation, assignment based on date of birth, ..."*

However, I dont think that any ordinary researcher believes that allocation by even and odd birthday leads to unbalanced groups, unless the researcher believes in some sort of strange astrology.

Thus, in Cochrane terminology, allocation by birth day is “non-randomized” study, but no-one uses “can be adjusted for [statistically] in the analysis” approach in such a study. Thus, Mark Jones confuses between quasi-random trials in which usually it makes no sense to “adjust” since the groups are usually very similar, and epidemiological studies in which adjusting is the standard approach. Both studies can be called “non-randomized” but they are very different kinds of studies.

In treatment trials, one of the most important, or the very most important, variable defining the average progress of the disease is the severity of the disease at the beginning. Kimbarowski (1967) describes that the severity of disease was similar at the baseline. We also write “The type of selection of the patients causes that the sex distribution is identical (all males), the cause of hospitalisation is identical (influenza A), and the age distribution is probably closely similar in the 2 study arms”. These all are important information when we consider whether there might be systematic differences between the trial groups. There is no evidence that there might be substantial differences in relevant variables.

“It is not so much baseline balance of the measured variables that is the issue in non-randomised studies (as these can be adjusted for in the analysis) but rather the baseline balance of the non-measured variables that can introduce a bias in a non-randomised study.”

HH: As described above, this comment confuses quasi-random trials with cohort studies. In cohort studies there can be dozens or hundreds of non-measured variables that can lead to residual confounding. In contrast, in ordinary quasi-randomized trials there is no basis to speculate that there necessarily are substantial differences between the study groups.

Cochrane Handbook (sect 8.9.1) states “The starting point for an unbiased intervention study is the use of a mechanism that ensures that the same sorts of participants receive each intervention.”

Thus, the fundmental issue is whether ”the same sorts of participants” are being compared. We are most interested in the variables that are most important. As mentioned above, one of the most important, or the very most important, variable predicting the progress of the disease is the severity of the disease. Given that the most important variable is balanced in the Kimbarowski (1967) study, I do not see justification to speculate that imbalance in numerous minor risk factors of pneumonia might be so great and in the same direction that it might explain the difference between the trial arms by confounding.

Mark Jones does not describe any reasoning how alternative allocation could lead to good balance in the most important risk factor but to great imbalance in a large number of minor risk factors.

“A critical reader may see the addition of this non-standard measure as a way to justify or enhance the status of non-randomised studies included in the review. I suggest it should be removed.”

HH: This is correct. It is important to look whether the severity of influenza in the Kimbarowski (1967) trial was balanced. Had we no information about the severity distribution in the two trial groups, we might question whether the groups consisted of “same sorts of participants”. Given that we do have information about the balance of influenza severity, we know that there are no meaningful differences on that baseline variable, which is, as mentioned above, one of the most important, or the very most important, variable predicting the progress of the disease.

Thus, the addition of baseline balance is intended for the critical reader who considers whether there is evidence that “same sorts of participants” were compared.

Pitt (1979) excluded 187 out of 864 randomised patients from the analysis but this trial was classified as low risk of bias for incomplete outcome data???

HH: It seems that Mark Jones has not read our review properly.

If there is a great proportion of drop-outs for which there are no good explanations, or if the dropouts are substantially un-balanced between the trial arms, then we should be concerned.

Pitt (1979) did not “exclude” patients. That is not a correct verb in the context.

In our Results section “Description of studies” we write:

> "*These 862 recruits were assigned randomly to either the vitamin C or placebo group from a list of consecutive numbers randomised in pairs. Randomisation was carried out by individual recruits within each platoon*" *(page 908). "Of the 862 recruits who began taking the pills, 64 recruits (34, vitamin C; 30, placebo) were removed from their platoons by the US Marine Corps for further training or for discharge during the eight-week study period. An additional 123 recruits (64: vitamin C; 59: placebo) were excluded from the final analysis because they did not continue to take their pills for the eight-week study period. One additional recruit was eliminated from the vitamin C group because of recurrent urticaria related to taking the tablets*" (page 909).

In our “Characteristics of included studies” table we write:

> *“The study was carried out in US marine recruits. Such a background and the descriptions given in the report does not suggest a considerable drop-out problem. Specifically, the authors described that 22% of the initial population were removed from their platoons or did not continue to take their pills and were not included in the final analysis, but the drop-outs were distributed evenly between the treatment arms”*

If Mark Jones considers that this reasoning is not reasonable, he should formulate specific reasons to express his dissatisfaction.

