These are reviewer comments on our Cochrane review “Vitamin C for pneumonia” update 2016.

The update is available at:

http://www.mv.helsinki.fi/home/hemila/CC/CochranePneu_Update_2016.pdf

Collated comments for A105 - Vitamin C for preventing and treating pneumonia			
	Referees	**Editor**	**Authors**
Title	Usually the title includes a population group “… in adults and children” (PR)	I’m not sure if it is necessary to be so specific.	**HH+PL:** If the search strategy is restricted to a specific population, or if the findings are positive only for a limited group, then it is informative to describe that group in title. Stating in title “adults and children” does not increase information since that covers all people. Thus, it would make the title longer without making a relevant restriction to any group.
Abstract	**Background:** what is meant by a ‘serious’ infection? – potentially life threatening?	OK	**HH+PL:** In our view, replacing “serious” by “potentially life threatening” does not improve the sentence. Eg. Tooth infections are potentially life threatening since sometimes they lead to septicaemia. However, we do not usually classify tooth infections as “severe” infections. Given that after “serious” we write “... causing two million deaths annually among young children in low-income countries” in our view that gives context to “serious” We are ready to change “serious” to some other word if the editor suggests, but “life threatening”

			does not seem any better, and that is already stated in the same sentence.
	Objectives: … people with pneumonia	I would say “children and adults with pneumonia” and this will cover concerns about the title.	We revised to “children and adults with pneumonia”
	Selection criteria: List the prophylactic studies first? And state: controlled trials that used a placebo in the control group or not. Sounds odd “with or without a placebo”	OK for consistency	**Done**
	Results: in last sentence: whatever the dose of vitamin C…	OK	“in last sentence: whatever the dose of vitamin C… “ We did not figure out where this is suggested. The last sentence is “The included studies did not find adverse effects of vitamin C” However, “whatever dose” is not a relevant issue, instead the lack of adverse effects by the maximal dose is much more informative. However, in the text we point out that very much larger single doses and years of 0.5 g/day are safe and those are even more informative as to safety.

	Authors' conclusions: Not sure why you say 'the therapeutic effects are more likely in people with low vitamin C levels' and limit the conclusions to these – certainly not backed up by the background section of review (this has not been tested in the present review and questionable that is so simple from the dose-response relationship in one of the included studies?) (CR1)	Please respond	"this has not been tested in the present review" This is not correct. We show that the benefit of vitamin C in the Hunt study was restricted to patients with more severe disease and they had lower vitamin C levels. Although this does not prove that vitamin C has effects only those with low vitamin C levels, it is consistent with such a notion. In the background section we describe that pneumonia and other severe infections decrease the level of vitamin C in body. Evidently there is variation in pre-disease vitamin C levels but there is also variation between diseases so that more severe diseases may lead to greater consumption of vitamin C and therefore vitamin C administration may be most beneficial for patients with low vitamin C levels. Most pneumonia cases are treated as outpatients and it does not seem likely to us, that the non-severe cases of pneumonia which are cured within a few days with antibiotics might have substantial benefits of vitamin C. We agree with CR1 that these questions have not been properly investigated, but that is the reason we prefer to be cautious and conservative in our conclusions.
	Do the authors recommend testing of vitamin C levels? (PR)	For discussion?	In our view, the by far highest priority should be to carry out further therapeutic studies on vitamin C. The possible modifying effect of vitamin C plasma level should be examined in those studies.

			Testing vitamin C plasma level is cheap and thus it can be easily tested from pneumonia patients in hospitals, but the real requirement in our view is further studies.
Plain language summary	A bit messy in the 'Key characteristics' and 'Results' sections as mixes up the prophylactic studies and therapeutic studies. Why not just say that the preventative studies all showed similar results with the different methodologies used…. Review question: …controlled studies Background: community-acquired pneumonia refers to where people develop pneumonia outside of hospitals – rather than talking about 'cases' … hospital-acquired pneumonia refers to where the infection is acquired within hospitals Leave out the reference to animal studies	This could be addressed by layout, add subheadings. Ok	**HH+PL:** We do not object subheadings, but eg under the Key results, the first paragraph has just one sentence and in our view the PLS would become more messy if we put a subtitle "treatment studies on community-acquired pneumonia" above that sentence and another subtitle "treatment studies on community-acquired pneumonia" below that sentence. We are ready to revise PLS, but please give more specific suggestions and take into account the short paragraphs on different issues. Added We stated "All three trials on prevention of community-acquired pneumonia…" We do not see the difference. We did not leave out the animal studies. That is essential background. We cannot extrapolate effects in animal studies directly to humans, but positive findings on numerous animal studies make

			it quite plausible that vitamin C has such effects also in humans. We can delete that sentence if the editor instructs so.
	Study characteristics: 1. The 3 preventative studies for community acquired pneumonia We also found one study which examined the prevention of ~~hospital-acquired~~ pneumonia in 37 hospitalised patients with burns in Japan -	Please address	We do not quite understand what this suggests. Is this suggestion that we add a subtitle “The 3 preventative studies for community acquired pneumonia” As described above, subtitles for short paragraphs make the PLS even more messy. We could add subtitle “prevention of pneumonia” to avoid too many subheadings, but hospital and community pneumonia are such difference diseases that they are not properly pooled under the same heading We are ready to revise PLS, but please give more specific suggestions and take into account the short paragraphs on different issues.
	using administration of vitamin C on one day only (corresponding to 110 grams per 70 kg in 24 hours) – and reported no effect of the vitamin C on hospital-acquired pneumonia (moderate quality evidence). 2. The 2 studies on treating community-acquired pneumonia. We did not identify trials reporting on the treatment of hospital-acquired pneumonia. Key results and quality of the	ok	We do not think that the dosage is relevant in PLS. Eg we do not list the doses of other five studies either.

	evidence: Prevention of community-acquired pneumonia - TheAll three trials on prevention of community-acquired pneumonia found an 80% or greater decrease in the incidence of pneumonia in the vitamin C supplemented group (high quality evidence). The preventative studies all showed similar results with the different methodologies used. Treating community-acquired pneumonia - The treatment study in the UK found a significant decrease in the severity of pneumonia in patients who were most severely ill at admission (moderate quality evidence), but vitamin C had no effect on those less ill at admission. There was one death in the vitamin C group, but five deaths in the placebo group (moderate quality evidence). The second treatment study in the former Soviet Union reported 4.0 days (21%) shorter hospital stay in pneumonia patients administered a high dose of vitamin C (high quality evidence). ~~The small study on hospital-acquired pneumonia in Japan found no effect of vitamin C on the incidence of pneumonia (moderate quality evidence). We~~	Ok Ok Ok	Done Revised

