Author’s response to reviews

Title: Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: the example of lung cancer

Authors:
Perrine Créquit (perrine.crequit@aphp.fr)
Ludovic Trinquart (ludovic.trinquart@aphp.fr)
Amélie Yavchitz (amelie.yavchitz@aphp.fr)
Philipppe Ravaud (philippe.ravaud@aphp.fr)

Version: 1 Date: 03 Dec 2015

Author’s response to reviews:

December 2nd 2015

Dear Mr Recchioni,

Thank you for the opportunity to resubmit our manuscript for further consideration by BMC Medicine. We thank the reviewers and yourself for your comments.

We believe that we have been able to address each of the comments raised, and include a detailed point-by-point response below. In particular, as recommended by the reviewer to strengthen the paper, we have expanded the discussion to elaborate on the methodological steps of live cumulative network meta-analysis, to address the challenges raised by the proposed novel methodology and some potential solutions; we incorporated a table and a figure for more clarity.

You will find enclosed a revised version of our manuscript incorporating all of the changes. We include 2 copies of the new version of the manuscript (clean version and version with changes tracked).

Thank you for your further consideration of our manuscript that we hope is now acceptable for publication in BMC Medicine.

Best regards,

Perrine Crequit, Ludovic Trinquart, Amelie Yavchitz, and Philippe Ravaud
Response to the reviewer Per Olav Vandvik

Reviewer #1:

I enjoyed reading this innovative study assessing to what extent the whole set of systematic reviews and conventional meta-analyses provide a comprehensive, up-to-date synthesis of evidence for all treatments in the case of lung cancer. The finding of substantially incomplete randomized evidence covered by systematic reviews demonstrates a disturbing waste in research and health care. The paper advocates well for "live cumulative network meta-analyses as a paradigmatic shift and potential solution within the context of "increasing value and reducing waste in research" being a hot topic these days.

Indeed I find the results and conclusion from this study quite compelling as available pairwise treatment comparisons do not allow for meeting clinicians' and patients' needs in decision making, as stated by the authors. Nor do they meet the needs of other target audiences such as developers of clinical practice guideline decision aids, funders and decision-makers in health care systems. The latter audiences could be included in the discussion of the paper to highlight the value of proposed solutions by the authors.

Answer:

Thank you for your comments. We agree that embracing the perspective of networks of trials of all alternative treatments for each condition and, in particular, developing live cumulative network meta-analyses could benefit various stakeholders, including physicians, patients and also guideline developers, funders, and decision-makers.

First, networks of trials and their synthesis through network meta-analysis could increase the value and reduce waste in research when treatment recommendations are based on an exhaustive up-to-date network of randomized evidence. Second, biomedical research funding agencies could improve the prioritization of research proposals if they had access to a mapping of existing and ongoing trials evaluating all the available treatments for a specific condition at the time a new trial is planned. Visualizing the network of trials and identifying which new trial maximizes the information can help stakeholders assess gaps in evidence and choose the next treatment comparison or trial that needs to be prioritized. Third, guideline developers may gain insights from the network meta-analyses. For instance, in the field of medical therapies for open angle glaucoma, Li et al. recently showed that if a network meta-analysis had been conducted earlier, prostaglandins could have been shown the most effective class in lowering intraocular pressure 7 years ahead of the guideline recommendation [Li, The Inaugural REWARD/EQUATOR Conference 2015].
We have modified the discussion section of the manuscript to highlight the value of the proposed solution.

Discussion section, page 19: “Nonetheless, embracing the perspective of networks of trials of all alternative treatments for each condition and in particular developing live cumulative network meta-analyses, could greatly benefit various stakeholders, including physicians, patients and also guideline developers, funders, and decision-makers[71]. Networks of trials and their synthesis through network meta-analysis could increase the value of research when treatment recommendations are based on an exhaustive up-to-date network of randomized evidence[72]. Guideline developers may gain insights from network meta-analyses, especially when considering recent developments to rate the quality of evidence supporting treatment effect estimates and rankings from network meta-analysis. For instance, in the field of medical therapies for open angle glaucoma, Li et al. recently showed that if a network meta-analysis had been conducted earlier, prostaglandins could have been shown as the most effective class in lowering intraocular pressure 7 years ahead of the guideline recommendation [73]. Moreover, biomedical research funding agencies could improve the prioritization of research proposals if they had access to a mapping of existing (and ongoing) trials evaluating all available treatments for a specific condition at the time a new trial is planned. Visualizing the network of trials and identifying which new trial maximizes the information can help stakeholders assess gaps in evidence and choose the next treatment comparison or trial that needs to be prioritized.”

