Replies to "Trials" journal reviewer comments on:
Subgroup analysis of large trials can provide valuable information: a case study of vitamin E and pneumonia
by
Harri Hemilä and Jaakko Kaprio

Reviewer's comments are in italics

Our replies are in the ordinary font

21 June 2010
Harri Hemilä and Jaakko Kaprio

Reviewer 2

http://www.trialsjournal.com/imedia/1246630932376444_comment.pdf

Reviewer's report
Title: Subgroup analysis of large trials can provide valuable information: a case study of vitamin E and pneumonia
Version: 1 **Date:** 6 April 2010
Reviewer: Victoria Kirsh
Reviewer's report:
This manuscript represents a stratified analysis within the Alpha-Tocopherol Beta-Carotene Cancer Prevention (ATBC) Study cohort. The authors concluded that depending simultaneously on the levels of age of smoking initiation (<=20, >=21), cigarettes per day (5-19, >=20), and exercise during leisure time (yes, no), vitamin E supplementation decreased, increased or had o effect on the incidence of pneumonia. Stratification led to 8 subgroups – in one of these subgroups the effect was protective, in a second – but only among those who did not take β-carotene—the effect was harmful, and in the remaining six subgroups there was no significant association. Furthermore, there was no clear pattern in either the direction or magnitude in the associations across all 8 subgroups, which greatly complicates the interpretation of these findings.
This is one of several analyses of pneumonia using ATBC data conducted by the authors, including:
- Vitamin E and beta-carotene supplementation and hospital-treated pneumonia incidence in male smokers (which also includes subgroup analyses by age of smoking initiation)
- Vitamin E supplementation and pneumonia risk in males who initiated smoking at an early age: effect modification by body weight and dietary vitamin C - Physical activity and the risk of pneumonia in male smokers administered vitamin E and beta-carotene
The authors indicate that while there are certain "conditions when subgroup analyses may be justified", "carrying out numerous subgroup comparisons leads to the multiple testing problem." Given the numerous subgroup analyses that have already been conducted within this study population, along with the lack of a biologic rationale for highlighting a protective effect in two cells of a third-order interaction (in the absence of significant first or second-order interactions), and the lack of a clear interpretation, the current analysis does raise questions around the multiple testing issue.

Major Compulsory Revisions:
1. The title is misleading and implies that this is a methodological paper when it fact it is a very specific analysis of effect modification within the ATBC trial. The findings have no implications outside of "vitamin E and pneumonia risk: effect modification by age of smoking initiation, cigarettes per day and exercise." The title should be edited accordingly (along with some of the content of the introduction).

HH+JK: We modified the title:
"Subgroup analysis of large trials can guide further research: a case study of vitamin E and pneumonia"
However, we do not agree with the comment: *The findings have no implications outside of.. "*
Our study gives a strong example of the divergence in the conclusions from the "overall estimate" from a trial and from a thorough subgroup analysis (extension of previous subgroup

analyses). In this respect we consider that our study is methodologically important example, although we do not suggest that the exact details can be extrapolated to other outcomes and interventions.

2. Conclusions, page 3: It is stated that "the role of vitamin E in susceptibility to pneumonia in physically active nonsmokers warrants further study." The ATBC trial, however, includes men who smoke 5 or more cigarettes per day, and the category for which the authors found a protective effect covers a range of 5-19 cigarettes per day. Additionally, results were unchanged when restricted to those in the lower smoking range. Thus, conclusions cannot be extrapolated to nonsmokers.

HH+JK: We are not extrapolating the estimates of effect to nonsmokers.
In Table 2, the point estimates for those who smoked 5-11 and 12-19 cigarettes per day were RR=0.31 and RR=0.31. Thus, there is no indication that the benefit would be restricted, say, to those who smoke 12-19 cigarettes per day, in which case it would be difficult to make any practical use of the findings of this subgroup analysis (not even as a justification for further research).
"Warrants further study" means that the issue should be studied in further trials. It does not mean that we are proposing vitamin E for nonsmokers. We are not proposing vitamin E for those who smoke 5-19 cigarettes per day either.
However, there is no justification to conclude that further research should simply be repeated with males smoking 5 cigarettes per day or more.
If the language of the sentence is not optimal, we would be glad for suggestions for improvement.

