

# The Association Between The Impact Factor of A Journal And The Trueness of Findings Published Therein: A Meta-Epidemiological Study

Andreas Heidenreich (✉ [andreas.heidenreich@uksh.de](mailto:andreas.heidenreich@uksh.de))

University of Luebeck <https://orcid.org/0000-0003-3279-416X>

**Nora Eisemann**

University of Luebeck

**Alexander Katalinic**

University of Luebeck

**Joachim Huebner**

University of Luebeck

---

## Research

**Keywords:** meta-epidemiology, Cochrane, Journal Impact Factor, bibliometrics

**Posted Date:** September 15th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-861273/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Systematic Reviews on January 18th, 2023. See the published version at <https://doi.org/10.1186/s13643-023-02167-8>.

# Abstract

**Background:** Scientists, physicians, and the general public legitimately expect scholarly publications to give true answers to study questions raised. We investigated whether findings from studies published in journals with higher Journal Impact Factors (JIFs) are more accurate than findings from studies in less-cited journals via a meta-epidemiological approach.

**Methods:** We screened intervention reviews from the Cochrane Database of Systematic Reviews (CDSR) and sought well-appraised meta-analyses. We defined the trueness of RCT study results as their point estimates' relative deviation from the pooled effect estimate. In addition to the JIF we considered the open-access policy of journals, the relative size of studies, and the studies' methodical quality.

**Results:** In 2,459 results from 446 meta-analyses, there was a positive association between a journal's JIF and the trueness of the study results published therein (Pearson's  $r$  for the deviation =  $-0.21$ , 95% CI  $-0.24$  to  $-0.17$ ,  $P < 0.001$ ). The mean relative deviation decreased from 1.11 (SD = 0.94) in the lowest JIF quartile to 0.80 (SD = 0.75) in the top quartile (95% CI: 0.21 to 0.40,  $P < 0.001$ ). Findings from larger and methodically sound studies were also more likely to be accurate. In a multiple regression analysis, the JIF proved to be the strongest determinant of trueness, independent of the covariates.

**Conclusions:** Our results indicate that higher-impact journals act as gatekeepers in the process of science more effectively than less-cited journals. However, the fact alone that a study result is reported in a journal with a high impact factor is only a weak and impractical indicator for its accuracy.

## Key Messages

- An approach to directly assess the trueness of medical study results using good-quality meta-analyses was applied
- RCT study results published in higher-impact journals lie closer to true values than results published in lower-impact journals
- As the association found is weak, the Journal Impact Factor is of limited practical value to predict the trueness of a study result

*"In 1955, it did not occur to me that "impact" would one day become so controversial. Like nuclear energy, the impact factor is a mixed blessing. I expected it to be used constructively while recognising that in the wrong hands it might be abused."* – Eugene Garfield, 2005

## Introduction

Academic publications are an important and much-noticed output of scientific work. The general public, policymakers, the scientific community, and physicians legitimately expect publications to give true answers to their study questions. However, there is increasing concern that the majority of published research findings are false.<sup>1</sup> In a constantly growing flood of publications, preferential trust is put in

journals with a high Journal Impact Factor (JIF).<sup>2</sup> Despite extensive criticism relating to its manipulability and widespread misuse, e.g. as a measure of the scientific performance of researchers and institutions,<sup>3-5</sup> existing research supports the notion of a relation between the impact factor of a journal and the quality of research published therein. Concerning clinical research, it has been shown that randomised controlled trials (RCTs) published in higher-impact journals are less prone to risk of bias compared with those published in lower-impact journals in their reported design, conduct, and analysis.<sup>6-8</sup> Furthermore, lower methodological quality is associated with larger, commonly beneficial effects.<sup>9, 10</sup> Taking these findings together, one could come to the conclusion that studies from lower impact journals tend to overestimate the true effect of an intervention. Contradicting this assumption, it has been found that trials published in the highest-impact journals show more favourable effects for experimental interventions than trials from other journals. As this difference has not been found for large trials, the exceeding effect size of small trials in high-impact journals provoked caveats regarding their reliability.<sup>11</sup>

To our knowledge, the direct association between the impact factor of a journal in which a study is published and the extent to which the study results tell us the truth has not been investigated to date. One possible reason for this is the obvious difficulty of determining the truth in empirical science. This is all the more the case when it is not a question of truth as a categorical statement, but of the magnitude of a true value. In medical science, this epistemological challenge can be met in a practically suitable manner. The idea is that the pooled effect estimates from distinctly reliable meta-analyses represent sufficiently true values. Cochrane reviews provide pertinent meta-analyses. Compared with other systematic reviews, Cochrane reviews use more rigorous methods,<sup>12-14</sup> laid down in a comprehensive, periodically updated handbook.<sup>15</sup> An integral part of these methods is to assess the quality of the underlying body of evidence (sometimes referred to as “certainty of evidence”) on a four-level scale. This results in a transparently structured judgment about the likelihood of trueness of the pooled effect estimate. The upper two levels (“high” and “moderate” quality) indicate that the authors are “very” or “moderately” confident that the true effect “lies close” (or “is likely to be close”) to the estimate of the effect.

