Author’s response to reviews

Title: Reporting of heterogeneity of treatment effect in cohort studies: a review of the literature.

Authors:
Meryl Dahan (dahanmeryl@gmail.com)
Caroline Scemama (scemamacaroline@gmail.com)
Raphael Porcher (raphael.porcher@aphp.fr)
David Biau (david.biau@aphp.fr)

Version: 1 Date: 17 Oct 2017

Author’s response to reviews:

Reviewer reports and our answers

Dru Riddle (Reviewer 1): Page 2, Line 9: this is not a systematic review design type; rather I would suggest literature review or narrative review

Answer:
The text has been changed accordingly.

Page 4 first paragraph: grammar (genetics and not genetic), BRAF needs to be fully spelled out for the reader please

Answer:
The text has been changed accordingly.

Page 4, Line 55: this statement seems to indicate case and effect relationships with cohort studies although this is not possible. Associations would be a better descriptor of the cohort study here

Answer:
The text has been changed accordingly.
Page 6, Line 7: was this protocol published? If so, this needs to be referenced.

Answer:

Prisma guidelines only recommend authors to register protocols for reviews/meta-analyses studying a treatment. Reviews evaluating a methodological question do not fall within this scope and cannot be registered. Therefore, although a protocol was written, it was not registered nor published. The protocol is provided as supplementary material for the reviewers’ appreciation.

Page 7, search: I'm not sure if this search string terminology is the best way to return all relevant cohort studies. Can you limit your search in other ways? A risk of bias is raised here

Answer:

We appreciate the reviewer's concerns. Our aim was to identify within specified journals all prospective cohort studies. In that respect, the search terms « Name of the journal »[TA] AND « prospective » AND « cohort study » were the most inclusive we could find after a few trials. There are certainly studies that were missed but it is unlikely that these errors were systematically associated with relevant studies’ characteristics. We therefore think that selection bias was minimized.

Methods: overall, I am unclear how you determined that the individual studies were indeed cohort studies. I see that you operationalized your search to point towards cohort studies, however, there appears to be a crucial missing step we hereby the review team actually confirm that the study was indeed a cohort study. I think this is a critical piece of the process and needs to be well articulated in the paper. I see reference to this in the PRISMA diagram, however, it seems to be missing from the language in the study.

Answer:

We agree with the reviewer that the process of study selection was not described sufficiently. We have improved this section. Studies were first assessed on title and abstract and in case selection criteria could not be approved or disproved, the full text was obtained and assessed. Of note, the number of articles reviewed for title and abstract (and full text) was dependent on when the required number of articles for each journal was obtained. For instance, say we had three studies to retrieve within a journal. A list of hits within this journal was obtained with the search phrase. Then, these hits were assessed from the earlier to the latest, in a top-down fashion, and the search was stopped when the three studies meeting the selection criteria were found. Therefore, we did not screen for the entire list of hits, only until necessary. The number of hits reported (n=2019) is the actual number of hits screened for title and abstracts (and for some full text).
Usually, in a review, the list of hits is definitive, determined at the onset of the search and these hits are all screened. Therefore it is easier to provide the number of hits screened for title and abstracts, the reasons for their exclusion, and the same for full text. Given the different process in this review, we did not provide exclusion criteria separately for title/abstracts and full text.

« Eligible studies were identified with the following search terms on PubMed « Name of the journal[TA] AND « prospective » AND « cohort study ». The term «Name of the journal[TA]» was replaced by the relevant journal name as randomly selected in the previous step. Hits (n=2019) were reviewed within each journal on title and abstract and then on full text for selection criteria until the adequate number of studies was reached. »

Page 9: Risk of bias screening. Did you omit any studies based on their risk of bias? I see the use of a tool to assess risk of bias, however, I cannot seem any reference to applying the results of that tool to the selection of studies to be included. I would recommend moving this section to the section preceding data because it would make more sense, methodologically, to screen for bias first before dealing with data points.

Answer:

We appreciate the reviewer's point. We did not provide a subgroup (sensitivity) analysis based on study quality. We have redone the analysis comparing studies addressing HTE with studies not addressing HTE with only studies rating good (7 or 8) on the Newcastle-Ottawa quality assessment scale. The new table is provided (table 5). We did not identify meaningful differences.

