# Publication bias impacts on effect size, statistical power, and magnitude (Type M) and sign (Type S) errors in ecology and evolutionary biology

Yefeng Yang[1,2,3*], Alfredo Sánchez-Tójar[4], Rose E. O'Dea[5], Daniel W.A. Noble[6], Julia Koricheva[7], Michael D. Jennions[6], Timothy H. Parker[8], Malgorzata Lagisz[1,9], Shinichi Nakagawa[1,9*]

[1] Evolution & Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia

[2] Department of Biosystems Engineering, Zhejiang University, Hangzhou 310058, China

[3] Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, China

[4] Department of Evolutionary Biology, Bielefeld University, Bielefeld, 33615, Germany

[5] School of Ecosystem and Forest Sciences, University of Melbourne, Parkville, Australia

[6] Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, ACT, Australia

[7] Department of Biological Sciences, Royal Holloway University of London, Egham, Surrey, TW20 0EX, U.K.

[8] Department of Biology, Whitman College, Walla Walla, WA 99362, U.S.A

[9] Equal contribution as senior authors.

*Correspondence: Y. Yang, e-mail: yefeng.yang1@unsw.edu.au; S. Nakagawa, e-mail: s.nakagawa@unsw.edu.au

**ORCID**

Yefeng Yang: 0000-0002-8610-4016; Alfredo Sánchez-Tójar: 0000-0002-2886-0649; Rose E. O'Dea: 0000-0001-8177-5075; Daniel W. A. Noble: 0000-0001-9460-8743; Julia Koricheva: 0000-0002-9033-0171; Michael D. Jennions: 0000-0001-9221-2788; Timothy H. Parker: 0000-0003-2995-5284; Malgorzata Lagisz: 0000-0002-3993-6127; Shinichi Nakagawa: 0000-0002-7765-5182

## Abstract

Collaborative assessments of direct replicability of empirical studies in the medical and social sciences have exposed alarmingly low rates of replicability, a phenomenon dubbed the ‘replication crisis’. Poor replicability has spurred cultural changes targeted at improving reliability in these disciplines. Given the absence of equivalent replication projects in ecology and evolutionary biology, two inter-related indicators offer us the possibility to retrospectively assess replicability: publication bias and statistical power. This registered report assesses the prevalence and severity of small-study (i.e., smaller studies reporting larger effect sizes) and decline effects (i.e., effect sizes decreasing over time) across ecology and evolutionary biology using 87 meta-analyses including 4,250 primary studies and 17,638 effect sizes. Further, we estimate how publication bias might distort the estimation of effect sizes, statistical power, and errors in magnitude (Type M or exaggeration ratio) and sign (Type S). We show strong evidence for the pervasiveness of both small-study and decline effects in ecology and evolution. There was widespread prevalence of publication bias that resulted in meta-analytic means being over-estimated by (at least) 0.12 standard deviations. The prevalence of publication bias distorted confidence in meta-analytic results with 66% of initially statistically significant meta-analytic means becoming non-significant after correcting for publication bias. Ecological and evolutionary studies consistently had a low statistical power (15%) with a 4-fold exaggeration of effects on average (Type M error rates = 4.4). Notably, publication bias aggravates low power (from 23% to 15%) and type M error rates (from 2.7 to 4.4) because it creates a non-random sample of effect size evidence. The sign errors of effect sizes (Type S error) increased from 5% to 8% because of publication bias. Our research provides clear evidence that many published ecological and evolutionary findings are inflated. Our results highlight the importance of designing high-power empirical studies (e.g., via collaborative team science), promoting and encouraging replication studies, testing and correcting for publication bias in meta-analyses, and embracing open and transparent research practices, such as (pre)registration, data- and code-sharing, and transparent reporting.

Keywords: open science; replicability; reproducibility; transparency; selective reporting; questionable research practices; *P*-hacking; registered report; many labs; generalizability; meta-research

## Introduction

Replicable prior findings are the foundation of cumulative scientific research. However, large-scale collaborative attempts to repeat studies have demonstrated that prior findings often fail to replicate in the medical and social sciences [1-3]. This raises concerns about the reliability of previously published studies (often referred to as the ‘replication crisis’; [4]). A similar issue of low replicability is likely to occur in ecology and evolutionary biology [6] (see also [5]). Yet, systematic assessments of replicability in this field are exceedingly rare [6, 7] perhaps because of the absence of strong incentives towards conducting replication studies [7, 8], and for logistical reasons (e.g., difficulties of conducting studies of rare species or remote ecosystems [9, 10]).

There are, however, two inter-related indicators that can be used to retrospectively gauge replicability in ecology and evolutionary biology: publication bias and statistical power. Publication bias and low statistical power increase the occurrence of unreliable effect size estimates that cannot be replicated. Publication bias commonly occurs when studies with statistically significant results are published more frequently than those with statistically non-significant findings (also referred to as ‘file-drawer problem’ [11]) or are published more quickly (‘time-lag bias’, respectively [12, 13]). More rapid publication of statistically significant results can also lead to a decline in reported effects over time (‘decline effect’ [12, 13]). When statistically significant effects are preferentially published, smaller studies will tend to report larger effect sizes (known as ‘small-study effects’; [14]). Statistical power is a proxy of ‘replicability probability’, as it is defined as the likelihood of detecting a given (true) effect as statistically significant (the complement to Type II error; [15]).