Following this assessment, we postulated that effect estimates from Cochrane meta-analyses that are based on evidence of “high” or “moderate” quality are sufficient proxies of true values. We then examined whether results of intervention studies published in higher-impact journals lie closer to these proxies of truth, compared with those published in lower impact journals. To better understand the hypothesised correlation, we further investigated the influence of the methodical quality of each study, the weight given to the study within the meta-analysis, and whether the journal is open access.

## Methods

### Selection of eligible Cochrane reviews and meta-analyses

We prospectively screened all intervention reviews published in the Cochrane Database of Systematic Reviews (CDSR) from July 2018 to January 2019 for meta-analyses. We restricted the selection to meta-

analyses generating evidence of “moderate” or “high” quality. In this respect, we adopted the review author’s assessment, based on the GRADE framework. Therein, “moderate” or “high” quality of evidence represents the upper two of four levels (the others being “low” and “very low”). These judgments imply that the review authors consider the pooled effect estimates to be close to the true values with “moderate” or “high” certainty, respectively.<sup>16</sup> We decided against a limitation to meta-analysis with “high” quality evidence, since only few meta-analyses are rated as such. In a pre-study, we found that of 986 examined meta-analyses published in 59 Cochrane intervention reviews, only 26 (2.6%) were of “high” quality of evidence and the proportion of “moderate”-quality meta-analyses was also small (7.7%). Thus, we considered pooled effect estimates of both “high” and “moderate” certainty to be suitable proxies for truth.

In case a review contained at least one meta-analysis of moderate or high quality of evidence, we obtained the accompanying data-file from the Cochrane website and extracted the corresponding data via Cochrane’s *Review Manager* software, version 5.3.<sup>17</sup> We entered the extracted data along with identifying information on the meta-analysis into our study database.

## **Matching study results with publication data**

For each study result included in eligible meta-analyses, we identified the source publication. When more than one publication was referenced, we screened the articles’ abstracts and full texts to identify the correct publication reliably, as not in every case the designated primary publication was the source of all data used in meta-analyses.

## **Journal’s impact**

For each publication included in eligible meta-analyses, we obtained information on the publishing journal’s JIF of the publication year from the Journal Citations Report (Clarivate Analytics). We preferred this indicator to others (e.g. SCImago Journal Rank and Source Normalized Impact per Paper (SNIP)), as it represents the de-facto standard of citation metrics in science and is most likely known outside the scientific community. Information on the JIF was available from 1997 to 2018. We excluded results reported in older publications from the analysis.

As the average reference list in the medical sciences and other fields became lengthier over time, JIFs are subject to inflation.<sup>18, 19</sup> We did not consider this increase as a real accretion in impact on the process of science. We, therefore, adjusted the JIF for publication date by calculating the ratio of the actual JIF and the mean value of all journals’ JIFs in our study sample of a given year.

## **Operationalisation of trueness**

We defined the agreement between study results and true values as closeness between each study’s point estimate and the pooled effect estimate of the related meta-analysis. We refer to the distance between both estimates, which is due to systematic error (bias) and unsystematic error (lack of precision), as “deviation”, indicating a lack of accuracy. Aiming at a meaningful comparison between results of

different study questions, expressed in various effect measures (e.g. risk ratio and odds ratio for dichotomous outcomes, mean difference and standardised mean difference for continuous outcomes), we computed a “relative deviation” as a quotient. The nominator was calculated as the absolute difference between the study’s point estimate and the pooled effect estimate. The denominator was the mean of the analogously calculated absolute differences of all studies included in the respective meta-analysis (see appendix for details).