The paper is well written, the methodology is sound and results are well presented. I find the discussion to be balanced and would think it would be of great interest to readers of BMC Medicine. I have no major comments but would like to raise some issues that the authors should consider in a revised version of the paper:

1. The resource-demands for doing systematic reviews are substantial and there are many unresolved questions about how to succeed with living systematic reviews both on the processes, methods and technology side, for authoring, publishing and updating. I would think this becomes even more challenging when moving to live cumulative network meta-analyses. The paper would be strengthened if authors address these challenges and also provide some examples of how this can be achieved, ideally how this has already been achieved within certain topics.

Answer:

We agree that live cumulative network meta-analyses will raise several challenges. In the discussion, we had proposed a novel methodology to perform knowledge synthesis that we call “live cumulative network meta-analysis”. We agree that we fell short on describing the methodological steps of such an approach. As suggested, we have now elaborated on the methodological steps of live cumulative network meta-analysis, some challenges to overcome at
each step and some potential solutions. For more clarity, we have chosen to present these different key points through a table and a figure.

The rigorous methodology of systematic reviews (exhaustive search of trials, minimization of subjectivity by independent duplicate assessments, assessment of risk of bias within trials) is inherently demanding of resources and time, especially for a systematic review incorporating network meta-analysis because in terms of the amount of information to process, it requires several conventional systematic reviews. Moreover, keeping a systematic review up to date requires processes closer to those of rapid reviews (ie, using accelerated and streamlined methods). Therefore, there is necessarily a trade-off between high-standard synthesis methods and real-time updating processes.

Automated technologies may help define this trade-off by alleviating the burden of manual tasks for systematic reviewers. Several teams are working on tools to improve, hasten, and ease the search for trials, trial selection, extraction of data, and assessment of risk of bias. The development and implementation of such tools will be feasible when tailored to a specific systematic review question; for instance, once search equations for a given clinical question have been defined, writing a script to automatically query the different information sources (eg, MEDLINE, EMBASE, CENTRAL) is relatively easy. Regarding the study selection process, any live cumulative network meta-analysis would start with an initial systematic review incorporating the network meta-analysis as a base, then supervised machine learning could be used to train a classifier based on the selection of reviewers during this initial phase to predict decisions regarding the eligibility of future records.

We also agree that live cumulative network meta-analysis may raise issues regarding the current authoring and publishing system. Online posting may be more adequate to report periodically the findings of such “real-time” syntheses.

Since Elliott et al. presented the theoretical framework of a living systematic review, some examples have been published [Badgett, JAMA Pediatrics 2015; Synnot, J Neurotrauma 2015]. To our best knowledge, these reviews partly addressed the aforementioned challenges. For instance, Badgett et al. published a report in JAMA Pediatrics with an accompanying website on GitHub where they give access to the update of their review.

We have modified the discussion section of the manuscript p 17-18:

“We propose to push further the shift towards a new paradigm by switching 1) from a series of standard meta-analyses focused on specific treatments (many treatments being not considered) to a single network meta-analysis covering all treatments and 2) from meta-analyses performed at a given time and frequently out-of-date to a cumulative network meta-analysis systematically updated as soon as the results of a new trial become available, an approach to synthesis we call
“live cumulative network meta-analysis. In Figure 5, we show the methodological steps we propose for live cumulative network meta-analysis.

