ISCB Subcommittee Statistics in Regulatory Affairs Comment to AHRQ: A Framework for Assessing the Strength of Methodological Recommendations for Systematic Review and Meta-analysis

Abstract

No comments

Background

It is unclear from this section on what empirical data it is being proposed to assess the methodological recommendation for its rigor, evidence base, applicability and feasibility.

We question the validity of the statement "The same is true …" on page 10, first line. Which guidance papers have clearly recommended the use of summary quality scores or the use of continuity corrections for meta-analyses of studies with rare events?

Developing the framework

It is unclear if the hypothetical recommendation statements were also evaluated in regards to their impact in the field if applied without empirical evidence. Would one's impact if applied be more than another (good or bad), regardless of its degree of specificity and/or technical sophistication?

Description, explanation and elaboration

Table 1: We cannot find the recommendations R2 and R3 in the cited papers [3] and [4]. The description that these recommendations are "reworded" is highly misleading. The given recommendations are simplified versions of the true recommendations which are much too complex than one can deal with them with the given framework. This is true especially for R3, which is so strongly simplified that the result has nothing to do with the true recommendations given in the reference [4].

We question the feasibility of the proposed framework in practice. For example, the definition and description of the available alternative choices is only possible for very clear and simple recommendations. In most cases, however, the description of the available alternative choices can only be done in a very rough way which is useless for practice or requires an own methodological review paper. How long would a table with available alternative choices be, e.g., for assessing consistency in mixed treatment comparisons?

The same holds true for the measures to optimize. Either one writes always the rough useless statement "Maximize SR/meta-analysis credibility" or you give 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 frequent problems that it is unclear which

measure is the most important one and that there is no method which is the best for all measures and features.

The proposed decomposition in testable and nontestable parts is only possible for very simple recommendations, which are rare in practice. This is quite obvious in the light of the simplified and misleading presentation of recommendation R3. The recommendations presented in [4] are not the nontestable statement to use random effects in meta-analysis of diagnostic studies. The use of random effects is not a "belief" of the data generation but a requirement to deal adequately with the heterogeneity of the data at hand. The true recommendations given in [4] are much more complex and the proposed framework is not helpful to assess the overall strength of these recommendations.

When formulating the testable and nontestable hypotheses, it is not clear why the authors consider only one formulated testable hypothesis from each statement and no other alternative but similar testable hypotheses that could be derived from the same statements. How will these or should these be considered in the framework?

The extensive discussion in the document on pages 23 - 26 (example R5 Funnel plot asymmetry, see also table 6) mirrors a long-lasting dissent between two groups of researchers in meta-analysis, the "US-American" group (Olkin, Ioannidis, Lau, Schmid, Terrin, Trikalinos, ...) rejecting funnel plot-based methodology and a UK-based group (Bristol (Sterne and coworkers) and Leicester (Sutton and coworkers)) proposing it. For the recommendations paper in BMJ (reference 34 of the draft), Jonathan Sterne, in an invaluable effort, finally succeeded in bringing all parties to the table after about six years of working. Obviously, the present draft document is written from the "American" point of view.

Nevertheless, it is true that things often happen as described in this context on page 23 ("A seminal paper in the British Medical Journal popularized a method for testing for funnel plot asymmetry. Its authors were very careful in explaining that publication bias is only one of many explanations of funnel plot asymmetry, but many others have chosen to ignore this advice"). The same was, e.g., true for the well-known I^2 (the authors were very careful in explaining that I^2 depends on the study size, but many others have chosen to ignore this advice) (Higgins &Thompson (2002): Statistics in Medicine 21:1539-1558; Higgins et al. (2003): BMJ 327:557-560).

We do not understand why the decomposition of R5 into R5.1 and R5.2 is helpful. It is not true that R5.1 is nontestable. Recommendation R5.1 is clearly testable, because it is wrong (as shown by [7]).

How do the authors define “expert opinion”? How is an expert defined in the context of this framework?

It is concerning that in 3.4.2 for Feasibility, that the reason why certain methods using statistical modeling and software are rated “low” is that the common public does not have access to these methods and software, but that those (R2) that only require a basic understanding of statistical concepts are rated with a “high” feasibility score. However, if reliance on sound empirical evidence and statistical comparisons are important, shouldn’t appropriate statistical modeling be the preferential choice?

In summary, we do not think that the proposed framework is useful because only very simple recommendations can be rated, which are very rare in practice.

Discussion

The authors indicate that the feasibility and ease of the framework is not clear and that is needs guidance on deriving scores, but that it is quite general so it can be applied to other methodological approaches. It seems that this is contradictory.

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 show very quickly that the framework is not feasible and not easy. For example, try to 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].

It seems highly aimed to publish recommendations for recommendations. We are not sure whether they are really needed and whether the crucial distinction of testable and nontestable statements is useful.

References

No comments

Tables

Se above

Figures

Se above

Minor comments

Table 3/Table 7-should be "Mantel-Haenszel" not "Manten"?