

## Supplementary Material

### Sensitivity analysis using other possible criteria

In the first stage, a cut-off point of 5% discrepancy between reported and reproduced results was set to screen meta-analyses. In Figure S1 this criterion is compared with other possibilities in absolute value. Figure S1 shows that the criterion used is one of the most liberal of those compared. Due to our design, the stage where the criterion was applied is only an initial screening, so the meta-analyses labelled by this criterion were reviewed at later stages in a qualitative way. Therefore, a more liberal criterion (minimizing false negatives, but increasing false positives) is more suitable. This criteria has the limitations of any relative index, but due to the mix of metrics included, it is also considered more appropriate.

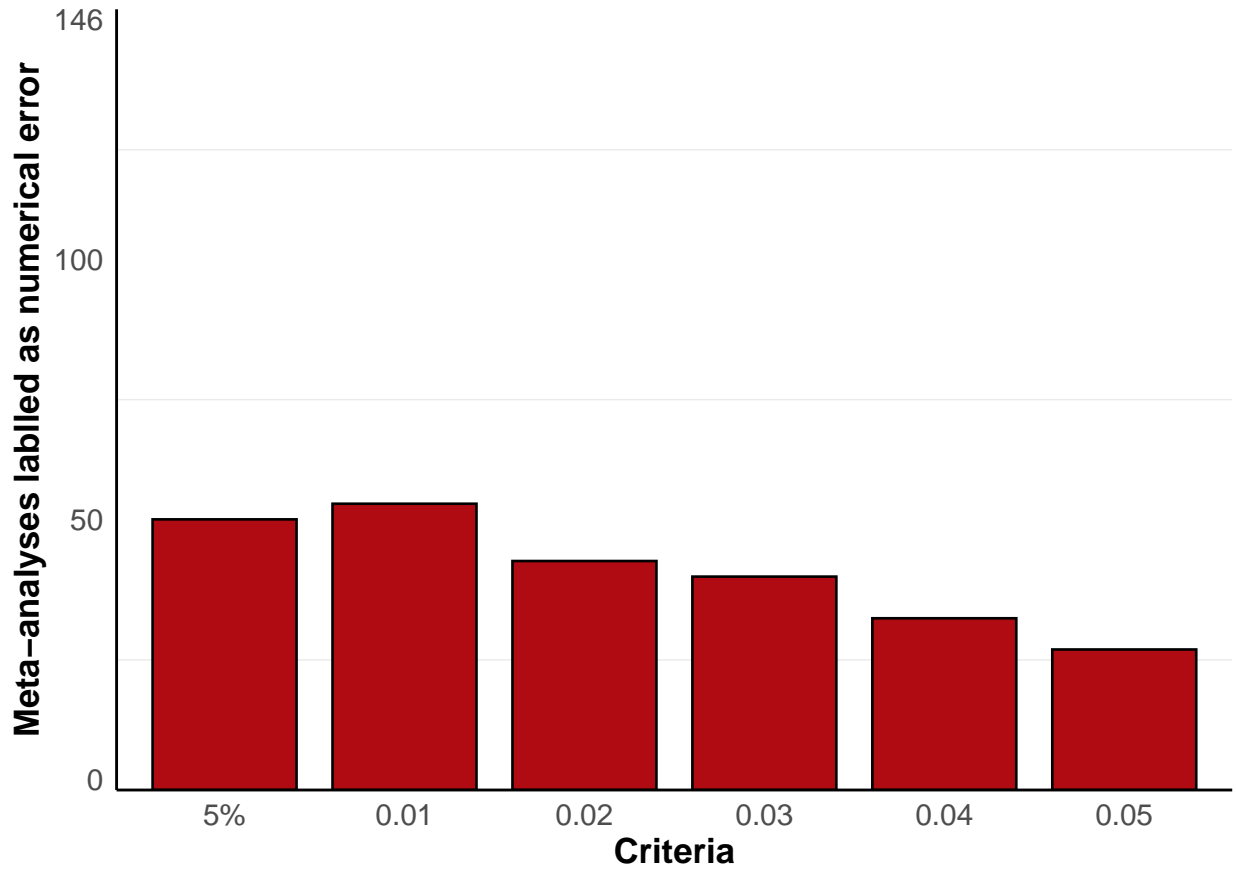


Figure 1: Barplot displaying the number of meta-analyses labelled as numerical error under different criteria. The value used in the original design is 5%, the rest are different values of difference in absolute value between reported and reproduced result.