We acknowledge that developing such methodology is challenging. In Table 2, we present some key challenges and potential solutions. The rigorous methodology of systematic reviews (exhaustive search of trials, minimizing subjectivity by independent duplicate assessments, assessing risk of bias within trials) is inherently demanding of resources and time, especially for a systematic review incorporating network meta-analysis. Moreover, keeping a systematic review up to date requires processes closer to those of rapid reviews (i.e., using accelerated and streamlined methods). Therefore, there is necessarily a trade-off between high-standard synthesis methods and real-time updating processes. Automated technologies may help define this trade-off by alleviating the burden of manual tasks for systematic reviewers. Several tools have been proposed to improve, hasten, and ease the search for trials, trial selection, extraction of data, and assessment of risk of bias[49, 50, 62, 63]. Live cumulative network meta-analysis may also raise issues regarding the current authoring and publishing system. Online posting may be more adequate to report periodically the findings of such “real-time” syntheses. Since Elliott et al. presented the theoretical framework of living systematic reviews, some examples have been published and have only partly addressed the aforementioned challenges[64-68], for instance, by using accompanying open-access websites to disseminate the updates of the systematic review.”

2. The quality of evidence assessment and presentation/ ranking of results in network meta-analyses represent key areas of development that will be crucial to make these analyses useful for clinical decision-making for example in the context of guidelines. Authors could refer to recent papers outlining advances also in these fields (e.g. GRADE paper in BMJ 2014) to demonstrate feasibility of their proposed solutions.

Answer:

We agree that our proposal of live cumulative network meta-analysis could benefit guideline developers. The recent GRADE approach to network meta-analysis allows for grading the quality of evidence supporting treatment effect estimates and the treatment ranking derived from network meta-analysis. Although this process is challenging, we agree that these developments should be incorporated in a live cumulative network meta-analysis.

We have added a comment in the Discussion to address this point p19.

“Networks of trials and their synthesis through network meta-analysis could increase the value of research when treatment recommendations are based on an exhaustive up-to-date network of randomized evidence[72]. Guideline developers may gain insights from network meta-analyses, especially when considering recent developments to rate the quality of evidence supporting treatment effect estimates and rankings from network meta-analysis.”
3. Also, the manuscript is a bit sparse on emphasizing the importance of including all patient-important outcomes in the systematic reviews and network meta-analyses and I would suggest they make this even more clear, with relevant references (e.g. GRADE papers on systematic reviews and network meta-analysis).

Answer:

The reviewer is correct, we did not emphasize the importance of addressing patient-important outcomes in network meta-analyses. Our analysis included trials regardless of the reported outcomes. In a sensitivity analysis, we excluded trials that did not report treatment effects on overall survival or progression-free survival, because the two endpoints are the most frequently used to assess the efficacy of second-line treatments of advanced NSCLC. However, there are other patient-important outcomes, in particular to measure the symptom burden of the disease and the quality of life of patients.

More generally, we agree that it is crucial to consider the network of trials according to the reporting of all patient-relevant outcomes. In fact, the geometry of the network of trials could vary across outcomes because of differential reporting of outcomes (e.g., efficacy and safety outcomes) across drugs and trials.

We expanded the Discussion to address this comment p18-19:

“Another challenge would be to consider all outcomes that are important or critical to patients for decision making in these live cumulative network meta-analyses [69, 70]. In our case study, we included trials regardless of reported outcomes; in a sensitivity analysis, we excluded trials that did not report treatment effects on overall survival or progression-free survival. However, there are other patient-important outcomes, in particular to measure the symptom burden of the disease and the quality of life of patients. More generally, it will be crucial to consider networks of trials according to the reporting of the different patient-important outcomes. In fact, the geometry of the network of trials could vary across outcomes because of differential reporting of outcomes (e.g., efficacy and safety outcomes) across drugs and trials.”

Response to the reviewer Sharon Straus

Reviewer #3: the authors have identified an interesting issue around comprehensiveness of systematic reviews. they have accurately presented their article that it is important for sys reviews to include relevant articles to fully inform clinical and policy decision making. their project then proceeds to assess whether published sys reviews included all relevant trials on a particular topic.

some questions that the authors could consider to further strengthen their review:
1. what was the rationale for the inclusion date of 2009?

Answer:

Our methodology consisted of repeatedly identifying all randomized trials and all systematic reviews available up to Dec 31, 2009, then up to Dec 31, 2010, and so forth up to March 31, 2015. As a consequence, 2009 was not defined as an inclusion date because any randomized trial results and any systematic reviews were eligible regardless of publication date.