	~~did not identify studies reporting on the treatment of hospital-acquired pneumonia. One preventive and one therapeutic study on community-acquired pneumonia were randomised placebo-controlled double blind trials. Two preventive and one therapeutic study on community-acquired pneumonia were methodologically less rigorous, but the observed differences between the vitamin C and control groups are not easily explained by potential biases.~~ All the five trials on community-acquired pneumonia were carried out in such extraordinary conditions that the results should not be extrapolated to the general population today or to average current pneumonia patients. Nevertheless, they suggest a biological effect of vitamin C against community-acquired pneumonia under certain conditions, but those conditions cannot be accurately defined from the five published controlled trials. None of the trials reported noteworthy adverse effects of vitamin C, regardless of the dose of vitamin C (low to high), although one person developed a		High dose studies are much more relevant about adverse effects compared with low dose studies. “regardless of the dose” does not seem informative therefore. We added that “even with a dose of 2

	skin rash (urticaria). More research is needed to define the populations and patient groups who may benefit from vitamin C administration. In the meantime, supplementing pneumonia patients who have particularly low plasma vitamin C levels may be a reasonable measure because of the safety and low cost of vitamin C.		g/day" which refers to the Pitt study
	At end of Plain Language summary: The studies did not report their funding.		Funding is mentioned under study characteristics and it seems more appropriate place than key findings section
	The interventions and comparisons /controls avoids whole issue of dose of vitamin C (CR1)		"avoids whole issue of dose of vitamin C" (CR1) We do not quite see what this comment means. In the Background section we discuss the dose-plasma level relationship implying that higher doses might lead to greater effects. In Additional Table 1 we show that Mochalkin reported dose-response in his therapeutic trial. However,

			that is a poorly reported study and the control group was not administered placebo. We cannot argue that the Mochalkin study demonstrates dose dependency, but we can state that it suggests or is consistent with dose dependency. Mochalkin study can be discussed in Results and Discussion, but we do not see there is basis to comment on the dose-relation of Mochalkin study in PLS.
	I think the plain language statement is clearer than the review conclusion (PR)		This comment contrasts to the comment "messy" by the other reviewer.
	Perhaps 'therapeutic' could be defined as 'treatment'.		In our title we use term treatment and following this comment we replaced most "therapeutic" in text with "treatment''
	I wondered why the authors saw fit to state: "The studies did not report the funding."	It is good practice to report funding	It seems possible that CR2 is not familiar with the

	Because it is well known that 'natural' substances cannot be patented. Thus no particular company could benefit from these trials. The comment is superfluous and misleading – possibly causing the reader to think there is something 'wrong' there. (CR2)	and leave interpretation to readers, so I don't agree this is misleading. Please comment.	problems of some drug companies carrying out numerous studies or/and measuring numerous outcomes, and reporting only the favorable findings. We do not consider that this is a problem with vitamin C, for the reason mentioned by CR2, however, in the context of a systematic review we consider it relevant to note that funding was not reported.
Backgr ound	This is not very easy to read as the authors try to present the many different actions of vitamin C, so tend to have bits of incomplete information that leaves many questions. Some of the information is from animal studies, making it of questionable importance Appendix 1 on the 'Diagnosis of pneumonia' is very interesting,	Please address and comment	**HH+PL:** "Some of the information is from animal studies, making it of questionable importance" We do not agree at all with this comment. When vitamin C has effects on many kinds of virus and bacterial infections in many species of animals, that background makes it highly plausible that vitamin C has effects on some human infections in some conditions. We cannot directly extrapolate effects on humans from animal studies, but biological plausibility is an essential background issue, and therefore animal studies are important when we consider plausibility. "Appendix 1 on the 'Diagnosis of pneumonia' is very interesting, but is somewhat hidden"

	but is somewhat hidden		In the first version of our review (2007) that section was within the main text. However, for some update the editor suggested moving the text to Appendix, since it is a bit long and not directly essential. We consider the issue important and mention in the main text and an interested reader can look at the Appendix.
	Some of my concerns are for example the 'Description of the condition' mentions the incidence of pneumonia in 'ORDINARY' middle-aged Western people, whatever that means;	Please address. Define "ordinary" and "trivial". These are value statements that are not appropriate in this context.	Completely value-free writing is not possible. Most comparisons depend on the context. If we say that a person is short, that can be a neutral statement (when an adult person is say 140 cm), or that statement can be biased because of our own height (if we are 190 cm, then a person of 170 cm is short), or for some reasons we may have negative emotions towards dwarfs in which case the comparison is not just about centimeters. Peoples height is a continuous variable and setting some cut-off on the basis of centimeters is arbitrary, but in ordinary talk and writings we need such descriptors. There are lots of medical philosophy publications about what "normal" means eg in laboratory values and in various characteristics of people. There is no universal definition for "normal" or "ordinary" etc.

			In the context of the left hand side, ordinary meant the majority or ”normal” people, which in the context is a statistical descriptor instead of value descriptor. However, incidence of pneumonia depends on age and several other factors and therefore it is not constant over middle-aged people, not even constant for people of the same age, since there are many other factors that determine the risk of pneumonia in addition to age. Above the reviewer states: “This is not very easy to read as the authors try to present the many different actions of vitamin C, so tend to have bits of incomplete information that leaves many questions”. We could give age-specific figures of pneumonia incidence from several studies (yet valid only for those populations at the time when the studies were carried out), but that would make the text even more “not very easy to read” without giving any real help to reader in his or her interpretations of our study. We did not add any age-specific estimates of pneumonia incidence. Although one to three cases per 1000 person years is a three-fold range (ie inaccurate), that is sufficient in the context, and it is a reasonable basis for comparing the Glazebrook and Pitt study control group incidences. In the context, “ordinary” means “normal people” or “average people”. Obviously we can ask what is

			normal or average. Military recruits is not “ordinary, normal or average”. Middle-aged people with long term medical conditions are not “ordinary”. We changed to average, but if that does not seem suitable, could you please suggest some better word. English is a second language for us and we may not easily know the ideal way to make that statement.
	marine and naval recruits are referred to without detailing what is ‘special’ about the e.g. diets that may lack fresh ingredients/high levels of physical activity/living in close contact with each other; whether the risk of pneumonia in the elderly is from both low and high-income countries?		This is not a review about pneumonia and we cannot cover thoroughly all interesting issues. We slightly rewrote the text.
	<u>Description of the intervention</u>: How short were trials on reducing blood pressure; how was endothelial function measured in patients with atherosclerosis;		We refer to the blood pressure and endothelial function systematic reviews as evidence that the effects of vitamin C are not limited to preventing scurvy. However, in our context the duration of the blood pressure studies and the details of the endothelial function studies are not relevant. They are relevant issues if we would consider the practical importance of those findings, eg whether some people should take vitamin C to lower their blood pressure. The reviewer comments above “This is not very

			easy to read as the authors try to present the many different actions of vitamin C”. That problem would become even greater if we would use much more space to describe the details of the blood pressure and endothelial function studies.
	in last sentence of this section what are the ‘special’ conditions with regard to susceptibility of patients?		Our Cochrane review identified three studies in which vitamin C group had significantly lower risk of pneumonia. The conditions of the three studies are special. The sentence to which the reviewer refers points out that large studies with negative overall effects on cancer and heart disease are not discordant with possible effects in special conditions and limited population groups. In the Discussion section we use more space to consider the characteristics of the “special” conditions, but that discussion seems better after describing the included studies.
	Pharmacokinetics and levels of vitamin C in society – or POPULATIONS?		We followed the suggestion and changed to populations

	Use of word ‘subjects’		We do not understand what this suggests.
	The following statements show the inconclusiveness of any evidence on low vitamin C levels and pneumonia: … changes in metabolism indicate that vitamin C might have a treatment effect on pneumonia patients, irrespective of their dietary intake. Thus, if pneumonia risk is increased by low dietary intake of vitamin C, this issue may be important in certain population groups of high-income countries. Low vitamin C levels are even more common in developing countries … it is possible that the variation in vitamin C intake does not influence the immune system in the ordinary Western population because of their relatively high dietary intake - a SUPPOSITION.		We do not see the point of this comment. Reviewer has cut and pasted three texts from different contexts. The first one (changes in metabolism) points out that when there is great change in vitamin C metabolism (during pneumonia), additional vitamin C might have effects that are different from additional vitamin C for healthy people. We cannot see that reviewer challenges this reasoning. The second (Thus, if pneumonia…) is part of text which refers to studies that have shown that very low levels of vitamin C have been measured in groups of Western populations and not just in developing countries. The third (it is possible) is part of text that considers the reported effects of vitamin C on the immune system. Although there are numerous laboratory studies that have found effects of vitamin C on the immune system, the effects may not be relevant in current average Western population. We cannot see what the Reviewer's conclusions from those excerpts are.

