

Biological science
practices

Cite this article: Pollo P, Lagisz M, Macedo-Rego RC, Mizuno A, Yang Y, Nakagawa S. 2025 Reliability of meta-analyses in ecology and evolution: (mostly) good news from a case study on sexual signals. *Proc. R. Soc. B* **292**: 20242782. <https://doi.org/10.1098/rspb.2024.2782>

Received: 20 November 2024

Accepted: 9 April 2025

Subject Category:

Evolution

Subject Areas:

ecology, evolution, behaviour

Keywords:

data integrity, methodological accuracy, open science, research synthesis, sexual selection

Author for correspondence:

Pietro Pollo

e-mail: p.pollo@unsw.edu.au

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.7772201>.

Reliability of meta-analyses in ecology and evolution: (mostly) good news from a case study on sexual signals

Pietro Pollo¹, Malgorzata Lagisz¹, Renato C. Macedo-Rego², Ayumi Mizuno^{1,3}, Yefeng Yang¹ and Shinichi Nakagawa^{1,3}¹Evolution & Ecology Research Centre, School of Biological, Earth & Environmental Sciences, University of New South Wales, Kensington, New South Wales, Australia²Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil³Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

IDs PP, 0000-0001-6555-5400; ML, 0000-0002-3993-6127; RCM-R, 0000-0002-2973-9912; AM, 0000-0003-0822-5637; YY, 0000-0002-8610-4016; SN, 0000-0002-7765-5182

Meta-analyses are powerful synthesis tools that are popular in ecology and evolution owing to the rapidly growing literature of this field. Although the usefulness of meta-analyses depends on their reliability, such as the precision of individual and mean effect sizes, attempts to reproduce meta-analyses' results remain rare in ecology and evolution. Here, we assess the reliability of 41 meta-analyses on sexual signals by evaluating the reproducibility and replicability of their results. We attempted to: (i) reproduce meta-analyses' mean effect sizes using the datasets they provided; (ii) reproduce meta-analyses' effect sizes by re-extracting 5703 effect sizes from 246 primary studies they used as sources; (iii) assess the extent of relevant data missed by original meta-analyses; and (iv) replicate meta-analyses' mean effect sizes after incorporating re-extracted and relevant missing data. We found many discrepancies between meta-analyses' reported results and those generated by our analyses for all reproducibility and replicability attempts. Nonetheless, we argue that the meta-analyses we evaluated are largely reproducible and replicable because the differences we found were small in magnitude, leaving the original interpretation of these meta-analyses' results unchanged. Still, we highlight issues we observed in these meta-analyses that affected their reliability, providing recommendations to ameliorate them.

1. Introduction

The literature of ecology and evolution, like of other fields of study, is expanding rapidly [1,2]. Consequently, synthesizing this growing body of work becomes increasingly necessary to identify patterns across individual studies. Meta-analyses, which are perceived as the gold standard for evidence synthesis, can contribute to this endeavour because they aim to detect and retrieve all relevant studies on a given topic, extract data from these studies and quantify an average effect of interest using the extracted data [3,4]. Nonetheless, many challenges arise during the execution of meta-analytical studies, requiring researchers to be vigilant to ensure their reliability.

Foremost, meta-analyses need to be transparent, which means that they must provide details on decisions and resources regarding all steps of their implementation (e.g. searches, screening, data extraction and analysis code). Yet, recent appraisals of the secondary literature in ecology and evolution show that meta-analyses in this field are often poorly transparent [5–7]. Reporting guidelines, such as MOOSE [8] and PRISMA [6,9], were developed to address transparency issues in evidence syntheses. Researchers who

adhere to these guidelines can thus enhance the quality of their meta-analyses. Still, transparency represents only the first step for high-quality syntheses as providing information does not guarantee its reliability. For instance, researchers can make mistakes when describing their methods, provide faulty code for their analyses or even falsify data, impairing the reproducibility and replicability of their findings. Therefore, it is imperative to evaluate studies beyond transparency.

Meta-analyses possess the advantage of having sources of information that are readily accessible (i.e. individual studies from which data are extracted). This means that meta-analyses can be more easily evaluated than other approaches regarding reproducibility and replicability aspects. More specifically, meta-analytical studies that report their decision criteria (i.e. studies and data considered valid for inclusion) and their effect sizes with information on their sources (i.e. studies they were extracted from) should allow evaluations regarding the reproducibility of their dataset and their results. However, to our knowledge, only computational reproducibility of meta-analyses (i.e. reproducibility of their code) has been examined in ecology and evolution [10]. Although meta-analyses with similar goals conducted by independent groups may represent examples of quasi-replications in ecology and evolution (e.g. [11] and [12]; [13] and [14]), reproducibility attempts of data extraction and general results have only been conducted in other fields of study (but see [15]; e.g. medicine [16]; psychology [17]). Furthermore, other elements connected to the reliability of meta-analyses, such as their efficacy in detecting relevant studies, remain virtually unexplored.

Here, using 41 meta-analyses related to sexual signals [18–58], we conduct, to our knowledge, the largest reproducibility and replicability effort for meta-analyses ever done in terms of number of re-extracted primary studies and effect sizes (5703 data points from 246 primary studies, representing almost a fifth of all primary studies in our dataset). We evaluate multiple reliability aspects related to distinct implementation stages of these meta-analyses (electronic supplementary material, figure S1). First, we re-analyse the dataset provided by these meta-analyses, comparing the mean effect sizes we obtained with those reported in them. Second, we extract data from their original sources (i.e. individual studies) and compare these extracted data points with the ones reported in these meta-analyses. Third, we assess how many data points from the original sources that we verified should have been extracted and included in these meta-analyses' datasets but were not (i.e. 'missed' data). Fourth, because these meta-analyses ask similar questions involving sexual signals, we estimate the minimum number of studies that contained relevant data but were not listed as sources in these meta-analyses (i.e. undetected studies). Fifth, we assess the extent that the results of these meta-analyses change when re-extracted data (along with missed data and undetected studies) is analysed in place of originally reported data.

2. Methods

This manuscript is part of a larger research project that uses data from specific meta-analyses (see subsections below). The dataset gathered and analysed in the present study was also used to conduct a secondary meta-analysis [59]. Our methodology, summarized in the electronic supplementary material, figure S1, was described in our pre-registration [60], and we adhered to it as much as possible (see changes in the electronic supplementary material, S1). We report author contributions using MeRIT guidelines [61] and the CRediT statement [62].

(a) Reported dataset

A recent systematic map identified the existence of 152 meta-analyses on topics related to sexual selection [7], 59 of them focusing on questions associated with 'pre-copulatory sexual traits' (i.e. sexual signals) [7]. In November 2023, P.P. selected 44 meta-analyses from this set, specifically the ones examining the relationship between sexual signals and distinct conditions, fitness proxies or individual traits (hereby *proxies*; see the electronic supplementary material, table S1). This means that P.P. excluded 15 meta-analyses that only explored other relationships of sexual signals, such as with species recognition (e.g. [63]), biotic factors (e.g. [64]), abiotic conditions (e.g. [65]), or other factors. P.P. also included another meta-analysis examining the relationship between ornament expression and parasite load [18], which was published after searches were conducted in [7]. This resulted in a total of 45 eligible meta-analyses to fulfil our objectives [18–58,66–69].

P.P. gathered the data reported by these meta-analyses from their tables, appendices, supplementary files and occasionally from direct correspondence with their authors. We could not obtain data for four eligible meta-analyses as their data were unreported [67–69] or poorly described [66]. Altogether, the remaining 41 meta-analyses from which P.P. extracted data [18–58] yielded 6773 data points (see §2e).

P.P. filtered the collated dataset for most analyses conducted here (see §2e and the electronic supplementary material, S2). This was done because not all data collected from meta-analyses were relevant to objectives of other parts of the larger research project (e.g. meta-meta-analysis of sexual signals; see [60]). Following this filtering process, the number of data points in the collated dataset decreased to 5496.

(b) Re-extractions

Many meta-analyses included in the collated dataset extracted data from the same primary studies. More specifically, after filtering, P.P. detected that a quarter of primary studies in the dataset (318 out of 1272) were used as sources by at least two different meta-analyses (hereby *duplicates*; electronic supplementary material, figure S1). P.P. selected a subset of these duplicates for data re-extraction to reduce sampling effort (generating a greater sample size of data points for each meta-analysis assessed), prioritizing certain primary studies (see details in the electronic supplementary material, S3). In total, this process

produced a set of 249 primary studies for re-extraction. However, we could not access the full-text of three of these studies, so our sample of studies for re-extraction was reduced to 246 (hereby *verified* primary studies; electronic supplementary material, S4).