Instead, we pre-specified the year 2009 as a starting point for our analyses. We acknowledge it was an arbitrary choice; the rationale was to allow for a sufficient amount of evidence (in terms of both randomized trials and available systematic reviews) to initiate the comparison between the available randomized evidence and that covered by systematic reviews, with randomized evidence comparing competing second-line treatments for advanced NSCLC.

We could modify the start date of 2009, considering that the first systematic review was published in Apr 2001; the second one was published in Dec 2005. However, there was no or little randomized evidence available between competing treatments at that time.

In the revised version of the manuscript, we have modified the Methods and Discussion sections to clarify the rationale and the limitation associated with our choice:

Methods section, page 6: “We first used a comprehensive strategy to repeatedly identify all randomized trials, with published and unpublished results, and all systematic reviews of second-line treatments for advanced NSCLC available up to the end of each year from 2009 to 2015. Second, we sequentially assessed the amount of randomized evidence that was covered by systematic reviews collectively: for the years 2009 to 2015, we assessed the articles published up to December 31 of each of those years for proportion of treatments, treatment comparisons, trials, and patients covered by systematic reviews on this topic, with comparison to the total randomized evidence available at each time.”

Methods section, Definition of randomized evidence available for inclusion in systematic reviews subsection, page 9: “We pre-specified the year 2009 as a starting point for our analyses in order to allow for a sufficient amount of evidence (in terms of both randomized trials and available systematic reviews) regarding the comparison between competing second-line treatments for advanced NSCLC to initiate a comparison between the available randomized evidence and that covered by systematic reviews.”

Discussion, page 20: “Fifth, we started our analysis for the year 2009; this pre-specified year was somehow arbitrary and we acknowledge that this starting point could have been earlier, considering that the first systematic review was published in April 2001; the second one was published in December 2005. However, there was no or little randomized evidence available between competing treatments at that time.”
2. was their lit search conducted by an information scientist? did a second librarian peer review it?

Answer:

Yes, the literature search was conducted by an information scientist at Cochrane France. The search strategy was developed according to the standards described in the Cochrane Handbook, the Methodological Standards for the Conduct of Cochrane Intervention Reviews (MECIR) and the methods described by An-Wen Chan (Chan, BMJ 2012).

However, we acknowledge that the literature search was not checked by a second librarian. We are not aware of such recommendation. The fact that we identified more trials than any systematic review on the topic is a form of external validation of the search strategy.

3. was a calibration exercise done before screening?

Answer:

Yes, according to the recommendations of the Cochrane Handbook, we pilot-tested the eligibility criteria on a sample of 100 records (for the selection on titles and abstracts) and 10 reports (for the selection on full-text). This pilot test was used to refine and clarify the eligibility criteria, train the two independent reviewers who applied them and ensure that the criteria were applied consistently by these two reviewers.

Methods section, Selection of studies and extraction of data subsection, page 8: “Two authors independently and in duplicate examined titles, abstracts and full-text articles to determine the eligibility of randomized trials and systematic reviews. We pilot-tested the eligibility criteria on a sample of 100 records (for the selection on titles and abstracts) and 10 reports (for the selection on full-text articles) to ensure that the selection criteria were applied consistently by the two authors.”

4. why were different drug administration schemes excluded?

Answer:

We pre-specified the exclusion of randomized trials and systematic reviews comparing two different administration schemes because our focus was a comparison of alternative treatments against each other. Randomized trials comparing only two different administration schemes of the same drug (eg, docetaxel 75 mg/m² once every 3 weeks versus docetaxel 35 mg/m² once a week) would not be informative regarding the comparison between alternative treatments.

Nonetheless, our analyses could be extended to randomized trials and systematic reviews comparing two different administration schemes. In such cases, each relevant node would have several subnodes that relate to different administration schemes.
We added this point as a limitation in the revised Discussion section, page 20:

“Similarly, we excluded randomized trials and systematic reviews comparing two different administration schemes because our focus was the comparison of alternative treatments against each other. Nonetheless, our analysis could be extended to such randomized trials and systematic reviews. In such cases, each relevant node would have several subnodes that relate to different administration schemes.”