Jo Leonardi-Bee (Reviewer 2): This paper describes how heterogeneity has been assessed within cohort studies assessing the effectiveness of a treatment, using systematic review methodology. This is an important area to focus on as cohort studies are being used more and more frequently to address the association between an exposure or treatment and health outcomes, especially where it is unethical or unfeasible to conduct RCTs.

Overall, this paper is well written; however, I have a major few comments that the authors may wish to address:

1. A systematic review includes all of the eligible studies addressing a single focused question—however, this systematic review included a random sample of journal articles from high impact and lower impact journals. It is not clear whether this would have introduced bias -
the selection was stratified by high versus lower impact journals - but this does assume that the instructions to authors for the journals within each strata are exactly the same and have not changed over time. This assumption is likely to be violated as the papers were sampled over the past 15 years - additionally, whilst we can assess the similarity of instruction to authors for the high impact journals, this can not be done for the lower impact journals due to the sheer number. Additionally, it is gleaned from the methods that the studies were included in a reverse chronological order - therefore it is not clear how this can be a random selection of eligible papers? Finally, as the number of high impact journals is so much smaller than the lower impact journals, you should find that the high impact journals are much more similar in their reporting - the authors need to address all these points, and describe how they could have impacted on the findings of the review.

Answer:

The reviewer raises numerous important points.

First the word « systematic review » was changed for « literature review » all throughout the paper in order to prevent any confusion among readers.

Second, we agree with the reviewer that instructions within each journal are different and likely influence manuscript submission and publication. We could not evaluate the effect of submission guidelines on the characteristics of studies. First because it is difficult to classify or rate submission guidelines and second because they have changed over time. Last we considered that the variable « journal impact factor » included « submission guidelines » and this latter variable is addressed within the former. Indeed journal impact factor stands for various other variables such as notoriety, expertise area (although these are general medical journals), submission guidelines, reviewing and editorial process and quality, etc. and differences between high and low impact journals could come from one or more specific characteristics. We have added a limitation at the end of the discussion to draw the attention of readers on this matter.

« Another limitation is that journal impact factor stands for various other variables such as notoriety, expertise area (although these are general medical journals), submission guidelines, reviewing and editorial process and quality, etc. By comparing low and high impact journal factors we could not identify precisely which variables have an effect on reporting heterogeneity of treatment effect. »

Third, we appreciate that the identification of studies in a reverse chronological order (or any other order) is not in strict accordance with a « random sample » of studies. However, it is the number of studies per journal that was randomly chosen (within high and low impact journals), not the studies themselves. Our method of selection was done by Sun and colleagues when they
studied the reporting of heterogeneity in randomized controlled trials [Sun et al, BMJ, 2011]. Ideally, to avoid this issue, we should have first identified all eligible studies for all 115 journals with no time limit, and then randomly draw the 150 studies from these sets (one set of high impact and one set of low impact journals). This was not possible to perform given the amount of work it entailed. By selecting studies from the earlier to the latest, our sample is probably biased with studies more likely to report heterogeneity than a random sample. We have added a limitation at the end of the discussion to address this point. Moreover, the term « random » is only associated with « journal », not with studies.

«One of the limitation is that if journals were randomly selected within their subgroup (low or high impact), studies themselves were identified in a reverse chronological order and constitute a biased sample of all prospective cohort studies within each journal. This bias probably favors a better reporting of heterogeneity in our sample. »

Last, there were only six high impact journal and 113 low impact. These journals represented all core medical journals defined by the National Library of medicine, known as the Abridged Index Medicus, covering all specialties of clinical medicine. If differences exist between journals within each category (low and high impact), by randomly selecting the journal from which the studies were then identified (see point above), we have samples representative of the entire category and we think that differences between journals are accounted for.

