Author's response to reviews

Title: Comparative effectiveness of antihypertensive medication for primary prevention of cardiovascular disease. Systematic review and multiple treatments meta-analysis.

Authors:

Atle Fretheim (atle.fretheim@nokc.no)
Jan Odgaard-Jensen (jan.odgaard-jensen@nokc.no)
Odd Brørs (ODBROE@ous-hf.no)
Steinar Madsen (steinar.madsen@noma.no)
Inger Njølstad (inger.njolstad@uit.no)
Ole Frithjof Norheim (ole.norheim@isf.uib.no)
Arne Svilaas (asvilaas@online.no)
Ivar Sønbø Kristiansen (ivarsk@c2i.net)
Hanne Thürmer (Hanne.Thurmer@sthf.no)
Signe Flottorp (signe.flottorp@nokc.no)

Version: 4 Date: 4 January 2012

Author's response to reviews: see over
To

The Editor

BMC Medicine

Revised manuscript:

Comparative effectiveness of antihypertensive medication for primary prevention of cardiovascular disease. Systematic review and multiple treatments meta-analysis

We thank the reviewers for generally positive comments and for their constructive critique!

We provide a point by point response to their comments on the following pages, and hope these are found satisfactory by you.

Our responses are written in italics.

Regards,

Atle Fretheim (on behalf of the authors)
Fretheim et al.

1st Reviewer: Milo Puhan

Fretheim and colleagues report about a systematic review and network meta-analysis. This review focuses entirely on methodological aspects as requested by the editor. The paper is well written except and the results appear to be valid overall. But providing some more detail on some issues would further improve the paper.

Major comments

- Some eligibility criteria are not clear. For example, it is unclear what it means that only trials with CVD mortality and morbidity were included. How was a primary outcome defined? Often, the primary outcome is not well defined in trials or there are, although not recommended, several primary outcomes or none.

  We have amended the Methods so it now reads:

  “To be considered for inclusion, a trial had to have cardiovascular morbidity or mortality as a primary outcome, either explicitly stated by the authors or based on our judgement.”

Also, why would a trial that had myocardial infarction as an outcome be excluded for the analysis on all-cause mortality? Just because it was not the primary outcome? This should be clarified.

  If a trial were included in the first place, all relevant outcomes available from the trial were included in our analyses. We have revised the wording slightly to make this clear:

  “We extracted, where possible, data for the following main outcomes from all the included studies: total mortality, myocardial infarction, and stroke.”

- Also, only including trials with CVD mortality and morbidity as primary outcome does not necessarily mean that smaller trials were excluded. Also, trial size is not generally regarded as a marker for study quality because it is not associated with bias per se but with greater sampling variability. Thus the authors should explain how they defined “smaller” trials and justify why they excluded them.

  We have re-worded the relevant paragraph in the methods section to clarify that study-size was indeed not an exclusion criterion – but that many smaller studies were in practice excluded since we excluded studies that were focused on “surrogate” outcomes:

  “To be considered for inclusion, a trial had to have cardiovascular morbidity or mortality as a primary outcome, either explicitly stated by the authors or based on our judgement. In practice this meant excluding many smaller studies, typically designed to evaluate effects on surrogate outcomes such as blood pressure. Some of these trials reported morbidity and mortality outcomes in addition to their primary “surrogate” outcome. We disregarded these data in the belief that the information was not likely to be important given the availability of findings from large-scale studies with morbidity and mortality as main outcomes. By doing this we also reduced the risk of introducing certain biases in our analyses, e.g. due to selective reporting of findings, which is likely to be a greater problem with smaller studies [20].”
Later in the results section the authors say that trials of “sufficient quality” were included in the review. It is not stated anywhere in the methods section that quality of trials was a criterion for inclusion and how this was assessed. This requires an explanation.