5. did they assess the quality of the systematic reviews? this would be an important piece of the project to include

Answer:

In the revised version of the manuscript, we assessed the methodological quality of the systematic reviews. We used AMSTAR, a measurement tool created to assess the methodological quality of systematic reviews [Shea, BMC Med Res Methodol 2007], which has been validated [Shea, PLOS One 2007; Shea, J Clin Epidemiol 2009]. Two authors independently assessed the AMSTAR items, with a formal consensus process in case of disagreement.

More specifically, we assessed the following items: Was there duplicate study selection and data extraction?; Was a comprehensive literature search performed?; Was the status of publication (ie, grey literature) used as an inclusion criterion?; Was a list of studies (included and excluded) provided?

We chose these items because they pertain to the only two methodological processes of systematic reviews (searching for studies and selecting studies) that are directly related to the potential gap between the amount of randomized evidence covered by systematic reviews and the amount of randomized evidence available for inclusion. For instance, two other questions are “Were the characteristics of the included studies provided?” and “Was the scientific quality of the included studies assessed and documented?” which are not especially relevant to our objective.

In our assessment of the methodological quality of the 29 systematic reviews, 45% of systematic reviews did not report independent study selection and data extraction, 31% did not report a comprehensive literature search (at least two electronic sources + one supplementary strategy among reviews, experts or reviewing the references), and 45% did not report a search of the grey literature (conference abstracts, non-industry trial registries and results databases, industry trial registries and results databases, regulatory agency online databases) or the exclusion of reports based on language. Of note, 17% did not search conference abstracts, 86% did not search non-industry trial registries and results databases (ClinicalTrials.gov), 97% did not search industry trial registries and results databases, and 97% did not search regulatory agency online databases.
In all, 79% of systematic reviews did not report duplicate study selection and data extraction, comprehensive literature search and searching for reports regardless of their publication type.

We have updated the manuscript to report this new analyses and interpretation.

Methods section, Selection of studies and data extraction subsection, page 8-9: “Two reviewers independently assessed the methodological quality of the systematic reviews, with a formal consensus process in case of disagreement. We used AMSTAR, a measurement tool created to assess the methodological quality of systematic reviews[13], which has been validated[14, 15]. We assessed the 4 items pertaining to duplicate study selection and data extraction, comprehensive literature search (at least two electronic sources and one supplementary strategy among reviews, experts or reviewing the references), searching for reports regardless of their publication type, and providing a list of included and excluded trials. Searching for trials regardless of their publication type was judged inadequate when authors did not report searching the grey literature (conference abstracts, non-industry trial registries and results databases, industry trial registries and results databases, regulatory agency online databases) or excluding reports based on language. We focused on these 4 specific items because the methods used for the identification and selection of studies are directly related to a potential gap between the amount of randomized evidence covered by systematic reviews and the amount of randomized evidence available for inclusion, and other domains are unrelated”

Results section, Systematic reviews of second-line treatments for NSCLC subsection, page 13: "Regarding the methodological quality of the 29 systematic reviews, 45% of reviews lacked information on independent study selection and data extraction, 31% a comprehensive literature search, and 45% a search for reports regardless of their publication type. Of note, 17% of reviews lacked information on a search for conference abstracts, 86% a search for non-industry trial registries and results databases, 97% a search for industry trial registries and results databases, and 97% a search of regulatory agency online databases. In all, 79% of systematic reviews did not report duplicate study selection and data extraction, comprehensive literature search and searching for reports regardless of their publication type. Finally, 7% of reviews provided a list of included and excluded trials.”

Discussion section, page 16:”Among the 29 systematic reviews, 79% could be considered at high risk of missing trials that would have met the inclusion criteria because they did not report duplicate study selection and data extraction, comprehensive literature search or searching for reports regardless of publication type”

6. can they outline whether the trials that they identified would have been eligible for inclusion in the sys reviews - it is not clear whether these were excluded because they weren't eligible or if the authors of the reviews didn't identify them. they mention the lack of including grey literature was a factor but it would be useful to consider eligibility as well

Answer:
The reviewer is correct; we considered eligibility in addition to the search methods. We confirm that the trials we identified would have always been eligible for inclusion in at least one systematic review (in terms of patients, interventions and comparators). Therefore, we can suppose that authors of the reviews did not identify them because of lack of screening grey literature.

