Revisiting the Fail-Safe Number in Meta-analysis: Insights from a Simulation Study

Vera Afreixo[1(✉)], Vanusa Rocha[1], Filipa Rocha[1], and Miguel Felgueiras[1,2]

[1] Center for Research and Development in Mathematics and Applications (CIDMA), Department of Mathematics, University of Aveiro, 3810-193 Aveiro, Portugal
{vera,vanusa,filipasarreira}@ua.pt, mfelg@ipleiria.pt

[2] ESTG, Polytechnic Institute of Leiria and CEAUL – Centro de Estatística e Aplicações, Universidade de Lisboa, Lisbon, Portugal

Abstract. Publication bias remains a major challenge in meta-analysis, posing a significant threat to the validity of synthesized evidence. The fail-safe number (FSN) is commonly employed to assess the robustness of meta-analytic findings by estimating the number of unpublished studies with null or opposing results needed to nullify an observed effect, that is, to shift a statistically significant result to a nonsignificant one. However, the performance capacity of different FSN estimators may vary across meta-analytic conditions, potentially raising concerns about their reliability and interpretability.

This study presents a simulation-based evaluation of the most commonly used FSN estimators: the Rosenthal FSN and the Rosenberg FSN and several modifications. We examine their performance availability by analyzing their behavior under controlled scenarios with null, weak, and moderate true effects within selective pressure. As a case study, we focus on the standardized mean difference using Cohen's d as the effect size measure. Additionally, we explore the usability of these methods, discussing their strengths and limitations in meta-analytic applications. The validity and applicability of commonly used rules of thumb for interpreting FSN values are also assessed.

Our findings provide a clearer understanding of when and how these FSN estimators yield meaningful insights, offering guidance about their appropriate use in evaluating the robustness of meta-analyses.

Keywords: Fail-safe number · Meta-analysis · Simulation · Performance · Standardized Mean Difference · Cohen's d

1 Introduction

A systematic literature review identifies, selects, and critically appraises research to answer a formulated question [1]. An important step in a systematic review is to carefully analyze whether it is suitable to combine the numerical findings of the studies [2]. Meta-analysis is a study design used to systematically evaluate previous research studies to

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
O. Gervasi et al. (Eds.): ICCSA 2025 Workshops, LNCS 15887, pp. 365–376, 2026.
https://doi.org/10.1007/978-3-031-97589-9_25

derive conclusions [3]. By doing a meta-analysis, a single overall estimate statistic can be obtained to determine whether an effect exists or to determine whether the effect is positive or negative [2, 3].

Although systematic reviews and meta-analyses are invaluable for synthesizing available evidence, they are susceptible to multiple forms of bias [4].

Publication bias arises when the possibility of a study being published is not independent of its results, which prevents the production of valid evidence. One type of this bias is selective reporting, defined as reporting results from only a selection of studies. It refers to the tendency only to publish studies with positive results or statistically significant findings [5], suggesting that studies with a p-value > 0.05 (the more commonly used significance level) may not get published. It can be caused by factors such as rejection by editors or reviewers, lack of interest in revising the study, competing interests, or simply a lack of motivation to write up the study [6].

When publication bias is present, the estimated average effect sizes obtained in meta-analysis tend to be inflated to an unknown and potentially significant extent. Additionally, this bias can skew the evidence favoring the dominant theory [7].

Several techniques can be used for evaluating the possibility of publication bias and its impact. To conduct these types of analyses, it is recommended to have a reasonable number of studies (minimum of 10), at least one study with a significant effect and relevant heterogeneity should be avoided [8].

The fail-safe number (FSN) is a procedure that allows the impact of potential publication bias to be partially discussed by answering the question, "How many studies would it take to 'nullify' the combined effect". It only applies to meta-analyses that lead to combined effects with a p-value < 0.05 or confidence intervals that do not contain the value that translates the null effect size. This procedure can be used as a sensitivity analysis of meta-analytical results to include studies with effects that somehow compensate for potential publication bias [9].

There are different ways of calculating FSN. In practice, authors propose calculating the number of additional studies with a nonsignificant effect (null/negative effects) that modifies the meta-analytical response. This procedure is limited as it does not allow the meta-analytical effect to be adjusted for potential publication bias or to measure this effect [9].