_The reviewer is right – we have forgotten to describe that! We have now added a sentence so that the paragraph on risk of bias-assessment goes like this:_

_“Studies fulfilling our eligibility criteria were assessed for internal validity at the study level by two reviewers independently using a standard check list [21]. Studies were excluded if the validity was judged as “low”.”_

The authors should be applauded that they are one of the very few groups that report about strength of evidence when it comes to network meta-analysis. I am very familiar with the GRADE approach but it is unclear how the evidence was graded. This is very challenging in network meta-analyses since the quality of direct and indirect evidence needs to be judged as well as their combination (for those comparisons where both is available). Thus there should be a more detailed explanation how the assessment of the strength of evidence was done.

_We have added a sentence in the following paragraph, for clarification:_

_“In line with recommendations from the GRADE Working Group by default we graded the included evidence as “high quality”, as all studies were randomised controlled trials. We then downgraded when deemed appropriate. For the comparisons where we had no direct evidence (i.e. the effect-estimates were only based on indirect comparisons) we rated the quality as “low” unless we found reasons to upgrade or to downgrade further. When the findings were based on a combination of direct and indirect evidence we elected to grade as “high quality” unless there were reasons to downgrade. The grading process was done by one reviewer (AF) and validated by a second person.”_

Please also note that the GRADE working group has released an approach for network meta-analysis yet although this is currently under development.

Minor comments:

- The authors had, as in any study, to take some arbitrary decisions as for example that one primary or mixed primary and secondary prevention populations. I do not challenge these decisions but since they are arbitrary some sensitivity analyses would be informative. For example, what happens to the result if just primary prevention populations are included?

_We have chosen not to conduct the proposed sensitivity analysis because there were practically no “pure” primary prevention studies. Additionally, getting access to the data for the relevant sub-groups from the different studies is likely to be very difficult. We considered this option at a very early stage, i.e. to conduct the analyses on the “primary prevention” sub-groups from each trial, but we found it unlikely to be feasible in practice, mainly due to lack of access to data. Also, we do address this issue to some extent when we discuss our choice of limiting the review to (mainly) primary prevention studies (Discussion, under Limitations)._
- Please explain how inconsistency was assessed. Was it assessed based on statistical significance or on some judgment on how much of a difference between direct and indirect evidence was acceptable? Just using statistical significance can be very misleading in network meta-analysis because the power to detect still relevant differences is often low.

  This was only based on a “qualitative” assessment of the comparison of the findings from direct vs. indirect effect estimates – i.e. by looking at the results in Additional file 6 (where the effect estimates with 95% CIs from direct and indirect analyses are tabulated). We consider that we report this sufficiently transparent in the current Results section, but we may expand this paragraph if the editor so wishes:

  “We also conducted traditional meta-analyses for all comparisons where head-to-head data were available, and compared the results with those from our MTM-analysis. We found few major discrepancies, see Additional file 6.”

- Please provide more details on the statistical model and how justifiable some assumptions were (distribution of random effects, priors).

  We have added the following paragraph in the methods-section:

  The statistical analysis is based on binomial likelihoods, with vague priors for the trial baselines, basic parameters (normal distribution with mean 0 and standard deviation 0.0001), and the random effects standard deviation (uniformly distributed i in the interval 0 to 2), and takes the correlation structure induced by multi-arm trials into account. We have used a random effects model which follows the Normal distribution with 0 as the mean.
The same point was made by reviewer 1, and we have revised the text accordingly (see above and/or in the revised manuscript itself).

3. Are the data sound and well controlled?
The study appears robust.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?
Yes

5. Are the discussion and conclusions well balanced and adequately supported by the data?
Yes, the discussion regarding study limitations and the grounding within the wider literature is helpful. I would only challenge the conclusion that the effects of publication bias and funding from industry are anticipated to be small. There is inadequate grounds upon which to draw this conclusion and it is also a controversial finding. It would need to be substantiated.