Several meta-research studies in ecology and evolutionary biology have investigated the prevalence of publication biases and low statistical power. Jennions and Moller [12] reported a statistically significant decline effect in a survey of 44 ecology and evolutionary biology meta-analyses that had been published in 2002. Using 52 meta-analyses published in 2000, Barto and Rillig [16] reached a similar conclusion. In a cumulative meta-analysis, Crystal-Ornelas and Lockwood [17] also identified a statistically significant decline in the magnitude of the effect of invasive species on species richness, using 240 papers published between 1999 and 2016. In their work, this decline effect was present consistently regardless of taxonomic groups, invasion time, or journal quality. Twenty years ago, statistical power in 10 ecology, evolution, and behaviour journals was estimated at 13% – 16% for small effects and 40% – 47% for medium effects (where small effects are $r = 0.1$ and medium effects are $r = 0.3$; *sensu* Cohen [18]). Even lower statistical power was estimated for the journal *Animal Behaviour* in 1996, 2003, and 2009 (7% – 8% and 23% – 26% to detect Cohen's small and medium effect sizes, respectively; [17]).

Despite earlier efforts in ecology and evolutionary biology, the field still lacks a systematic overview of the extent to which different forms of publication bias would distort the estimation of true effects. Further, no studies have evaluated how such distorted effect sizes prevent us from correctly estimating statistical power. The statistical power of a given study depends on sample size and the estimate of corresponding "true" effect size (e.g., a larger effect size leads to a higher power; see Fig. 1A). Therefore, to avoid overestimating the statistical power of a given study, an unbiased proxy of the "true" effect size should be used. Contrastingly, previous attempts in ecology and evolution often used Cohen's benchmarks to quantify statistical power for a given study [19, 20]. Yet, these benchmarks were derived from Cohen's qualitative intuitions for studies in the social sciences rather than a quantitative synthesis of the representative literature [21]. Cohen's benchmarks are arbitrary, and not necessarily applicable to ecological and evolutionary studies. As with exemplar studies in other fields [22], "true" effects can be estimated via meta-analytic approaches and preferably corrected for potential publication bias [23, 24]. Using publication bias-corrected effect size estimates as "true" effects would, more accurately, quantify statistical power as well as the two related, yet underappreciated, statistical errors: Type M and S errors (Fig. 1B and

1C; [25]). Type M error, also known as exaggeration ratio (magnitude error), represents the ratio between an estimated effect and a “true” effect, whereas Type S error represents the probability of attaining statistical significance in the direction opposite to the true effect [26]. No study has yet quantified these two quantities systematically across the field of ecology and evolutionary biology.

**Fig. 1**

Statistical power, Type S and M errors as a function of the “true” effect size (the alpha level is fixed at 0.05). The generic form of effect sizes (e.g., SMD, lnRR, *Zr*) are simulated from 0 to 1 with a fixed standard error (0.25). These panels (A – C) show that studies investigating larger true effects have higher power (A) and lower rates of Type M (B) and S (C) errors. If a study suffers from publication bias, the effect size is likely to be exaggerated, and consequently, the corresponding statistical power, Type M and S errors would be underestimated.

Here, we capitalize on the rapid growth of ecological and evolutionary meta-analyses to systematically assess the extent to which patterns consistent with publication biases are common across the fields of ecology and evolutionary biology, and, if attributed to actual publication bias, their impacts on the estimates of effect size, statistical power, Type M and S errors [27]. First, we test for the presence and severity of two indices of publication bias (i.e., small-study effect and decline effect) at two levels: (i) the within-meta-analysis level using a newly proposed multilevel meta-regression method; and (ii) the between-meta-analysis level using second-order meta-analyses (i.e., meta-meta-analyses). Second, we correct for these publication biases and quantify the degree of decline in bias-corrected effect-size magnitude. Finally, we use uncorrected and bias-corrected mean effect sizes as proxies of the “true” effect to assess the statistical power, Type M and S errors in ecology and evolutionary biology both at the primary study (effect-size) and the synthesis (meta-analysis) level.

## Materials and Methods

In this registered report, we have already finished collection (Section ***Data collection***), retrieval, and cleaning (Section ***Data retrieval and cleaning***) of data from a pre-existing dataset [28]. We have not yet commenced the statistical analyses process (Section ***Statistical analysis***).

## Database

### Data retrieval and cleaning

By checking the main text, supplementary materials, and/or online data repositories (e.g., Dryad, GitHub, Open Science Framework) of the 102 meta-analytic papers, and emailing corresponding authors, if necessary, we were able to include 80 papers that reported essential information for our statistical analyses. These 80 papers contained 108 independent meta-analyses. Among these 108, 36 meta-analyses used standardised mean difference (SMD) which includes some well-known estimators such as Hedges’ $g$ or Cohen’s $d$; [29]; 20 of these meta-analyses provided raw data (i.e., descriptive statistics: mean, standard error or deviation, and sample size) whereas the remaining 16 cases provided only effect sizes and variance. Twenty meta-analyses used the log response ratio (lnRR; [30]; also known as the ratio of means, ROM): 10 cases with raw data, and 10 cases without raw data. Thirty-one cases used the correlation coefficient or its Fisher’s transformation, $Zr$ (given that the variance of $Zr$ and sample size is convertible, all cases of $Zr$ were with raw data). All correlation coefficients were converted to $Zr$ to better approximate normal errors [31]. The remaining 20 meta-analyses used other effect size metrics, such as heritability ($h^2$; [32]), regression slope (e.g., reaction norm or selection gradient; [33, 34]), 2-by-2 binary data (e.g., log odds and risk ratios; [35]), raw mean difference [36], and non-standard metrics (proportion; [37]).

