

# Comment to AHRQ:

---

A Framework for Assessing the Strength of Methodological Recommendations for Systematic Review and Meta-analysis

## Summary

We are very glad to have the opportunity to respond to the draft of this framework and also wish to emphasize that we appreciate the importance of this endeavor.

In summary, we find that the proposed framework has only limited applicability because only simple recommendations, which are rare in practice, can be rated. Overall, the conduct of systematic reviews is a multidisciplinary endeavor, and this should be reflected in such a framework. We suggest considering this aspect throughout the whole draft. Our review explicitly refers to these different perspectives.

## Background

On page 9, the authors state that certain methods are “often applied dogmatically due to popularity rather than methodological appropriateness”. We would suggest deleting the words “due to popularity”, as we believe that dogmatic recommendations are the reason for, rather than the consequence of, the respective method’s popularity. The statement “The same is true ...” (page 10, first line) is not supported by the evidence, in our view. It would be useful to know which guidance papers have clearly recommended the use of summary quality scores or of continuity corrections for meta-analyses of studies with rare events.

## Description, explanation and elaboration

On page 28, the relevance of empirical data for methodological recommendations is stressed. In many cases, however, such data will be lacking. Therefore, we suggest mentioning that methodological recommendations should also address the research needs surrounding the methodology of systematic reviews. For example, if two different methodological options are available, but there are no good data for favoring one option over the other, one could recommend one of the two options based on expert opinion. An alternative and probably more suitable methodological recommendation would be to recommend both options, with a subsequent comparison of the results. This would improve the empirical evidence for future systematic reviews.

In section 3.4.3 the “expected practical impact of implementation” is addressed. However, it remains unclear what impact is referred to in this section: the impact on the results of the meta-analysis, or the impact on the research process (e.g. the need for additional funding or specific software, delays in the completion of the meta-analysis)? For clarity, we suggest adding this information to the first

sentence of the chapter: “The practical impact that methodological recommendations may have on the results of a systematic review can be large or marginal, ...”.

Regarding insufficient practices and the question of whether to assign a third category, we believe that insufficient practices alone strongly support the creation of a third category. (If practices are insufficient or unhelpful, it is very important to indicate this.)

Table 1: The recommendations R2 and R3 in the cited papers [3] and [4] could not be found. The information that these recommendations have been "reworded" is, in our opinion, rather inaccurate. The recommendations provided are simplified versions of the original recommendations, which are far too complex to be dealt with within the given framework. R3 in particular has been simplified to such an extent that it no longer represents the original recommendations given in reference [4].

In our opinion the feasibility of the proposed framework is limited in practice. For example, a definition and description of the available alternative choices is only possible for very clear and simple recommendations. In most cases, however, these choices can only be roughly described, which is inadequate for application in practice or requires a separate methodological review paper. For instance, how long would a table with available alternative choices have to be, e.g., for assessing consistency in mixed treatment comparisons?

The same holds for performance measures. Either one provides only a rough (and inadequate) statement, e.g., "Maximize SR/meta-analysis credibility" or one provides a long list of statistical performance measures and features (bias, standard error, mean square error, power, coverage probability, technical complexity, computation time, etc.), with the common problems that (a) it is unclear which measure is the most important one, and (b) there is no method which is best suited for all measures and features.

The proposed decomposition of recommendations into testable and nontestable parts is only possible for very simple recommendations, which are rare in practice. This is demonstrated by the simplified recommendation R3. The recommendations presented in [4] do not include the nontestable statement using random effects in the meta-analysis of diagnostic studies. In our view, the use of random effects is not a "belief" in the generation of data but a requirement to deal adequately with the heterogeneity of the data at hand. The original recommendations given in [4] are much more complex and the proposed framework is not suitable to assess the overall strength of these recommendations.

It is unclear to us what can be gained by decomposing R5 into R5.1 and R5.2. We do not consider R5.1 to be nontestable; it is clearly testable, because in our opinion it is wrong (as shown by [7]).

## **Minor comments**

“Manten-Haenszel” should be changed to the correct “Mantel-Haenszel”. This typo appears in the text several times.

## Discussion

Page 41, line 6, statement: "Finally, the feasibility and ease of using the framework itself is not clear." The attempt to assess a typical methodological recommendation would very quickly show the limitations of the framework. For example, it would be very challenging to try and create a table with the background context for the recommendation "To obtain a summary sensitivity and specificity use the theoretically motivated bivariate meta-analysis models" given in [4].

## References

We suggest adding a guidance report on the methodology of systematic reviews:

Centre for Reviews and Dissemination (CRD). Systematic Reviews. Guidance for Undertaking Reviews in Health Care. CRD, University of York. January 2009, 3<sup>rd</sup> ed., pages 1-292. Available online: [http://www.york.ac.uk/inst/crd/index\\_guidance.htm](http://www.york.ac.uk/inst/crd/index_guidance.htm)