	Safety of vitamin C: Use of: ‘trivial’ adverse effect....	Please address. Define “ordinary” and “trivial”. These are value statements that are not appropriate in this context.	"Trivial" has a value component but we find the word appropriate here . An adverse effect that goes away quite rapidly is not serious. English is a second language for us and we do not know which would be the best word in the context. We changed it to “minor”, but we will change it to some other word if the editor suggests.
	Vitamin C and pneumonia These were actual people rather than ‘cases’ of scurvy; and they were people with “the general condition....” (CR1)	Use of “case” instead of patient is common in medical writing. Please comment.	Case is not a negative word, but it is the standard word to refer to persons with a disease. The sentence “numerous cases of scurvy which appeared at about the same time” is direct copy of the original sentence which thus had the word “case”. “Case” may be unsuitable word in some contexts such as when writing to Readers digest, but this review is intended as a medical text.
	Definition of pneumonia – while I don’t dispute the inclusion of such studies, I would suggest that clinically diagnosed pneumonia should be called “lower respiratory tract infection” to	Very good point. Please address.	Cecil's textbook (2014, p. 587) defines: “It is a useful distinction to separate pneumonias, which are infections of the lung parenchyma and thus distinct from infections limited to the trachea

	make the difference explicit. Much of the evidence that the authors cite for the rationale for clinical pneumonia is from developing countries and in children, and the fact that a chest xray is imperfect doesn't make clinical diagnosis any better. This is particularly the case in adults, where a variety of other respiratory conditions (particularly in patients with pre-existing respiratory disease) may mimic pneumonia. (PR)		or large bronchi ..." However, such pathology-based definition leads to the problem that in many cases the exact location of the lung infection cannot be defined on clinical grounds and the CXR can give false negative findings as we point out in Appendix. Reviewer suggests that "clinically diagnosed pneumonia should be called "lower respiratory tract infection" to make the difference explicit." However, we cannot see any logic in this suggestion. In our Background section we describe that pneumonia is caused by numerous viruses and bacteria, and even though the pathological definition is quite clear, the clinical diagnosis is not. In Appendix we point out that HRCT finds more pneumonias than Chest X-ray, and therefore the latter is not a gold standard. "Lower respiratory tract infection" (LRTI) is a standard word, but it is even more ambiguous than pneumonia. When a patient coughs because of respiratory virus infection, sometimes GPs classify that as bronchitis (LRTI) without any direct evidence of infection at the level of lungs. The reviewer suggests that "clinically diagnosed pneumonia should be called ..." but why not call a "clinically diagnosed" pneumonia simply "pneumonia" as that is even part of the term "clinically diagnosed pneumonia". If we want potential readers to understand what is the focus of our review, it is best to use the standard words and pneumonia describes the

			diseases of the included studies more accurately than even-more-ambiguous term LRTI. We state clearly that “Although the pathological definition of pneumonia is clear, the clinical diagnosis is sometimes ambiguous (Appendix 1).” That should point out to medically less experienced people that the term is not exact. Medically experienced people know that pneumonia is an umbrella for diverse viral and bacterial infections with very difference severity ranging from mild symptoms such as persistent cough which is explained by changes in Thorax X-ray, to severe disease for which people die for. Finally, we do not see to which part of our review, or to which specific included study, the reviewer refers with the term “ clinically diagnosed pneumonia”. In our text we point out that all included studies stated that they used CXR in diagnosis even thought they did not explicitly define a “case of pneumonia” on the basis of changes in CXR.
Objectives			
Methods	Clearly explained and very useful (CR1) **Types of studies** – while it may be true that there is no plausible effect of a placebo on pneumonia, blinding is important particularly where the diagnosis of pneumonia is based on subjective clinical findings, rather than an objective	Please address.	**HH+PL:** We do not argue that the role of placebo in placebo-controlled trials is just to control for the “placebo effect” of the patients. We describe the essential features of the included studies. We describe Kimbarowski study as follows: “CXR ("Röntgenoscopie") was explicitly mentioned in the paper as a method that was used. It is probable

	test eg CXR. Hospital acquired and community-acquired pneumonia should be defined (the usual definition is arising within or after 48 hours of admission)		that the diagnosis of bronchopneumonia was based on the CXR but this was not explicitly stated in the paper." Glazebrook et al. wrote: "When a youth felt ill he was admitted to Sick Quarters unless his complaint was very mild. . . The admission to and discharge from the hospital was not under our control" (pages 13 to 14). [As to pneumonia:] "These cases were subjected to special investigations by us (X-rays, etc.) to establish certain criteria for the diagnosis" (page 16)." Mochalkin used CXR as shown for example the reporting of recovery of CXR changes. Although the studies did not define a "case of pneumonia" all included studies used CXR and we do not see what is the basis for reviewer to state " the diagnosis of pneumonia is based on subjective clinical findings" Reviewer writes: "Hospital acquired and community-acquired pneumonia should be defined (the usual definition is arising within or after 48 hours of admission)" There are different types of definitions. Pragmatic and unambiguous definition is needed for a new RCT, and the RCT protocol has to define hospital-acquired pneumonia that is used in the particular study. Then we have general definitions at the level of

			concepts which often are not exact. Hospital-acquired pneumonia refers to cases which occur because of the high exposure of viruses and bacteria in hospitals and/or the particular decreases of the immune system of the patients because of the disease that took the patient to the hospital. Although that concept is clear, setting exact border between community and hospital pneumonia is not possible. Although define pragmatic outcomes and inclusion criteria for a single RCT is possible, it is not possible in the same way in systematic review. Lets say we define in a systematic review that adults are 18 years and older, and children are younger, and we intend to carry out a subgroup analysis by age groups. If we find an RCT that has patients in the age range 15 to 21 years, what should be done with the study. Is it a study on adults or children, or should it be ignored (which would lead to bias). Thus, even though it is reasonable to carefully plan a systematic review, the definitions can be useless, because the original RCTs had their own definitions. Cecil's textbook of Medicine (2014) defines hospital-acquired and other categories of pneumonia as follows (p 588): “**DEFINITION** Community-acquired pneumonia includes cases of infectious pneumonia in patients living independently in the community. Patients who have been hospitalized for other reasons for less than 48 hours before the development of respiratory symptoms are also considered to have community-acquired pneumonia because it is likely that the inoculation had occurred before admission. However, patients who have

			previously been hospitalized for at least 2 days within the 90 days before infection; patients from nursing homes who received intravenous antibiotic therapy, chemotherapy, or wound care within the past 30 days; and patients from hemodialysis centers are considered to have health care–associated pneumonia and are therefore excluded from the case definition of community-acquired pneumonia. Patients contracting pneumonia more than 48 hours after the institution of endotracheal intubation and mechanical ventilation are also excluded inasmuch as they are considered to have ventilator-associated pneumonia. These distinctions are important because they help define the most likely infectious agents and hence strongly influence appropriate choices for the initial antibiotic therapy.” Thus, reviewer writes “(the usual definition is arising within or after 48 hours of admission)” However, the standard definition is pneumonia cases occurring after 48 hours, and not “within” 48 hours. Although we consider that definitions are usually important, we do not consider that in the context of our review, detailed definitions are useful for the average reader. In the case of the Tanaka study, all patients were burn patients and therefore they should not be classified as community-acquired pneumonia. In Table 3, Tanaka reports “No. of patients with pneumonia (within 2 wk)” Thus, if we would use a definition that hospital-acquired pneumonia needs at least 48 hours hospital stay, we could not draw any conclusions of Tanaka's pneumonia cases, since they do not describe anything else than “within 2