We decided to only include meta-analytic cases using SMD, lnRR, and *Zr* in our datasets because, in addition to being the most commonly used effect sizes in ecology and evolutionary biology [38, 39], they share statistical properties necessary to fit a formal meta-analytic model: (i) they are “unit-less,” which allows comparisons of studies originally using different units, (ii) they are (asymptotically) normally distributed, and (iii) they have readily computable (unbiased) sampling variance [31]. To keep our datasets independent, we only used the effect sizes in their original forms, although data augmentations (e.g., conversions between *Zr* to SMD) could maximise the statistical power of the following statistical analyses by maximising the number of sample sizes per dataset (in this case, the number of effect sizes). Therefore, our final three datasets consisted of (1) 36 meta-analytic cases of SMD, (2) 20 cases of lnRR, and (3) 31 cases of *Zr* (Fig. 2). For each primary study included in the final dataset, we retrieved four key variables: (i) effect sizes reported (i.e., SMD, lnRR, or *Zr*), (ii) standard errors (or sampling variances) of each effect size (to test for small-study effects), (iii) sample sizes per condition where possible (i.e., experimental group *versus* control group for SMD and lnRR); sample sizes are used to create a predictor to test and correct for small-study effects (i.e., ‘effective sample size’; see Section ***Second-order meta-analysis*** for details), and (iv) publication year (to test for a decline effect).