We extracted all data on the relationship between sexual signals and proxies from verified primary studies (see details in the electronic supplementary material, S2), blinded to which exact data points were extracted for each of the meta-analyses that included a given study. More specifically, P.P. extracted data from 59.3% of the selected primary studies, while S.N., Y.Y., A.M., R.C.M.R. and M.L., respectively, extracted data from other 15.5%, 7.7%, 6.5%, 5.7% and 5.3% studies. P.P. then cross-checked all data extractions done by other authors (40.7% of studies). Importantly, authors never re-extracted data from primary studies they participated in. However, we cannot rule out that we experienced some unconscious bias when dealing with the meta-analyses of co-authors. We extracted data required to calculate effect sizes from text, tables, supplementary material and figures (using the package *metaDigitise* 1.0.1 [70]). When primary studies reported similar results in various forms, we prioritized extractions in the following order: (i) raw data (calculating estimates directly) from sources other than figures; (ii) raw estimates (i.e. means and correlation coefficients) from sources other than figures; (iii) raw data or raw estimates from figures; and (iv) other estimates (e.g. t , β , χ^2) regardless of their origin. Nonetheless, we first prioritized data sources that showed more details. For instance, if a correlation was given for all individuals in text but a scatter plot showed the same data with dots separated by age or sex, we collected data from the latter. In total, we extracted 5703 valid data points. We note that we do not claim that our re-extracted data points are more or less correct than the data originally reported by meta-analyses, yet we do expect them to converge, meaning that mismatches should be taken seriously given our transparent procedures.

(c) Matching reported data with re-extracted data

P.P. carefully examined the inclusion criteria reported in meta-analyses to verify which re-extracted data points should have been included by them. However, P.P. found several issues with these inclusion criteria. First, the proxies and sexual signals that meta-analyses included were often vague or ambiguous. For instance, Dougherty [19] was interested in behavioural sexual signals but their dataset also included extended phenotypes (e.g. domes built by crabs). Even though these extended phenotypes can be considered behavioural products, it was unclear whether other similar structures (e.g. bowers, ornamented nests) were deemed valid for inclusion by Dougherty [19]. Second, we detected apparent inclusion criteria patterns in datasets of some meta-analyses' that were not mentioned in text. For example, Nakagawa *et al.* [40] stated that they included data on reproductive success but the only reproductive success measure in their dataset was the number of fledglings, even though the studies they used for effect size extraction also contained other measures (e.g. number of eggs, number of hatchlings). Third, some meta-analyses outwardly contradicted their own information. For example, Weaver *et al.* [56] stated that they included standardized colour metrics (hue, chroma or composite measures of those) for carotenoid-based colours in adult birds, describing specific proxies in their table 2. Yet, they seemed to have included data points in which (i) the sexual signal was the size of a colourful plumage (e.g. patch size), (ii) individuals were juveniles (including when data points were separated by age), and (iii) proxies other than the ones reported in text were used (e.g. offspring size). We summarized all meta-analyses' originally reported inclusion criteria, the ambiguities, omissions and contradictions we detected in them, and how we dealt with these issues for matching purposes in the electronic supplementary material, S5.

We then attempted to match data points reported in meta-analyses with the ones we re-extracted from primary sources. P.P. mainly used the description of sexual signals and proxies of each data point to match them with re-extracted data. When multiple data points from the same primary study had similar descriptions, we also used sample size and other additional information (e.g. statistics reported, if given) for matching purposes.

There were three possibilities for each matching attempt. First, when both original and re-extracted data points had a similar description, P.P. linked them by labelling the latter with the identity (ID) of the former (i.e. successfully matched them). Yet, it was common to find multiple data points in our re-extracted dataset that matched one or many data points from the originally reported dataset (or vice-versa), so this matching was not necessarily exact (see examples in the electronic supplementary material, S5). Second, there were cases in which we could not find original data points with an equivalent description to relevant re-extracted ones. We assumed that these data were missed or undetected by meta-analyses' authors. This allowed us to obtain two aspects related to reliability: (i) the number of relevant data points that were absent in meta-analyses' datasets despite being present in primary studies reported as sources (hereby *missing data*); and (ii) the number of primary studies that contained relevant data points that should have been used as sources but were not (hereby *undetected studies*). To clarify the latter, consider a hypothetical meta-analysis that investigated the relationship between X and Y, reporting data from 10 primary papers. We then noticed two other studies containing relevant data (relationship between X and Y) that were used as data sources by other meta-analyses in our dataset but not the hypothetical one, even though the data from these two primary studies fit the inclusion criteria reported by the hypothetical meta-analysis. We thus deem that the hypothetical meta-analysis failed to detect at least two relevant articles. Third, there were cases in which data points were shown in meta-analyses' reported datasets but were absent in our re-extractions. We re-checked all of these latter cases: although some data points were not extracted by us because of our criteria (e.g. invalid proxy) or because we considered them repeated data, most of them could simply not be found in primary studies (see details in the electronic supplementary material, table S2). This could have happened if meta-analyses' authors contacted primary authors and were thus able to obtain more data than what was shown in the primary articles (i.e. unpublished data). Alternatively, meta-analyses' authors may have made mistakes during data extraction, even though we cannot ascertain when this was truly the case.

After P.P. finalized the matching process, A.M., M.L., R.C.M.R., S.N. and Y.Y. cross-checked matching decisions for five different primary studies each. This resulted in matching decisions for approximately 10% of all verified studies being cross-checked, partially assessing the reliability of our process.

(d) Effect sizes

Originally reported data points were given as the following effect size types: Cohen's d [22,24], logarithm of response ratio (logRR) [32], and Fisher's Z_r or correlation coefficients (r) (remaining meta-analytical studies). One meta-analysis in particular [42] only provided p -values and sample sizes, so P.P. calculated effect sizes from this information. We transformed effect sizes reported to Z_r for all analyses except the replication of general results (see §2e). We also calculated Z_r (along with its sampling variance) from all re-extracted data points. Additionally, effect sizes reported by Koch *et al.* [32] did not reflect the raw data that they provided (means and standard deviations), so P.P. re-calculated their effect sizes. All equations for calculation and conversion of effect sizes are given in the electronic supplementary material, S7, while the direction rationale applied to effect sizes is detailed in the electronic supplementary material, S8.

(e) Analyses

First, we attempted to replicate meta-analyses' reported mean effect sizes using their original datasets. To do so, we conducted a meta-analytical model for each meta-analytical study using all of their reported effect sizes together (i.e. global model, *sensu* [7]). However, for meta-analytical studies that only performed subgroup analyses, we only re-analysed the largest or first reported subgroup. For instance, Nolasco *et al.* [41] only analysed the relationship between plumage colour and proxies for each sex separately, so we re-analysed only their data related to females. Effect sizes for these replication analyses were of the same type as results reported by meta-analyses (see §2d). Meta-analytical models for all of our analyses contained multiple random factors (see the end of this section) but, for this replication analysis, we also included an additional random factor if provided by authors. For example, both Robinson & Creanza [46] and Sánchez-Tójar *et al.* [49] used population ID as an additional random factor in their models. Yet, we were unable to include these additional random factors in meta-analytic models when these variables were not provided by authors with the rest of the data (e.g. experiment ID in [19] and population ID in [50]) or when they were redundant (population ID was different for every source in [20]). Moreover, two meta-analytical studies from our collated dataset were excluded from this specific analysis: Parker & Ligon [42] did not provide a confidence interval (CI) for their mean effect size and only data points of interest were extracted from Thornhill *et al.* [55] (i.e. their dataset was not fully extracted). We also tested for signs of publication bias (small-study effect) in re-analysed datasets by adding the inverse of the effective sample size as a moderator in meta-analytical models (alternative Egger's regression) [71].

Second, we attempted to replicate part of individual effect sizes reported in meta-analyses. To do so, we used linear models to compare effect sizes originally reported by meta-analyses with the effect sizes we re-extracted from primary studies. For this analysis, we only used exactly matched data, in which each originally reported data point was matched to a single re-extracted data point (see the electronic supplementary material, S6). In these linear models, the intercept was forced to be zero, originally reported effect sizes were the response variable, and re-extracted effect sizes were the predictor variable. We evaluated whether the 95% CI of the estimated slope in each linear model included the value 1, which would represent that reported and re-extracted effect sizes are very similar. We complemented this analysis by computing the repeatability of individual effect sizes, i.e. the proportion of the variance from their origin (reported versus re-extracted) from the total variance observed. We could not compare originally reported effect sizes with re-extracted effect sizes from 11 meta-analyses because they contained less than six exactly matched data points (the arbitrary threshold we established for this analysis), so we only reported results related to the remaining 30 meta-analyses in our dataset.

Third, we assessed meta-analyses' reproducibility by comparing the results of four slightly distinct meta-analytical models for each meta-analytical study. In model 1, we analysed the filtered dataset, which contained only reported effect sizes. In model 2, we used the same data as in model 1, but we replaced reported effect sizes that came from verified studies with their equivalent re-extracted effect sizes (i.e. the ones that matched in description). In model 3, we used the same data as in model 2, but we also added other relevant re-extracted data points from verified studies that were originally missed by meta-analyses' authors (i.e. missed data). In model 4, we used the same data as in model 3, but we also added all relevant effect sizes from undetected studies. Differently from our analysis attempting to replicate mean effect sizes (first described in this section), we conducted global models for all meta-analytical studies, even those that had done only subgroup analyses. We note that the datasets used for most meta-analyses remained similar across the four models described above because we did not re-extract data from all primary studies reported as sources (electronic supplementary material, figure S2). For example, Dougherty [19] extracted data from 197 primary studies, but we verified only 15 of these studies. This means that results from these distinct models were unlikely to change for meta-analyses with proportionally few verified sources. Thus, we additionally conducted the same four models described above using only data from verified studies (instead of using data from both verified and unverified studies) for 15 meta-analyses with at least 15 verified sources. Furthermore, we also evaluated whether the inclusion of missing data and undetected studies changed the generalizability of results by examining the σ (a measure of heterogeneity) from each random factor in meta-analytical models.