This approach becomes questionable when a single study dominates a meta-analysis. The pooled effect estimate is calculated as a weighted average of the intervention effects estimated in the individual studies.<sup>20</sup> The more a pooled effect estimate is influenced by a particular study, the less it may serve as an appropriate benchmark for the accuracy of this very study. This is obvious when a meta-analysis comprises only one individual study. This study cannot fail to meet the truth according to our approach. Therefore, we considered all individual studies with a weight of 50% or more as non-informative regarding our research question. We did not use these results for calculating the relative deviation and excluded them from further analysis. We also excluded study results with no point estimates, e.g. results with no events in either study arm for dichotomous outcome measures or missing standard deviations in case of continuous outcome measures.<sup>20</sup>

## Covariates

To control for confounders, we considered three variables that might also be associated with both the JIF and the accuracy of the reported results.

First, we suspected that larger studies with a correspondingly smaller variance of the effect measure are more likely to be published in higher impact journals than smaller studies with larger variance. As smaller variance translates to bigger weight in meta-analyses, the suspected association would result in the observation that studies from higher impact journals show systematically smaller relative deviations. As the risk of confounding has been mitigated, but not eliminated, by the exclusion of studies with a weight of 50% or more, we further considered the study weight as a covariate.

Second, we accounted for the methodical quality of studies by considering compliance with the methodical requirements of a journal as a potential explanatory factor. For this purpose, we used the risk of bias assessment conducted by the review authors. For every study included in Cochrane reviews, the results of a detailed and mostly standardised procedure are reported.<sup>20</sup> Usually, seven key items are assessed for each study (“random sequence generation”, “allocation concealment”, “selective reporting”, “blinding of participants and personnel”, “blinding of outcome assessment”, “incomplete outcome data”, and “other”). We transformed the qualitative verdicts (“low risk of bias,” “unclear risk of bias,” and “high risk of bias”) for each key item to numerical values (0.0, 0.5, and 1.0, respectively) and calculated the mean value. The result is an aggregated risk of bias between “0” (low risk of bias in every item) and “1” (high risk of bias in every item). As this approach truncates the qualitative dimensions of the risk of bias assessment, it is not suitable for comprehensive critical appraisals of an RCT. Here, we pursue a general statistical relationship in this study, and a loss of qualitative information appears acceptable.

Third, we considered the open-access policy of a journal. We defined a journal as an Open Access journal only if it published all content under a corresponding license at the time of publication (i.e. excluding hybrid journals). Information on the open-access policy of a journal was obtained from the Journal Citations Report and the Directory of Open Access Journals (DOAJ). For bivariate and regression analyses, we coded “no open access” as “0” and “open access” as “1.”

## Statistical analysis

The main outcome was the *relative deviation* between the estimated effect measure of a study and the true value represented by pooled effect estimates from Cochrane meta-analyses.

The mean relative deviation in JIF quartiles was examined using one-way ANOVA and pairwise posthoc tests with Bonferroni-Holm correction. For correlation and regression analyses, we applied a log-transformation to the relative deviation in order to deal with its limited range (zero to infinity). The pairwise correlations between the logarithm of the relative deviation, JIF, open-access policy, risk of bias, and study weight were analysed with Pearson product-moment correlation coefficients and the corresponding 95% confidence intervals (CI). Then, the effect of the JIF on the log relative deviation was analysed with multiple linear regression, adjusting for all considered covariates. We conducted a sensitivity analysis of the multiple regression analysis by adding an interaction term between JIF and study weight.

All statistical analysis was performed with the *R* software in version 3.6.3.<sup>21</sup>

## Results

We obtained 4,226 study results from 619 eligible Cochrane meta-analyses, presented in 148 reviews (Fig. 1). Of these meta-analyses, 114 (18.4%) were of “high” quality of evidence, and 505 (81.6%) were of “moderate” quality of evidence. They only included RCTs. Two hundred eighty-two (6.7%) results were reported in open-access journals.

The application of exclusion criteria resulted in a sample of 2,827 study results usable for the pairwise comparisons of the deviation by JIF (no vs. any JIF). Of those, 368 (13%) were published in journals without an assigned JIF in the respective publication year. The difference in the relative deviation was 1.04 for no JIF compared to 0.98 for any JIF (95% CI: -0.04 to 0.15) and not significant ( $P = 0.271$ ).

For the group of study results from journals with an associated JIF ( $N = 2,459$ ), the pairwise comparisons of the relative deviation by JIF quartiles are shown in Fig. 2. The mean relative deviation varied by the magnitude of the JIF, ranging between 1.11 in the lowest quartile to 0.80 in the top quartile (95% CI: 0.21 to 0.40,  $P < 0.001$ ). The mean relative deviation was higher in studies published in open-access journals compared to studies published in non-open-access journals (1.09 compared to 0.97, 95% CI: -0.02 to 0.25,  $P = 0.082$ ).