			weeks". Thus, it is theoretically possible that many pneumonia cases started "within 48 hours" in which case the pneumonias would be a combination of community and hospital-acquired. However, the clinical context of severely burned patients made us to classify the Tanakas cases as hospital-acquired pneumonia, even though we do not know the timing of the pneumonias in relation to hospital admission. We slightly modified our text on community and hospital pneumonias.
	Types of interventions – might be worth defining high dose and low dose vitamin C **Baseline balance** – I don't think this is a substitute for randomisation. Baseline balance can be achieved in a well matched cohort, but the point of randomisation is to balance both known and unknown covariates.	Agree	In our Methods, subgroup analyses, we write: "We set the limit of subgroup analysis to 100 mg/day in the preventive trials, since it is close to the dosage leading to maximum vitamin C plasma levels in healthy people. We set the limit of subgroup analysis to 1000 mg/day in the treatment trials, since there is evidence of changes in vitamin C metabolism in infections and larger doses might be needed for significant therapeutic effects." Thus, the low and high limits depend on the clinical context and in both cases the limits are arbitrary. We do not propose that baseline balance is a substitute for randomization. Randomization can lead to substantial unbalance in small studies,

			which leads to invalid comparison, because of the lack of baseline balance. Allocation by birthday leads to baseline balance in large studies in the same way as randomization. Thus, randomization is no guarantee for baseline balance and baseline balance can be achieved by methods other than randomization. “Baseline balance can be achieved in a well matched cohort, but the point of randomisation is to balance both known and unknown covariates.” This is not relevant issue in the context. If we construct a comparison of smokers vs non-smokers, it is true that we can select the same number of males and females to both groups, and we can match the age of the participants in the groups. However, in epidemiology we usually collect much data and use statistical adjustments for the differences between the groups, since it is more and more more difficult to construct the groups, the more variables one wants to match. But that issue is irrelevant for our own review. “the point of randomisation is to balance both known and unknown covariates.” That is the goal or randomization, but as noted above, when the studies are small, random allocation leads to baseline imbalances purely by definitions. Thus, randomization in small studies does not guarantee balance of known or unknown variables, but they must be measured if they are considered to be relevant. We do not see that the reviewer comment

			challenges the validity of our analyses.
	The statistical editor should review the measures of treatment effect (particularly the issue of continuity corrections). While the Peto method is an accepted method of analysis, I'm not sure if it can be applied to such small numbers.		The statistician (below) does not criticize the use of Peto method. We do not claim that the Peto method is ideal, but if we wish to present the findings as a forest plot, we must select from the options that are available in RevMan and Peto option seemed the best option for our data. Reviewer does not suggest any other option instead.
	Unit of analysis issues – given that randomisation was by unit, clustering effects need to be considered. The statistical editor should review this (I'd suggest Jo MacKenzie at the Cochrane Centre, who has worked in this area). I don't accept that just because an outbreak of tonsillitis affected all units in the past, there is not likely to be clustering. (PR)	Please respond	First, we do not see any evidence in the Glazebrook report that the units were allocated by randomization. Second, we discuss the unit of analysis issue in Methods. In addition to tonsillitis incidence, the authors wrote (see our text): "The youths of one division worked as a unit, and occupied certain tables in the dining hall. To some extent each division occupied particular dormitories but this separation was not absolute and there was a fair amount of mixing of divisions in the sleeping quarters." Thus, the administration units cannot be considered as groups that had no connections to each other. Third, in the Methods section we also write:

			“we also analysed the data by administrative units as a sensitivity analysis (see Results).” and describe that approach in Results.
Results	Well covered and explained (CR1) Pitt 1979 – I note that this trial was not analysed on an intent to treat basis.	Please address	**HH+PL:** We do not understand what the reviewer means. ITT has two components. 1) all randomized participants should be analyzed 2) participants should be analyzed in the groups to which they were randomized. Item 1 can often be satisfied with studies that are small and short, and with large and long studies if they use outcome which is available from registers, such as death or diseases in hospital registers. The latter are available even if a participant drops out. Large and long RCTs essentially never can satisfy item 1, if the participant needs to be contacted, since essentially always some participants drop out during a long trial. Item 2 means that eg a patient who is randomized to surgery, but does not go to surgery because of very poor condition, should be analysed in the group in which he or she was randomized. Item 2 is relevant eg studies with surgery, but much less frequently in trials with drugs, since in drug-trials there rarely is change of group during the trial. If there is an adverse effect, the drug is simply discontinued. Thus, item 2 is not a relevant concern eg in vitamin C studies.

			In the Pitt study, there were participants who moved out from the unit and they were not available according to item 1. However, all participants were analyzed according to item 2. We do not understand what the reviewer means with the comment about ITT in the Pitt study.
	Kimbaraowski 1967 – How was bronchopneumonia defined? This is usually a pathological or radiological diagnosis.		Reviewer: “Kimbaraowski 1967 – How was bronchopneumonia defined?” We do not understand what this comment means. In our text we describe the Kimbarowski study and write ”CXR ("Röntgenoscopie") was explicitly mentioned in the paper as a method that was used. It is probable that the diagnosis of bronchopneumonia was based on the CXR but this was not explicitly stated in the paper.” Also, we state that an English translation of the report is available. Thus, we describe that it seems to use that the diagnosis was based on CXR, but that was not explicitly stated.
	Glazebrook – is it plausible that there was blinding? (doesn’t vitamin C have a distinctive taste?).		We do not understand this comment. A pure vitamin C as ascorbic acid has acidic taste. However, when a small amount of vitamin C is added to cocoa or milk, the cocoa and milk neutralizes the acidity (of the small amount of vitamin C). We write that the authors described:

	This trial is clearly cluster-randomized.		"Pure ascorbic acid powder was added to . . . the morning cocoa, and an evening glass of milk. The mixing was done in bulk in the kitchens before issue. The powder dissolved quickly and easily, and did not alter the appearance or taste of the vehicle" (page 7) Thus, we do not understand whether the reviewer did not read that text, or whether the reviewer disagrees with the original authors statement about taste. Reviewer writes: “This trial is clearly cluster-randomized.” We do not understand this comment either. When we read the report, we cannot see any basis to state that the study was randomized. The groups could be allocated by other means as well. We explicitly note that intervention was administered in units. In Methods we write: “Unit of analysis issues The Glazebrook 1942 study reported the number of pneumonia cases per seven administrative groups of the school. Thus, the unit of analysis was the group of schoolchildren in the administrative division...” Thus, we do not understand if the comment is intended as a suggestion for some changes or what is the meaning of the reviewer comment about clusters.

	Hunt – 57 patients over 3 years seems very low, raising the possibility of selection bias Tanaka – This appears to be cluster randomized by month,		Reviewer: “Hunt – 57 patients over 3 years seems very low, raising the possibility of selection bias” Here the reviewer seems to confuse between concepts. Selection bias occurs when there are systematic differences between the compared groups at baseline, so that the baseline differences explain the differences (or part of them) in the outcome. For example, if we allocate many more males (than females), or more sick people (than less sick people) to one group, then we have selection bias. Low frequency of including participants makes a study long, but has nothing to do with selection bias. Reviewer: “Tanaka – This appears to be cluster randomized by month, raising obvious problems as the incidence of pneumonia and influenza is strongly seasonal.” We do not understand this comment. First, a season covers several months so that higher and lower incidence in different seasons is balanced by taking shorter periods (months) for treatment and control within the seasons. Second, Tanaka study was about hospital-acquired pneumonia, which should not have any relation to the seasonality of influenza and

	raising obvious problems as the incidence of pneumonia and influenza is strongly seasonal. Excluded studies – should studies that didn't have any events be included for completeness? (they obviously won't contribute to the outcome)	Studies meeting	community-acquired pneumonia. Third, in our text we write: "the methods section states that "randomisation was performed according to the month of admission" suggesting that randomisation may have been used as a synonym for allocation." Thus, we do not agree with reviewer that the months were "randomized" for the different treatments. There may have been simply alternative allocation by month. Fourth, Tanaka did not find effect of vitamin C on the incidence of hospital-acquired pneumonia. Thus, we do not understand the reasoning of the reviewer that seasonality of influenza might explain the lack of effect of vitamin C on hospital-acquired pneumonia. How could seasonality explain lack of effect? If there is less pneumonia in one group and more in another, then we can ask whether the difference is explained by unbalance of seasons between the groups, but we cannot see that seasonality could explain lack of effect, in particular we cannot see that seasonality in influenza could have anything to do with the occurrence of pneumonia in severe burn patients. We do not quite agree. First, as the reviewer states, inclusion of studies