1.1 Rosenthal's FSN

Rosenthal's method discusses if the meta-analytical effect is robustly significant, assuming that a positive and significant meta-analytical overall effect is estimated [10].

Rosenthal's FSN considers the null hypothesis.

$H_0 : \theta_1 = \theta_2 = \cdots = \theta_k = 0,$

and the test statistic,

$$Z_i = \frac{T_i}{\sqrt{v_i}} \sim N(0,1) \Rightarrow \frac{\sum Z_i}{\sqrt{k}} \sim N(0,1), \tag{1}$$

with T_i the effect size obtained from study i ($i = 1, \ldots, k$ and k the number of studies in the meta-analysis) and $\sqrt{v_i}$ the corresponding standard error.

Assuming that there are N_α additional studies with null effect that were not considered, then the previous test statistic has the following formulation,

$$\frac{\sum Z_i}{\sqrt{k+N_\alpha}} \sim N(0,1), \tag{2}$$

with $i = 1, \ldots, k$.To find the minimum number of null effect studies to change the significant effect into a non-significant, solve the following equation in order to N_α

$$\frac{\sum Z_i}{\sqrt{k+N_\alpha}} = z_{1-\alpha} \tag{3}$$

with $z_{1-\alpha}$ the quantile of order $1-\alpha$ of the standard normal distribution. The solution of the previous equation is Rosenthal's FSN (**Rt**),

$$N_\alpha = \left[\frac{\sum Z_i}{z_\alpha}\right]^2 - k \tag{4}$$

Rosenthal proposed a cut-off value (one rule of thumb), $5k + 10$, to determine the robustness of a meta-analysis result. A smaller N_α may indicate that the effect being analyzed might not be robustly significant. A larger N_α suggests the effect is significant.

1.2 Rosenthal's FSN Based on Counting

Rosenthal also proposed an alternative procedure, described as "probably less powerful," which estimates the FSN by considering the number of studies with statistically significant results (n_s) and those with non-significant results ($n_{ns} = k - n_s$) [10]. This approach assumes that, under the normal assumption and the null effect hypothesis, the proportion of significant results vs the non-significant results ratio is known. Figure 1 presents a graphical illustration of both proportions, given the alpha level (α) and considering the effect size distribution. If the number of true non significative studies is $N_\alpha + n_{ns}$, the number of observed non significative studies is ns and the number of significative studies is s the following equality is obtained,

$$\frac{N_\alpha + n_{ns}}{n_s} = \frac{1-\alpha}{\alpha} \tag{5}$$

Then, Rosenthal's FSN based on counting is then estimated using the formula (**Rtc**):

$$N_\alpha = n_s\frac{1-\alpha}{\alpha} - n_{ns} = \frac{n_s}{\alpha} - \mathrm{k} \tag{6}$$

1.3 Rosenthal's FSN Modification

Rosenthal's proposal reinforces that the missing studies observe null effect values, which may be an inappropriate assumption. With the same type of approach, Iyengar and

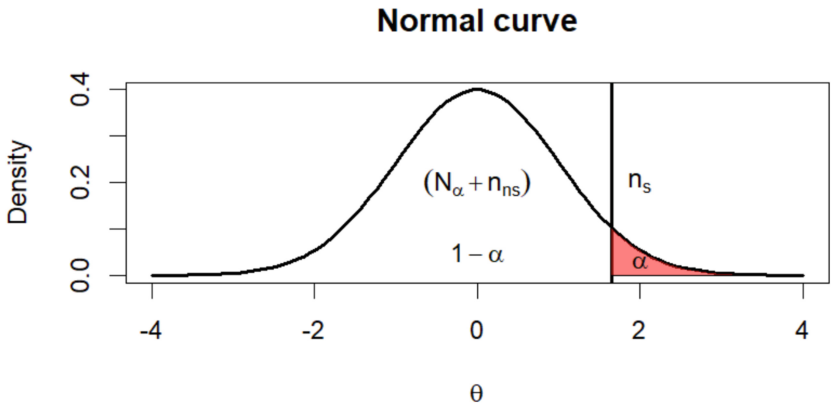

Fig. 1. Graphical illustration of the proportion of significant vs non-significant results, under the null effect hypothesis, for α significance level.