We have revised the text which now reads:
“The majority of the included trials in this review were sponsored by companies with a vested interest in the study results. Such sponsorship has been associated with bias in favour of the product made by the funding company [63]. Possible explanations include publication bias and use of inappropriate comparators. Limiting our review to large-scale studies should reduce risk of publication bias or other forms of selective reporting [20]. Whether the most appropriate comparator has been selected is more difficult to assess. Biased analyses were minimized in our review because we based our effect-estimates on actual figures presented in the various articles, rather than relying on the analyses conducted by the study-authors and/or sponsors.”

6. Do the title and abstract accurately convey what has been found?
Yes
This paper presents a network meta-analysis of available trials of anti-hypertensive drugs to attempt to answer questions about the relative efficacy of different drug classes in preventing different disease endpoints. The methods are appropriate and well described, the selection of trials is generally defensible (see below), the conclusions are generally cautious and the writing is clear. The paper updates previously published findings, including some using similar approaches and some using conventional meta-analysis, and the presentation places their findings in contrast to other similar analyses. While there are no particularly notable new conclusions from this review, the thoroughness of the review, inclusion of several new trials and inclusion of GRADE strength of evidence ratings probably make it worth publishing. Finally, a comprehensive set of supplementary documents provides a wealth of other information to place the results in context, including a comparison of the data from the network meta-analysis to results limited to meta-analysis of direct comparisons.

Minor Essential Revisions:

The primary concern is how much it adds to our understanding and whether the use of indirect comparisons enabled by a network meta-analysis improves on estimates based only on direct comparisons. The primary value is allowing inferences about relative comparisons between classes for which no direct data exist. It is notable that in Table 6, comparing results of direct meta-analysis vs. the network model, the confidence intervals are actually wider in several cases (more uncertainty) for the network model. This seems counterintuitive since the network model should incorporate all the direct comparisons as well as indirect comparisons and thus should produce a somewhat more precise estimate due to including more data. The authors should verify these results and consider explaining them.

We consider it reasonable that that in some cases the network-analysis resulted in greater uncertainty, since the additional (indirect) data might point in a different direction than the direct data. We are reluctant to addressing this in the text because we consider it is a too detailed methodological issue for this paper. We do agree that the reviewer’s observation is relevant to a general discussion about the value of network meta-analysis, but contributing to that debate is not one of the main intended objectives of the current manuscript.

Network meta-analyses and conventional direct meta-analyses share the same limitation that they are only useful if the assumption of homogeneity is correct. In a conventional meta-analysis one can test that assumption several ways: one can assess homogeneity of results using I squared and other statistics; one can graphically assess in a forest plot whether estimates are generally similar across studies or vary widely; and finally one can assess the basic elements of each trial (patient characteristics, duration, outcome rates, treatment responses) to assess whether studies really are comparable enough to justify pooling. I am not an expert on network meta-analyses but since the authors do not present any statistical measures of homogeneity I assume these aren’t possible. Nonetheless some discussion of the patient populations, baseline risk estimates, etc. would provide some reassurance that the trials really were assessing roughly comparable populations in terms of
risk. (These are in the supplemental tables but some key elements could be brought into the discussion).

We agree that this is an important point, and the reviewer is right in assuming that assessment of homogeneity across trials is slightly different in a network-analysis than in a traditional meta-analysis.

We have used the traditional I-squared approach to assess the degree of heterogeneity across trials that have directly compared the same two drugs, and this has been used to assess “inconsistency” in our GRADE-assessment. In our manuscript this was only apparent in the footnotes in the GRADE-tables in Additional File 5, so we have added clarification towards the very end of the Methods-section (new second last sentence in Methods, under “Quality assessment of the evidence”):

“The inconsistency of results-dimension in GRADE was only assessed for direct comparisons (using the I²-statistic).”