		inclusion criteria should be included. Please comment.	that reported no events has no influence on the analysis. Second, it would be confusing to the ordinary reader when authors discuss studies that have no contribution to the topic of the study. Third, we do not see which specific studies the reviewer has in mind, ie which excluded studies are consistent with our inclusion criteria but which are excluded for lack of events. The border between listing a study to the excluded list instead of simply ignoring the study is quite arbitrary. In our view the excluded list should include studies that are close enough to the included studies so that a reader can see the reason for their exclusion if he or she might know one of more of the excluded studies.
	As it is likely that there is an interaction according to baseline vitamin C status, should baseline deficiency be a subgroup of interest?	Comment	We do not quite understand what the reviewer means in practical terms. In our protocol (see methods section of the current review) we planned that we will carry out subgroup analysis by dosage if there are suitable studies available. Thus, vitamin C dosage, and also dosage from diet, is a relevant issue. Low baseline vitamin C intake in diet is obviously interesting since greater effects of vitamin C may

			be expected. Low dietary vitamin C intake is our explanation for the effect seen in the Glazebrook study. However, low baseline vitamin C is not the only difference between eg Glazebrook and Pitt studies, so that we cannot see any rationale to present them as a high diet and low diet vitamin C subgroups, ignoring all the numerous other differences. Measuring dietary or plasma vitamin C in future RCTs is well justified, but we cannot do any within-trial subgroup analyses by dietary vitamin C. Neither can we do any meaningful between-study subgroup analysis by baseline deficiency.
	The statistician should review the use of the “mid-p” method – I’m not familiar with this method.	Please explain	In brief, if we have a 2x2 table which has 1 observation in each cell, the 1-tailed p-value should be 0.5 to reflect that the distribution does not favor benefit or harm. However, the Fisher test P-value for such a table gives P(1-tailed) = 5/6 = 0.833. This discrepancy between no difference and the P-value is corrected by counting only half of the probability of the observed table. This is the mid-P approach which gives 0.5 for such a table. In the Methods we give references, but we do not think a description of the method is relevant in the context.
	Secondary analysis (PDF page 15) – decrease of 4.6 days – does this refer to the difference in mean duration of stay, or difference in the mean time to resolution of illness or other? (PR)	Please explain	Slightly rewritten

Discussion	Very well addressed, and related back well to issues raised in Background section (CR1) **Overall completeness** – I would only consider the Pitt and Hunt trials to be even vaguely reliable.	Please address	**HH+PL:** We do not understand this comment. The Pitt and Hunt studies were randomized double-blind and placebo controlled studies. We do not see the basis to state that they are only "vaguely reliable". We discuss the shortcomings of the other studies and whether they might explain the reported differences. Although we are skeptical of the GRADE system as a mechanistic way to evaluate studies, the items of the scale are reasonable. One of the items is large magnitude of an effect. When RR<0.2, the GRADE suggests that quality of evidence "may increase 2 levels". That applies to Glazebrook and Kimbaroski studies. Curtis Meinert is a famous and experienced clinical trials and in one of his books he commented on "valid and invalid criticism" He wrote that: "A criticism, to be valid, should: •Have some basis in fact •Be buttressed with supporting evidence •Make a difference in the interpretation of the results. All three tests should be met. 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"

			http://www.mv.helsinki.fi/home/hemila/M/Meinert_1986_p276.pdf This comment is consistent with the GRADE item about the large magnitude of the effect. Thus, when there is truly dramatic difference between two groups, there should be truly dramatic difference in important baseline variables. Reviewer does not formulate any arguments why the findings in the Glazebrook and Kimbaroski studies should be ignored.
	Was vaccination (against influenza or pneumococcus) considered as a potential confounder?		We do not understand this comment. In epidemiology, potential confounder means a third factor that is associated with both exposure and outcome. In controlled trials the goal is that all baseline variables are equal and therefore there cannot be a problem of “potential confounders” if the groups are balance on baseline. Vaccination might modify the effect of vitamin C, however, the studies do not give any information on effect modification. Testing subgroup effects would need much larger groups of people compared with testing an overall hypothesis. According to Medscape “In 1942, a bivalent vaccine was produced after the discovery of influenza B”. http://www.medscape.com/viewarticle/812621 According to “historyofvaccines” page, “A

			pneumococcal vaccine that protected against 14 different strains was licensed in 1977, and expanded to protect against 23 strains in 1983." http://www.historyofvaccines.org/content/articles/pneumococcal-disease-0 Pitt (1979) vaccinated participants against adenovirus 4 and influenza, as we describe in our review, but pneumococcal vaccinations were not available widely at that time. In any case, there was no difference in influenza vaccination between the vitamin C and placebo groups and therefore vaccination cannot have been a confounder. Glazebrook could not have used influenza or pneumococcus vaccination. Kimbarowski could not have used pneumococcus vaccination; in the text we do not find any mention of influenza vaccination.
	Diagnosis of pneumonia – an acute exacerbation of chronic bronchitis does not imply a bacterial aetiology.		The reviewer does not give any justification for that statement and many authors have a different opinion. For example, "The syndrome of COPD consists of chronic bronchitis (CB), bronchiectasis, emphysema, and reversible airway disease that combine uniquely in an individual patient. Older patients are at risk for COPD and its components--emphysema, CB, and bronchiectasis. Bacterial and viral infections play a role in acute exacerbations of COPD and in acute exacerbations of CB without features of COPD" http://www.ncbi.nlm.nih.gov/pubmed/20398122 "Daily azithromycin decreases acute exacerbations of chronic obstructive pulmonary disease, but

	Publication bias – the fact that pneumonia was a secondary outcome does not preclude the possibility of publication bias – were the primary analyses positive?		long-term side effects are unknown.” http://www.ncbi.nlm.nih.gov/pubmed/24779680 “There is increasing evidence for the role of bacterial infection in causing acute exacerbations of COPD, particularly in patients with chronic bronchitis who present with all three cardinal symptoms defined by Anthonisen et al. (in Introduction, not in Abstract) http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2104610/ Therefore we disagree with reviewer. It seems to us that the reviewer misses our point. Publication bias means that a finding is reported because the finding is positive, but the researcher would not publish it if the finding was negative. If researchers publish a finding just because it is positive, they want to inform other researchers of that positive finding. It is not logical to propose that secondary findings that have no relation to the main message of the report would be explained by publication bias. For example, Kimbarowski were interested in a laboratory method (in which we have no interest), but pneumonia was a nuisance for them. They did not report the number of pneumonia cases in the study groups with an intention to convince readers that vitamin C was effective against pneumonia. They did not even analyze the difference between the groups by the chi-square or the Fisher test. When researchers do not calculate the P--value to find out whether it is more or less than 0.05 and they are fully uninterested in the difference, it does not make sense to speculate that there is

			publication bias. Glazebrook reported that vitamin C did not influence the incidence of the common cold and in that respect their primary-interest analysis was negative. Pitt reported no effect of vitamin C on the incidence of the common cold and in that respect their primary-interest analysis was negative. Pitt did not report the P-value for the pneumonia difference and thus the P-value could not be the basis for reporting or not pneumonia.
	Low interest in hypothesis – this is probably true, but I don't think it is relevant to the study question. (PR)		We do not see the point of this comment. In our comments on potential publication bias, we note that all three study groups of prophylactic trials were not interested in their pneumonia findings. We state that this is also an argument against publication bias. Given that publication bias means that authors publish findings on the basis that they are positive, low interest in the findings is inconsistent with such reasoning. We are ready to revise our text, but we cannot see any point in this comment.
	Since IV vit C has been the focus of study by 2-time Nobel Prize winner Linus Pauling, a		CR2: "The doses shown in these studies are quite low compared to what Dr. Pauling studied, wrote about and published." Except for the Pitt-study, that comment is correct