When applicable, we compared estimates from meta-analytic models both qualitatively and quantitatively. We first ascertained whether an estimate was positive, negative, or not different from zero (if its 95% CI overlapped zero), so that estimates with distinct classifications represented a qualitative difference. By contrast, a quantitative difference occurred when the

absolute difference between two estimates produced a z-score of more than 1.96 (i.e. statistically significant, two-tailed $\alpha = 0.05$), calculated as:

$$z = \frac{X_{\text{reported}} - X_{\text{re-analysed}}}{\sqrt{\text{se}_{\text{reported}}^2 + \text{se}_{\text{re-analysed}}^2 - 2r\text{se}_{\text{reported}}\text{se}_{\text{re-analysed}}}},$$

where X represents the estimated mean effect size, se represents its standard error, and r represents the correlation coefficient between these two groups (set as 0.8 for all main results as we assume this correlation is high but not perfect; but see the electronic supplementary material, figures S5–S7 for results using alternative methods). Although we showed all mean effect sizes generated in the last set of analyses, where we conducted up to four models per meta-analysis, we only discuss comparisons between the first and last models for simplicity.

We conducted all analyses described above in R 4.4.0 [72]. Multilevel meta-analytical models were fitted using the *rma.rv* function from the package *metafor* 4.6–0 [73]. All meta-analytical models fitted contained primary study ID as a random factor. Species ID (non-phylogenetic effect) and phylogenetic relatedness were also added as random factors in meta-analytical models to analyse data from meta-analytical articles that involved multiple species [74]. However, we removed phylogenetic relatedness from meta-analytical models related to certain articles [22,28,35,36], otherwise some of them would not converge. Phylogenetic trees were built using the packages *ape* 5.8 [75] and *rotl* 3.1.0 [76]. We conducted repeatability analyses using the package *rptR* 0.9.22 [77]. Unless stated otherwise, estimates reported in the manuscript represent mean \pm s.e.

3. Results

(a) Reproducibility of mean effect sizes

We found qualitative differences in 15.4% (6 out of 39) of the comparisons between mean effect sizes reported by meta-analyses and mean effect sizes resulting from re-analyses of those meta-analyses' reported datasets (figure 1). More specifically, five of six meta-analyses reported a mean effect size not different from zero, but its re-analysed counterpart was positive, while another meta-analysis reported a mean effect size not different from zero, but its re-analysed counterpart was negative. We also detected quantitative differences between originally reported and re-analysed mean effect sizes (i.e. surpassed z-score of 1.96) in approximately a tenth of comparisons made (4 out of 39), although none of them were qualitative differences (figure 1; but see the electronic supplementary material, figure S3). Despite this, the absolute difference in magnitude between originally reported and re-analysed mean effect sizes was always lower than 0.2, even for cases with detected qualitative or quantitative differences (figure 2). We also found that effect sizes were positively predicted by effective sample size (i.e. evidence of publication bias) in 23.1% (9 out of 39) of meta-analyses. Two-thirds of these meta-analyses (6 out of 9) diligently reported that they detected publication bias [27,28,30,31,43,53], while the remaining third did not assess publication bias whatsoever [21,24,58] (electronic supplementary material, table S3).

(b) Reproducibility of individual effect sizes

We found that effect sizes originally reported by meta-analyses were, on average, statistically identical to their equivalent re-extracted ones in less than half (14 out of 30) of the meta-analyses evaluated (figure 3; electronic supplementary material, table S4). For the remaining meta-analyses, the slope from linear regressions between reported and re-extracted effect sizes was always lower than 1 (0.55 ± 0.07). While we observed, on average, a moderate repeatability of individual effect sizes (0.68 ± 0.05), meta-analyses whose re-extracted effect sizes we deemed statistically identical to the ones reported (i.e. slope not different from 1) showed higher repeatability in effect sizes than other meta-analyses (0.79 ± 0.05 versus 0.55 ± 0.07 ; electronic supplementary material, figure S4).

(c) Missing data and undetected studies

We found that meta-analyses missed (i.e. failed to extract and report), on average, $16 \pm 2.6\%$ of relevant effect sizes from primary studies they used as sources for data extraction (figure 4A). Additionally, we found that meta-analyses were unsuccessful in detecting, on average, at least 10% of primary studies they should have included as data sources (figure 4B). Furthermore, we found no association between the number of missing effect sizes and the minimum proportion of undetected studies across meta-analyses (figure 4C).

(d) Replicability of results

Using data from both verified and unverified primary studies (i.e. those for which we did or did not re-extract data, respectively), we found that mean effect sizes from analyses of originally reported datasets and those from analyses incorporating all relevant re-extracted data (matched, missed data, and undetected studies) were qualitatively and quantitatively distinct for 12.2% (5 out of 41) and 9.7% (4 out of 41) of meta-analyses, respectively (both types of differences occurred for one meta-analysis; figure 5; but see the electronic supplementary material, figure S5). Similar comparisons using only data from

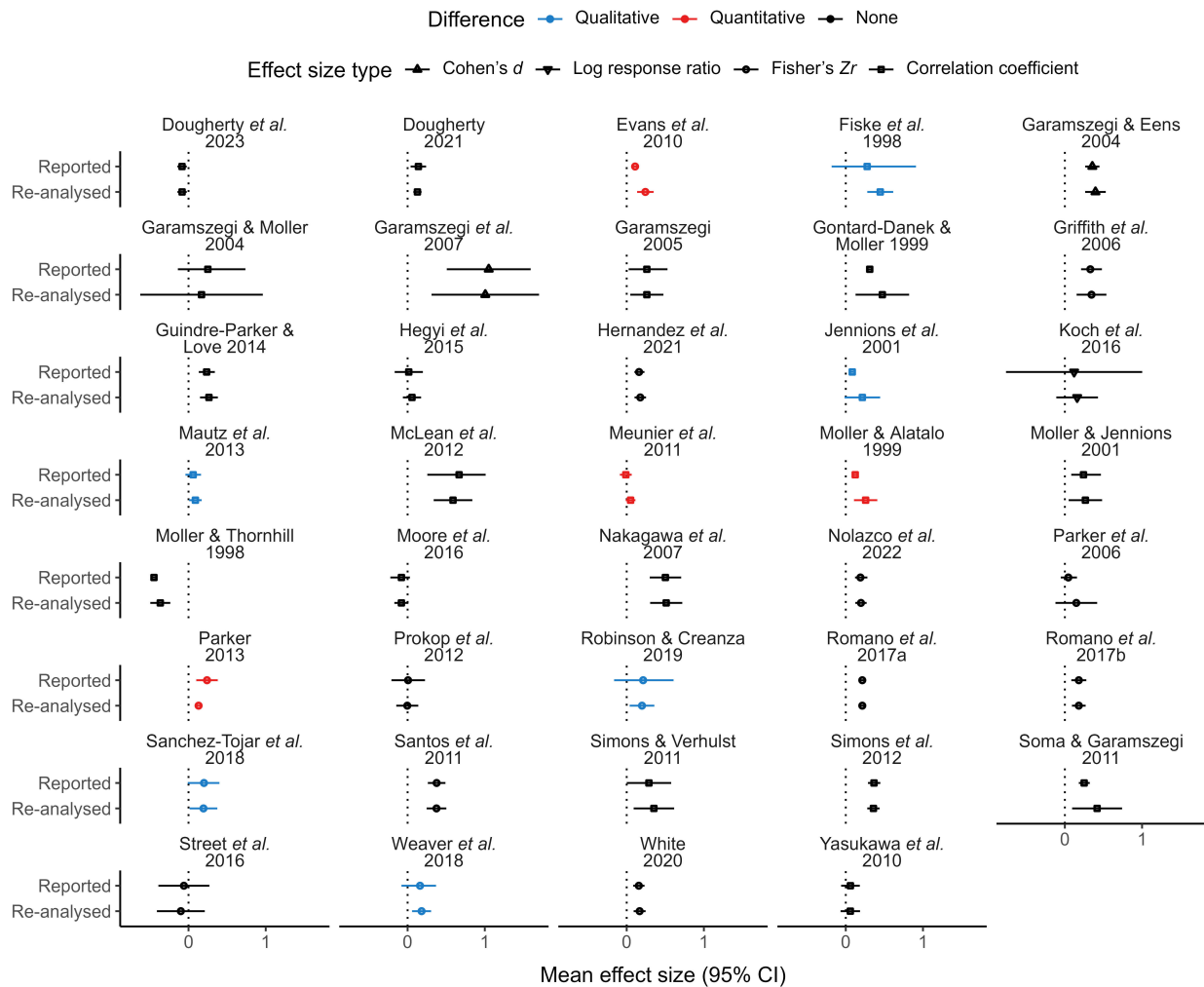


Figure 1. Originally reported and re-analysed mean effect sizes of 39 meta-analyses. Qualitative differences represent a change in interpretation between pairs (positive versus not different from zero or vice-versa), while quantitative differences represent statistical differences (absolute z-score greater than 1.96).