Cochrane Handbook sect 8.12.2 states:

> *"The risk of bias arising from incomplete outcome data depends on several factors, including the amount and distribution across intervention groups, **the reasons for outcomes being missing**, the likely difference in outcome between participants with and without data, what study authors have done to address the problem in their reported analyses, and the clinical context. Therefore it is not possible to formulate a simple rule for judging a study to be at low or high risk of bias. The following considerations may help review authors assess whether incomplete outcome data could be addressed in a way that protects against bias, when using the Collaboration's tool (Section 8.5*).
>
> ***It is often assumed that a high proportion of missing outcomes**, or a large difference in proportions between intervention groups, **is the main cause for concern over bias**. However, **these characteristics on their own are insufficient to introduce bias.** Here we elaborate on situations in which an analysis can be judged to be at low or high risk of bias. **It is essential to consider the reasons for outcomes being missing** as well as the numbers missing."*

Thus, the Handbook emphasizes that the reasons for missing data are essential and should be considered. Mark Jones did not consider the reasons for the drop-outs, which in the Pitt (1979) trial were symmetrical for both trial groups and therefore cannot introduce systematic bias.

The Handbook emphasizes that "high proportion of missing outcomes" by itself is not any issue for concern, but that is the only focus of Mark Jones' comments, see above. Evidently, I may also disagree whether 22% is a large proportion in absolute terms. In my view it is not, but that is not even the major question here, see the Handbook sections above.

Thus, Mark Jones has not read the Cochrane Handbook thoughtfully.

Also none of the non-randomised studies are classified as high risk of bias for any item including random sequence generation or allocation concealment??? These decisions seem highly dubious.

HH: This comment is not scientifically sound. The fact that a study is non-randomized (in our case the studies are quasi-randomized, not cohort studies) does not mean there necessarily is high risk of bias. For example, if participants are allocated to two groups by their even and odd birthday, we expect baseline balance, unless we believe in some sort of strange astrology.

For the Kimbarowski study we describe that there was baseline balance in disease severity, which is probably the most important factor influencing the risk of pneumonia.

Stating that there is "high risk" of bias would require specific justification. For example, if Kimbarowski reported that there was substantial bias in the baseline

severity, we would have justification to classify it “high risk” of bias, but that is not the case.

For the Glazebrook, Kimbarowski, Mochalkin and Tanaka, we have set “unclear risk of bias” for random sequence allocation and allocation concealment.

Here we need to distinguish between two questions: (1) was there allocation concealment and (2) did the lack of allocation concealment lead to bias if there was allocation concealment. The lack of allocation concealment can or cannot lead to high risk of bias depending on the study. If there is a prevention study in which no important risk factors for the outcome are not known, it is not possible for the researchers to divide participants to imbalanced groups even if they would like to. In contrast, in treatment trials the severity of disease is a highly influential variable predicting progress. If there is no allocation concealment, then the researcher might put less severe patients to one trial group and more severe to another trial group. However, if the researcher publishes baseline balance, that kind of imbalance would be seen by the readers. On the other hand, if there is substantial baseline balance in the severity of disease, we know that the researcher did not put less sick and more sick to different trial groups even if there was no allocation concealment. Thus, the real issue in the interpretation of trials is question (2), and not question (1), although they are related.

Curtis Meinert is a highly experienced clinical trialist in the USA.
In one of his books he wrote about constructive and non-constructive criticism about RCTs. I scanned a few pages since they formulate the problem very well:
http://www.mv.helsinki.fi/home/hemila/M/Meinert_1986_p276.pdf

> Meinert writes:
> *A criticism, to be valid, should:*
> *1.Have some basis in fact*
> *2.Be buttressed with supporting evidence*
> *3.Make a difference in the interpretation of the results*
> *... Among the three, the third is the most difficult one to satisfy. For example, it is fairly easy to criticize a trial because of differences in the baseline composition of the treatment groups. However, it is quite another thing to show how those differences might have accounted for the results observed.*

Thus, it is easy to state that there was no allocation concealment (1) but that does not directly lead to any explanation for observed differences. The real question is (2) above and that needs consideration of the context.

Meinert uses term "universal criticism" to describe criticism of controlled trials that are nonconstructive.

Kimbarowski (1967) is classified under community acquired pneumonia however on reading the English translation of the original manuscript it appears the patients of the study were admitted with influenza then treatment with vitamin C was provided to around half the patients. The outcome of bronchopneumonia occurred at day 6 or 7 presumably while the patients were in-patients (as patients were admitted for at least 9 days according to Table 2). This appears to suggest the pneumonia should be classified as “hospital acquired” which is typically defined as “Hospital-acquired

pneumonia (HAP) or nosocomial pneumonia refers to any pneumonia contracted by a patient in a hospital at least 48–72 hours after being admitted”? Also it appears vitamin C was given as a treatment for influenza hence should this study be included as a prevention trial?