	nod ought to be made to the idea of IV studies. The doses shown in these studies are quite low compared to what Dr. Pauling studied, wrote about and published.		but the comment is not relevant on our review. First, another Cochrane review on vitamin C and the common cold has shown that 1-3 g/day vitamin C does not prevent colds, which refutes Pauling's most optimistic proposals. Pauling promised protection against the common cold in particular, and made much less comments on other infections. Second, the included studies that used the low doses found benefit, ie Glazebrook, Kimbarowski, Hunt. Thus, those studies may have been low compared to what Pauling wrote about, yet the findings were positive.
	A good point was made that vit C needs vit E to potentiate it. There is such a strong tendency to isolate and examine 'active' elements/ingredients' but this is a Pharma methodology and does NOT lend itself to 'natural ' substances. I have made this particular point many times publicly at cancer	Please comment I don't think this debate is very relevant here, but please comment.	CR2: "A good point was made that vit C needs vit E to potentiate it." That is false. An analysis of large scale ATBC Study found that in middle-aged people vitamin E increased mortality in participants who had dietary vitamin C intake high. http://www.ncbi.nlm.nih.gov/pubmed/19218294 Also, in two subgroups the combination of vitamin E and high dietary vitamin C increased pneumonia risk. http://www.ncbi.nlm.nih.gov/pubmed/19019244 In general, in a systematic review we are restricted to studies that have been published so far. In our Authors conclusion we can and we do propose paths for future studies, but the empirical basis for our Cochrane review is studies that have been published so far.

	meetings worldwide. **We NEED new methods to view traditional or 'natural' treatments. (CR2)**		We agree with the comment that RCTs may not be a good method if we are interested in the effects of life-style differences including nutrition on health over several decades from childhood to old age. However, we do not consider that our review is a correct place to discuss such issues.
Conclusions	This is very 'carefully' and 'correctly' written – I like the end of the discussion where it states: "… vitamin C might decrease the risk of pneumonia by 80% in some population groups that have a very high incidence of pneumonia. The dominant theory of vitamins in mainstream medicine is that the function of vitamins in the body is to prevent deficiency diseases [and therefore, other uses of vitamins belong to the domain of alternative medicine]. Thus, it seems possible that the low level of interest in the reported effects of vitamin C on pneumonia is caused by general prejudices against vitamins as medicines." (CR1) This is a thorough and careful review, but I disagree with the	Ok Please address As above	**HH+PL:** This criticism of our GRADE-values is unfair. In a previous version of the update-draft, we used classification that seemed verbally more appropriate for us. However, we were strongly instructed to follow the letter of the GRADE. We are not fond of the GRADE and in our previous responses to editorial office comments we referred to our Finnish colleague's paper in which he criticized GRADE. However, Cochrane chief editorial office responded that GRADE is the policy of

	conclusions. I simply don't accept that a small handful of cases in niche populations in only a small number of well conducted trials constitutes "high quality" evidence. While the authors may well argue that they are following the letter of the GRADE rules, to quote a well-known Australian movie, "it's the vibe" that doesn't match. Summary of findings – I do not accept that a single reasonably well conducted trial with 8 events constitutes high quality evidence. I also do not accept that the results of non-randomized studies can be regarded as high quality evidence,		Cochrane and we do not have an option not to follow it. If the reviewer points out that we have read and followed the GRADE recommendations incorrectly, we will do our best to correct our errors. However, if the reviewer personally disagrees with the GRADE system, he should focus his criticism to the Cochrane editorial office, or to the GRADE authors, or to J Clinical Epidemiology in which the GRADE system was published. Furthermore, reviewer confuses between different issues. "niche populations" is an irrelevant issue for the validity of studies and conclusions based on the studies, but "niche" is related to generalization, which is a fully different issue. Thus, studies with niche populations can firmly show that a treatment is beneficial in those niche populations. Such findings should not be extrapolated outside, however. In our summary of findings table we clearly describe the population characteristics of the included studies. Reviewer also has a misunderstanding what surrogates mean in clinical trials: "the results of a surrogate outcome (symptom score)" There are lots of controlled trials that measure laboratory outcomes or radiological findings. They are not always related to clinically related outcomes. Therefore, in the analysis of controlled trials, we should focus on outcomes directly relevant for the patients, and not on surrogate outcomes.

	and that the results of a surrogate outcome (symptom score) in subgroup analyses (severely ill patients) can be regarded as moderate quality evidence. Implications for practice – agree that it is reasonable to treat vitamin C deficiency but is it expensive to test vitamin C levels? (PR)		Clinically relevant outcomes mean for example symptoms; eg whether the patient has pain, difficult cough, is tired, has difficult to breathe etc. These are symptoms. Thus, a symptom score is the exactly the kind of outcome that is clinically relevant. Furthermore, the findings on the symptom score were consistent with the findings on mortality though the latter difference was not significant. Testing vitamin C level in hospital setting is very cheap compared with all other expenses. Screening of vitamin C level in ordinary people is a different question from cost benefit view; although the test is cheap its implications are not clear.
Declarations of interest			**HH+PL:**
Summary of findings Table and	I'm glad that someone else is reviewing these tables, because the tables should be presenting baseline risk and risk under intervention. Presenting only odds	Please address	**HH+PL:** We do not understand this comment. If we assume, say, that a treatment has a fixed effect, say 70% reduction in the outcome in all

GRADE	ratios doesn't add to the Data Summary tables. Also, in Table 1, the estimated risk can be summarised across the studies (as they did in the meta-analysis), and different scenarios for baseline risk considered, mirroring the heterogeneity of participants. These tables need a thorough revision. (SE)	Please address	population groups, we can simply calculate the proportion of people who benefit (or NNT) for a variety of assumed baseline risk levels. However, we cannot see any scientific basis to assume that the 100% effect in Glazebrook study in 1940s, the 80% effect in Kimbarowskis influenza patients behind the iron curtain in the 1960s, or the 85% effect in the Pitt study with US marines can be taken as valid figures which we could use to multiply any selected baseline risks. Furthermore, even if we would consider that some of the estimates is generalizable, all of them have such wide 95%CIs that the resulting estimates for NNT would be highly inaccurate. The other reviewer used term "niche populations" and that seems quite appropriate term to describe all the three studies. For vitamin E there is direct evidence of non-uniformity of effect on pneumonia. http://www.ncbi.nlm.nih.gov/pubmed/21386974 When the over 20000 participants of the ATBC Study were divided to eight group on the basis of smoking and exercise, there was very strong evidence that the effect of vitamin E was not uniform (chi-sq[7 df] = 26.6, P = 0.0004). Therefore we have good basis to assume that the effect of vitamin C on pneumonia is not uniform so that there is a fixed effect which we could use to multiply a diversity of baseline risks to reach the incidence of pneumonia when on vitamin C. Therefore the request that "the tables should be presenting baseline risk and risk under

			intervention.” is not valid.
	Kimbarowski – more detail about bronchopneumonia definition is required – it would be hard to tell clinically if influenza was complicated by bronchopneumonia.		“Kimbarowski – more detail about bronchopneumonia definition is required” We do not understand this comment. We are restricted to the data that have been published. We have arranged a translation of the report and thus the reviewer can read the original report to see whether there is something that we could add. We write “CXR ("Röntgenoscopie") was explicitly mentioned in the paper as a method that was used. It is probable that the diagnosis of bronchopneumonia was based on the CXR but this was not explicitly stated in the paper.” CXR was a standard method at hospitals at that time and from the reading of the report we conclude that it is highly likely that CXR was used in the diagnosis of pneumonia. In particular, Kimbarowski writes about “bronchopneumonia” which is an X-ray diagnosis also implying that they used CXR..
	Mochalkin – if the review is of vitamin C vs no vitamin C, then it would be preferable to compare both vitamin C to placebo. This would seem a post hoc justification.		In the protocol we planned to compare prophylactic vitamin C vs placebo/no-placebo, but we required placebo for the therapeutic comparison. We feel that in the Cochrane context there is sometimes paranoiac attitude towards post hoc changes in reviews and therefore we do not propose a change to allow therapeutic comparison against no-placebo. If the editor considers that such a revision improves the review, we are ready to revise the presentation of the Mochalkin study.

