Harri Hemilä 23 Dec 2009 (minor changes in June 3, 2010)
harri.hemila@helsinki.fi

Dear Dr Geneviève Chêne,

Please, find my counter-comments to the two comments.
My comments are below the copy of your email.

Yours

Harri Hemilä

-------- Original Message --------
Subject:PLoS ONE Decision [09-PONE-RA-14368]
Date: Mon, 21 Dec 2009 16:01:13 UT
From:PLoSONE@plos.org
Reply-To:genevieve.chene@isped.u-bordeaux2.fr
To:harri.hemila@helsinki.fi

Dear Dr. Hemila,

Thank you very much for submitting your manuscript "Zinc lozenges may shorten the duration of colds: a systematic review" for review by PLoS ONE. After careful consideration, we have decided that your manuscript does not meet one or several of our criteria and must be rejected for publication.

Specifically:
1. the outcome, i.e "common cold" is never defined, nor discussed, though a very subjective one.

2. statistical analysis of this meta-analysis of randomised trial is not standard (standard methods include: use of random-effect models, estimation of I^2 statistic, funnel plot to search evidence of publication bias).
Moreover, given the scientific question of interest and the subjective nature of the outcome, a meta-analysis on individual data would be preferable.

Unfortunately, due to the concerns described above, we cannot accept a revised version of your manuscript. I am sorry that we cannot be more positive on this occasion, but hope that you appreciate the reasons for this decision.
Sincerely,
Dr Geneviève Chêne
Academic Editor, PLoS ONE

1. the outcome, i.e "common cold" is never defined, nor discussed, though a very subjective one.

HH: There is no accurate definition for “the common cold”. Therefore this is not valid criticism for my manuscript. I can add a short discussion describing the common cold to the manuscript, if you consider relevant.

I am an author of the Cochrane review on vitamin C and the common cold
http://dx.doi.org/10.1002/14651858.CD000980.pub3
The Cochrane collaboration is specifically focused on systematic reviews; however, the editor and the reviewers of our review on vitamin C and the common cold have never criticized us for the lack of definition of the common cold.
Furthermore, the editor of the Acute respiratory infection group of the Cochrane collaboration is all the time working with various respiratory infections so that ignoring the exact definition for the common cold cannot be caused by his lack of familiarity with the respiratory infection field.
We have recently updated our Cochrane review on vitamin C and the common cold. In the update we are not giving any definition, but we briefly describe what the concept “common cold” means (see below a copy). This description also shows why it is impossible to give an exact and biologically meaningful definition for the common cold.
However, as noted above, I can add a short description of the common cold to the manuscript if you consider that it would be useful for readers.

*****

Description of the condition (from the Cochrane review on vit C and colds):

The term ‘the common cold’ does not denote any precisely defined disease, but this illness is familiar to practically everybody. Typically symptoms of the common cold consist of some combination of nasal discharge and obstruction, sore throat, cough, lethargy, and malaise, with or without fever. The common cold is the leading cause of acute morbidity and of visits to a physician in Western countries, and a major cause of absenteeism from work and school.

The common cold is usually caused by respiratory viruses (rhino, corona, adeno, parainfluenza, influenza, respiratory syncytial), which overall have some 200 serotypes (Eccles 2005; Gwaltney 2005; Heikkinen 2003). Thus, the term ‘the common cold’ does not refer to a single entity but to a group of diseases caused by numerous unrelated aetiological agents. The most frequent agent causing the common cold is rhinovirus, which is found in 30% to 50% of sufferers. In a third of subjects with cold symptoms, the aetiology remains undefined even when extensive virological tests are used. It is not clear to what extent this latter group is explained by the low sensitivity of the tests, unidentified viruses, or similar symptoms arising from non-viral aetiology, such as allergic or mechanical irritation of the airways. Different respiratory viruses have different symptom profiles, but the patterns are not consistent enough to validate aetiological conclusions from the patients’ symptoms.

Although the great majority of common cold episodes are caused by the respiratory virus group, the symptom-based definition of the ‘common cold’ also covers some diseases caused by other viruses (varicella, measles, rubella, cytomegalo, Epstein-Barr) and some bacterial infections. For example, since streptococcal pharyngitis cannot be differentiated from viral pharyngitis on clinical grounds, it can also be included within the broad definition of the common cold. Symptoms of illnesses caused by *Mycoplasma pneumoniae (M. pneumoniae)* and *Chlamydia pneumoniae (C. pneumoniae)* may also be similar to the symptoms caused by the respiratory viruses.

The manifestations of the common cold are so typical that usually the clinical diagnosis of the common cold can be made reliably by adult patients themselves. Allergic and vasomotor rhinitis can sometimes mimic the common cold, but usually these conditions can be easily differentiated (Heikkinen 2003).

In common cold trials an explicit definition of the common cold is used for logistic reasons; for example, based on the duration and the set of symptoms to yield an explicitly defined outcome. However, such limits are biologically arbitrary. There is no exact minimum duration or combination of symptoms which is meaningful when drawing a conclusion as to whether the symptoms should be explained by a viral infection, or by allergic or mechanical irritation of nasal airways or throat.