Minor Essential Revisions:
3. Results, page 10, last paragraph: The results are not, in fact, as straightforward or consistent with the earlier finding of weight and vitamin C as modifiers of the effect of vitamin E on pneumonia risk as the authors imply. In this paragraph, the authors overlooked the fact that the findings for the lower-right corner of Table 1 are modified by β-carotene and the significant harmful effect if only seen in the sub-group that did not receive β-carotene.

HH+JK: In Table 1 we divide the participants who started smoking at an early age by baseline smoking and exercise. In our previous analysis, we divided the "early smoking" participants by weight and vitamin C (ref. 14). Therefore the consistency or inconsistency of these two approached is important. In our previous analysis, beta-carotene did not modify the effect of vitamin E in the low- and high-weight participants (tables 3 and 4 of Ref. 14). Thus, there is no justification to restrict this weight+vit C analysis to the no-beta-carotene participants. We find harm of vitamin E (of ref. 14) in three cells outside the lower-right corner, which means that our earlier subgroup findings are not inconsistent with the new ones. It also means that there may be groups of participants within the six intermediate cells in which vitamin E might have an effect, depending on factors other than smoking and physical activity.

4. Table 1: The heading for the rightmost two columns should indicate "effect of vitamin E" with a subheading for exercise during leisure time.

HH+JK: Done

5. Discussion, page 12: It is stated that among those who had the least exposure to smoking and exercise, the vitamin E effect was more pronounced among coffee drinkers. The p-interaction, however, was not significant and the statement in the discussion should be tempered accordingly.

HH+JK: On page 12 we wrote: "except that the effect was more evident among heavy coffee drinkers compared with those who drank less."

This referred to Table 2, in which the effect of vitamin E was in the coffee groups:
<500 ml/day: vit E 95% CI: 0.21-1.05
>=500 ml/day: vit E 95% CI: 0.06-0.48
Test for interaction: 0.12

We cannot conclude that there is different effect in the coffee groups, because the p-value for interaction is quite large. However, neither can we conclude that the effect is the same for both coffee groups because the estimates differ substantially. We can conclude that there is strong evidence that vitamin E has effects on those who drank much coffee, because the confidence interval is far from no-effect. However, we cannot know whether the "real effect" of the less drinking people might be a) no effect or b) the same as for the heavy coffee drinkers or c) something between.
In this kind of ambiguous situation it is usually stated that the common estimate is the most reliable. Nevertheless, based on Table 2, the evidence of effect in heavy coffee drinkers is stronger than in those who drink less.
We deleted the sentence in Discussion but kept the comments on coffee in Results, because the consistency or inconsistency with our earlier subgroup findings is important.

6. Discussion, page 12, 3rd paragraph: The second sentence is a misstatement and should indicate that this 79% reduction is restricted to the sub-group that the not take β-carotene.

HH+JK: We wrote: "In this group, vitamin E increased pneumonia risk by 79%"
This implied the effect of vitamin E per se. We do not consider that this was a misstatement, but we rewrote the sentence.

7. Discussion, page 14: The credibility of the heterogeneity seen in Table 1 is not adequately explained. While biologic mechanisms are given for the role of smoking and exercise, not explanation is provided for the very particular associations highlighted in this paper.

HH+JK: Proponents of evidence-based medicine emphasize that conclusions about intervention effects should be based on controlled trials with clinically relevant outcomes, and not on biological mechanisms. Biological arguments do have value, but we do not consider that this kind of paper should use more space to the discussion of biology, or that the validity of Table 1 analysis would be affected by more detailed proposals for the interaction between exercise and different measures of smoking.

Discretionary Revisions:

8. Methods, baseline characteristics, page 6: It is not clear why age at smoking initiation and cigarettes per day were used as the measure of level of cigarette smoke exposure. What about duration of smoking at baseline? Would it have been possible to assess pack-years and incorporate this into the analysis instead of simply cigarettes per day at baseline?