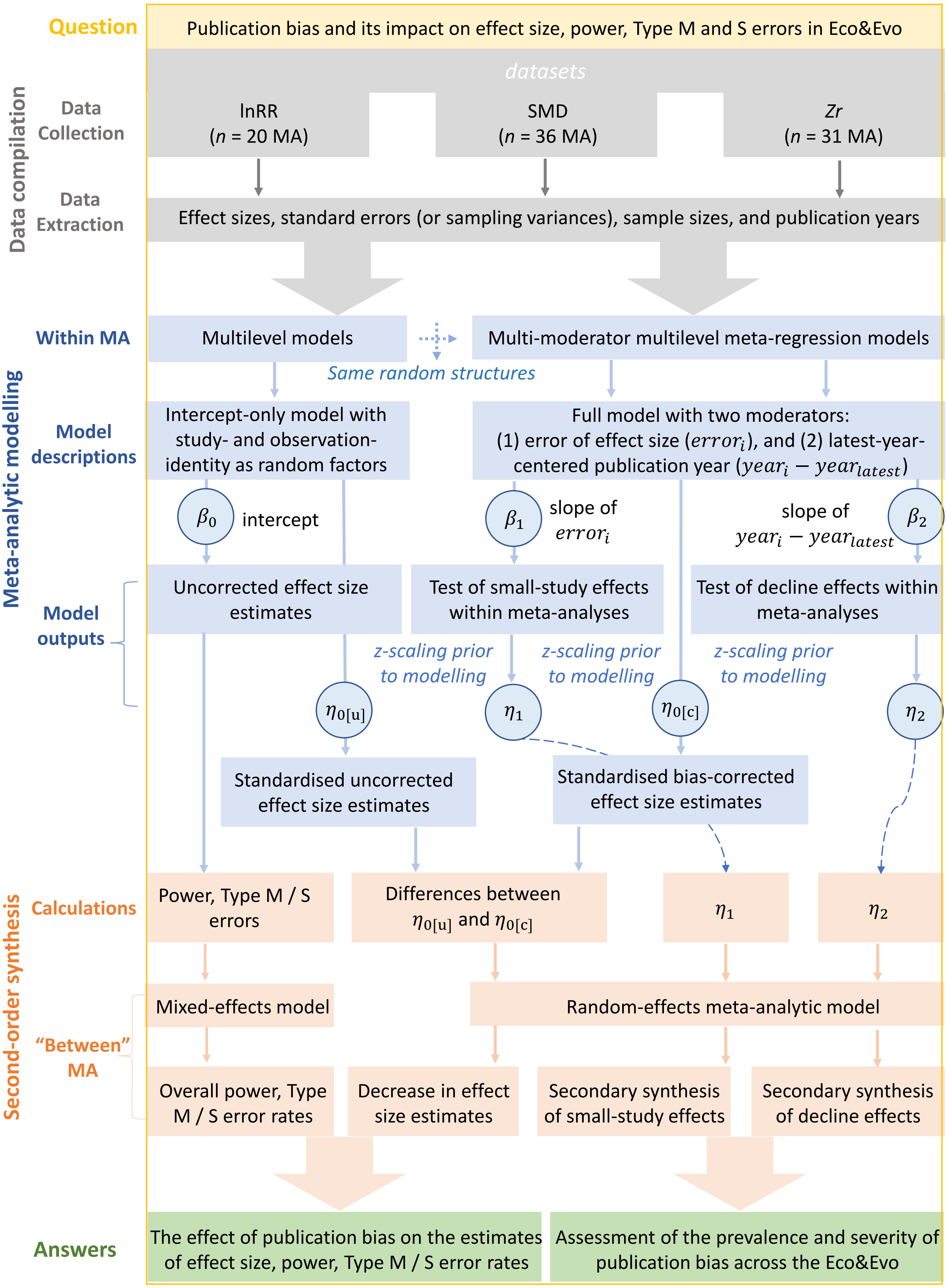

**Fig. 2**

The workflow showing the data compilation, statistical modelling processes, and our aims. Using the datasets containing 88 independent meta-analytes (36 SMD, 20 lnRR, and 31 *Zr* cases, respectively), we used a two-step modelling procedure to assess (i) the estimated prevalence and severity of publication bias across the fields of ecology and evolutionary biology, and (ii) how such publication bias affects the estimates of effect size, statistical power, Type M and S errors. In the first step (i.e., within-meta-analysis level), multilevel meta-analytic approaches will be used to estimate the overall mean (used for power and errors calculations), and test and adjust for publication bias for each meta-analytic case. In the second step (i.e., between-meta-analysis level), the estimates from the first step were statistically aggregated using either mixed-effect models or random-effects meta-analytic models (i.e., secondary meta-analysis). $\beta_{\mathrm{o}}$ is the meta-analytic overall mean (i.e., $\beta_{\mathrm{o[overall]}}$ in Equation 1), which signifies the uncorrected effect size estimate if publication bias exists but is not corrected. $\beta_1$ and $\beta_2$ are the indicators of small-study effect and time-lag bias (equivalent to $\beta_{1[\mathrm{small-study}]}$ and $\beta_{1[\mathrm{time-lag}]}$ in Equation 2). $\eta_{0[\mathrm{u}]}$ is the standardised $\beta_{\mathrm{o}}$. (i.e., $\eta_{0[\mathrm{overall}]}$). $\eta_{0[\mathrm{c}]}$ is the standardised bias-corrected meta-analytic overall mean (i.e., $\eta_{0[\mathrm{bias-corrected}]}$ in Equation 6). $\eta_{1[\mathrm{small-effect}]}$, $\eta_{2[\mathrm{time-lag}]}$ are standardised model coefficients corresponding to $\beta_{\mathrm{o}}$, $\beta_1$, and $\beta_2$ (i.e., $\eta_{1[\mathrm{small-effect}]}$ and $\eta_{2[\mathrm{time-lag}]}$ in Equation 6).

## Statistical analysis

### Data collection

We used a recent meta-analytic database that had been collected to evaluate the reporting quality of systematic reviews and meta-analyses published in ecology and evolutionary biology [28]. The inclusion and screening criteria identified meta-analyses that were broadly representative of meta-analyses published in ecology and evolutionary biology journals from 2010-2019. In brief, the database creators compiled a list of ‘Ecology’ and/or ‘Evolutionary Biology’ journals via the categories of the ISI InCites Journal Citation Reports®. Within the included journals, they searched Scopus using the string “meta-analy*” OR “metaanaly*” OR “meta-regression”. They restricted the search to articles published from January 2010 to 25 March 2019. Search results were then filtered to the 31 journals most frequently publishing meta-analyses. By taking a random sample of studies within each journal, a total of 297 papers was returned. After screening (search records, and inclusion and screening criteria are available at [28]), the database included a representative sample of 102 ecological or evolutionary meta-analyses.

## Multilevel meta-analytic modelling

We used multilevel meta-analytic approaches to (i) estimate the meta-analytic overall mean (i.e., uncorrected effect size estimates), (ii) detect potential publication bias (i.e., test small-study and decline effects), and (iii) correct for publication bias for each meta-analysis included in our datasets (Fig. 2).

### *Estimating uncorrected effect sizes*

To obtain uncorrected effect sizes for each meta-analysis (i.e., within-meta-analysis level), we fitted intercept-only multilevel meta-analytic models with SMD, lnRR, and *Zr* as our response variables, as in Equation 1 [39]. Equation 1 can account for dependent data by modelling both between-study variance (heterogeneity) and within-study variance (residual). It was written as:

$$ES_{ji} = \beta_{\mathrm{o[overall]}} + s_j + o_{ji} + m_{ji}, (1)$$

where $ES_{ji}$ is the extracted effect size, either SMD, lnRR, or *Zr*; $\beta_{\mathrm{o[overall]}}$ is the intercept, representing the estimate of overall effect (i.e., meta-analytic estimate of effect size); $s_j$ = the study-specific (between-study) effect of study *j*; $o_{ji}$ = the observation-level (within-study) effect for the effect size *i* (used to account for residual heterogeneity); $m_{ji}$ = the measurement (sampling) error effect for the effect size *i*. Between- and within-study effects are normally distributed with mean 0 and variance, $\sigma^2$ (i.e., $\mathcal{N}(0, \sigma^2)$). In Equation 1, effect size ($ES_{ji}$) and sampling variance ($m_{ji}$) can be calculated from the meta-analytic data. Using the restricted maximum likelihood (REML) method, we can obtain (approximately) unbiased estimates of variance parameters $\sigma^2$ for between- and within-study effects ($s_j$ and $o_{ji}$) [40]. With the REML estimate of $\sigma^2$, we can obtain the maximum likelihood estimate of the model coefficients (i.e., $\beta_{\mathrm{o[overall]}}$). These estimated model coefficients represent the (uncorrected) overall meta-analytic means for SMD, lnRR, or *Zr*. The model fitting was implemented via the *rma.mv* function from the *metafor* R package (version 3.4-0) [41].

*Detecting publication bias*

To test for patterns consistent with publication bias within each meta-analysis, we used a multi-moderator multilevel meta-regression model (an extended Egger's regression; cf. [42]). This approach deals with two common issues in ecological and evolutionary datasets: (i) using a multilevel model to control for data dependency [43], and (ii) using a regression method with multiple moderators to account for between-study heterogeneity [44]. We adopted this approach to test the presence of small-study and decline effects, respectively. This was written as:

$$ES_{ji} = \beta_{0[\text{bias-corrected}]} + \beta_{1[\text{small-study}]} error_i + \beta_{2[\text{time-lag}]}(year_i - year_{latest}) + s_j + o_{ji} + m_{ji}, (2)$$

where $\beta_{0[\text{bias-corrected}]}$ is the intercept, representing bias-corrected overall effect/meta-analytic estimate of effect size (see more details below); $error_i$ is the uncertainty index of effect size (i.e., sampling error of effect size, $se_i$), and $\beta_{1[\text{small-study}]}$ is the corresponding slope and an indicator of small-study effect; $year_i$ is the publication year, $year_{latest}$ is the latest year of published papers, and $\beta_{2[\text{time-lag}]}$ is the corresponding slope and an indicator of decline effect (i.e., time-lag bias).