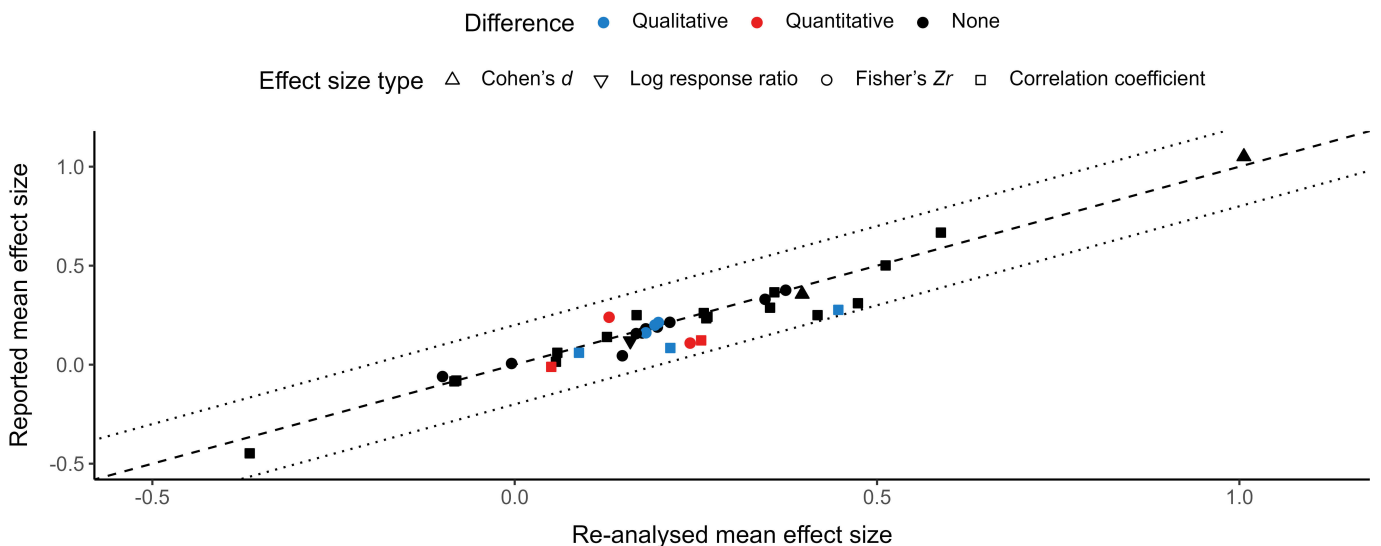


Figure 2. Comparison between mean effect sizes originally reported by meta-analyses and mean effect sizes resulting from the re-analysis of the reported dataset of the same meta-analyses. The dashed line highlights a perfect correspondence between variables, while dotted lines highlight a difference of 0.2 and -0.2 between reported and re-analysed mean effect sizes.

verified primary studies increased the occurrence of both qualitative differences (20%, 3 out of 15) and quantitative differences (26.7%, 4 out of 15; figure 6; but see the electronic supplementary material, figure S6). Nonetheless, differences in magnitude between mean effect sizes were often small (electronic supplementary material, figure S7). Furthermore, heterogeneity varied only slightly among fitted models (with perhaps one exception; electronic supplementary material, figure S8).

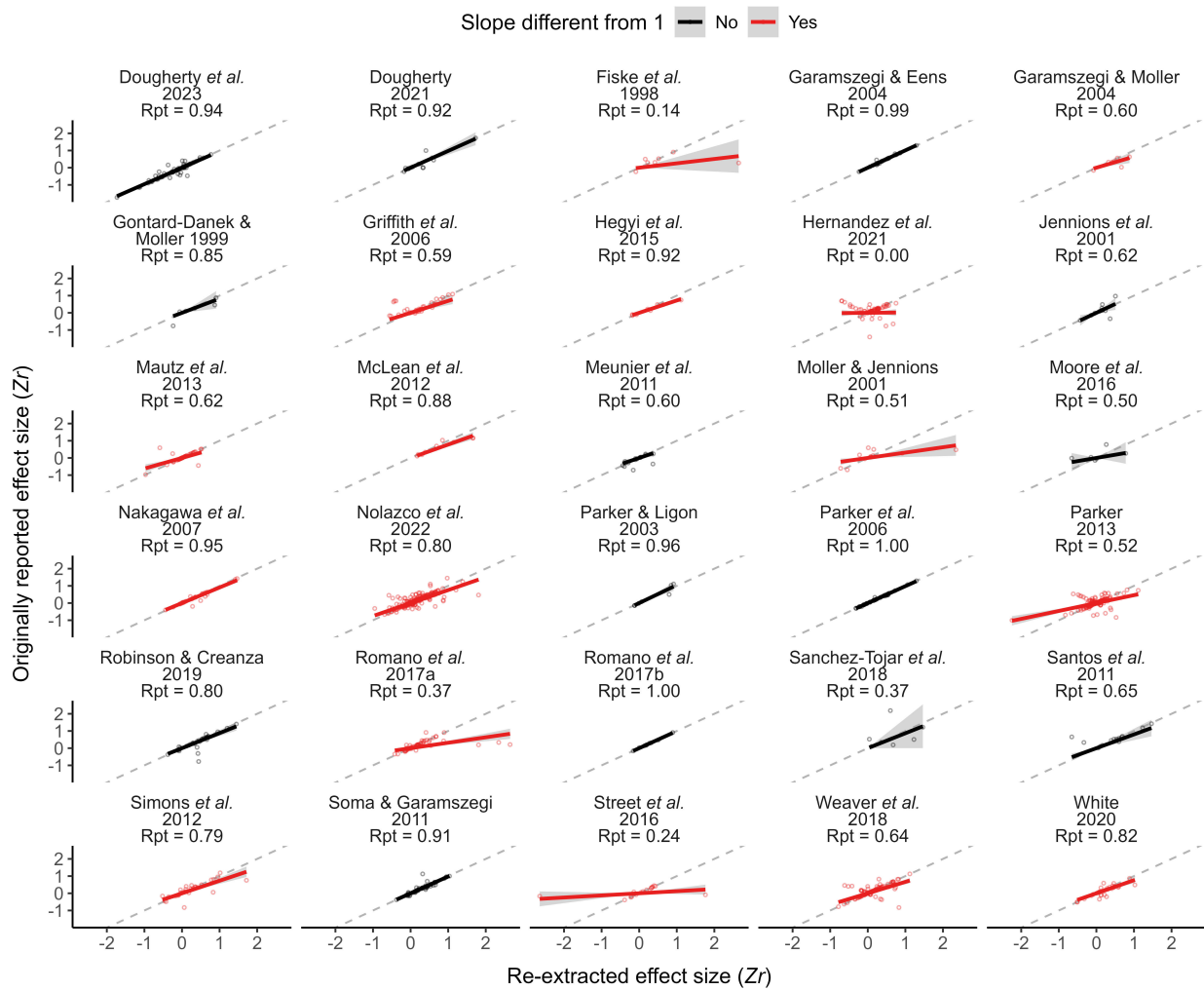


Figure 3. Relationship between effect sizes originally reported by meta-analyses and effect sizes re-extracted from primary studies. Solid lines represent a linear regression between reported and re-extracted effect sizes (with the intercept forced to zero), while shaded areas represent their 95% confidence interval. Red lines indicate slopes that significantly differ from 1. Dashed lines highlight a perfect relationship between reported and re-extracted effect sizes. ‘Rpt’ stands for repeatability of individual effect sizes.

4. Discussion

In the first case study of the reliability of meta-analyses in ecology and evolution, we assessed a sample of meta-analyses that focused on the relationship between sexual signals and various measurements relevant to sexual selection. First, we were able to precisely replicate meta-analyses’ reported mean effect sizes in most but not all cases. Second, we found that effect sizes reported by meta-analyses and those we re-extracted did not overlap in multiple cases (i.e. the slope from linear regressions between reported and re-extracted effect sizes was statistically different from 1). Third, we observed that some relevant data, both from included and undetected studies, were missing from certain meta-analyses. Fourth, incorporating such missed relevant data into further analyses, in addition to replacing reported data from verified primary studies with re-extracted data, sometimes produced distinct meta-analytical results. However, the discrepancies we found tended to be small in magnitude and rarely generated distinct interpretations from the ones originally reported. Thus, we argue that the meta-analyses we evaluated are fairly reliable, although this is subjective to the degree of accuracy expected. Below we discuss the impacts of our findings and provide recommendations to further improve the reliability of meta-analyses in ecology and evolution.