HH: Usually “hospital-acquired pneumonia” means cases when people have been hospitalized because of surgery or burns etc. and they happen to get pneumonia while on hospital. As an example, the Tanaka (2000) study was focused on burn patients and we do not expect that burns themselves should lead to pneumonia in the same way as pneumonia is a well-known complication of influenza in the community and the relation between influenza and pneumonia is similar in the hospital.

Given that pneumonia is a well-known complication of influenza, it is reasonable to classify the Kimbarowski cases as community-acquired pneumonia, even though they did occur after a short stay at the hospital.

Vitamin C was administered as treatment for influenza. However, in our review we are not examining influenza by itself. The patients did not have pneumonia at the start of the trial, when vitamin C administration was started, and therefore the effects on pneumonia are prevention effects. Of course, we might interpret that the same effects are treatment effects for influenza by making them more severe due to the complication. However, when we are specifically interested in pneumonia, the Kimbarowski study tests the prevention effects of vitamin C against pneumonia.

The grading of some of the studies is also highly dubious. For example the non-randomised studies should have a baseline grade of low before and up or down-grading occurs.

HH: This comment is not scientifically sound. The fact that a study is non-randomized (in our case quasi-randomized, not a cohort study) does not mean there necessarily is high risk of bias. For example, if participants are allocated to two groups by their even and odd birthday, we expect baseline balance, unless we believe in some sort of strange astrology.

For the Kimbarowski study we describe that there was baseline balance in disease severity, which is probably the most important factor influencing the risk of pneumonia.

Stating that there is “high risk” of bias would require specific justification. For example, if Kimbarowski reported that there was substantial bias in the baseline severity, we would have justification to classify it “high risk” of bias, but that is not the case.

Also the (lack of) directness of the results doesn’t appear to have been accounted for in analysis 1.1???

HH: The meaning of the above sentence is obscure.

In the Cochrane Handbook section 12.2.2. ”Indirectness of evidence” concept is defined as follows

> ”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.”

In our review, there is no problem of the first type of indirectness. Also, when we point out that the populations are special and we are not extrapolating findings to the general population, we cannot see any problems of the latter type of indirectness.

Mark Jones does not describe what he means with “the (lack of) directness of the results” in the above sentence.

The 3 studies are all in very specific subgroups in the population undertaken many years ago. Therefore they appear to be somewhat irrelevant today as diets are presumably better balanced in terms of vitamins and minerals? And generalising results to other populations appears questionable.

HH: In our review we emphasize that the 3 prevention trials were done with very specific subgroups. We are cautious in extrapolating their findings. In the beginning of the Discussion we write e.g. “However, we urge great caution when extrapolating the findings to the general population...”

Mark Jones does not give any evidence to support his belief that those studies “appear to be somewhat irrelevant today as diets are presumably better balanced in terms of vitamins and minerals”.

In our review we write

> “*Low vitamin C levels are not rare in Western hospital patients nor in the Western community populations (Mosdøl 2008; Raynaud-Simon 2010; Schleicher 2009). Thus, if pneumonia risk is increased by low intakes of vitamin C, then this issue may be important in certain population groups of high-income countries and not just in low-income countries.*”

Mark Jones does not challenge these references and he does not give any references to support his belief that there are no deficiencies of vitamin C nowadays.

In another paper on vitamin C and pneumonia we summarized evidence about the prevalence of vitamin C deficiencies more extensively as follows:

> “*Surveys have found that plasma vitamin C levels below 11 µmol/L are found in 14% of males and 10% of females from the USA,[23] in 19% of males and 13% of females from India,[24] in 14% of elderly people living at home and 40% of elderly people living in institutions in the UK,[25] in 23% of children and 39% of women in*

Mexico,[26] and in 79-93% of men from Western Russia.[27] In a cohort of pregnant women from rural India, 45% had plasma vitamin C level below 4 µmol/L.[28] In a cohort of pregnant or lactating women from Gambian villages in the rainy season, the average plasma vitamin C level fell to 10 µmol/L.[29] Although vitamin C status varies with time, geography and population groups, these figures indicate that low vitamin C levels are not rare, even though frank scurvy is. Thus, if low vitamin C levels increase the risk of pneumonia, this effect may be of wide interest globally. The possible effect of vitamin C on the risk of respiratory infections in physically stressed people is also highly relevant."
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2099400/

Thus, it seems that Mark Jones is not familiar with the current prevalence of low vitamin C intakes.

Furthermore, if the benefit in the Pitt (1979) study was caused by the particularly high dose of vitamin C, as we propose, the deficiency of vitamin C is not an issue in that trial. The plasma levels were not low.

Thus, none of Mark Jones' comments challenges the scientific validity of our review update. Evidently, we are ready to revise our review, but the comments above do not help in our improvements.

Dr Mark Jones