When assuming there is no small-study effect (i.e., $error_i = 0$) and decline effect (i.e., $year_i - year_{latest} = 0$), the intercept $\beta_{\text{o[overall]}}$ in Equation 2 becomes a conditional estimate that can be interpreted as the bias-corrected overall effect (i.e., the estimate of "true" effect which is distinct from the unconditional estimate of $\beta_{\text{o[overall]}}$ in Equation 1). We centred the 'year' variable by subtracting each year ($year_i$ ) from the latest $year_{latest}$ to set the latest year as the intercept, $\beta_{0[\text{bias-corrected}]}$. This process allowed the estimate of true effect (i.e., $\beta_{0[\text{bias-corrected}]}$ in Equation 2) to be conditional on $year_i = year_{latest}$ so that $\beta_0$ was least affected by a decline effect if it existed. Further, we used a sampling error equivalent $\sqrt{1/\tilde{n}_i} = \sqrt{(n_e + n_c)/n_e n_c}$) to replace $se_i$ when fitting SMD and lnRR where possible ($4\tilde{n}_i$ is referred to as an effective sample; $n_e$ is the sample size of the experimental group, $n_c$ is the sample size of the control group; [42]). This can correct for the ‘artefactual’ correlation between $ES_{ji}$ and $error_i$ as the point estimate of SMD and lnRR are inherently correlated with their sampling variances (see Table 3 in [31], and Equation 10 in [45]).

A small-study effect is statistically detected if Equation 2 has a statistically significant $\beta_{1[\text{small-study}]}$ (i.e., $p$-value < 0.05). Similarly, the decline effect (i.e., time-lag bias) is indicated by a statistically significant $\beta_{2[\text{time-lag}]}$. Depending on the specific phenomenon tested, $\beta_{1[\text{small-study}]}$ and $\beta_{2[\text{time-lag}]}$ might be expected to be positive or negative when publication bias exists. For example, for an effect that is expected to be positive, a small-study effect and decline effect would be expressed in a positive value of $\beta_{1[\text{small-study}]}$ (i.e., small-size non-statistically significant effects and small-size statistically significant negative effects are underrepresented)) and negative value of $\beta_{2[\text{time-lag}]}$ (i.e., overall effect size declines over time), respectively. In such a case, a slope ($\beta_{1[\text{small-study}]}$ or $\beta_{2[\text{time-lag}]}$) with opposing direction (unexpected sign) indicates no detectable publication bias and subsequently does not require correction for such a bias. The magnitude of the slope represents the severity of the small-study effect or decline effect. Therefore, using Equation 2, we were able to detect the existence of publication bias and identify its severity for each meta-analysis and each effect size statistic.

*Correcting overall estimates for publication bias*

To avoid the biased estimate of $\beta_{0[\text{bias-corrected}]}$, we fitted Equation 3 when detecting a statistically significant $\beta_{0[\text{bias-corrected}]}$ in Equation 2. Equation 3 was written as:

$$ES_{ji} = \beta_{0[\text{bias-corrected}]} + \beta_{1[\text{small-study}]} error_i^2 + \beta_{2[\text{time-lag}]}(year_i - year_{latest}) + s_j + o_{ji} + m_{ji}, (3)$$

In contrast to Equation 2, Equation 3 used a quadratic term of uncertainty index (i.e., sampling variance $v_i$ or $1/\tilde{n}_i$) to alleviate the downward bias of an effect size estimate (for explanations see

[42, 46]). Theoretically, this procedure provided an easy-to-implement method to correct for publication bias for each meta-analysis (i.e., the conditional estimate of intercept in Equation 3). In practice, however, there were two different types of $\beta_{0[\mathrm{bias-corrected}]}$ estimates to consider. This is because high heterogeneity [44] can lead the signs of the slopes ($\beta_{1[\mathrm{small-study}]}$ and $\beta_{2[\mathrm{time-lag}]}$) to be opposite from that expected from publication bias [42]. We would subsequently misestimate $\beta_{0[\mathrm{bias-corrected}]}$ if slopes with unexpected signs are included in Equations 2 and 3.