Most of our results are based on comparisons between mean effect sizes (figures 1, 5 and 6) using categorical interpretations (i.e. negative, not different from zero or positive; indicating qualitative differences) and statistical calculations (indicating quantitative differences). We detected qualitative differences in only a handful of these comparisons, most of them resulting from differences in the width of mean effect sizes’ CIs rather than in their value. Similarly, we detected few quantitative differences when comparing mean effect sizes that ideally would have remained identical across analyses. Differences between re-analysed and reported mean effect sizes regarding the width of their CI can be explained by distinct analytical decisions from those originally conducted (e.g. inclusion of additional random factors). On the other hand, differences in mean effect size estimates could have also stemmed from errors made by the original authors in data analysis or result reporting. Regardless of the origin of the differences observed, these discrepancies are only relevant if they impact the interpretation of results from original meta-analyses. Ecology and evolution researchers rarely mention the magnitude of mean effect sizes from meta-analyses, focusing their attention on the existence or absence of effects or relationships [78]. This means that quantitative differences alone would probably be unimportant for most researchers (unless hypotheses rely on effect size comparisons, as

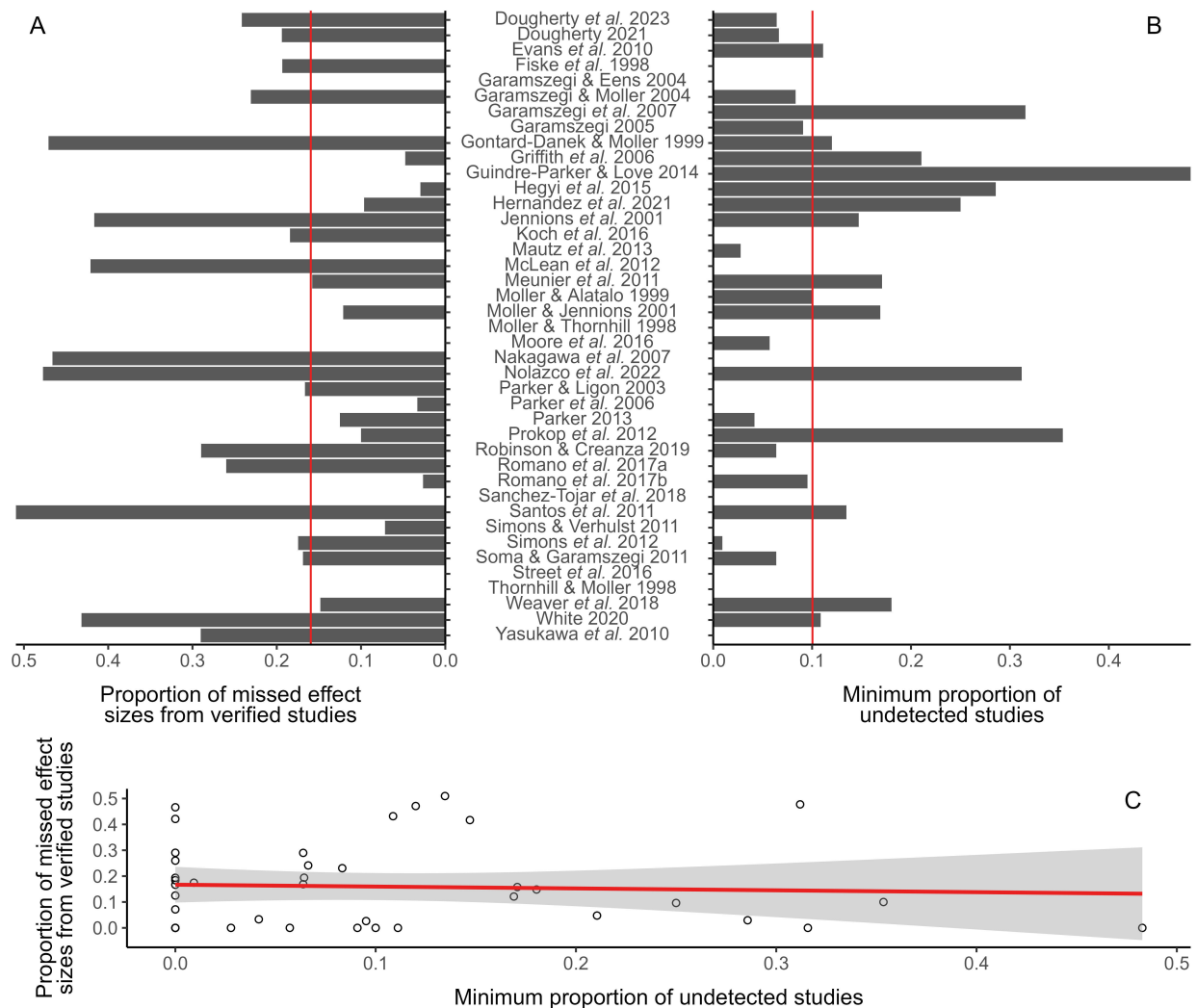


Figure 4. Proportion of missed effect sizes from all relevant effect sizes re-extracted from verified primary studies (A), proportion of undetected studies from the minimum number of primary studies that each meta-analysis should have included (B), and the relationship between these variables (C). Vertical red lines in panels A and B represent the average proportion of missed effect sizes from all relevant re-extracted effect sizes and the minimum proportion of undetected studies, respectively. The red line in panel C represents the fit of linear regression between variables with its 95% CI as the shaded area.

in [13]), while qualitative differences would affect how results are perceived. However, even though interpreting results based on the inclusion of zero by CIs is a common practice, it can generate misleading conclusions [79]. For instance, the lower confidence bound for the mean effect size of Sánchez-Tójar *et al.* [49] was reported as -0.01 but was slightly above zero in all our analyses, while the magnitude of this study's mean effect size was reported as 0.2 and ranged from 0.19 to 0.21 in our analyses (figures 2 and 5). Therefore, we conclude that results from Sánchez-Tójar *et al.* [49], as well as those from meta-analyses with similar minute discrepancies, should be deemed replicable despite being different to the ones we obtained when comparing the inclusion of zero by CIs. Importantly, statistical definitions of replicability and reproducibility are highly debatable, meaning that conclusions based on result comparisons become inevitably subjective to a certain degree [80,81].

Some effect sizes reported in meta-analyses were very different from the ones we re-extracted (figure 3). A portion of these discrepancies could be explained by the use of distinct effect size calculations, the extraction of data from distinct sources within primary studies (e.g. raw data from a figure versus statistics reported in text), or even the erroneous use of standard error instead of standard deviation in effect size calculations [82]. Yet, we noticed that numerous effect sizes reported were similar in value but opposite in direction to the ones we re-extracted. Although this did not strongly affect differences among estimated mean effect sizes, effect sizes with the wrong direction can be particularly dangerous as they should be more impactful than simply imprecise ones, at least for large effect sizes. Furthermore, errors in data extraction or effect size calculation can especially affect meta-regressions, which are often performed with subsets of meta-analytical datasets. This emphasizes the importance in meta-analyses of (i) establishing a coherent rationale to ascertain direction of effect sizes, (ii) reporting directionality decisions in detail, and (iii) remaining vigilant during data extraction. Additionally, we recommend cross-checking data extractions (i.e. independent verification of the data by someone who did not extract them) to increase the chances of spotting and correcting mistakes, including those related to the direction of effect sizes.

We observed that not all relevant data from verified primary studies were used by meta-analyses (figure 4). Although we cannot ascertain how or why meta-analyses' authors missed relevant data from studies they extracted data from, the reason that some primary studies remained undetected might be more easily explained. While it is possible that meta-analyses could have employed suboptimal searches, primary studies commonly neglect to inform their entire scope and results in their title, abstract and keywords, which are used to retrieve and screen studies [83]. Thus, even though authors of meta-analyses should follow

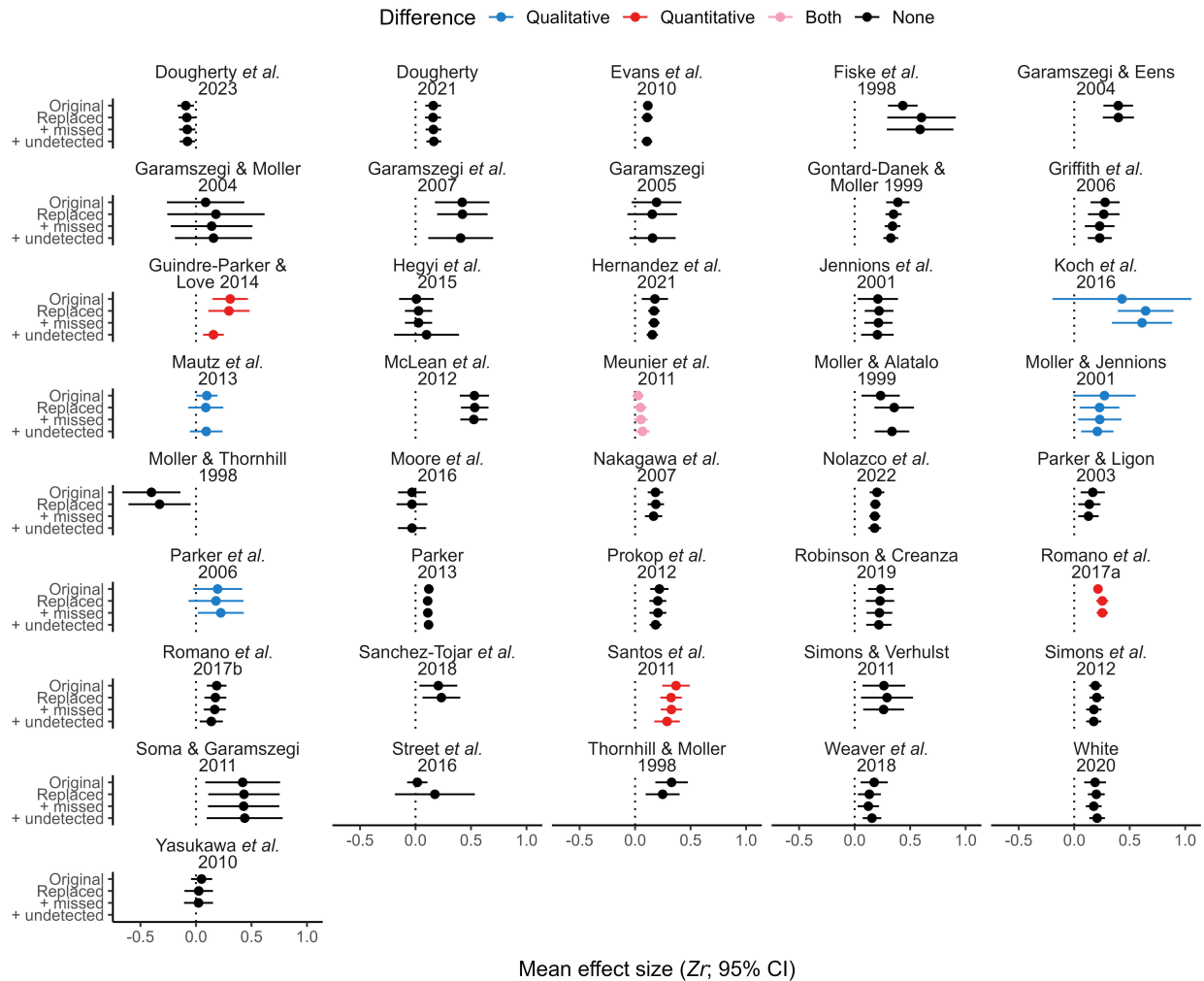


Figure 5. Mean effect sizes from up to four distinct meta-analytical models for each of 41 meta-analyses, using data from all primary studies (both verified and unverified by us, see details in text). Dotted lines highlight zero. Comparisons were made between the first and the last result shown within each subplot, with qualitative differences representing a change in interpretation (positive versus not different from zero or vice-versa) and quantitative differences representing statistical differences (absolute z-score greater than 1.96).