Greenhouse proposed an alternative FSN (N_α), here denominated by Rosenthal's FSN modified (**Rtm**)

$$\frac{\sum Z_i + N_\alpha M_\alpha}{\sqrt{k + N_\alpha}} = z_{1-\alpha} \tag{7}$$

where the statistics from additional studies are not considered null but the mean value of the standard normal distribution truncated at $z_{1-\alpha}$ [11]. Recall that the density function of the standard normal distribution truncated at $z_{1-\alpha}$ is given by

$$f(x) = \begin{cases} \frac{\phi(x)}{\mathbf{\Phi}(z_{1-\alpha})}, x \leq z_{1-\alpha} \\ 0, x > z_{1-\alpha} \end{cases} \tag{8}$$

with ϕ as the standard normal density function and $\mathbf{\Phi}$ as the standard normal distribution function, and the corresponding mean value is given by

$$M_\alpha = E[X] = \int_{-\infty}^{+\infty} x f(x) = -\frac{\phi(z_{1-\alpha})}{\mathbf{\Phi}(z_{1-\alpha})} = -\frac{\phi(z_{1-\alpha})}{1-\alpha}. \tag{9}$$

In practice, N_α is computed as the smallest non-negative integer that satisfies the condition $\frac{\sum Z_i + N_\alpha M_\alpha}{\sqrt{k+N_\alpha}} \leq z_{1-\alpha}$,

$$N_\alpha = argmin\{N \in \mathbb{N}_0 \mid \frac{\sum Z_i + N_\alpha M_\alpha}{\sqrt{k + N_\alpha}} \leq z_{1-\alpha}\} \tag{10}$$

1.4 Rosenberg's FSN

Rosenberg's FSN is similar to Rosenthal's FSN with differences in the test statistic [12]. In Rosenberg's FSN approach, test statistic's uses the overall meta-analytic estimate,

$\hat{\theta} = \frac{\sum_{i=1}^{k}\left(w_i \times \hat{\theta}_i\right)}{\sum_{i=1}^{k} w_i}$, considering the individual studies weights (w_i), and considers the global null hypothesis $H_0 : \theta = 0$. The test statistic for Rosenberg's FSN is

$$\hat{\theta}/SE \sim N(0,1), \tag{11}$$

with $SE^2 = \frac{1}{\sum_{i=1}^{k} w_i}$.

Let us consider that additional null-effect studies are missing. These studies are capable of reducing the value of the test statistic to the critical value and their weight assumes the average value of the studies published. Therefore, the FSN can be obtained from:

$$z_{1-\alpha}^2 = \frac{\left(\sum_{i=1}^{k}\left(w_i \times \hat{\theta}_i\right)\right)^2}{N_\alpha \frac{\sum_{i=1}^{k} w_i}{k} + \sum_{i=1}^{k} w_i} \tag{12}$$

Explicitly, N_α is given by the following equation:

$$\mathrm{N}_\alpha = \frac{\left(\sum_{\mathrm{i}=1}^{\mathrm{k}}\left(\mathrm{w_i} \times \hat{\theta}_\mathrm{i}\right)\right)^2}{\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{w_i}} - \mathrm{k}. \tag{13}$$

2 Methods

In this section we present a simulation procedure where the new functions were implemented in RStudio (2022.07.1 + 554) using R (version 4.2.2). Several meta-analytic scenarios were simulated and the different FSN were computed. Rosenthal's FSN and Rosenberg's FSN were implemented in *metafor* package through the fsn function. The functions to calculate the modified Rosenthal's FSN and Rosenthal's FSN based on counting were programed. Association between several features of meta-analytic studies (heterogeneity level, I^2, number of studies, effect size strangeness) and the FSN levels were explored graphically.

2.1 Simulation Procedure

Different conditions were implemented for each meta-analysis study. Table 1 summarizes these conditions.

Standard Deviation Between Studies
To generate the set of effect sizes to each meta-analytic study we randomly generated n values from a normal distribution with a fixed mean from the set $\{0, 0.1, 0.2, 0.5\}$ and one fixed standard deviation from the set $\{0.05, 0.1, 0.2, 0.5\}$. This simulation allows for the evaluation of scenarios with null, weak, and moderate effect sizes.