HH+JK: As discussed in our earlier paper (ref. 6), there is much earlier evidence indicating that starting to smoke at early age may cause permanent changes in pulmonary functions. This effect is dependent on the age a person starts smoking and it is a different question than the duration.
Duration is a complex variable, because many people have periods when they do not smoke. Furthermore, in ref. 9 the duration of smoking did not modify the effect of vitamin E (p=0.9 for interaction). Pack-years is an even more complex variable, because it is the combination of duration and intensity of smoking. Proper estimation of pack-years would require that there is data on the variation of smoking over the history of the participants, and we do not have such data.
Furthermore, the existence of other measures for smoking exposure does not challenge the evidence of heterogeneity in our Table 1.

9. Statistical methods: In analyzing results from 2x2 factorial design, the first step should be to rule out an interaction between the two treatments (vitamin E and β-carotene, in this case). Only in the absence of interaction are the analysis of vitamin E (AT and AT+BC) versus no vitamin E (placebo and BC) valid. While stratified analyses by β-carotene were done for the upper-left and lower-right corners of Table 1, it is unknown whether the RR estimates for the remaining six cells of Table 1 are valid.

HH+JK: In our first paper on pneumonia (ref. 6), we tested vitamin E and beta-carotene interactions for the whole ATBC cohort and the subgroups of that paper, and there were none. In Table 1, the only significant interaction between vitamin E and beta-carotene is in the lower right corner. A note was added to the footnotes of Table 1 that there are no vitamin E-beta-carotene interactions in other cells.
Given the interaction between vitamin E and beta-carotene in the lower-right cell, we had been thinking whether the presentation of Table 1 is misleading or not. The primary purpose of Table 1 is to show the heterogeneity. In this respect the point estimates are not highly relevant. We are describing the vitamin E-beta-carotene interaction in the Results soon after Table 1 and therefore reader will see the more appropriate vitamin E estimates for the lower-right corner after a few lines of after Table 1 is discussed.
It does not seem reasonable to limit Table 1 to the no-beta-carotene participants, because in the upper-left corner and in all 6 middle cells there is no evidence of interaction (vit E-beta-car). It would be possible to present 2 different sets of vitamin E effect values in the lower-right corner (beta and no-beta), but that would be confusing to the reader. Thus, presenting the interaction in Results text soon after the overall heterogeneity in Table 1 seemed a suitable approach.

10. Statistical methods, page 8, 3rd paragraph: In a cohort analysis, categories should be created based on the distribution of the cohort at baseline (i.e. equal numbers of men in each category), not by the number of cases in the placebo group. In the absence of a placebo effect, the categories here should be similar, but this may not always be the case.

HH+JK: If the baseline risk is uniform, then cutting the groups at the median of the baseline distribution leads to similar number of cases in the subgroups. However, if the baseline risk substantially varies, then cutting the baseline distribution at the median can lead to substantial divergence in the number of cases in the subgroups. That leads to low statistical power, because one of the groups has very small number of cases.
Thus, our approach does not mislead the reader, but leads to narrower CI:s compared with the situation when one of the two subgroups has a much lower number of cases.

11. Results, page 9, second paragraph: It is not clear what prompted the analyses in Tables 2 and 3. It is highly unusual to select out two of 8 subgroups in which estimates were significant and embark on a search for additional effect modification by 7 additional variables (fourth-order interactions) (stratification by β-carotene in exception, since this is warranted given the trial design).

12. Tables 2 and 3: Given that none of the fourth-order interactions were significant, the authors may wish to consider not including these results in a table, but simply reporting the null findings in the text. The stratified results by β-carotene (particularly for table 3) should, however, be reported.

HH+JK:

Rationalization behind Tables 2 and 3.

1) participants were administered vitamin E and beta-carotene (2x2) and the interaction between these two interventions should be tested if the difference between vit E and no-vit E groups is attributed to vitamin E alone.

2) The age of smoking initiation and the level of baseline smoking are continuous variables, and information is wasted when the variables are dichotomised. One approach to take into account more information from the continuous variables, is to divide the dichotomised groups to smaller subgroups, because that shows whether the findings are consistent in smaller ranges.