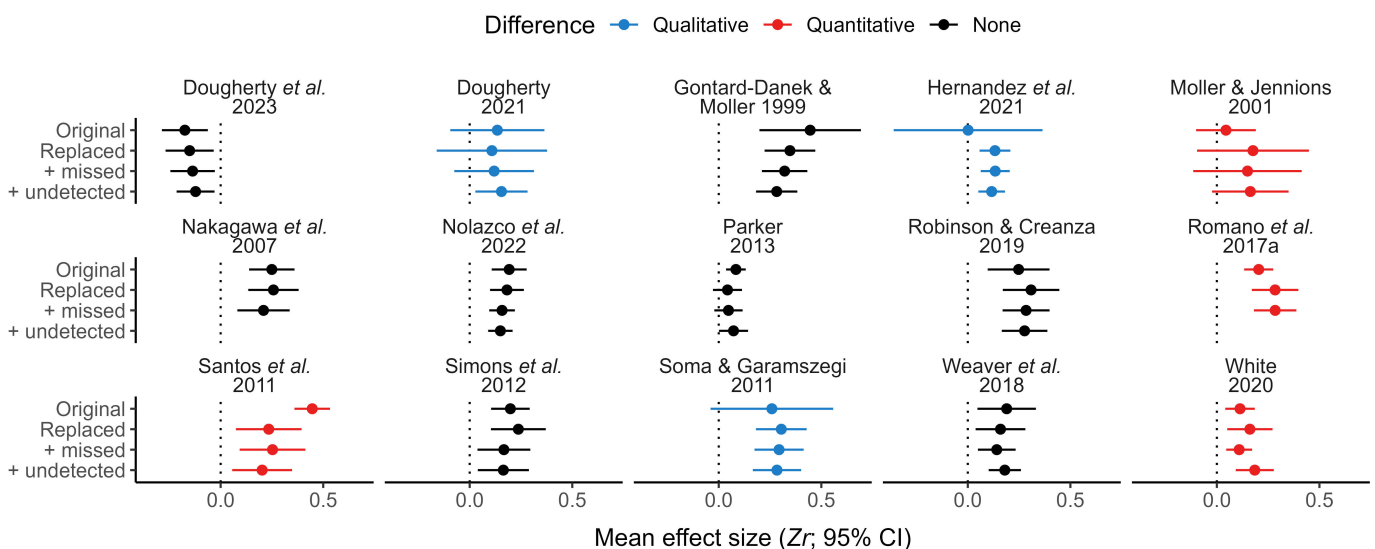


Figure 6. Mean effect sizes from up to four distinct meta-analytical models for each of 15 meta-analyses, using only data from primary studies that we verified (see details in text). Dotted lines highlight zero. Comparisons were made between the first and the last result shown within each subplot, with qualitative differences representing a change in interpretation (positive versus not different from zero or vice-versa) and quantitative differences representing statistical differences (absolute z-score > 1.96).

certain guidelines to build effective searches (e.g. [84]), meta-analysts may unfortunately miss relevant studies despite their best efforts. Conversely, authors of primary studies should be mindful of how search engines work, crafting their title, abstract and keywords to enhance the findability of their work [83,85].

The goal of our study was to evaluate the reliability of meta-analyses beyond transparency, yet our results might be affected by transparency issues found in the meta-analyses we assessed. First, matching reported and re-extracted data points proved to be a difficult task because meta-analyses usually lack details on their extracted data. For instance, the sexual signal and proxy for each data point were only vaguely described in most meta-analyses. Furthermore, no meta-analysis in our dataset reported the location of the extracted data within primary studies (e.g. which page, table and figure). Second, meta-analyses were often poorly transparent with their inclusion criteria. We tried to comprehend vague information, incorporate omitted criteria and ignore contradictions when matching re-extracted to reported datasets (see the electronic supplementary material, S5). Nonetheless, our decisions might have affected the amount of missing data and of undetected studies by each meta-analysis. Ultimately, this could have influenced our findings related to the reproducibility of mean effect sizes.

In addition to the recommendations we already mentioned (e.g. cross-checking of data extracted), we urge meta-analysts to provide all possible details on the data they collect. For instance, authors must clarify which exact measurements they sought instead of simply mentioning umbrella terms (e.g. condition-dependence by [28]). The location of the information is also crucial: readers should not have to examine datasets to find important details. Instead, summarized details should be in the manuscript or, less preferably, in the supplementary material, but always in a readable format (figures, simplified tables, in text, not in spreadsheets). Furthermore, we reiterate recommendations by Ivimey-Cook *et al.* [5], such as providing the within-text source for each data point extracted and the equations used to calculate effect sizes (along with assumptions and transformations used). We highlight that multiple factors can reduce the replicability of meta-analyses, such as large researcher degrees of freedom (i.e. decisions made by authors; *sensu* [86]), unclear or vague hypotheses, mistakes in data extraction, reporting errors or outright fraud. It is possible that meta-analyses in ecology and evolution can be particularly prone to suffer from some of these problems (e.g. large researcher degrees of freedom), given that researchers in this field explore broad questions with diverse taxa. Still, we argue that most of these issues can be improved, corrected, or at least detected by following our recommendations, summarized in the electronic supplementary material, table S5.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. All data and code used in this study are available at [87].

Supplementary material is available online [88].

Declaration of AI use. The authors declare that they occasionally used GPT-4-turbo (OpenAI) to improve the clarity and readability of this work. After using these tools, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Authors' contributions. P.P.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, validation, visualization, writing—original draft, writing—review and editing; M.L.: conceptualization, investigation, methodology, writing—review and editing; R.C.M.R.: conceptualization, investigation, methodology, writing—review and editing; A.M.: conceptualization, investigation, methodology, writing—review and editing; Y.Y.: conceptualization, investigation, methodology, writing—review and editing; S.N.: conceptualization, funding acquisition, investigation, methodology, supervision, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. P.P., M.L., Y.Y., and S.N. were supported by ARC (Australian Research Council) Discovery Project grants (DP210100812 and DP230101248). S.N. was also supported by the Canada Excellence Research Chair Program (CERC-2022-00074). R.C.M.R. was supported by a grant from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) as part of the Programa de Desenvolvimento da Pós-Graduação (PDGP 88887.803974/2023-00).

Acknowledgements. We thank László Zolt Garamszegi, Sarah Guindre-Parker, América Hernández Martínez, and Tim Parker for sending us data from their meta-analyses. We also thank Yong Zhi Foo for contributing to the conceptualization of this project and the editor, Tim Parker, Alfredo Sánchez-Tójar, and an anonymous reviewer for feedback on the manuscript.

References

- McCallen E, Knott J, Nunez-Mir G, Taylor B, Jo I, Fei S. 2019 Trends in ecology: shifts in ecological research themes over the past four decades. *Front. Ecol. Environ.* **17**, 109–116. (doi:10.1002/fee.1993)
- Nunez-Mir GC, Iannone BV III, Pijanowski BC, Kong N, Fei S. 2016 Automated content analysis: addressing the big literature challenge in ecology and evolution. *Methods Ecol. Evol.* **7**, 1262–1272. (doi:10.1111/2041-210X.12602)
- Koricheva J, Gurevitch J, Mengersen K. 2013 *Handbook of meta-analysis in ecology and evolution*. Princeton, NJ: Princeton University Press. (doi:10.1515/9781400846184)
- Borenstein M, Hedges LV, Higgins J, Rothstein H. 2021 *Introduction to meta-analysis*. Hoboken, NJ: John Wiley & Sons.
- Ivimey-Cook ER, Noble DWA, Nakagawa S, Lajeunesse MJ, Pick JL. 2023 Advice for improving the reproducibility of data extraction in meta-analysis. *Res. Synth. Methods* **14**, 911–915. (doi:10.1002/jrsm.1663)
- O'Dea RE *et al.* 2021 Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. *Biol. Rev.* **96**, 1695–1722. (doi:10.1111/brv.12721)
- Pollo P, Lagisz M, Yang Y, Culina A, Nakagawa S. 2024 Synthesis of sexual selection: a systematic map of meta-analyses with bibliometric analysis. *Biol. Rev.* **99**, 2134–2175. (doi:10.1111/brv.13117)
- Stroup D *et al.* 2000 Meta-analysis of observational studies in epidemiology. *J. Am. Med. Assoc.* **283**, 2008. (doi:10.1001/jama.283.15.2008)
- Page MJ *et al.* 2021 The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Br. Med. J.* **372**, n71. (doi:10.1136/bmj.n71)
- Kambouris S, Wilkinson DP, Smith ET, Fidler F. 2024 Computationally reproducing results from meta-analyses in ecology and evolutionary biology using shared code and data. *PLoS ONE* **19**, e0300333. (doi:10.1371/journal.pone.0300333)
- Jones BC, DuVal EH. 2019 Mechanisms of social influence: a meta-analysis of the effects of social information on female mate choice decisions. *Front. Ecol. Evol.* **7**, 1–14. (doi:10.3389/fevo.2019.00390)

