# Comment to AHRQ:

## Assessing Confounding, the Risk of Bias and Precision of Observational Studies of Interventions or Exposures: Further Development of the RTI Item Bank

### Introduction

We are pleased to have the opportunity to comment on the draft of this research report and wish to emphasize that we acknowledge the importance of this work. We greatly appreciate this proposal to develop a framework and believe that it will be a helpful tool for other systematic reviewers to use or build upon when assessing observational studies, especially cohort and case-control studies. However, in our opinion the following issues deserve attention.

It seems that only case series, case-control studies, cohort studies, and cross-sectional studies are considered. Why is there no consideration of non-randomized controlled trials (according to Appendix C)?

P. 3: Why are only 3 conditions for causality mentioned in the introduction? Several more conditions are listed on page 14 (criteria by Hill).

The disadvantages listed for RCTs are not a feature of randomization per se. It should be underlined that problems related to length of follow-up, insufficient sample sizes for rare events and subgroup analyses are general problems independent of study type, and not problems of randomization. As explained in Appendix A, such problems may be even more pronounced in observational studies.

P. 5: It is correctly stated that "the risk of bias will always be greater for non-randomized studies than for randomized studies". We propose avoiding the term "risk of bias" for non-randomized trials, as non-randomized trials always have a high risk of bias. It is therefore only possible to further classify non-randomized trials within the class of studies with a high risk of bias. For this purpose an alternative terminology would be useful.

P. 6: Project Objectives: Throughout the report and appendices, the relevance of the assessment of precision seems to vary. The title of the report indicates that all topics named are addressed equally, but most parts of the text deal with bias and confounding. The title and some passages in Appendix A (p. A-1, e.g. "these risks of bias" after describing "threats to validity and precision") even seem to suggest that precision is part of the assessment of bias.

## Methods

P. 6: The previously developed taxonomy of observational studies is suggested for reviewers "to guide the choice of questions needed for risk of bias assessments", although the corresponding classification tool given in Appendix C has only moderate reliability and low accuracy (Hartling et al., 2011). Furthermore, the "questions … by observational study design type" are grouped by only four definitions according to Appendix B (case series, case-control, cohort or cross-sectional study). The further diversity of study design features and the corresponding bias are covered by the item catalogue. Other guidance, for example the Cochrane guidance, explicitly does not advocate using labels and instead recommends that review authors only use explicit (multiple) study design features. This guidance seems to be in accordance with the decision to eliminate the question "Is the study design prospective, retrospective, or mixed?", which was judged to be "problematic and uninformative" by the Working Group (p. 9). Against this background, we suggest putting the relevance of the classification tool into perspective.

Reference:

Hartling, L., Bond, K., Santaguida, P.L., Viswanathan, M. & Dryden, D.M. (2011): Testing a tool for the classification of study designs in systematic reviews of interventions and exposures showed moderate reliability and low accuracy. *J. Clin. Epidemiol.* 64, 861-871.

## Results

P. 9: Why is the determination of evidence of reporting bias outside the scope of the risk-of-bias assessment? Reporting bias usually represents an important aspect of bias.

P. 12 and Appendix B: The text and Appendix B seem to suggest that "risk of bias" and "confounding" can be clearly distinguished. However, Table 3 on p. 12 shows for example that recruiting strategies differing across groups may result in both "selection bias" and "confounding" (see items Q2 and Q3). Furthermore, item Q6 is related to "questions to assess the risk of bias" and "questions to assess confounding". We suggest explaining in more detail how the assessment of "risk of bias" and "confounding" should relate to each other.

The use of the term "confounder" or "potentially confounding variable" instead of "confounding" could be considered, e.g. in item Q6: *"Were valid … measures … used to assess inclusion/exclusion criteria, intervention/exposure outcomes … and (potential) confounders"*.

P. 14: The authors refer to the GRADE approach. In describing GRADE's method of grading evidence, the description of study designs is inaccurate. In the GRADE approach, "randomized controlled trials (RCTs) start as high-quality evidence and observational studies as low-quality evidence supporting estimates of intervention

effects" (Guyatt et al., 2011), which means that the GRADE approach differentiates between RCTs and (all other) non-randomized study designs.

Reference:

Guyatt, G., Oxman, A.D., Akl, E.A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., DeBeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P. & Schünemann, H.J. (2011): GRADE guidelines: 1. Introduction – GRADE evidence profiles and summary of findings tables. *J. Clin. Epidemiol.* 64, 383-394.

## Discussion

No comment

## References

No comment

## Tables

Tables 2 and 3, Q3: This item proposes that suboptimal comparison groups have to be accepted if "feasibility and ethical considerations" are taken into account. We believe that the risk of bias should be considered separately from technical and ethical aspects. We fear that this item opens a back door to biased studies becoming acceptable as trustworthy evidence, only because a better study design was considered unfeasible. If the assessment of feasibility also includes the consideration of the organizational and funding circumstances of research, the door towards bias is opened very wide here.

Tables 2 and 3, Q5: We suggest the inclusion of a similar question regarding the blinding of exposure assessors to disease status in case-control studies.
This item on blinded outcome assessment can be dropped as non-applicable, "when clinical evaluators cannot be blinded to exposure status". However, if the assessment of risk of bias is the aim of the item bank, it makes no difference for what reasons the outcome assessment was not blinded. If a blinded outcome assessment is impossible, the study quality may be the best possible under the specific circumstances, but the study results will still be biased. We suggest deleting this possibility of "justified" bias.

Tables 2 and 3, Q8: Q8 addresses missing values due to loss to follow-up and is therefore only relevant for cohort studies. A corresponding question regarding other types of missing data in other study types would be useful (e.g. missing values of exposure in case-control studies).

We suggest providing a reference for the statement that "Cochrane standard for attrition is 20 percent for shorter term (<1 year) and 30 percent for longer term (≥ 1 year)".

No threshold is provided in the instructions for the assessment of differences in loss to follow-up. We suggest that bias should be suspected if attrition rates differ by more than 10% between groups (Kristman et al., 2004).

Reference:

Kristman, V., Manno, M. & Côté, P. (2004): Loss to follow-up in cohort studies: How much is too much? *Eur. J. Epidemiol.* 19, 751-760.

Table 3, Q12: If no adjusted analysis was performed in the study, this item should be scored as negative. However, if the comparison groups in a study are similar with regard to all known confounders (this may be due to pure chance), there is no need to adjust the analyses for these confounders. If the comparison groups are similar, this item should therefore also be scored as positive.

Table 3, Q15 and Q16: Inadequate statistical methods may not only lead to reduced precision, but also to flawed (biased) results. The last sentence of the instructions to Q15 implies that "risk ratio" and "relative risk" are different effect measures, although they in fact mean the same thing. Furthermore, it is not stated that the risk ratio should be calculated instead of (or in addition to) what measure (presumably the odds ratio)? It also remains unclear why especially this aspect of calculating the risk ratio in cases where prevalence is greater than 10 percent is explicitly addressed (out of many other possible methodological aspects that are important for the adequate application of statistical methods), and why there is no corresponding statement for Q16.

P. 14: Please write "Schünemann" (instead of Shunemann).

# Figures

See above

# Appendices

See above