The results of the correlation analyses confirmed the hypothesised relation between the JIF and the outcome variable, the log-transformed relative deviation ( $r = -0.21$ , 95% CI:  $-0.24$  to  $-0.17$ ,  $P < 0.001$ , see Table 1).

Table 1  
Correlation matrix including the outcome variable and all covariates

	log(rel. dev.)	Adjusted JIF	Study weight	Risk of bias	Open access
<b>log(rel. dev.)</b>	1.00				
<b>Adjusted JIF</b>	-0.21 [- 0.24, - 0.17]	1.00			
<b>Study weight</b>	-0.17 [- 0.20, - 0.13]	0.17 [0.13, 0.21]	1.00		
<b>Risk of bias</b>	0.09 [0.50, 0.13]	-0.18 [- 0.22, - 0.14]	-0.10 [- 0.14, - 0.06]	1.00	
<b>Open access</b>	0.03 [- 0.01, 0.07]	-0.14 [- 0.17, - 0.10]	-0.02 [- 0.06, 0.02]	-0.08 [- 0.11, - 0.04]	1.00
Pearson product-moment correlation coefficients with corresponding 95% CI					

The weight of a point estimate in its respective meta-analysis also correlated negatively and significantly with the outcome variable ( $r = -0.17$ , 95% CI:  $-0.20$  to  $-0.13$ ,  $P < 0.001$ ) and correlated positively with the adjusted JIF ( $r = 0.17$ , 95% CI:  $0.13$  to  $0.21$ ,  $P < 0.001$ ). There was also a significant correlation between the methodical quality of a study and the impact factor of the publishing journal: the higher the aggregated risk of bias, the less influential the journal ( $r = -0.18$ , 95% CI:  $-0.22$  to  $-0.14$ ,  $P < 0.001$ ). The correlation between the aggregated risk of bias and the outcome variable was less pronounced, yet significant ( $r = 0.09$ , 95% CI:  $0.05$  to  $0.13$ ,  $P < 0.001$ ). Open access was not associated with the outcome variable ( $r = 0.03$ , 95% CI:  $-0.01$  to  $0.07$ ,  $P = 0.196$ ).

The results of the regression analysis show that the adjusted JIF was a notable and significant predictor of the log relative deviation (beta =  $-0.174$ , 95% CI:  $-0.214$  to  $-0.134$ ,  $P < 0.001$ ) when adjusting for the covariates (see Table 2).

Table 2  
Results from linear regression using the log-transformed relative deviation as outcome

	Simple linear regression			Multiple linear regression		
	Coefficient (b) [95% CI]	Std. coefficient (beta) [95% CI]	P-value	Coefficient (b) [95% CI]	Std. coefficient (beta) [95% CI]	P-value
<b>(Intercept)</b>	-0.339 [- 0.395, - 0.282]		< 0.001	-0.288 [- 0.392, - 0.184]		< 0.001
<b>Adjusted JIF</b>	-0.067 [- 0.080, - 0.054]	-0.206 [- 0.244, - 0.167]	< 0.001	-0.057 [- 0.070, - 0.044]	-0.174 [- 0.214, - 0.134]	< 0.001
<b>Study weight</b>				-0.014 [- 0.018, - 0.010]	-0.131 [- 0.170, - 0.091]	< 0.001
<b>Risk of bias</b>				0.333 [0.069, 0.597]	0.050 [0.010, 0.089]	= 0.013
<b>Open access</b>				-0.019 [- 0.160, 0.197]	-0.004 [- 0.035, 0.043]	= 0.838
	<i>P</i> < 0.001, adjusted <i>r</i> <sup>2</sup> = 0.04			<i>P</i> < 0.001, adjusted <i>r</i> <sup>2</sup> = 0.061		

The study weight also affects the outcome (beta = - 0.131, 95% CI: -0.170 to - 0.091, *P* < 0.001). The interaction between the JIF and the study weight in the multiple linear regression model was weak and not significant (b = 0.001, 95% CI: -0.000 to 0.002, *P* = 0.202). Methodical quality contributed weakly, yet significantly (beta = 0.050, 95% CI: 0.010 to 0.089, *P* = 0.013) to the overall regression model (adjusted *r*<sup>2</sup> = 0.061, *P* < 0.001).