12. Davies AD, Lewis Z, Dougherty LR. 2020 A meta-analysis of factors influencing the strength of mate-choice copying in animals. *Behav. Ecol.* **31**, 1279–1290. (doi:10.1093/beheco/araa064)
13. Pollo P, Nakagawa S, Kasumovic MM. 2022 The better, the choosier: a meta-analysis on interindividual variation of male mate choice. *Ecol. Lett.* **25**, 1305–1322. (doi:10.1111/ele.13981)
14. Dougherty LR. 2023 The effect of individual state on the strength of mate choice in females and males. *Behav. Ecol.* **34**, 197–209. (doi:10.1093/beheco/arac100)
15. Stubblefield JW, Orzack SH. 2013 Resource transfers and evolution: helpful offspring and sex allocation. *Theor. Popul. Biol.* **83**, 64–81. (doi:10.1016/j.tpb.2012.09.004)
16. Page MJ *et al.* 2021 The REPRISE project: protocol for an evaluation of reproducibility and replicability in syntheses of evidence. *Syst. Rev.* **10**, 112. (doi:10.1186/s13643-021-01670-0)
17. Maassen E, van Assen MALM, Nuijten MB, Olsson-Collentine A, Wicherts JM. 2020 Reproducibility of individual effect sizes in meta-analyses in psychology. *PLoS ONE* **15**, 1–18. e0233107. (doi:10.1371/journal.pone.0233107)
18. Dougherty LR, Rovenolt F, Luyet A, Jokela J, Stephenson JF. 2023 Ornaments indicate parasite load only if they are dynamic or parasites are contagious. *Evol. Lett.* **7**, 176–190. (doi:10.1093/evlett/qr4017)
19. Dougherty LR. 2021 Meta-analysis reveals that animal sexual signalling behaviour is honest and resource based. *Nat. Ecol. Evol.* **5**, 688–699. (doi:10.1038/s41559-021-01409-z)
20. Evans SR, Hinks AE, Wilkin TA, Sheldon BC. 2010 Age, sex and beauty: methodological dependence of age- and sex-dichromatism in the great tit *Parus major*. *Biol. J. Linn. Soc.* **101**, 777–796. (doi:10.1111/j.1095-8312.2010.01548.x)
21. Fiske P, Rintamäki PT, Karvonen E. 1998 Mating success in lekking males: a meta-analysis. *Behav. Ecol.* **9**, 328–338. (doi:10.1093/beheco/9.4.328)
22. Garamszegi LZ, Eens M. 2004 Brain space for a learned task: strong intraspecific evidence for neural correlates of singing behavior in songbirds. *Brain Res. Rev.* **44**, 187–193. (doi:10.1016/j.brainresrev.2003.12.001)
23. Garamszegi LZ, Møller AP. 2004 Extrajoint paternity and the evolution of bird song. *Behav. Ecol.* **15**, 508–519. (doi:10.1093/beheco/arh041)
24. Garamszegi LZ, Török J, Hegyi G, Szöllösi E, Rosivall B, Eens M. 2007 Age-dependent expression of song in the collared flycatcher, *Ficedula albicollis*. *Ethology* **113**, 246–256. (doi:10.1111/j.1439-0310.2007.01337.x)
25. Garamszegi LZ. 2005 Bird song and parasites. *Behav. Ecol. Sociobiol.* **59**, 167–180. (doi:10.1007/s00265-005-0041-8)
26. Gontard-Danek MC, Møller AP. 1999 The strength of sexual selection: a meta-analysis of bird studies. *Behav. Ecol.* **10**, 476–486. (doi:10.1093/beheco/10.5.476)
27. Griffith SC, Parker TH, Olson VA. 2006 Melanin- versus carotenoid-based sexual signals: is the difference really so black and red? *Anim. Behav.* **71**, 749–763. (doi:10.1016/j.anbehav.2005.07.016)
28. Guindre-Parker S, Love OP. 2014 Revisiting the condition-dependence of melanin-based plumage. *J. Avian Biol.* **45**, 29–33. (doi:10.1111/j.1600-048x.2013.00190.x)
29. Hegyi G, Kötél D, Laczi M. 2015 Direct benefits of mate choice: a meta-analysis of plumage colour and offspring feeding rates in birds. *Sci. Nat.* **102**, 62. (doi:10.1007/s00114-015-1311-6)
30. Hernández A, Martínez-Gómez M, Beamonte-Barrientos R, Montoya B. 2021 Colourful traits in female birds relate to individual condition, reproductive performance and male-mate preferences: a meta-analytic approach. *Biol. Lett.* **17**, 20210283. (doi:10.1098/rsbl.2021.0283)
31. Jennions MD, Møller AP, Petrie M. 2001 Sexually selected traits and adult survival: a meta-analysis. *Q. Rev. Biol.* **76**, 3–36. (doi:10.1086/393743)
32. Koch RE, Wilson AE, Hill GE. 2016 The importance of carotenoid dose in supplementation studies with songbirds. *Physiol. Biochem. Zool.* **89**, 61–71. (doi:10.1086/684485)
33. Mautz BS, Møller AP, Jennions MD. 2013 Do male secondary sexual characters signal ejaculate quality? A meta-analysis. *Biol. Rev.* **88**, 669–682. (doi:10.1111/brv.12022)
34. McLean MJ, Bishop PJ, Nakagawa S. 2012 Male quality, signal reliability and female choice: assessing the expectations of inter-sexual selection. *J. Evol. Biol.* **25**, 1513–1520. (doi:10.1111/j.1420-9101.2012.02533.x)
35. Meunier J, Figueiredo Pinto S, Burri R, Roulin A. 2011 Eumelanin-based coloration and fitness parameters in birds: a meta-analysis. *Behav. Ecol. Sociobiol.* **65**, 559–567. (doi:10.1007/s00265-010-1092-z)
36. Møller AP, Alatalo RV. 1999 Good-genes effects in sexual selection. *Proc. R. Soc. B* **266**, 85–91. (doi:10.1098/rsob.1999.0607)
37. Møller A, Jennions M. 2001 How important are direct fitness benefits of sexual selection? *Naturwissenschaften* **88**, 401–415. (doi:10.1007/s001140100255)
38. Møller AP, Thornhill R. 1998 Bilateral symmetry and sexual selection: a meta-analysis. *Am. Nat.* **151**, 174–192. (doi:10.1086/286110)
39. Moore FR, Shuker DM, Dougherty L. 2016 Stress and sexual signaling: a systematic review and meta-analysis. *Behav. Ecol.* **27**, 363–371. (doi:10.1093/beheco/arv195)
40. Nakagawa S, Ockendon N, Gillespie DOS, Hatchwell BJ, Burke T. 2011 Assessing the function of house sparrows' bib size using a flexible meta-analysis method. *Behav. Ecol.* **18**, 831–840. (doi:10.1093/beheco/arm050)
41. Nolazco S, Delhey K, Nakagawa S, Peters A. 2022 Ornaments are equally informative in male and female birds. *Nat. Commun.* **13**, 5917. (doi:10.1038/s41467-022-33548-7)
42. Parker TH, Ligon JD. 2003 Female mating preferences in red junglefowl: a meta-analysis. *Ethol. Ecol. Evol.* **15**, 63–72. (doi:10.1080/08927014.2003.9522691)
43. Parker TH. 2013 What do we really know about the signalling role of plumage colour in blue tits? A case study of impediments to progress in evolutionary biology. *Biol. Rev.* **88**, 511–536. (doi:10.1111/brv.12013)
44. Parker TH, Barr IR, Griffith SC. 2006 The blue tit's song is an inconsistent signal of male condition. *Behav. Ecol.* **17**, 1029–1040. (doi:10.1093/beheco/arh041)
45. Prokop ZM, Michalczyk Ł, Drobniak SM, Herdegen M, Radwan J. 2012 Meta-analysis suggests choosy females get sexy sons more than 'good genes'. *Evolution* **66**, 2665–2673. (doi:10.1111/j.1558-5646.2012.01654.x)
46. Robinson CM, Creanza N. 2019 Species-level repertoire size predicts a correlation between individual song elaboration and reproductive success. *Ecol. Evol.* **9**, 8362–8377. (doi:10.1002/ece3.5418)
47. Romano A, Costanzo A, Rubolini D, Saino N, Møller AP. 2017 Geographical and seasonal variation in the intensity of sexual selection in the barn swallow *Hirundo rustica*: a meta-analysis. *Biol. Rev.* **92**, 1582–1600. (doi:10.1111/brv.12297)
48. Romano A, Saino N, Møller AP. 2017 Viability and expression of sexual ornaments in the barn swallow *Hirundo rustica*: a meta-analysis. *J. Evol. Biol.* **30**, 1929–1935. (doi:10.1111/jeb.13151)
49. Sánchez-Tójar A *et al.* 2018 Meta-analysis challenges a textbook example of status signalling and demonstrates publication bias. *eLife* **7**, 1–26. 37385. (doi:10.7554/elife.37385)
50. Santos ESA, Scheck D, Nakagawa S. 2011 Dominance and plumage traits: meta-analysis and metaregression analysis. *Anim. Behav.* **82**, 3–19. (doi:10.1016/j.anbehav.2011.03.022)
51. Simons MJP, Verhulst S. 2011 Zebra finch females prefer males with redder bills independent of song rate—a meta-analysis. *Behav. Ecol.* **22**, 755–762. (doi:10.1093/beheco/arr043)
52. Simons MJP, Cohen AA, Verhulst S. 2012 What does carotenoid-dependent coloration tell? Plasma carotenoid level signals immunocompetence and oxidative stress state in birds—a meta-analysis. *PLoS ONE* **7**, e43088. (doi:10.1371/journal.pone.0043088)