	Lafargue – what is the proposed sample size, and has this trial concluded? (PR)		The clinicaltrials.gov record does not give the number of planned participants Not published as far as we know. The clinicaltrials.gov record describes “February 2016 (final data collection date for primary outcome measure)”
Additional comments	Would be really good if the Background section could be tightened up as the remainder of the paper is excellent and very clear about the limitations of the existing identified studies but also their usefulness	Please review with this in mind	**HH+PL:** We are ready to revise the Background if there are more specific instructions. However, Description of the condition is 156 words. Given the frequency and importance of pneumonia, and the complexity of the disease as an umbrella concept for various viral and bacterial diseases, with greatly varying severity, and the problems in the clinical diagnoses, in our view this section is too short rather than too long. However, in the context we do not consider it worthwhile to extend it. People find easily reviews of pneumonia elsewhere. Description of the intervention is 1813 words. This is quite long, we agree. On the other hand 1800 words is not a lot on a complex topic. Given that many/most people, including physicians, believe that the only physiological effect of vitamin C is to prevent and cure scurvy, we need to give proper background for our topic: why is it plausible that vitamin C might have an effect on pneumonia. If a reader reads our review with a mind set that “vitamin C is effective only on scurvy”, he or she

			may be biased against other possible effects of the vitamin. For research topics that do not have conceptual problems, the description of intervention may be very short. For example, if a Cochrane review investigates whether 3 days vs 5 days treatment of otitis media with antibiotics makes any difference, there is not much need to cover the antibiotics background in detail. To make it easier for the reader, we have used several subtitles to Description of intervention. The introductory paragraph is 102 words Physiology 432 words Pharmacokinetics 347 Heterogeneity in vit C effects 455 Safety 159 Specific history of vit C and pneumonia 283 All these sections are written briefly. Personally we are satisfied with the current version. However, it is possible to shorten one or more of them, but we would like to have more specific suggestions. How the intervention may work has 633 words, and thus it is much shorter than the preceding section. This section has the same rationale as the preceding, ie to show that the effect of vitamin C on pneumonia is a plausible concept. Here too, it is possible to shorten, but we would like to have more specific suggestions.

	The plain language summary does not help as it follows the layout of the review rather than the important messages (CR1)		The structure of the PLS is closely defined by the instructions that we have received. Personally we would prefer a more liberal format so that we could write about what we consider most important messages ourselves.
	My suggestions about providing information regarding how the other sources were discovered was an attempt to improve the ability to replicate the systematic review. Someone else being able to replicate the results of a SR is why the reporting of methodology used is so important. In a case where expert knowledge, either the author/s or an outside expert/s, is used to add studies to a SR it is worthwhile to let people know, as this is very hard to reproduce unlike a literature search or checking reference lists. Therefore if someone goes to attempt to reproduce the SR they know which studies can be found using reproducible methods and which cannot. To this end it could be an idea to provide information regarding which	Can you be more specific about who contributed?	

	experts contributed their knowledge to locating information for the SR. So although the explanatory note accompanying the PRISMA flowchart is excellent, it could possibly be strengthened by noting the experts. For instance by changing it from: Flow diagram. **Other sources indicates literature searches other than those carried out for this particular review, reading reference lists of various journal articles and books, and reading the publications themselves. The Cook 2007 and Sesso 2008 trials are listed as potentially relevant, because they were particularly large and long studies on vitamin C, the authors were contacted to ask whether any data on pneumonia might be available.* To Flow diagram. **Other sources indicates the knowledge of the following experts, (expert 1, expert 2 etc.) who provided additional studies from literature searches other than those carried out for this particular review, reading reference lists of various journal articles and books, and reading the publications themselves. The Cook 2007 and Sesso 2008 trials are listed as potentially relevant, because they were particularly large and long studies on vitamin C, the authors were contacted to ask whether any data on pneumonia*		We understand the comment by trials search coordinator. However, he does not seem to understand how passive field this is. There are hundreds of studies being published all the time on diabetes type II, hypertension, and cholesterol. In such a case it is a good idea to contact the most active researchers and ask what is going on and whether there are studies that we have not identified, eg might be unpublished by those researchers etc. Vitamin C and pneumonia is very quiet field. Our first version was published in 2007. No studies falling under our inclusion criteria have been published after that, ie during one decade. Compared with the first version, we have added one study (Tanaka 2000) with rational described in the differences between protocol and review, but that was published 7 years before our first version of the review. PubMed literature search for “vitamin C and pneumonia” gives 57 hits since the very beginning of the PubMed. Over the last five years there are 11 hits with the search, and none of them are relevant for our review, except our own Cochrane review in 2013. Thus, we do not know any experts on this field whom we could contact with the goal of writing thereafter: “Other sources indicates the knowledge of the

	might be available. I should reiterate that the whole point is to be as transparent as possible regarding the provenance of information in a SR to aid with the ability to reproduce the results of that SR. (TSC) I found this review to be quite interesting and thorough in its exploration of the included studies. I will restrict my comments to the statistical points. **On baseline balance:** I appreciated seeing the additional information on baseline balance, and it adds to the overall picture of the study. As long as it is clear that baseline balance does not substitute for randomisation, or is used to alleviate concerns about the lack of randomisation, it seems to a good addition to the		following experts, (expert 1, expert 2 etc.)...” The other sources which we show in the flow diagram are studies that we have ourselves found from different directions. We understand the goal of “to be as transparent as possible” This does not seem to need a response. These responses to statistical reviewer are by Harri alone:

	review. **Subgroup and Sensitivity Analysis**: The authors claim they considered a sensitivity analysis by taking into account the potential clustering effects in Glazebrook(1942). However, they assume a Poisson distribution with no over-dispersion,	Please make sure this is clear in the review. Please address. Please address	“they assume a Poisson distribution with no over-dispersion” That is correct. However, compared with the theoretical Poisson-distribution, the actual distribution in the Glaebrook study is under-dispersed, and not over-dispersed as the statistical reviewer suggests. For the theoretical Poisson distribution, variance = mean. For the control groups of Glazebrook study (observations in five units: 5, 3, 2, 4, 3) mean= 3.400 variance=1.300 Thus, for that data, variance is substantially smaller than mean. If the actual distribution is “over-dispersed” then the Poisson approach would lead to over-optimistic results. However, when the actual distribution is “under-dispersed” the Poisson approach leads to conservative results. Thus, in our case the evidence of difference between the two vitamin C units and the five control units is even stronger than calculated by the assumption of the Poisson-distribution. We considered two further tests for the difference

			between the two vitamin C units and the five control units.
			