## Discussion

### Main findings

We found an association between the impact factor of medical journals and the trueness of research findings published therein: The higher the JIF, the more likely the reported results are accurate. Adjusted JIFs in the third and fourth quartile (higher than 1.09) are associated with above-average accuracy. In our sample of medical journals, an adjusted JIF of 1.09 roughly translates to an actual JIF of 2.5 in the year 2000 or 4.9 in the year 2018. The increase of the adjusted JIF by 1 yields a decrease of the relative

deviation by about 6.5%. However, the predictive ability of the (adjusted) JIF for the accuracy of a study result is low. Putting trust in a specific study result due to the journal's impact is therefore precarious. Just as it is problematic to deduce from the journal's impact to the impact of an individual study.<sup>22</sup>

We also found that higher study weight and lower aggregated risk of bias were associated with more accurate results. Is then the hypothesised and observed association between the journals' impact and trueness of study findings a mere reflection of the strengths of well-powered and well-designed studies? The results from the multiple regression analysis suggest that this is not the case. The JIF remained to be the most powerful determinant of trueness, followed by study size. The influence of methodical quality, however, nearly vanished. This finding was somewhat surprising initially, as we deemed high methodical standards to be a major journal characteristic to convey the hypothesised advantage of high-rank journals. The low impact we found may, in part, be due to the pragmatic manner by which we assessed the methodical quality. A more elaborate assessment might result in a stronger influence of methodologic strength and vice versa in a weaker influence of the JIF. However, it should be noted that this consequence would impinge on the explanation of the observed correlation between JIF and trueness, but it would not affect the predictive power of the JIF as such.

Our results do not positively confirm the reservations towards the reliability of small studies, early published in the highest-impact medical journals. They arise from the observation that such studies tend to show large effects that would not be replicated in subsequent studies ("decline effect").<sup>11,23</sup> In our data, this phenomenon should emerge as an interaction between JIF and study weight in the regression analysis. Such interaction was not observed. We consider three reasons for that. First, we were interested in the general relationship between the journals' rank and the trueness of results. We, therefore, conducted no extra evaluation of only a few top-rank journals. Second, our conceptualisation of trueness is indifferent towards the magnitude of reported effects. Hence, the comparability of our results and existing research is inherently limited. Third, small studies with large, unreplicated effects may be underrepresented in our sample. According to the GRADE-framework, those may weaken the quality of evidence in several domains, e.g. the inconsistency of results, imprecision of results, and the probability of publication bias. If related concerns led the review authors to downgrade the quality of evidence to less than "moderate," the meta-analysis and the included studies would not be eligible for our analysis.

## Strengths and Limitations

By putting study results in relation to effect estimates from good-quality Cochrane meta-analyses, we applied a direct and feasible approach to measure the trueness of study results. Within this concept, large effects are not suspicious per se, even if they have not been replicated by more extensive studies. Large effects may occur not only by chance but also due to highly-selected study populations or optimised study conditions. The latter reasons would jeopardise the generalisability of the study results rather than their trueness. The downside of this approach is that the chosen measure for trueness is only available for a small proportion of all studies. Studies not meeting the inclusion criteria of Cochrane reviews (i. e. observational studies) and large RCTs, which are excluded due to their inherent correlation with the

pooled effect estimate, were not considered. Observational studies, the results of which are less likely to be accurate, might be published rather in lower-impact journals, while large RCTs are published in higher-impact journals. We, therefore, assume that if our study is not representative of the medical literature as a whole, our approach underestimates the association between the JIF and the trueness of study results.

Our study relies largely on the judgments of review authors regarding the inclusion of studies into the review, the risk of bias assessment, and the subsequent valuation of the quality of evidence. These judgments are ultimately at the review authors' discretion but are not arbitrary, as they are guided by transparent rules. Furthermore, flawed assessments, which are likely to put the validity of specific review findings into question, are less likely to bias our results based on several thousand observations.

## Implications

As the hypothesised association was confirmed, we state that higher-rank journals' manuscript selection process is effective in identifying research results that are more likely to be true. This is the more meritorious as publishing true results may not always be the most promising way to achieve a maximum of citations, a goal appealing to both authors and journal editors. An alternative, potentially incompatible strategy might be to focus on studies in "hot" scientific fields, but the findings of which are supposed to be less likely true.<sup>1</sup> We conclude that more prestigious journals serve as gatekeepers in the process of science more effectively than less-cited journals. Within a fierce competition for limited publication space, these journals can choose from a larger number of better-quality submissions, and they can use the support of a larger number of experienced and diligent peer reviewers, helping in both separating the wheat from the chaff and improving the quality of research found worthy to be published.<sup>24</sup> The criteria used lie widely in darkness. Our results suggest that they go far beyond the factors we considered in our study. This fits the notion that the peer review process, which critically informs the editorial decision on acceptance, is highly complex and subjective.<sup>25</sup> This is not necessarily a disadvantage. A modelling study showed that exercising some subjectivity in reviewer decisions can be beneficial for the scientific community in processing available information to estimate truth more accurately.<sup>26</sup>

In all this, we should keep in mind that giving true answers to specific study questions is an essential but not the only function of useful research. False findings caused by intelligent mistakes, unorthodox and immature methods, and improper samples may broaden the spectrum of scientific thinking more effectively than adding certainty to what is already known to be true.