53. Soma M, Garamszegi LZ. 2011 Rethinking birdsong evolution: meta-analysis of the relationship between song complexity and reproductive success. *Behav. Ecol.* **22**, 363–371. (doi:10.1093/beheco/arq219)
54. Street SE, Cross CP, Brown GR. 2016 Exaggerated sexual swellings in female nonhuman primates are reliable signals of female fertility and body condition. *Anim. Behav.* **112**, 203–212. (doi:10.1016/j.anbehav.2015.11.023)
55. Thornhill R. 1998 The relative importance of size and asymmetry in sexual selection. *Behav. Ecol.* **9**, 546–551. (doi:10.1093/beheco/9.6.546)
56. Weaver RJ, Santos ESA, Tucker AM, Wilson AE, Hill GE. 2018 Carotenoid metabolism strengthens the link between feather coloration and individual quality. *Nat. Commun.* **9**, 73. (doi:10.1038/s41467-017-02649-z)
57. White TE. 2020 Structural colours reflect individual quality: a meta-analysis. *Biol. Lett.* **16**, 20200001. (doi:10.1098/rsbl.2020.0001)
58. Yasukawa K, Enstrom DA, Parker PG, Jones TC. 2010 Male red-winged blackbirds with experimentally dulled epaulets experience no disadvantage in sexual selection. *J. Field Ornithol.* **81**, 31–41. (doi:10.1111/j.1557-9263.2009.00258.x)
59. Pollo P, Lagisz M, Macedo-Rego R, Mizuno A, Yang Y, Nakagawa S. 2024 Synthesis of nature's extravaganzas: an augmented meta-meta-analysis on (putative) sexual signals. *EcoEvoRxiv* (doi:10.32942/X2F045)
60. Pollo P, Foo YZ, Lagisz M, Macedo-Rego RC, Mizuno A, Yang Y, Nakagawa S. 2024 Pre-registration for a methodological evaluation and a meta-meta-analysis of sexual signals. (doi:10.17605/OSF.IO/HR2YK)
61. Nakagawa S *et al.* 2023 Method reporting with initials for transparency (MeRiT) promotes more granularity and accountability for author contributions. *Nat. Commun.* **14**, 1788. (doi:10.1038/s41467-023-37039-1)
62. McNutt MK *et al.* 2018 Transparency in authors' contributions and responsibilities to promote integrity in scientific publication. *Proc. Natl Acad. Sci. USA* **115**, 2557–2560. (doi:10.1073/pnas.1715374115)
63. Parker TH, Greig EI, Nakagawa S, Parra M, Dalisio AC. 2018 Subspecies status and methods explain strength of response to local versus foreign song by oscine birds in meta-analysis. *Anim. Behav.* **142**, 1–17. (doi:10.1016/j.anbehav.2018.05.023)
64. White TE, Latty T, Umbers KDL. 2022 The exploitation of sexual signals by predators: a meta-analysis. *Proc. R. Soc. B* **289**, 20220444. (doi:10.1098/rspb.2022.0444)
65. Parris KM, McCarthy MA. 2013 Predicting the effect of urban noise on the active space of avian vocal signals. *Am. Nat.* **182**, 452–464. (doi:10.1086/671906)
66. Moller AP, Christe P, Lux E. 1999 Parasitism, host immune function, and sexual selection. *Q. Rev. Biol.* **74**, 3–20. (doi:10.1086/392949)
67. Hamilton WJ, Poulin R. 1997 The Hamilton and Zuk hypothesis revisited: a meta-analytical approach. *Behaviour* **134**, 299–320. (doi:10.1163/156853997x00485)
68. Leung B, Forbes MR. 1996 Fluctuating asymmetry in relation to stress and fitness: effects of trait type as revealed by meta-analysis. *Écoscience* **3**, 400–413. (doi:10.1080/11956860.1996.11682357)
69. Macedo-Rego R, Jennions M, Santos E. 2020 Larger females and larger males produce more offspring, regardless of the social mating system: a meta-analysis on animals. In *When and how do episodes of sexual selection occur in animal species? A meta-analytical investigation* (ed R Macedo-Rego), pp. 148–187. São Paulo, Brazil: Universidade de São Paulo.
70. Pick JL, Nakagawa S, Noble DWA. 2019 Reproducible, flexible and high-throughput data extraction from primary literature: the metaDigitise R package. *Methods Ecol. Evol.* **10**, 426–431. (doi:10.1111/2041-210x.13118)
71. Nakagawa S, Lagisz M, Jennions MD, Koricheva J, Noble DWA, Parker TH, Sánchez-Tójar A, Yang Y, O'Dea RE. 2022 Methods for testing publication bias in ecological and evolutionary meta-analyses. *Methods Ecol. Evol.* **13**, 4–21. (doi:10.1111/2041-210x.13724)
72. R Core Team. 2024 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
73. Viechtbauer W. 2010 Conducting meta-analyses in R with the metafor. *J. Stat. Softw.* **36**, 1–48. (doi:10.18637/jss.v036.i03)
74. Nakagawa S, Santos ESA. 2012 Methodological issues and advances in biological meta-analysis. *Evol. Ecol.* **26**, 1253–1274. (doi:10.1007/s10682-012-9555-5)
75. Paradis E, Schliep K. 2019 ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528. (doi:10.1093/bioinformatics/bty633)
76. Michonneau F, Brown JW, Winter DJ. 2016 rotl: an R package to interact with the Open Tree of Life data. *Methods Ecol. Evol.* **7**, 1476–1481. (doi:10.1111/2041-210x.12593)
77. Stoffel MA, Nakagawa S, Schielzeth H. 2017 rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol. Evol.* **8**, 1639–1644. (doi:10.1111/2041-210x.12797)
78. Pollo P *et al.* 2025 Harnessing meta-analyses' insights in ecology and evolution research. *EcoEvoRxiv*. (doi:10.32942/X2PW5P)
79. Amrhein V, Greenland S, McShane B. 2019 Scientists rise up against statistical significance. *Nature* **567**, 305–307. (doi:10.1038/d41586-019-00857-9)
80. Hedges LV, Schauer JM. 2019 More than one replication study is needed for unambiguous tests of replication. *J. Educ. Behav. Stat.* **44**, 543–570. (doi:10.3102/1076998619852953)
81. Schauer JM, Hedges LV. 2020 Assessing heterogeneity and power in replications of psychological experiments. *Psychol. Bull.* **146**, 701–719. (doi:10.1037/bul0000232)
82. Kadlec D, Sainani K, Nimphius S. 2023 With great power comes great responsibility: common errors in meta-analyses and meta-regressions in strength & conditioning research. *Sport Med.* **53**, 313–325. (doi:10.1007/s40279-022-01766-0)
83. Pottier P *et al.* 2024 Title, abstract and keywords: a practical guide to maximize the visibility and impact of academic papers. *Proc. R. Soc. B* **291**, 20241222. (doi:10.1098/rspb.2024.1222)
84. Foo YZ, O'Dea RE, Koricheva J, Nakagawa S, Lagisz M. 2021 A practical guide to question formation, systematic searching and study screening for literature reviews in ecology and evolution. *Methods Ecol. Evol.* **12**, 1705–1720. (doi:10.1111/2041-210x.13654)
85. Schilhan L, Kaier C, Lackner K. 2021 Increasing visibility and discoverability of scholarly publications with academic search engine optimization. *Insights UKSG J.* **34**, 1–16. (doi:10.1629/uksg.534)
86. Simmons JP, Nelson LD, Simonsohn U. 2011 False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22**, 1359–1366. (doi:10.1177/0956797611417632)
87. Pollo P, Lagisz M, de Macedo Rego RC, Mizuno A, Yang Y, Nakagawa S. 2025 Supplementary material for a methodological evaluation and a meta-meta-analysis of sexual signals. See <https://doi.org/10.17605/OSF.IO/6NJEM>.
88. Pollo P, Lagisz M, Macedo-Rego RC, Mizuno A, Yang Y, Nakagawa S. 2025 Supplementary material from: Reliability of meta-analyses in ecology and evolution: (mostly) good news from a case study on sexual signals. Figshare. (doi:10.6084/m9.figshare.c.7772201)