2. The background only glosses over the potential biases introduced from using cohort studies to look at treatment effects - more recognition and discussion is needed regarding confounding by indication - and how this links to conducting full explorations of heterogeneity

Answer:

We appreciate the reviewer's concern about confounding by indication. Confounding by indication is a bias encountered in observational studies because the allocation of treatment in non randomized studies results in some imbalance in the underlying risk profile between treated and nontreated groups. Since this risk profile may be related to future health outcomes it can generate biased estimation of the effect of treatment. When looking at heterogeneity in the effect of treatment, one usually looks at subgroups where the treatment is more or less likely to be effective. However, significant heterogeneity may also be the effect of confounding by indication, with subgroups of patients being identified not because the effect of treatment differs, but because their risk profile differ. We have modified the introduction to alert to this idea.
This can lead to apparent heterogeneity in treatment effect when there is none or conceal a true difference of effect between categories of patients. Cohort studies are inclined to confounding by indication [Kyriacou DN, Lewis RJ. Confounding by indication in Clinical Research. JAMA. 2016 Nov 1;316(17):1818-1819.], therefore subgroups of interest may be identified not because the effect of treatment differs, but because their risk profile differ.

3. How generalisable are the results from this systematic review given that only 119 journals from the Abridged Index Medicus are included in the review, and are very likely to be from specific areas of the world. Discussion regarding this needs to be provided

Answer:

The Abridged Index Medicus only refers 119 core clinical journals and all journals, within low and high impact factor categories, had a similar probability to provide a study. Therefore we do not see a specific risk of bias here.

4. At the bottom of Page 6 - beginning of Page 7 - "High impact journal studies were identified by generating a random list of 75 journal names..." - this is contradictory from the other text which states only 6 high impact journals were included

Answer:

We appreciate the confusion. What we meant is that 75 journal names were randomly generated but among 6 possible categories. We have rephrased this part.

“High impact journal studies were identified by generating a random vector of 75 names of journals among the six possible categories; each of the six journals had the same probability of being selected per draw”.

5. How did the authors deal with prospective controlled cohort studies which were matched for specific characteristics - for example, age and sex? Did they exclude these studies from the review or just exclude them from the age and sex predictive variables analysis?

Answer:

Studies matched for variables addressed in the « predictive variables studied », such as age or sex were treated in the same way as other studies because matching does not preclude a predictive analysis.
6. The fact that such a very high proportion of the included studies were funded by industry is
great cause for concern - can the authors provide more information as to whether the industry
funders were involved in the design, analysis and interpretation stages of the paper?

Answer:

We agree with the reviewer that funding by the industry is a cause for concern. However, how
funding was done was not reported consistently and we cannot analyze this variable further

Minor comments include:

1. What was the rationale for including 150 cohort studies rather than a different number - it is
not fully clear what the precision of a 50% probability means - this needs to be explained in
sufficiently more detail

Answer:

The number of studies included is arbitrary. Our sample size was computed so that if a variable
had a 50 % probability of occurring (50 % of studies reporting heterogeneity for instance) the
precision of the estimate at the 95 % confidence interval would be 42 % to 58 %. We have added
a precision in the text to avoid confusion.

However, the precision of a 50% probability with 150 trials is +/-8% at the 95 % level which is
reasonable.

2. As the outcomes reported are typically quite common, then estimating the association based
on an odds ratio is likely to over-estimate the true association

Answer:

We appreciate the reviewer's concern that if one interprets odd's ratios as risk ratios, they would
overestimate the true effect. There is little we can do about this appreciation.

3. The protocol seems to have been documented before the review was conducted - however,
the protocol has not been attached/uploaded with the review, and hasn't been published

Answer:

Prisma guidelines only recommend authors to register protocols for reviews/meta-analyses
studying a treatment. Reviews evaluating a methodological question do not fall within this scope.
Therefore, although a protocol was written, it was not registered nor published. The protocol if provided as supplementary material for the reviewer’s appreciation.

4. The search terms for identifying a prospective controlled cohort study are very brief and it is not clear that this would yield all eligible studies from the particular journals - this is likely to result in selection bias

Answer:

We appreciate the reviewer’s concerns. Our aim was to identify within specified journals all prospective cohort studies. In that respect, the search terms « Name of the journal »[TA] AND « prospective » AND « cohort study » were the most inclusive we could find after a few trials. There are certainly studies that were missed but it is unlikely that these errors were systematically associated with relevant studies’ characteristics. We therefore think that bias was minimized.