3) Age, dietary vitamin E, dietary vitamin C and coffee are well motivated factors either because of basic biology (age, vit E) or because of our previous subgroup findings (vit C, coffee).

The order above 1), 2) and 3) shows the priority order of the subgroup analyses in Tables 2 and 3. Thus there is the greatest justification for 1) and the weakest for 3).
We reordered the factors in Tables 2 and 3 to follow this priority order.

4) The point estimates in the subgroup analyses of Tables 2 and 3 are informative.
If we assume that the benefit in the upper-left cell and the harm in the lower-right cell of Table 1

are caused by random fluctuation over the 8 cells of Table 1, then we should expect an even greater random variation when we compare subgroups within these two cells. We do not see substantial variation in the point estimates between the subgroups in Tables 2 and 3.
This lack of great variability in the point estimates in the subgroups of Tables 2 and 3 gives more credibility to the divergence in the opposite corners of Table 1.

In particular, when the age of smoking initiation and baseline smoking are divided into smaller subgroups, the point estimates within these smaller subgroups are not conflicting with the findings in Table 1.
Thus, for example, if the benefit of vitamin E in Table 2 would be restricted to participants who smoked 12-19 cigarettes per day (or started at 21-25 years), so that there would be no difference between vit E and no-vit E in those who smoked 5-11 cigarettes per day (or started at 26+ years), then the findings would be confusing when related to the Table 1 findings.
Tables 2 and 3 give 4 comparisons of smoking exposure subgroups and as a set these do not suggest great variation in the point estimate between the subgroups.

In Table 1 footnote we show that the variation in vitamin E effect in the 6 middle groups is fully explained by chance. All the important variation in vitamin E effect is located at the two corners of Table 1. On this basis, the two corners are particularly important for a detailed analysis.

The similarity of the point estimates in various subgroup comparisons suggests that the estimates are rather robust (the 69% decrease in Table 1 upper-left corner and the 79% increase in Table 1 lower-right corner).

****
We consider that the reviewer has missed our paper's main point in the arguments above.

At the end of our Discussion section, we describe that medicine has “two different goals” as Vandenbroucke described.
First, pragmatic testing: whether an intervention works or not.
Second, trying to find new paths for research, which means curiosity, imagination and generation of new hypotheses.
Although the first type of goal has huge practical importance, the second goal is a more fundamental goal in science in general since it gives us the new questions and explanations and paths forward.

In our Discussion we point out that subgroup analysis can give important information falling under the second goal.

Based on our current analysis, there is definite motivation to study the effect of vitamin E on nonsmoking males who exercise at leisure, whereas the overall result of the ATBC Study suggests that there is no justification to study the effect of vitamin E on any people. Restricting the analysis of a large trial to the overall estimate may be inefficient use of collected data. Furthermore, if there really is heterogeneity in the effect, it is misleading to imply that a single estimate of effect is valid for all people corresponding to the selection criteria of a trial.

Our interpretation of the reviewer's comments are that they are primarily based on the evaluation of research of the first kind (pragmatic testing). Nevertheless, we are not proposing vitamin E for any people on the basis of the current subgroup analysis. We emphasize even in the Abstract that the estimates should not be extrapolated to other populations.

Thus, our conclusion is that subgroup analyses should not be discouraged so intensely as they often are. There are many reviews on subgroup analyses, some of them are cited in our manuscript, which give guidance on carrying out subgroup analyses (e.g. Rothwell's paper is 11 pages long (ref. 20)). We do not consider that we should summarize such guidelines in this paper.
Nevertheless, our study gives a strong example of the divergence in the conclusions from the “overall estimate” of a large trial, and of a thorough subgroup analysis (in this case an extension of previous subgroup analyses). In this respect we consider that our study is methodologically important example, although we do not suggest that the exact details can be extrapolated to other outcomes and interventions. But our example suggests that in other large trials cautious subgroup analyses should be encouraged.

The reviewer has no comments on our comments which are based on Vandenbroucke's paper, although that is a most important issue at the end of our Discussion.

Level of interest: Reject as not of sufficient priority to merit publishing in this journal
Quality of written English: Acceptable
Statistical review: No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:
I declare that I have no competing interests.