## Declarations

### ETHICS APPROVAL

This study is research on research. Ethical approval was not required.

### AUTHOR CONTRIBUTIONS

**Andreas Heidenreich** participated in the design of the study, selection and acquisition of data, conducted the statistical analysis, interpreted the data, and wrote the manuscript.

**Nora Eisemann** participated in the design of the study, provided statistical consulting, and critically reviewed the manuscript.

**Alexander Katalinic** participated in the design of the study and critically reviewed the manuscript.

**Joachim Hübner** generated the idea, participated in the design of the study, interpretation of the data, and writing of the manuscript.

## DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

## FUNDING

The study was supported by the **Federal Ministry of Education and Research, Germany** within the funding line “Quantitative research on the science sector” [grant number 01PU17004].

## CONFLICT OF INTEREST

None declared.

## References

1. Ioannidis JP. Why most published research findings are false. *PLoS Med.* 2005;2:e124.
2. Saha S, Saint S, Christakis DA. Impact factor: a valid measure of journal quality? *Journal of the Medical Library Association: JMLA.* 2003;91:42–6.
3. McKiernan EC, Schimanski LA, Munoz Nieves C, Matthias L, Niles MT, Alperin JP. Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations. *Elife* 2019; 8.
4. Paulus FM, Cruz N, Krach S. The Impact Factor Fallacy. *Front Psychol.* 2018;9:1487.
5. Ioannidis JPA, Thombbs BD. A user's guide to inflated and manipulated impact factors. *Eur J Clin Invest.* 2019;49:e13151.
6. Bala MM, Akl EA, Sun X, et al. Randomized trials published in higher vs. lower impact journals differ in design, conduct, and analysis. *J Clin Epidemiol.* 2013;66:286–95.
7. Dechartres A, Trinquart L, Atal I, et al. Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: research on research study. *BMJ.* 2017;357:j2490.
8. Lee KP, Schotland M, Bacchetti P, Bero LA. Association of journal quality indicators with methodological quality of clinical research articles. *JAMA.* 2002;287:2805–8.
9. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet.* 1998;352:609–13.

10. Nuesch E, Trelle S, Reichenbach S, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. *BMJ*. 2009;339:b3244.
11. Siontis KC, Evangelou E, Ioannidis JP. Magnitude of effects in clinical trials published in high-impact general medical journals. *Int J Epidemiol*. 2011;40:1280–91.
12. Olsen O, Middleton P, Ezzo J, et al. Quality of Cochrane reviews: assessment of sample from 1998. *BMJ*. 2001;323:829–32.
13. Moseley AM, Elkins MR, Herbert RD, Maher CG, Sherrington C. Cochrane reviews used more rigorous methods than non-Cochrane reviews: survey of systematic reviews in physiotherapy. *J Clin Epidemiol*. 2009;62:1021–30.
14. Jorgensen AW, Hilden J, Gotzsche PC. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. *BMJ*. 2006;333:782.
15. Cumpston M, Li T, Page MJ, et al. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. *Cochrane Database Syst Rev*. 2019;10:ED000142.
16. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64:401–6.
17. The Nordic Cochrane Center. Review Manager (RevMan). 5.3 ed. Copenhagen: The Cochrane Collaboration; 2014.
18. Althouse BW, Bergstrom JD, Bergstrom CT. T. Differences in Impact Factor Across Fields and Over Time. *Journal of the Association for Information Science Technology*. 2009;60:27–34.
19. Neff BO. JD. Not So Fast: Inflation in Impact Factors Contributes to Apparent Improvements in Journal Quality. *Bioscience*. 2010;60:455–9.
20. Higgins JPTTJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. 2 ed. Chichester: John Wiley & Sons; 2019.
21. R Core Team. *R: A language and environment for statistical computing*. 3.6.3 ed. Vienna: R Foundation for Statistical Computing; 2020.
22. Cagan R. San Francisco Declaration on Research Assessment. *Disease Models & Mechanisms* 2013; dmm.012955.
23. Brembs B, Button K, Munafo M. Deep impact: unintended consequences of journal rank. *Front Hum Neurosci*. 2013;7:291.
24. Paine CET, Fox CW. The effectiveness of journals as arbiters of scientific impact. *Ecol Evol*. 2018;8:9566–85.
25. Smith R. Peer review: a flawed process at the heart of science and journals. *J R Soc Med*. 2006;99:178–82.
26. Park IU, Peacey MW, Munafo MR. Modelling the effects of subjective and objective decision making in scientific peer review. *Nature*. 2014;506:93–6.