Standard Error of Each Individual Effect Size (Sample Unity of Each Meta-analysis)

For each individual meta-analytic study, we generated for each effect size one corresponding standard error. The standard error (SE) was generated assuming the effect size under analysis is a Cohen's d (d) for the mean difference of two independent sample, which is obtained by

$$SE = \sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{d^2}{2(n_1 + n_2)}} \tag{14}$$

With n_1 and n_2 the samples sizes of each group from one fixed study. Values n_1 and n_2 are randomly generated from 20 to 200.

Other Simulation Parameters

Each different scenario was simulated 100 times, with the number of studies per meta-analysis ranging from 15 to 60. When the number of studies with null or weak effect sizes was lower than 15, the meta-analytic overall effect in high number of scenarios was non-significant; therefore, these results are not presented.

After the data generation procedure, three scenarios of selective pressure (publication bias) were considered:

- without selective pressure - a control scenario;
- total pressure through the non-significative studies; in this case some meta-analytic scenarios are null meta-analytic scenarios (scenarios without studies);
- partial pressure through the non-significative studies; 1/3 of studies without positive results and p-values > 0.05 were deleted (1/3 of studies with highest p-values).

Table 1. Different conditions for each simulation/meta-analysis unit.

Simulation runs	100
Number of studies by each meta-analysis study (n)	15, 30, and 60
Effect size (ES)	0, 0.1, 0.2, and 0.5
Standard deviation between studies	0.05, 0.1, 0.2, and 0.5
Proportion of deleted studies whose p-values > 0.05	0, until 1/3, and until 1
Standard error of each study - Eq. (11), n_1 and n_2	Integer between 20 and 200

Table 2 synthesizes the values that were extracted from the simulation that was carried out. For each simulation we perform a meta-analysis, and the overall effect was estimated (effect size and corresponding confidence interval) with heterogeneity estimated by I^2 index. When the overall effect was significant the fail-safe numbers under analysis were extracted. The number of studies eliminated by pressure selection through non-significative studies was also registered (number of deleted studies).

We also count in each simulation scenario the number of meta-analytic studies with non-significant overall effect size or null meta-analytic scenarios. Since FSNs can only be

Table 2. Extracted results.

Classic methods	Rosenthal's FSN (**Rt**) Rosenberg's FSN (**Rb**)
Method based on counting	Rosenthal's FSN counting (**Rtc**)
Methods with modification from Iyengar & Greenhouse 1988	Rosenthal's FSN modified (**Rtm**)
Other results	Number of deleted studies Heterogeneity I-squared (I^2) Overall ES and CI

calculated in meta-analyses with a significant overall effect, it is important to determine in how many simulation scenarios the FSNs calculation was applied.

2.2 Performance Evaluation

To evaluate the performance of each FSN approach, the Root Mean Square Error (RMSE) was used to measure the difference between the number of deleted studies and the estimated value by the FSNs. So, the bias of FSN estimators was quantified using the square root of the mean squared error.

The cut-off criteria, proposed by Rosenthal, for determining the robustness of meta-analytic findings were examined under described scenarios.

The distribution of FSN values was analyzed, considering variations in the number of studies and different levels of selective pressure. For each selective pressure condition, key metrics were reported, including the number of true eliminated studies, the median FSN values, the proportion of meta-analyses yielding a p-value below 0.05 and the number of cases where FSN met or exceeded the true number of true removed studies.

To provide a straightforward guideline for interpreting the robustness of null hypothesis rejection in meta-analysis, several cut-off rules were evaluated, offering insights into their applicability across different scenarios.

3 Results

Figure 2 shows that the different FSNs present similar patterns even with different levels of selective pressure and different numbers of studies involved in each meta-analysis/simulation unit. We can observe greater dispersion and higher median values for the FSNs without modification (classical methods, **Rt** and **Rb**). Even though there isn't much difference in the patterns of the different FSNs, there are differences in the scales depending on the number of studies per meta-analysis ($n = 15$, $n = 30$, and $n = 60$). As expected, more studies under meta-analysis imply a higher FSN.

The following results set the scenario in which all unities of the simulation have 60 initial studies, where it is expected more stability of the meta-analytic results.