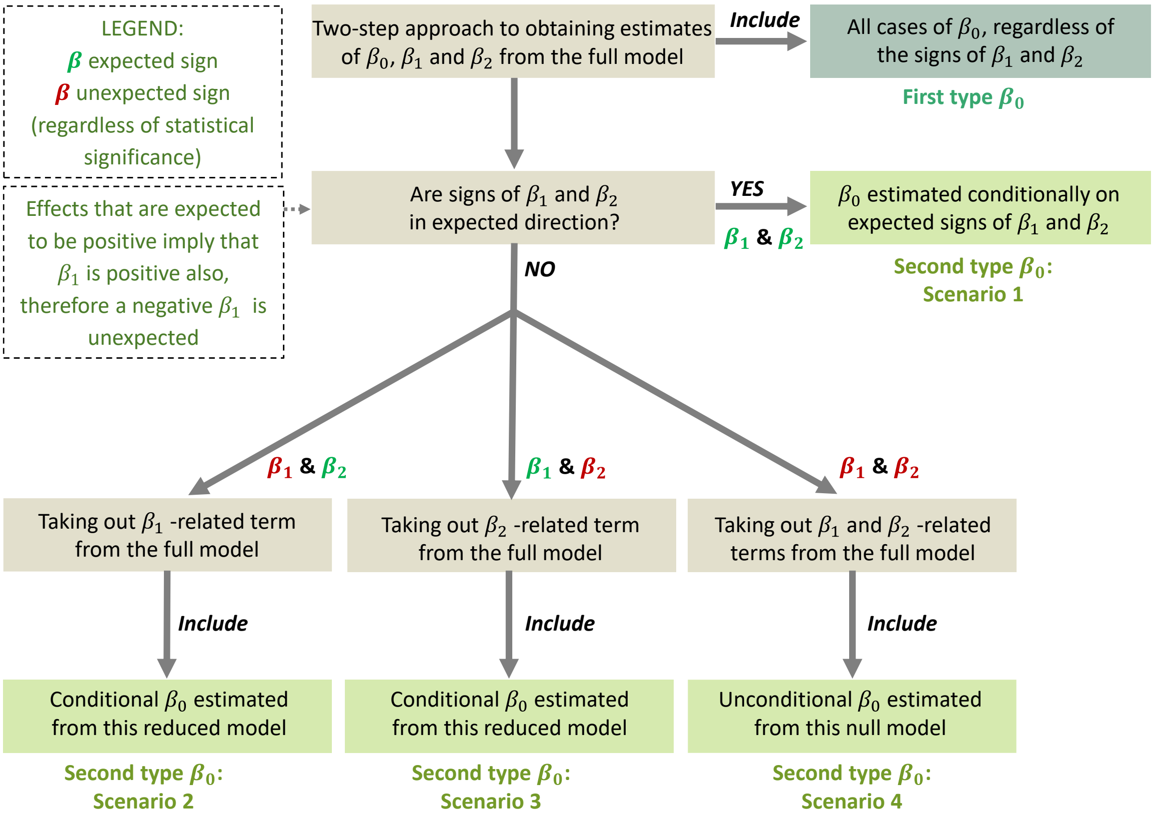

**Fig. 3**

The decision tree used to obtain the estimate of the ‘unbiased’ effect (i.e., conditional $\beta_0$). First, use a two-step procedure to estimate $\beta_0$, $\beta_1$ and $\beta_2$ from the full model (Equations 2 or 3). Then, depending on whether the signs of slopes ($\beta_1$ and $\beta_2$) are opposite from what will be expected from publication bias (caused by a high amount of unaccounted heterogeneity), there are two types of estimates of $\beta_0$. The first type includes all $\beta_0$ regardless of their signs ($\beta_1$ and $\beta_2$); the second type of estimated $\beta_0$ has four scenarios. Scenario 1 = only select $\beta_0$ with expected signs of $\beta_1$ and $\beta_2$ from the full model; Scenario 2 = employ reduced model 1 (Equation 4) to re-estimate $\beta_0$ where $\beta_1$ has an unexpected sign, while $\beta_2$ has an expected sign; Scenario 3 = employ reduced model 3 (Equation 5) to re-estimate $\beta_0$ if

$\beta_1$ has an expected sign, while $\beta_2$ has an unexpected sign; Scenario 4 = use $\beta_0$ from the null model (Equation 1) when both $\beta_1$ and $\beta_2$ have unexpected signs (i.e., without the small-study effects or decline effects). The symbols ($\beta_0$, $\beta_1$, and $\beta_2$) are as in Fig.2.

Depending on the signs of the slopes ($\beta_{1[\text{small-study}]}$ and $\beta_{2[\text{time-lag}]}$), there were two types of estimated $\beta_{0[\text{bias-corrected}]}$. We used a decision tree (Fig. 3) to obtain the estimate of each type of $\beta_{0[\text{bias-corrected}]}$ for each meta-analytic case. The function of the decision tree was that, if the slopes ($\beta_{1[\text{small-study}]}$ and $\beta_{2[\text{time-lag}]}$) had unexpected signs, we took out the corresponding slope-related term(s) from the full models to form reduced models (Equations 4 and 5) to better estimate $\beta_0$. The reduced models were written as Equations 4 and 5, respectively:

$$ES_{ji} = \beta_{0[\text{bias-corrected}]} + \beta_{1[\text{small-study}]} error_i + s_j + o_{ji} + m_{ji}, (4)$$

$$ES_{ji} = \beta_{0[\text{bias-corrected}]} + \beta_{2[\text{time-lag}]} (year_i - year_{latest}) + s_j + o_{ji} + m_{ji}, (5)$$

Specifically, the first type of estimate of $\beta_{0[\text{bias-corrected}]}$ was obtained by fitting Equation 2 or 3 (termed as full models). That included all cases of $\beta_{0[\text{bias-corrected}]}$ without consideration of the signs of $\beta_1$ and $\beta_2$ (i.e., conditional $\beta_{0[\text{bias-corrected}]}$ estimated from the full model; see Fig. 3). The second type of estimate of $\beta_{0[\text{bias-corrected}]}$ was obtained under the following four scenarios: (i) $\beta_{0[\text{bias-corrected}]}$ estimated under expected signs of $\beta_{1[\text{small-study}]}$ and $\beta_{2[\text{time-lag}]}$ (i.e., conditional $\beta_{0[\text{bias-corrected}]}$ estimated from the direction-controlled full model; see Fig. 3), which meant a co-occurrence of a small-study effect and a decline effect, (ii) $\beta_{0[\text{bias-corrected}]}$ estimated under the unexpected sign of $\beta_{1[\text{small-study}]}$ and the expected sign of $\beta_{2[\text{time-lag}]}$, which signalled the existence of a small-study effect but no decline effect (i.e., conditional $\beta_{0[\text{bias-corrected}]}$ estimated from reduced model 1; see Equation 4 and Fig. 3), (iii) $\beta_{0[\text{bias-corrected}]}$ estimated under the expected sign of $\beta_1$ and the unexpected sign of $\beta_2$, which indicated the occurrence of a decline effect but no small-study effect (i.e., conditional $\beta_{0[\text{bias-corrected}]}$ estimated from reduced model 2; see Equation 5 and Fig. 3), and (iv) $\beta_{0[\text{bias-corrected}]}$ estimated under unexpected signs of

$\beta_{1[\text{small-study}]}$ and $\beta_{2[\text{time-lag}]}$, which suggested little concerns about a small-study effect or a decline effect.