# Figures

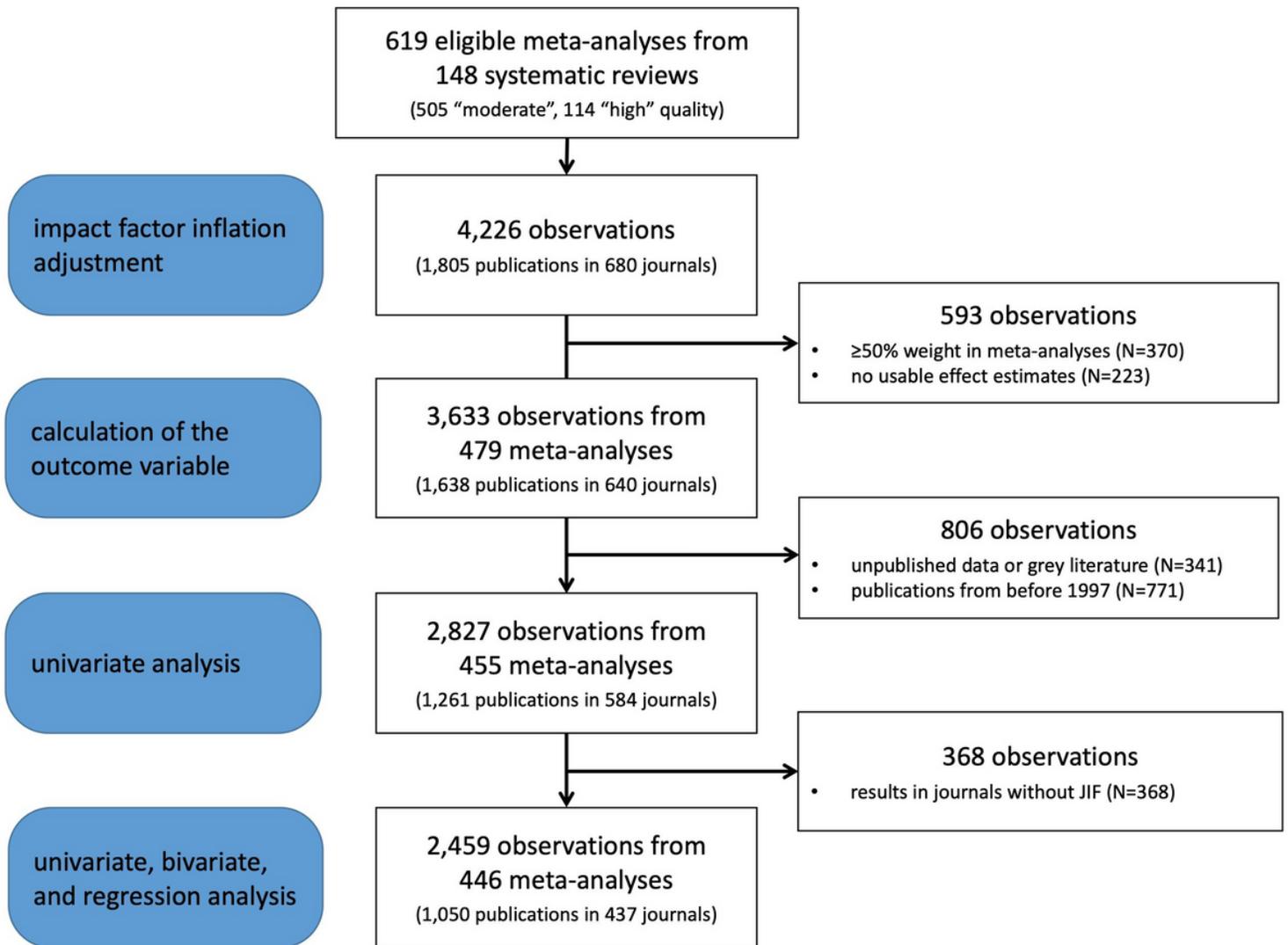
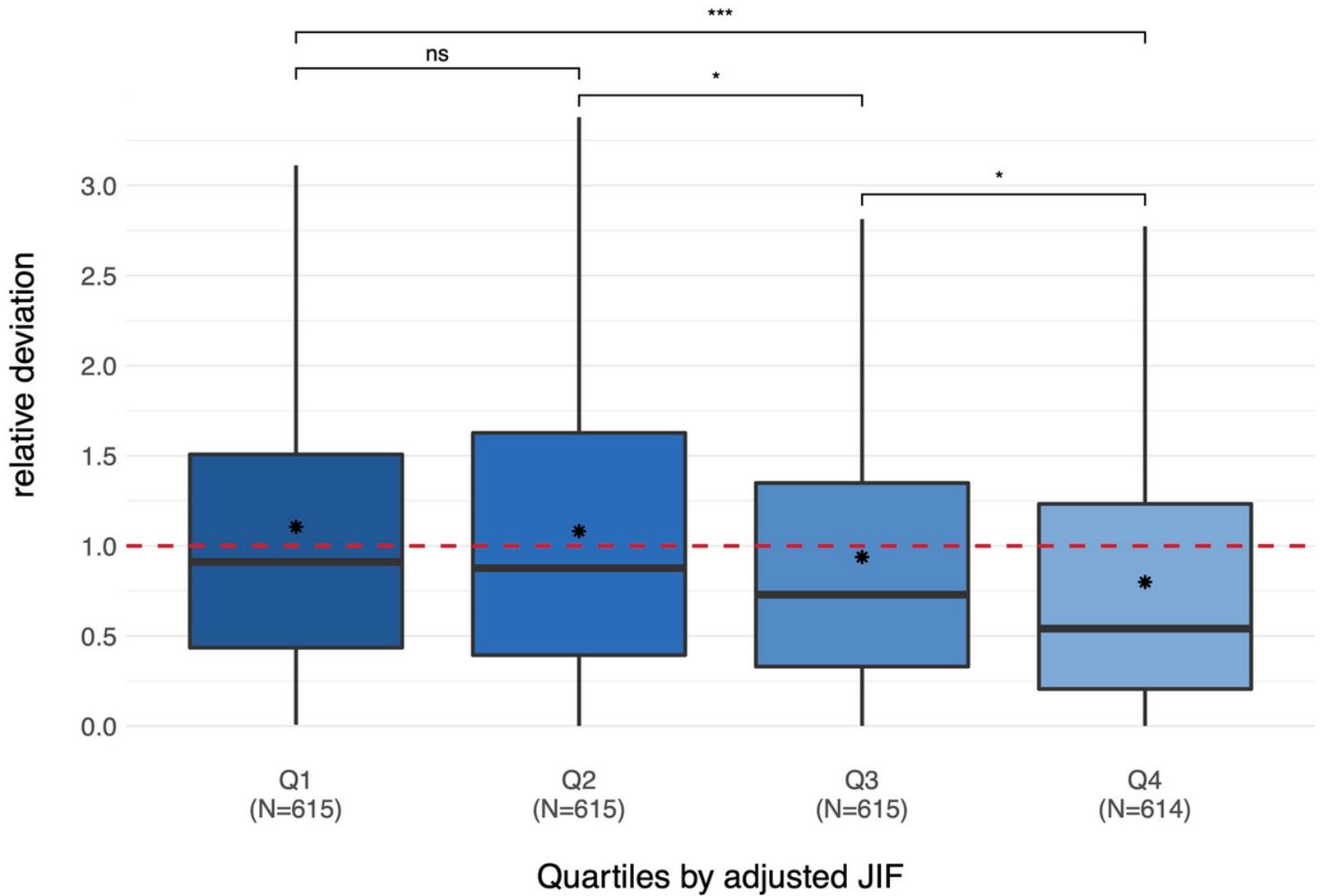


Figure 1

Inclusion and exclusion criteria at various steps of analysis.



**Figure 2**

Comparison of the mean relative deviation by adjusted-JIF quartiles. The asterisks in the boxes denote the mean. The whiskers' length is 1.5 times the interquartile range. The horizontal, dashed line at 1.0 represents the mean value across all observations in the study sample (including results with no associated JIF). P-values are Bonferroni-Holm-adjusted for multiple testing.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [20210831Systematicreviewsappendix.docx](#)