

Supplementary Material

Inter-coder agreement

Out of the 25 papers selected to carry out independent double-coding, 21 had primary data available for one or more meta-analyses selected under the criteria explained in the main manuscript. Some disagreements or coding errors in the primary data were found in 8 (38%) cases. The intraclass correlation coefficient between the values coded by each coder was computed for each of those 21 datasets. The ICC values varied between 0.988 and 1 (mean = 0.999). Full results of the inter-coder agreement are presented in Table S1.

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

Table 1
Full results of the inter-coder agreement

Paper	ICC	Full agreement
1	0.988	FALSE
2	1.000	TRUE
3	1.000	FALSE
4	1.000	FALSE
5	1.000	TRUE
6	1.000	TRUE
7	0.997	FALSE
8	1.000	TRUE
9	1.000	FALSE
10	1.000	TRUE
11	1.000	TRUE
12	0.999	FALSE
13	1.000	TRUE
14	1.000	TRUE
15	1.000	TRUE
16	1.000	TRUE
17	1.000	TRUE
18	1.000	FALSE
19	0.997	FALSE
20	1.000	TRUE
21	1.000	TRUE

due to the mix of metrics included, it is also considered more appropriate.

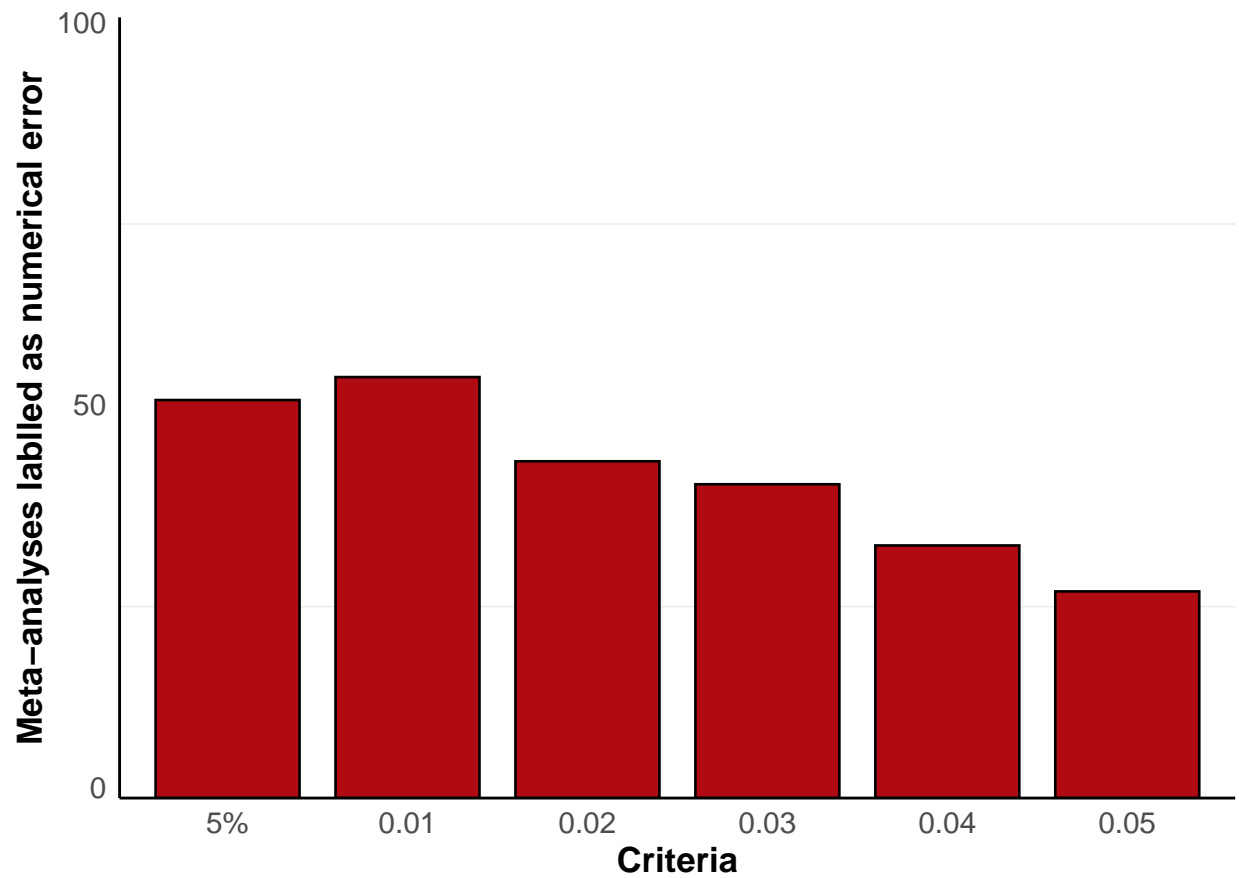


Figure 1

Barplot displaying the number of meta-analyses labelled as numerical error under different criteria. The value used in the original design is 5%, the other bars correspond to alternative criteria of the difference in absolute value between reported and reproduced result.