The use of antibiotics for a typical acute common cold episode is useless since the vast majority of colds are caused by viruses. Nevertheless, according to some surveys about 50% of common cold patients in the USA received antibiotics (Gonzales 1997; Mainous 1996). In this respect, the alternative treatment options for the common cold are of substantial public health interest.

****

I wrote a chapter on vitamins and minerals to a book “Common cold” which was published last Spring.

I am copying the introduction of the chapter because that discusses the same issue: the nature of the common cold. This is partially overlapping with the text above, but there are additional aspects which are relevant when considering common cold studies.

Hemilä H. Vitamins and minerals [book chapter]. In: "**Common cold** " (Eccles R, Weber O, eds.) Birkhauser Verlag, 2009:275-307

****

Introduction (of the book chapter mentioned above)

The term ‘the common cold’ does not denote any precisely defined disease, yet the symptoms of this illness are personally familiar to practically everybody. Although the great majority of common cold episodes are caused by the group of respiratory viruses, the symptom-based definition of the ‘common cold’ also covers some diseases caused by non-respiratory viruses and even some bacterial infections and allergies. The large number of etiological agents, the benign character of the disease, and the high cost of the virologic tests (*e.g.*, $ 700 per patient in one study [1]) mean that a functional everyday definition of the ‘common cold’ cannot be based on laboratory tests, but must be based on symptoms.

Furthermore, a chest x-ray has no relevance in excluding pneumonia when the patient is not seriously ill.

The liberal definition of 'the common cold' has implications for research in the general community. First, it is much cheaper to count the number of respiratory-symptom episodes and the days of illness compared with searching for the etiologic agent. Second, the general community does not have access to rapid tests that reveal the cause of the disease. Therefore a treatment that is focused on a specific agent cannot be efficiently used in the community anyway. Third, the rationale for vitamin and mineral supplementation is based on the assumption of non-specific effects on the immune system and against diverse infections. Thus, the symptom based definition is particularly appropriate when examining whether vitamins or minerals have non-specific effects relevant at the public health level.

The primary focus of this chapter is on the common cold type of symptoms; however, the border between upper respiratory infections (URI) and lower respiratory infections (LRI) is ambiguous. For example, computer tomography identifies many more cases of pneumonia compared with a chest x-ray [2], and thus a patient may have an URI simply because he or she has not been studied with sophisticated methods. In some trials all respiratory infections or all infections were combined. Those trials are not excluded from this chapter, because the great majority of infections in the general community are URI so that the wide definitions primarily measure the URI and the common cold.

Taking vitamins to improve health and the immune system is popular in the western countries. About half of the elderly in the USA take some vitamin or mineral supplements [3]. Therefore it is important to find out whether they have effects on respiratory infections. If vitamins or minerals are shown to be effective, their use may be encouraged. If they are ineffective, their use should be discouraged. I focus on the findings of controlled trials and describe the biological rationales only to a minor extent.

2. statistical analysis of this meta-analysis of randomised trial is not standard (standard methods include: use of random-effect models, estimation of I^2 statistic, funnel plot to search evidence of publication bias).
Moreover, given the scientific question of interest and the subjective nature of the outcome, a meta-analysis on individual data would be preferable.

HH: Standard means that the method is popular.

Standard does not guarantee that the method is scientifically sound.

Neither does non-standard imply that the method is scientifically unsound.

Funnel plot

I agree that “funnel plot” is a standard method (popular).
However, it is scientifically unsound.

As an early criticism to the Egger et al. (1997) paper [describing the funnel plot], Vandenbroucke wrote in the correspondence section of the BMJ (1998;316[7 Feb];469):

“If we accept the test, or any similar test of heterogeneity on meta-analyses, what should we conclude from it? The main message from it is that there might be a problem because the funnel plot is asymmetrical— which we also see on the plot. The real questions to which we would like an answer are: what is the cause of the asymmetry and, more importantly, which trials should we believe? The cause of the asymmetry can be anything, from publication bias, “willingness to please” during data collection, data massage in the analysis, unclear rules for stopping the trial, or downright fraud (as indicated by Egger et al); it can also be a mix of all these things. Alternatively, the source of heterogeneity might be a true difference in underlying populations. Most difficult to live with is the overall conclusion of the test that the literature is biased. If the test is positive, should we dismiss all randomised trials on the subject? This means that we discard one trial by one group of investigators because of the results of another trial by a completely unrelated group...”
Vandenbroucke BMJ
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2665608

In the same BMJ issue there were three further pieces of correspondence pointing out that the “funnel plot” is not a sound approach for considering the possibility of publication bias.

Lau et al. formulated valid counter-comments to funnel plot in BMJ in 2006 (The case of the misleading funnel plot)
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1570006

I do not repeat their arguments here.

A further paper which empirically examined the usefulness of funnel plot concluded “Researchers who assess for publication bias using the funnel plot may be misled by its shape. Authors and readers of systematic reviews need to be aware of the limitations of the funnel plot.”
Terrin et al. (2005) J Clin Epidemiol (In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias)
http://dx.doi.org/10.1016/j.jclinepi.2005.01.006

Thus, the lack of using the “standard” funnel plot is no indication that my manuscript is methodologically unsound.