			Given that the variance is much smaller than mean, and the distribution in control groups is quite symmetric around the mean, we also used: t-test, which gives P = 0.003 for:
			****
			`> t.test(Glaz7$pneu ~ Glaz7$treat, var.equal = F)`
			`Welch Two Sample t-test`
			`data: Glaz7$pneu by Glaz7$treat`
			`t = 6.668, df = 4, p-value = 0.00263`
			****
			
			However, even though the distribution is quite symmetric in the control units, the t-test is not ideal since the potential number of cases in the units is restricted by the requirement that they cannot be less than zero, whereas t-test is most reasonable for genuinely continuous variables
			
			Therefore, we also used the Fisher-Pitman permutation test to compare the two vitamin C units against the five control units.
			This permutation test also gives a significant difference between the vitamin C and control units:
			****
			`> oneway_test(Glaz7$pneu ~ Glaz7$treat, distribution="exact")`
			`Exact Two-Sample Fisher-Pitman Permutation Test`
			`data: Glaz7$pneu by Glaz7$treat (0, 1)`
			`Z = 2.136, p-value = 0.0476`
			****
			
			We added the permutation test to our review, since it is the most robust method to compare the pneumonia counts/incidences between the units. Nevertheless, we kept the Poisson-distribution

	so it would follow mathematically that the “cluster” analysis would give the same result as the original analysis.	Please address	analysis, since assuming an average number of counts per unit of observation time in a population group is statndard approach in epidemiology. “it would follow mathematically that the “cluster” analysis would give the same result as the original analysis” That is not correct. The RCTs are usually analyzed assuming binary outcomes per participant and the SE for RR calculated eg in RevMan follows the formula: SE(ln[RR]) = sqrt(1/a + 1/c – 1/n1 -1/n2) see page 2 in http://tech.cochrane.org/revman/documentation/Statistical-methods-in-RevMan-5.pdf Poisson RR values assume Poisson outcomes per unit of person years, which leads to a different SE-estimate. SE(ln[RR]) = sqrt(1/a + 1/c) This is standard formula in epidemiologic textbooks, eg Rothman-Greenland (1998, p. 238) Furthermore, since the number of cases in the Glazebrook vitamin C units is zero, in our Analysis 1 forest plot we are using the Peto-method for estimating effect and its 95%CI, which uses a third type of SE calculation: SE(ln[OR]) = sqrt(1/V) and there is a complex formula for V, which can be seen through the link given above. Since pneumonia is rare in the population, RR and OR are essentially identical.

			Finally, since the number of cases in the vitamin C units is zero, the exact method for calculating the 95% CI for the OR is even better than the Peto approach. Thus, it is quite inaccurate to claim that: “it would follow mathematically that the “cluster” analysis would give the same result as the original analysis” although all the described approaches asymptotically lead to identical results when the groups are large and the number of cases in the study groups differ much less than in the Glazebrook study.
		Already mentioned above	Our primary “original analysis” is the calculation of the exact 95% CI and mid-P, and our analysis of the seven units shows that even if they are considered as separate statistical units of analysis, the difference between vitamin C and control is not easily explained by random variation. We added the permutation test, which is the most robust comparison of the units.
		Please address	We do not understand what statistical reviewer means by considering between-study variation in rates. If the reviewer means the two other prophylactic studies, they are so very different contexts that we cannot take their pneumonia rates as some kind of comparison what to expect in a boarding school during WWII.

			We are sorry for the confusions. We had thought that the R programs are unambiguous and we assumed that in the google-world it is easy to trace the programs. We corrected to “or.midp” and we added the package information. “fisher.exact” is part of the exact2x2 package
	It would make better sense for the review authors to comment on the between-study variation in rates. If this variation is similar to what one would expect under a Poisson distribution (which is appears to be), then enough said.		The package name was added to the Methods section.
	I could not reproduce the meta-analysis for the primary endpoint in the prevention studies: incidence of pneumonia. I found a function called “or.midp” in the package “epitools” in R, and assumed that the function that reported “or.mid” was a typo. I could not find the function “fisher.exact” in R. The authors should report all information for other programs so that the analyses can be reproduced. In		The function or.midp **gave different OR estimates** and 95% confidence intervals for 2 of the studies. We do not understand this comment In our review we wrote “for Kimbarowski 1967, OR = 0.19 (95% CI 0.03 to 0.77; mid-P = 0.018)” Statistical reviewer gives 0.02677116 and 0.77345475 These are rounded to the 95%CI values which we

	the case of R, they should report the name of the package, and the name of the function. The function or.midp gave different OR estimates and 95% confidence intervals for 2 of the studies. My estimates were: > kimbarowski<-or.midp(c(2,10,112, 102)) > kimbarowski $x [,1] [,2] [1,] 2 10 [2,] 112 102 $estimate [1] 0.1940669 $conf.int [1] 0.02677116 0.77345475 $conf.level [1] 0.95		give “for Pitt 1979, OR = 0.16 (95% CI 0.01 to 0.95; mid-P = 0.044).” Statistical reviewer gives 0.006388236 and 0.950009546 These are rounded to the 95%CI values which we give

	attr(,"method") [1] "median-unbiased estimate & mid-p exact CI" > pitt<-or.midp(c(1,7,330,336)) > pitt $x [,1] [,2] [1,] 1 7 [2,] 330 336 $estimate [1] 0.163255 $conf.int [1] 0.006388236 0.950009546 $conf.level [1] 0.95 attr(,"method") [1] "median-unbiased estimate & mid-p exact CI"		We cannot understand what statistical reviewer means by statement: "The function or.midp gave different OR estimates and 95% confidence intervals for 2 of the studies." "It was disappointing that they didn't report an overall estimate." We disagree with this comment. It seems possible that the statistician does not see how much heterogeneity there is in pneumonia, so that the statistician just looks at the estimates and CIs. The 95%CIs are all wide for the three prevention studies, and widely overlapping CIs mean no statistical heterogeneity. Therefore, on the basis of statistical reasoning the three studies could be pooled to reach a single narrower estimate of effect. However, the clinical contexts of the three studies are so very different, that we do not see any basis

	It was disappointing that they didn't report an overall estimate. I think that one can use the inverse-variance method to get an overall estimate and confidence interval using log-odds ratio estimate, and an estimate of the SE of the log-odds ratio using the 95% CI for the log-odds ratio. Perhaps they could speak to their statistician about this possibility. (SE)		to consider that they measure the same phenomenon. Although "pneumonia" is the same word, pneumonia in US marine recruits, and in influenza patients behind the former iron curtain, and in UK schoolchildren during WWII are very different contexts. In our Methods section we write: "There is no statistical heterogeneity in Analysis 1.1, but the studies were clinically so divergent that pooling was inappropriate." In our Results we write: "The confidence intervals (CIs) in the three trials were wide and overlapped substantially and there is no evidence of statistical heterogeneity (I^2 statistic = 0%; Chi^2 [2 df] = 0.03). However, the trials were clinically so heterogeneous that we did not calculate a pooled estimate of the effect because we did not consider that such a pooled estimate was meaningful." We already responded to the reviewer's comment on GRADE-levels above.

	This is a thorough and careful review, but I disagree with the conclusions. I simply don't accept that a small handful of cases in niche populations in only a small number of well conducted trials constitutes "high quality" evidence. While the authors may well argue that they are following the letter of the GRADE rules, to quote a well-known Australian movie, "it's the vibe" that doesn't		That text section is relevant when considering the field, why there are no newer studies on vitamin C and pneumonia since 2000 though half a dozen studies found benefit. However, that text part is irrelevant to the scientific validity of our analyses.

	match. (PR) **Please note this statement from the paper is the one I am most fully in agreement with**: "Thus, it seems possible that the low level of interest in the reported effects of vitamin C on pneumonia is caused by general prejudices against vitamins as medicines. Goodwin 1998 gave several examples that illustrated the prejudices of the mainstream medicine towards vitamins."(CR2)		

CR = Consumer Referee
SE = Statistical Editor
PR = Peer Referee
TSC = Trials Search Co-ordinator

CR1 – Janet Wale
CR2 – Ann Fonfa
SE – Terry Neeman
PR – Allen Cheng
TSC – Justin Clark