**Put under “risk of bias” section**

**SoF tables in Review**

**Overall quality of evidence for main outcomes**

Quality of evidence

We planned to assess the quality of evidence for the main comparison at the outcome level using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (Guyatt 2011a). This methodological approach considers evidence from randomised controlled trials as high quality that may be downgraded based on consideration of any of five areas: design (risk of bias), consistency across studies, directness of the evidence, precision of estimates and presence of publication bias. (Guyatt 2011a). The GRADE approach results in an assessment of the quality of a body of evidence in one of four grades: 1) High: We are very confident that the true effect lies close to that of the estimate of the effect; 2) Moderate: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different; 3) Low: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect; 4) Very Low: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect (Schünemann 2013).

The review authors planned to independently assess the quality of the evidence found for outcomes identified as critical or important for clinical decision making. These outcomes include: Please include the outcomes that would have been included in SoF tables.

In cases where we planned to consider the risk of bias arising from inadequate concealment of allocation, randomised assignment, complete follow-up or blinded outcome assessment to reduce our confidence in the effect estimates, we planned to downgrade the quality of evidence accordingly (Guyatt 2011b). We evaluated consistency by similarity of point estimates, extent of overlap of confidence intervals and statistical criteria including measurement of heterogeneity (I2). We planned to downgrade the quality of evidence when large and unexplained inconsistency across studies results was present (i.e. some studies suggest important benefit and others no effect or harm without a clinical explanation) (Guyatt 2011d). We planned to assess precision based on the width of the 95% confidence interval (CI) and by calculating the optimal information size (OIS). If the total number of patients included in the pooled effect estimation was less than the number of patients generated by a conventional sample size calculation for a single adequately powered trial, we planned to downgrade for imprecision (Guyatt 2011c). When trials were conducted in populations other than the target population, we planned to downgrade the quality of evidence because of indirectness (Guyatt 2011e).

We planned to enter data (i.e. pooled estimates of the effects and corresponding 95% confidence Interval) and explicit judgments for each of the above aspects assessed into the Guideline Development Tool, the software used to create Summary of Findings (SoF) tables (GRADEpro 2008). We planned to explain all judgements involving the assessment of the study characteristics described above in footnotes or comments in the SoF table.

**Differences between protocol and review**

We added the methodology and plan for Summary of findings tables and GRADE recommendations, which were not included in the original protocol. These will be applied to future updates.