Appendix 2 (as supplied by the authors): The GRADE approach to evidence synthesis

The quality of evidence will be categorized as follows:

- **High:** Further research is very unlikely to change the confidence in the estimate of effect.
- **Moderate:** Further research is likely to have an important impact in the confidence in the estimate of effect.
- **Low:** Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
- **Very Low:** Any estimate of effect is very uncertain.

The evidence was graded on the domains in the following manner:

1. Study design
Randomized controlled trial
Controlled clinical trial, not randomized
Crossover trial

The quality of the evidence was downgraded:

- by ONE level: if >25% of participants were from a controlled clinical trial that was not randomized or if >25% of participants were from a crossover trial that did not report the first phase results separately to reduce potential carryover effect.

2. Risk of bias
Limitations in the study design and implementation may bias the estimates of the treatment effect. Our confidence in the estimate of the effect was reduced if studies suffer from major limitations.
The quality of the evidence was downgraded:

- by ONE level: if >25% of participants were from studies with low quality methods (e.g. PEDro score <7)

3. Inconsistency
Inconsistency refers to an unexplained heterogeneity of results. Results are considered consistent when the direction and effect size are sufficiently similar to lead to the same conclusion, and by the I^2 test. Inconsistency may arise from differences in: populations (e.g. drugs may have larger relative effects in sicker populations), interventions (e.g. larger effects with higher drug doses), or outcomes (e.g. diminishing treatment effect with time).

The quality of the evidence was downgraded:

- by ONE level: if the statistical heterogeneity or variability in results was large (e.g. I^2 above 75%)
- by TWO levels: if the statistical heterogeneity or variability in results was large AND there was inconsistency arising from populations, interventions or outcomes

There was no downgrade if only one study is present.

4. Imprecision
Results are imprecise when studies have wide confidence intervals around the estimate of the effect. In this case, we judge the quality of the evidence lower than it otherwise would because of resulting uncertainty in the results. Results of small trials were judged precise if there was no effect but the narrow 95% confidence interval did not cross predetermined

thresholds (see below). We focused directly on the precision of the effect estimate rather than the number of studies which we consider a surrogate measure of imprecision.

The quality of the evidence was downgraded:
For continuous outcomes:

- by ONE level: if the 95% confidence interval included no effect and upper or lower limit crossed an effect size (mean difference) of 10% of the applied outcome scale in either direction
- by ONE level: if the 95% confidence interval included no effect and upper or lower limit crossed an effect size (standardised mean difference) of 0.5 in either direction

For dichotomous outcomes:

- by ONE level: if the 95% confidence interval included no effect and upper or lower limit crossed an appreciable benefit or appreciable harm of a relative risk reduction (RRR) or relative risk increase (RRI) greater than 25%

5. Indirectness

Indirection refers to a discrepancy between the population, intervention, comparator, or outcome for the included studies. Indirectness was not assessed as the population, intervention, comparator, and outcome being addressed in this systematic review were selected within the research question and eligibility criteria.

The quality of evidence was downgraded:

- by ONE level: if there was indirectness in only one area of population, intervention, comparator or outcomes.
- by TWO levels: if there was indirectness in two or more areas of population, intervention, comparator or outcomes.

6. Publication bias

Publication bias is a systematic underestimate or an overestimate of the underlying beneficial or harmful effect due to the selective publication of studies. The quality of evidence was downgraded by ONE level if a funnel plot could be constructed and the funnel plot suggested publication bias.