According to the expected, Table 3 reveals that with larger effect sizes, there is an increase in the number of studies that are needed to 'nullify' it. For the true null effect

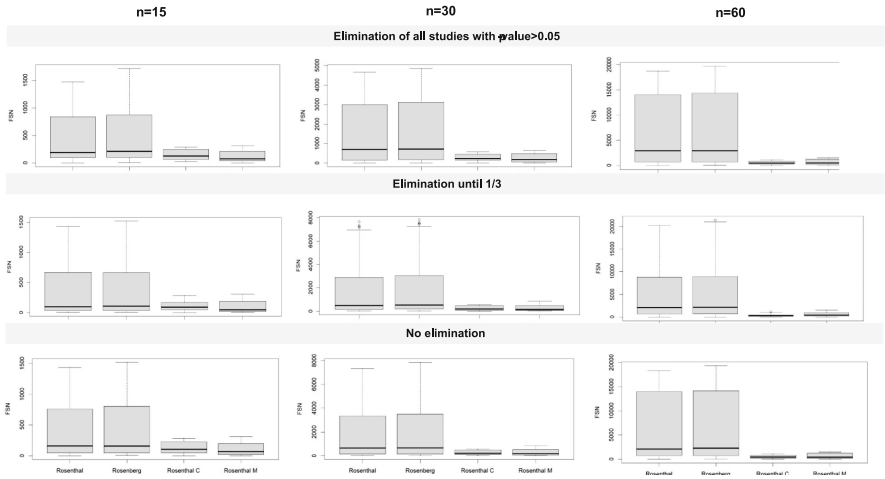

Fig. 2. Distribution of the four different FSNs considering the number of included studies in each meta-analysis ($n = 15$, $n = 30$, and $n = 60$) and different selective pressures (deleting all studies with p-values >0.05, deleting up to 1/3 of studies with p-values >0.05 and no deletion of studies).

sizes (ES $= 0$), the **Rt**, **Rtc** and **Rb** generate values that are higher than the modified FSNs under analysis (**Rtm**), which is one limitation of classic procedures.

Regarding the scenario where all studies with p-values >0.05 are deleted, only nine meta-analyses presented p-values <0.05 and included an average of nine studies per meta-analysis. The scenario without the elimination of studies represents a control scenario. Therefore, it is natural that when the effect is null, the combined effect of the meta-analyses tends to be null, and when the overall effect is not significant the calculation of FSNs is not suitable to be applied. However, for $n = 15$ (low true meta-analysis samples sizes) the control scenario reveals for ES $= 0$, 37 meta-analyses with positive results and p-values <0.05 with FSNs given by **Rt** $= 4$, **Rb** $= 12$, **Rtc** $= 25$ and **Rtm** $= 3$, showing better results the classic and the modified Rosenthal FSNs. A low FSN suggests that minimal effort is required to nullify the effect, implying that the overall significance is weak or not robust.

Because of the elimination of huge simulation unities in these two previous scenarios, we suggest exploring the results of simulation studies with a selective pressure through studies with non-positive effect size and p-values >0.05 up to 33.3%.

To discuss the bias of the FSNs under analysis, meta-analyses with 60 studies and selective pressure up to 33.3% were considered. Table 4 describes the square root of the mean squared error for each FSN and the percentage of times each FSN was equal or higher than the number of deleted studies, regarding the different effect sizes. The classic FSNs (**Rt** and **Rb**) present higher error values than the other FSNs under analysis (**Rtc** and **Rtm**). As expected, the larger the effect sizes, the greater the error.

For ES $= 0$, all the FSNs under analysis reveal a high percentage of results (more than 80%) where the FSN is higher than the number of studies that were eliminated (maximum selective pressure of 33.3% on studies with p-values < 0.05).

Table 3. Description of the median number of studies eliminated, the median of the FSNs, and the number of meta-analyses with positive results and p-value <0.05 with different selective pressures for simulation scenarios with 60 initial studies.

n = 60						
Elimination of **all** studies without positive results and without p-values < 0.05	Medians					
	No. of deleted studies	**Rt**	**Rb**	**Rtc**	**Rtm**	No. of meta-analyses with positive results and p-values < 0.05
ES = 0	51	8	45	120	4	9
ES = 0.1	47	581	575	200	160	301
ES = 0.2	35	2915	2974	440	476	400
ES = 0.5	4	14930	15447	1060	1297	400
Elimination of studies without positive results and without p-values < 0.05 until 1/3	Medians					
	No. of deleted studies	**Rt**	**Rb**	**Rtc**	**Rtm**	No. of meta-analyses with positive results and p-values < 0.05
ES = 0	18	268	305	98	92	244
ES = 0.1	17	710	733	117	187	400
ES = 0.2	13	2126	2188	373	390	400
ES = 0.5	1,5	15726	16230	1062	1338	400
No deletion of studies	Medians					
	No. of deleted studies	**Rt**	**Rb**	**Rtc**	**Rtm**	No. of meta-analyses with positive results and p-values < 0.05
ES = 0	-	**-**	**-**	**-**	-	0
ES = 0.1	0	599	641	100	163	400
ES = 0.2	0	2054	2214	360	380	400
ES = 0.5	0	15144	15682	1060	1308	400