Data availability over years

The publication year range of the included meta-analyses in our study is quite extensive. The initial pool, from which we randomly selected a sample, comprised meta-analytic reports published between 2000 and 2020, encompassing a span of two decades. For our analysis, we focused on a random sample of 100 of these meta-analytic reports. From the 100 included reports, 217 independent meta-analyses were selected following the criteria explained in the ‘Identification and selection of articles and meta-analyses’ section of the main paper. These meta-analyses were published between 2001 and 2020 (mean = 2015.04; sd = 4.05; median = 2016; interquartile range = 2012-2016). As an unrestricted random sample, the publication year distribution is clearly left-skewed, with 92.2% of the included meta-analyses published between 2010 and 2020 (see Figure 2 of the main paper for the full distribution).

As shown in the main paper (see Figure 3a), there appears to be an improvement in the rate of data availability over the years. The overall data availability rate for the full sample was found to be 67%. However, when examining meta-analyses published within specific time periods, the rates varied. For meta-analyses published between 2000 and 2010, the data availability rate was 41%. For those published between 2011 and 2015, the rate increased to 59%. Notably, meta-analyses published between 2016 and 2020 exhibited the highest data availability rate of 80%. These findings suggest a positive trend of improved data availability in more recent years.

The association between data availability and publication year was explored by fitting binary logistic regression models with publication year as predictor and process-reproducibility as dependent variable. We quantified the strength of the association by calculating odds

ratios and 95% confidence intervals based on the profile likelihood. The analyses of the main manuscripts were mostly carried out at meta-analysis level. However, our dataset has a nested structure with meta-analyses nested within papers, which could compromise the assumption of independence of the regression model. Hence, we fitted the regression model at both meta-analysis and paper levels. Since we selected more than one meta-analysis from some of the papers, in 7 cases primary data only could be retrieved for certain of the meta-analyses in that paper, but not for all the selected meta-analyses in that paper. To avoid misclassifications, these cases were excluded from the paper-level analyses. Furthermore, as we work on an unrestricted random sample, the publication year distribution of the sample is clearly left-skewed. Therefore, the information provided by the data at the bottom range of the predictor is limited. For this reason, the analyses were complemented by fitting logistic regression models on a subset of the data excluding the papers published before 2010. In summary, we fitted four binary logistic regression model at both meta-analysis and paper levels and using the full dataset or a subset of meta-analyses published after 2010. Despite multiple contrasts performed, we did not introduce any corrections for multiple comparisons due to the exploratory nature of our analyses. Table S2 summarises the results of the models.

Table 2

Binary logistic regression models results

Level	Before 10s exclusion	Slope	OR	OR LL	OR UL	p	Pertecenage change
Meta- analysis	No	0.155	1.168	1.086	1.262	0.000	16.769
Paper	No	0.112	1.119	0.994	1.260	0.059	11.852
Meta- analysis	Yes	0.296	1.345	1.212	1.503	0.000	34.464
Paper	Yes	0.250	1.284	1.026	1.633	0.032	28.351

Based on the results of the four models, there seems to be an association between the publication year and the possibility of retrieving primary data from a meta-analysis. We found that all 4 ORs computed were > 1 , indicating a higher probability of retrieving data the more recent the publication year of the meta-analysis. Specifically, the odds increased from 11.85% to 34.46% per year. Additionally, in only one of the cases (paper-level model without excluding cases published before 2010) did the 95%CI of the OR include a value < 1 .

Qualitative check results

Table 3

Reasons found during the qualitative assessment

Case	Reason
1	No clear reason was found. Labelled as numerical error due to discrepancy in upper confidence limit. This difference (0.029) was considered not relevant .
2	Inverted signs of results. Can be explained by authors choosing to report absolute values in the main text.
3	Inverted signs of results. Can be explained by authors choosing to report absolute values in the main text.
4	No clear reason was found. Labelled as numerical error due to discrepancy in upper confidence limit. This difference (0.017) was considered not relevant.
5	Reproduced values rounded to two decimal places match with the original results (reported rounded to two decimal places).
6	Original results extracted from a subgroup analysis. Results match using a pooled between-studies variance estimation instead of a separate one.
7	No clear reason was found. Labelled as numerical error due to discrepancy in lower confidence limit. This difference (0.023) was considered not relevant.
8	No clear reason was found. Labelled as numerical error due to discrepancy in upper confidence limit. This difference (0.016) was considered not relevant.
9	Reproduced values rounded to one decimal place match with the original results (reported rounded to one decimal place).
10	Inverted signs of results. Can be explained by authors choosing to report absolute values in the main text.
11	Inverted signs of results. Can be explained by authors choosing to report absolute values in the main text.
12	No clear reason was found. Labelled as numerical error due to discrepancy in lower confidence limit. This difference (0.026) was considered not relevant.
13	No clear reason was found. Labelled as numerical error due to discrepancy in upper confidence limit. This difference (0.026) was considered not relevant.
14	Reproduced values rounded to two decimal places match with the original results (reported rounded to two decimal places).
15	The results match if the sign of the upper confidence interval is reversed. It was considered a minor reporting error.