The other part of the issue is how well we have assessed the homogeneity across all trials included in the network-analysis – not just across the trials of the same pairwise comparisons. As the “art” of network-analysis is still in its infancy, guidance as to how this should be done is still in development (ref. comment from 1st reviewer on expected guidance from the GRADE working group). We have addressed the issue by qualitatively comparing our effect-estimates with those we got when only including direct comparisons, and based on that we concluded “We found few major discrepancies, indicating that our model is trustworthy”, which basically is the same as saying that the trials in the network seem to be reasonably homogeneous. We have added this point to the final paragraph in the Results-section:

“We also conducted traditional meta-analyses for all comparisons where head-to-head data were available, and compared the results with those from our MTM-analysis. We found few major discrepancies, indicating that the included trials are reasonable homogeneous and our model is trustworthy. The comparisons are shown in Additional file 6.”

A key issue which would invalidate assumption of homogeneity would be if mean blood pressure reduction was different across trials, especially in the 6 placebo controlled trials, since Law has argued that the risk reduction is dependent directly on level of BP change. Underlying patient risk may or may not make a difference, depending on whether the relative risk reduction is constant across risk groups, but since the paper has decided to select out a primary prevention population (excluding those with prior CVD events) they should discuss whether the risk levels were comparable across the different trials.

We have added the following paragraph to address these highly relevant concerns from the reviewer:

“In a network analysis like ours, it is assumed that all the included trials are sufficiently homogeneous to allow for the combining of all the study findings into one analysis. This assumption is difficult to validate. Differences in study populations can in particular distort estimates for the effects on total mortality, since these are related to the proportion of deaths that are due to cardiovascular diseases, for each study. We did not formally assess how
comparable the various populations were. However, our use of relatively strict inclusion criteria, e.g. including only studies where the majority of participants had no prior cardiovascular event and excluding studies of specific high-risk groups, substantiates that the populations were somewhat similar. Also, the finding that the effect-estimates from the network analysis were similar to the estimates from the direct comparisons provides some evidence that the trials were reasonably homogeneous.”

All cause mortality is an especially problematic outcome to pool across primary prevention trials when an intervention is only acting on some causes of death. The ability for a drug that only prevents CVD to lower all cause mortality depends on what proportion of all deaths are due to CVD. This may vary across the trials included. Thus some agents may appear more effective against all cause mortality than others only because they have been tested in populations at highest risk of CVD (e.g. diabetics). This should be mentioned.

True. In retrospect it would have been a good idea to extract the proportion of deaths that were due to CVD, for each trial – as an indicator for how high-risk (for CVD) the various study populations were. We have added this point in the Discussion (same paragraph is in previous point):

“Differences in study populations can in particular distort estimates for the effects on total mortality, since these are related to the proportion of deaths that are due to cardiovascular diseases, for each study.”

2. Discretionary revisions

I have some possible minor revisions aimed primarily at expanding the discussion of the limitations of network meta-analysis and some further attempts to place the findings in context. Specific comments:

The authors state that all the trials used atenolol as the beta blocker. This is technically true but several included metoprolol and atenolol as treatment options presumably at discretion / preference of physician.

We disagree that this is worth mentioning in the text since our reason for bringing it up in the Discussion was simply to explain why we had not conducted a separate analysis of non-atenolol trials. We don’t interpret the reviewer’s comment to mean that we should have considered a separate analysis for trials in which also non-atenolol drugs were “allowed” (we too don’t think that is likely to be informative).

Table 2 – the description of the stop 1 trial is confusing as to what the different arms were – the use of or and + make it unclear what are options within arms and which are distinct arms.

We agree, and we have revised the text to make it clear.

The authors should discuss the existence of different drug classes within the CCB class. It appears that most of the CCBs in these trials were dihydropyridines which may justify treating them as one
class but that might not be true of all of them. Certainly different agents differ in their vasodilatory activity.

The comment is fair enough. We have added these two sentences to the Discussion:

Similarly, our handling of calcium channel blockers as an entity can be questioned. The pharmacological properties vary across these drugs, and it could be argued that they should be grouped according to property and not as an entity. [62].