In an effort to clarify that trials were not excluded because they were not eligible, we have edited the sentences in the Methods section and the Discussion section that already addressed this question. They now read as follows.

Methods section, Definition of randomized evidence available for inclusion in systematic reviews subsection, page 9: “From 2009 to 2015, we identified the cumulative list of trials eligible for inclusion in systematic reviews; we checked that each trial identified would have been eligible for inclusion in at least one systematic review (i.e., corresponded to the selection criteria in terms of patients, interventions and comparators).”

Discussion section, page 15: “However, all trials we identified would have been eligible for inclusion in at least one systematic review; missing trials were not excluded from systematic reviews because they were not eligible.”

7. how would the results have changed by including the various trials? would these have made a clinically important difference to the results? this is the ultimate issue and would be terrific if the authors could provide this information.

Answer:

Our understanding is that the reviewer asks if the relative effects between treatments would have changed if systematic reviews had covered 100% of the randomized evidence available at each time.

In this study, our objective was to assess the gap between the amount of available randomized evidence and the amount covered by systematic reviews over time. As a consequence, we did not perform any outcome data synthesis at this stage.

The question of the reviewer concerns 10% to 17% of treatment comparisons that were partially covered by systematic reviews. In these cases, we acknowledge that one may compare the effect sizes as estimated in the original systematic reviews to that obtained with meta-analyses of the complete evidence.

However, it would be very difficult to address this comparison because the systematic reviews addressed a variety of comparisons by frequently lumping together similar but not identical treatments. Moreover, we think such analyses would go beyond the scope of our work and would put the focus back on conventional meta-analyses, whereas our objective was to promote the adoption of a network perspective to broaden the synthesis scope to all treatments and to all
evidence. In this framework, a network meta-analysis would allow for estimating all treatment comparisons, whereas the direct evidence would inform 54 comparisons (only 5%). Adding 40% of missing evidence (8,000 patients) to the network would likely lead to clinically important differences, in particular for treatment comparisons partially covered by systematic reviews.

We have addressed this point in the Discussion p20:

“Finally, we did not perform any outcome data synthesis at this stage. One may ask if covering all the randomized evidence available would have led to clinically important differences for the 10% to 17% of treatment comparisons partially covered by systematic reviews. In the framework we are promoting, a network meta-analysis would allow for estimating all treatment comparisons. Adding up to 40% of missing evidence (about 8,000 patients) to the network would likely lead to clinically important differences, in particular for treatment comparisons partially covered by systematic reviews.”

minor comments - there are a couple of typos so suggest a careful read.

8. suggest in the 'inclusion criteria' that the material be separated into search and eligibility criteria to make it a bit easier to read

Answer:

We corrected the typos:

page 7: 6) regulatory agency online databases (US Food and Drug Administration and European Medicines Agency);

page 10: (ie, an edge connected two nodes when at least one randomized trial compared the two corresponding treatments).

page 13: In 2014, 27 reviews still did not cover 18 treatments (40%), 20 treatment comparisons (38%), 34 trials (46%), and 8,486 patients (30%).

page 17: 1) from a series of standard meta-analyses focused on specific treatments (many treatments being not considered) to a single network meta-analysis covering all treatments

We separated the “inclusion criteria” into two parts as requested, page 6 and 7.

9. since the authors did a sys review, did they include a prisma checklist? i couldn't find it but perhaps i missed it.

Answer:
We have not included the PRISMA checklist. We are aware of the PRISMA extension statement for systematic reviews incorporating network meta-analysis. However, we are not reporting a systematic review and (network) meta-analysis. Instead, we are reporting a piece of methodological research in which we assessed the gap between evidence covered by systematic reviews and available RCTs on a specific topic. We used a novel approach by using a series of systematic overviews and networks of randomized controlled trials (RCTs). As a consequence, the PRISMA checklist would not be adequate to appraise the content, clarity, and transparency of our article.