Random effects models

I agree that "random effects models" is a standard method (popular).
However, it is scientifically unsound approach.

When there is strong evidence of heterogeneity, it is far better to look at the trials separately than combining the results to a fictional pool that is not valid to any people.

Let us consider an example.
Let there be two trials, one with men and another with women.
Let the effect of intervention be considerably different for men and women and therefore the results of the two trials are considerably heterogeneous.
In this case, the random effects model means that we are combining the results so that we get a single estimate which is valid for the "average of men and women" rather than for either sex alone. In my opinion it is more reasonable to present the estimate for men and women separately than presenting a single estimate for people who are somewhere in the middle (?) of the two sexes.

Random effects models can lead to lack of refuting the null hypothesis even though all individual trials individually refute the null hypothesis. This may seem quite strange, but here I show it by an example:

Let us assume we have populations A and B, and let the two trials find a statistically significant benefit of the intervention, but there is big difference in the size of the effect.
The fixed and random effect models for this example were calculated by using the RevMan program.

	Intervention Deads/Popul	Control Deads/Popul	RR (95% CI)	P (2-t)
A	2850/10000	3000/10000	0.95 (0.91- 0.99)	0.02(Z=2.33)
B	15/100	30/100	0.50 (0.29-0.87)	0.01(Z=2.45)

Fixed effect model: RR=0.95 (0.91 – 0.99) P(2-t) = 0.01
Random effects model: RR=0.73 (0.40 – 1.36) P(2-t) = 0.32

Consistent with both trials individually, the fixed-effect model finds significant benefit.

However, the random effects model tells us that there is no evidence that the intervention has any effect (P=0.3).
Thus, the random effects model can indicate that there is no evidence of effect, whereas the real problem is that the size of effect is inaccurate. 5% benefit and 50% benefit are, of course, substantially different, but this poor accuracy of estimate should not lead to rejection of the evidence of benefit – which happens with the random effects model.

I have been planning to write a short note about the problems of random-effects models to some epidemiological journal, but so far I have not had time.

Thus, the lack of using the "standard" random effects models is no indication that my manuscript is methodologically unsound.

estimation of I^2 statistic

I do not have objections to I-square. It is sound method and I have been using it in other papers.

However, it needs the value of standard deviation that is not available for all Zinc-common cold trials.

I could impute the standard deviation for those trials in which SD is not available.
In our Cochrane review on vitamin C and the common cold, we have imputed SD value for a few trials, so that we could include them to the RevMan program. We found that on average SD was 0.7 times the mean duration. We decided to use 1.0 as the multiplier in imputation because that leads to lower weigh of those trials in which the SD was imputed.

If you consider that the manuscript would be better with imputed SD values and calculation of I-square, that is easily done.

Fisher's method of combining P-values

This is the method I use in the manuscript
This is not a standard method (i.e. not popular).
However, it is a scientifically sound approach.
See e.g.
http://en.wikipedia.org/wiki/Fisher%27s_method
http://digital.library.adelaide.edu.au/coll/special//fisher/224A.pdf
http://dx.doi.org/10.1016/j.csda.2003.11.020 (the chi-square is the Fisher's method)

The particular benefit of the Fisher method is that it is not sensitive to the variations in original studies.
For example, one study may report the duration of colds in days (+/- SD) and another may report the duration as the number of persons who have colds lasting longer than 1 week. Such results cannot be combined to a pooled estimate of effect because the outcome definitions are so different. However, both studies test the effect of intervention on the duration of colds and the resulting P-values can be combined by the Fisher method.
When there are missing SD values for duration, that is no problem for the Fisher method, if we have available categorical data that gives e.g. the proportion of participants with a given duration. Therefore the Fisher method is practical for the review on zinc trials.

Moreover, given the scientific question of interest and the subjective nature of the outcome, a meta-analysis on individual data would be preferable.

This is not a valid comment.

When we try to understand the role of, say, sex on the effect of treatment, it is highly unreliable to analyze trials by the proportion of sex in different trials - if the variation between trials is small (i.e. in the proportion of sexes).
In such cases, study-level analyses can lead to different conclusions than do corresponding individual-level analysis, a difference called the "ecological fallacy". This issue is discussed e.g. in
Berlin et al. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Stat Med. 2002;21:371-387.

However, the situation with the Zn trials is different. All people in the particular zinc trial are getting the same dose of Zn according to the same protocol. Thus, the trials are not a mix of different Zn treatments (compare with different proportions of men/women if we try to understand the role of sex).
Therefore we can draw conclusions of the zinc trials without concern of the ecological fallacy.

Futhermore, individual level data is rarely available. In our Cochrane review on vitamin C and the common cold we are using the study-level data and not individual level data.
As far as I know, essentially all Cochrane reviews are on study-level data. Also, essentially all meta-analyses that I have read in the regular journals are on study-level data. Thus, my Zn review does not deviate from the ordinary systematic reviews in this respect.

the subjective nature of the outcome

All trials in the review are double-blind trials.
Therefore, the subjective nature of the outcome cannot explain a significant difference between the study groups.