Table 5 describes the obtained results regarding the possible establishment of cut-off rules (rule of thumb) in FSN to evaluate the robustness of the significance of the overall meta-analytic effect detected. In these results we kept the scenario with 60 studies and selective pressure up to 33.3%. Firstly, we present the results for the rule proposed by Rosenthal [10]. When ES = 0, we observe that the classical FSNs (Rt and Rb) point to over 50% of meta-analyses with robustness in rejecting the null effect, whereas ideally, all studies with ES = 0 should have FSN values of 5k + 10 or lower. The 5k + 10 cutoff rule was expected to work well for the Rosenthal´s FSN (Rt) because it was created for this procedure. We observe that it is unsuitable in the null effect scenario, but it seems

Table 4. The square root of the mean squared error for each FSN regarding the different effect sizes and the percentage of times each FSN was equal or superior to the number of deleted studies (n = 60).

Square root of the mean squared error				
	Rt	Rb	Rtc	Rtm
ES = 0	462	556	165	116
ES = 0.1	2551	2627	255	356
ES = 0.2	2658	2771	363	425
ES = 0.5	15600	16098	1009	1325
FSN > = No. of deleted studies				
ES = 0	86%	96%	82%	82%
ES = 0.1	100%	100%	77%	100%
ES = 0.2	100%	100%	100%	100%
ES = 0.5	100%	100%	100%	100%

adequate for the modified positive effect sizes. Additionally, for the modified Rosenthal's FSN, in a global way, the 5k + 10 cutoff works well.

Table 5. Evaluation of two decision rules for the FSNs (n = 60 and selective pressure up to 33.3%).

Rosenthal's FSN evaluation/decision rule	FSN > 5k + 10			
	Rt	Rb	Rtc	Rtm
ES = 0	66%	70%	34%	0%
ES = 0.1	100%	100%	26%	25%
ES = 0.2	100%	100%	91%	100%
ES = 0.5	100%	100%	100%	100%

It is observed that **Rt**, **Rb** and **Rtm** procedures exhibit an increasing trend as the standard deviation between studies (and heterogeneity) increases (see for example Fig. 3 the results for **Rtm**).

4 Discussion and Conclusions

This study conducted a detailed simulation-based comparison of four FSN estimators—Rosenthal's original estimator (**Rt**), Rosenberg's estimator (**Rb**), a counting-based variant (**Rtc**), and a modified version of Rosenthal's estimator incorporating the Iyengar & Greenhouse correction (**Rtm**). By systematically removing studies with non-significant

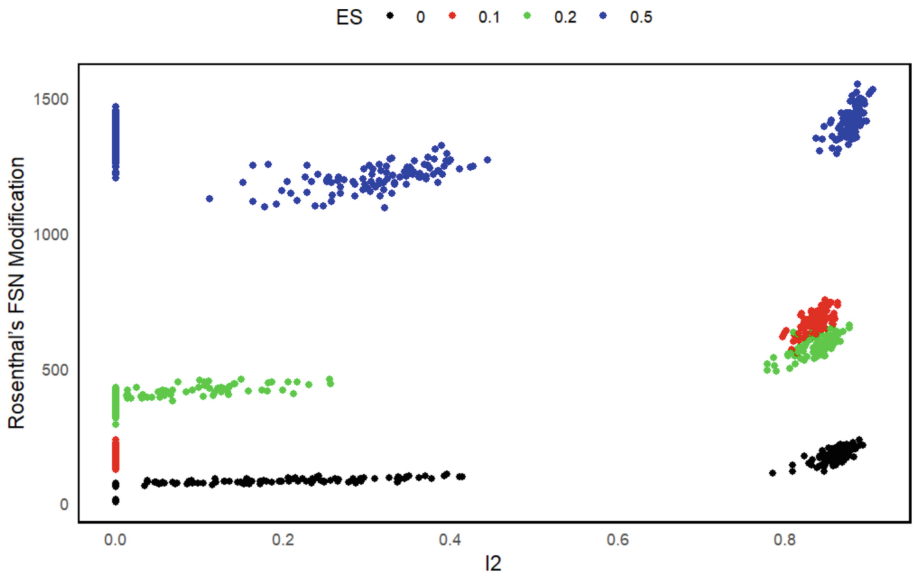

Fig. 3. Scatter plot Rosenthal's FSN Modification (**Rtm**) vs I^2 heterogeneity measure, with different sizes represented by different colors (n = 60 and selective pressure up to 33.3%).

results ($p > 0.05$), we explored how selective reporting and effect size strength influence FSN estimates.

Our findings suggest that **Rtm** consistently offers more reliable and conservative estimates under varying conditions. Specifically, it produces lower FSNs in the presence of true null effects, reflecting higher sensitivity to the absence of a real effect and reduced risk of false robustness. When compared with classical estimators, **Rtm** showed smaller errors across different effect size magnitudes and levels of selective pressure.

In conclusion, these results highlight that **Rtm** should be preferred when there is concern about publication bias or small effect sizes. In particular, researchers should interpret low FSNs (especially under **Rtm**) as indicators of fragility in the evidence and consider them alongside other bias-detection tools. The overestimation observed with **Rt** in null scenarios underscores the risk of relying solely on classical thresholds such as Rosenthal's "5k + 10" rule, which may give a false sense of confidence.

This work provides a practical guide for meta-analysts choosing between FSN estimators and evaluating the robustness of their results. Future research could extend these findings by integrating FSN estimation into more complex bias-adjustment frameworks or by validating them on empirical datasets with known levels of publication bias. Another promising direction is the development of FSN approaches that adapt to heterogeneity or use Bayesian priors to better model uncertainty in missing studies.

To sum up, while all FSNs offer insight into meta-analytic robustness, the **Rtm** estimator appears better suited to detecting fragility and mitigating bias in realistic scenarios.

Acknowledgments. This work is partially supported by Portuguese funds through CIDMA, The Center for Research and Development in Mathematics and Applications of University of Aveiro, and the Portuguese Foundation for Science and Technology (FCT – Fundação para a Ciência e a Tecnologia), within project UIDB/04106/2020 (https://doi.org/10.54499/UIDB/04106/2020), project UIDP/04106/2020 (Thematic Line BIOMATH) (https://doi.org/10.54499/UIDP/04106/2020), project PALOP/BD/155077/2024 and under the projects UID/00006/2025 and UIDB/00006/2020. https://doi.org/10.54499/UIDB/00006/2020.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Literature Review: Systematic literature reviews. https://libguides.csu.edu.au/review/Systematic. Accessed 28 April 2025
2. Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). https://training.cochrane.org/handbook. Accessed 28 Apr 2025
3. Haidich, A.: Meta-analysis in medical research. Hippocratic **14**(Suppl 1), 29–37 (2010)
4. Drucker, A.: Research techniques made simple: assessing risk of bias in systematic reviews. J. Investig. Dermatol. **136**(11), e109–e114 (2016)
5. DeVito, N.: Catalogue of bias: publication bias. BMJ Evidence-Based Med. **24**(2), 53–54 (2019)
6. Mlinarić, A.: Dealing with the positive publication bias: why you should really publish your negative results. Biochem. Med. (Zagreb) **27**(3), 030201 (2017)
7. Ropovik, I.: Neglect of publication bias compromises meta-analyses of educational research. PLoS ONE **16**(6), e0252415 (2021)
8. Dalton, J.: Publication bias: the elephant in the review. Anesth. Analg. **123**(4), 812–813 (2016)
9. A brief history of the Fail-Safe Number in Applied Research. https://arxiv.org/abs/1010.2326. Accessed 28 Apr 2025
10. Rosenthal, R.: The "file drawer problem" and tolerance for null results. Psychol. Bull. **86**(3), 638–641 (1979)
11. Iyengar, S.: Selection models and the file drawer problem. Stat. Sci. 109–117 (1988)
12. Rosenberg, M.: The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. Evolution **59**(2), 464–468 (2005)