## Second-order meta-analysis

In this section, we statistically aggregated the above-mentioned regression coefficients (i.e., $\beta_{0[\text{bias-corrected}]}$, $\beta_{1[\text{small-study}]}$ and $\beta_{2[\text{time-lag}]}$) to (i) reveal the patterns of potential publication bias across the fields of ecology and evolutionary biology, and (ii) quantify the extent to which publication bias might cause a reduction in effect-size magnitude across meta-analyses (Fig. 2).

### *Estimating the overall extent and severity of publication bias*

To allow for aggregations of $\beta_{1[\text{small-study}]}$ (i.e., an indicator of small-study effect) and $\beta_{2[\text{time-lag}]}$ (i.e., an indicator of decline effect) over different effect size metrics (i.e., SMD, lnRR, and *Zr*), we standardized coefficients to eliminate scale-dependency [47]. This was achieved by *z*-scaling (i.e., mean-centring and dividing by the standard deviation) $error_i$, $year_i - year_{latest}$, and standardizing the response variable $ES_{ji}$ by dividing by the standard deviation without mean-centring, prior to modelling, as given by Equation 6:

$$c(ES_{ji}) = \eta_{0[\text{bias-corrected}]} + \eta_{1[\text{small-effect}]} z(error_i) + \eta_{2[\text{time-lag}]} z(year_i - year_{latest}) + s_j + o_{ji} + m_{ji}, (6)$$

Equation 6 indicates that one standard deviation change in $error_i$ and $year_i - year_{latest}$ would change $ES_{ji}$ by $\eta_{1[\text{small-effect}]}$ and $\eta_{2[\text{time-lag}]}$ standard deviations, respectively. Further, to interpret $\beta_0$ as a bias-corrected overall effect, $\beta_0$ was set conditional on $error_i = 0$ (i.e., without small-study effect) and $year_i - year_{latest} = 0$ (i.e., without decline effect). As such, we replaced $z(error_i)$ by $z(error_i) - z(error_0)$ and replace $z(year_i - year_{latest})$ by $z(year_i) - z(year_{latest})$, as shown in Equation 7:

$$c(ES_{ji}) = \eta_{0[\text{bias-corrected}]} + \eta_{1[\text{small-effect}]}(z(error_i) - z(error_0)) + \eta_{2[\text{time-lag}]}(z(year_i) - z(year_{latest})) + s_j + o_{ji} + m_{ji}, (7)$$

where $z(error_0)$ denotes the *z*-score when $error_i = 0$, which is equal to $\frac{0-\text{mean}[error_i]}{\text{SD}[error_i]}$; $z(year_{latest})$ is the *z*-score when $year_i$ is the latest year. Likewise, to obtain the best estimate of standardized bias-corrected effects, we introduced Equation 8 where a quadratic error term was used:

$$c(ES_{ji}) = \eta_{0[\text{bias-corrected}]} + \eta_{1[\text{small-effect}]}\,(z(error_i) - z(error_0))^2 + \eta_{2[\text{time-lag}]}\,(z(year_i) - z(year_{latest})) + s_j + o_{ji} + m_{ji}, (8)$$

Therefore, fitting 8 created two datasets: (1) the full dataset containing $\eta_{0[\text{bias-corrected}]}$, $\eta_{1[\text{small-effect}]}$ and $\eta_{2[\text{time-lag}]}$ without consideration of their signs (standardised slopes of the first type estimate), and (2) the reduced dataset containing $\eta_{0[\text{bias-corrected}]}$, $\eta_{1[\text{small-effect}]}$ and $\eta_{2[\text{time-lag}]}$ with expected directions (standardised slopes of the second type estimate: scenarios 1 – 4, Fig. 3). We then conducted a series of second-order meta-analyses to statistically aggregate these standardised regression coefficients across meta-analyses [48, 49]. We employed a random-effects meta-analytic model with the inverse square of each coefficient's standard error as weights to fit such second-order meta-analyses [41]. For both the full and reduced databases, we obtained a weighted average of the regression coefficient $\eta_{1[\text{small-effect}]}$ (or $\eta_{2[\text{time-lag}]}$) to indicate the occurrence of small-study effects (or decline effects) across the fields of ecology and evolutionary biology. To compare the severity of publication bias between different types of effect size, we further incorporated effect-size types as a moderator (i.e., a fixed factor or predictor with three levels: SMD, lnRR, and *Zr*) in these random-effects models.

*Quantifying the reduction in effect-size magnitude after controlling for publication bias*

Likewise, to quantify the differences between uncorrected effect sizes and their bias-corrected estimates for the different types of effect-size metrics, we required standardised estimates of these effect sizes to draw comparisons. The term $\eta_{0[\text{bias-corrected}]}$ in the full dataset provided a standardised bias-corrected effect size (i.e., an intercept estimated using the full model, where all cases of $\eta_{1[\text{small-effect}]}$ and $\eta_{2[\text{time-lag}]}$ were included regardless of their directions). Also, $\eta_{0[\text{bias-corrected}]}$ in the reduced dataset provided standardised bias-corrected effect sizes, which were obtained using expected directions of $\eta_{1[\text{small-effect}]}$ and $\eta_{2[\text{time-lag}]}$. In contrast, the standardised uncorrected effect sizes were obtained by standardizing $ES_{ji}$ by dividing by standard deviation before fitting Equation 1 (that is, standardised intercept in the null model: $\eta_{0[\text{overall}]}$). We then used the absolute mean difference as a metric to quantify the reduction in effect-size magnitude following correction for publication bias, where the point estimate and sampling variance was written as:

$$D = |\gamma^{s}_{\text{uncorrected-effect}} - \gamma^{s}_{\text{corrected-effect}}|, (9)$$

$$\text{Var}(D) = \text{SE}^2_{\gamma^{s}_{\text{corrected-effect}}} + \text{SE}^2_{\gamma^{s}_{\text{uncorrected-effect}}} - 2r\text{SE}_{\gamma^{s}_{\text{corrected-effect}}}\text{SE}_{\gamma^{s}_{\text{uncorrected-effect}}}, (10)$$

where $\gamma^{s}_{\text{corrected-effect}}$ and $\gamma^{s}_{\text{uncorrected-effect}}$ are the values of standardised uncorrected effect size (standardised $\eta_{0[\text{overall}]}$ in the null model) and its bias-corrected version (standardised $\eta_{0[\text{bias-corrected}]}$ in the full or reduced models), respectively; $\text{SE}_{\gamma^{s}_{\text{corrected-effect}}}$ and $\text{SE}_{\gamma^{s}_{\text{uncorrected-effect}}}$ are associated standard errors; $r$ is the correlation between standard errors ($\text{SE}_{\gamma^{s}_{\text{corrected-effect}}}$*vs.* and $\text{SE}_{\gamma^{s}_{\text{uncorrected-effect}}}$), which is assumed to be 1 because the two estimates should be strongly correlated.

Given that $D$ is an absolute variable, it follows a ‘folded’ normal distribution because taking the absolute value will force probability density on its left side ($x$-axis < 0) to be folded to the right [50, 51]. The corresponding folded mean and variance could be derived from its ‘folded’ normal distribution as Equations 11 and 12:

$$D_f = \sqrt{\frac{2}{\pi}\text{Var}(D)e^{-D^2/2\text{Var}(D)}} + D(1 - 2\Phi(\frac{-D}{\sqrt{\text{Var}(D)}})), (11)$$

$$\text{Var}(D_f) = D^2 + \text{Var}(D) - (\sqrt{\frac{2}{\pi}\text{Var}(D)e^{-D^2/2\text{Var}(D)}} + D(1 - 2\Phi(\frac{-D}{\sqrt{\text{Var}(D)}})))^2, (12)$$

where Φ is the standard normal cumulative distribution function (see more details in [50, 52]). Equations 9 to 12 enable us to calculate $D_f$ and $\mathrm{Var}(D_f)$ for both full and reduced databases. We used a random-effects meta-analytic model (*rma.uni* function; [41]) to synthesise these $D_f$ with $\mathrm{Var}(D)_f$ as sampling variance across meta-analyses. Also, we incorporated effect size type as a moderator to compare the differences in effect size reduction between SMD, lnRR, and *Zr*.

## Estimating statistical power, and Type M and S errors

We assessed the statistical power and Type M and S errors in the primary studies with experimental effects that were approximated by uncorrected and bias-corrected effect sizes [24, 53]. Although meta-analyses can increase power over primary studies [54], they might still be underpowered to detect the true effect (i.e., *p*-value > 0.05). Therefore, we also calculated the statistical power, Type M and S errors for each meta-analysis. To obtain averaged statistical power, and Type M and S errors at the primary study level, we used a mixed-effects model to aggregate over the estimates of power, and Type M and S errors from primary studies. We used the *lmer* function in the *lme4* R package (version 1.1-26) to fit these mixed-effects models [55], which incorporated the identity of the primary study as a random factor to account for between-study variation. Similarly, we used a weighted regression to aggregate meta-analysis level power, and Type M and S errors, with the number of effect sizes (*k*) within each meta-analysis as weights. We implemented the weighted regression via the *base R* function (version 4.0.3), *lm*.

## Deviations and additions

The Stage 2 of this registered report has three deviations from the Stage 1 protocol. First, in the section on *Correcting for overall estimates for publication bias*, the best estimate of the bias-corrected overall effect (i.e., model intercept $\beta_{0[\mathrm{bias-corrected}]}$) was initially planned to be obtained by a two-step procedure where when a zero effect exists (i.e., statistically non-significant $\beta_{0[\mathrm{bias-corrected}]}$), uncertainty index (i.e., sampling error $error_i$ or $\sqrt{1/\tilde{n}_i}$) was used (Equation 2) to estimate $\beta_{0[\text{bias-corrected}]}$, while when a non-zero effect exists (i.e., statistically significant $\beta_{0[\text{bias-corrected}]}$), a quadratic term of uncertainty index (i.e., sampling variance $v_i$ or $1/\tilde{n}_i$) was used (Equation 3) to estimate $\beta_{0[\text{bias-corrected}]}$ [56, 57]. We decided to only use Equation 3 to estimate $\beta_{0[\text{bias-corrected}]}$ because there is no need to estimate $\beta_{0[\text{bias-corrected}]}$ when no genuine effect exists (Equation 2).

Second, in the section *Estimating the overall extent and severity of publication bias*, we changed *z*-scaling (i.e., mean-centring and dividing by the standard deviation) response variable $ES_{ji}$ prior to model fitting to standardizing response variable $ES_{ji}$ by dividing by the standard deviation without mean-centring. This is because centring the response variable would make estimating model intercept ($\beta_{0[\text{bias-corrected}]}$) unfeasible [47]. The same change was made in the section on *Quantifying the reduction in effect-size magnitude after controlling for publication biases*.

Third, we added a post-hoc analysis where we removed the meta-analyses with statistically non-significant mean effects and subsequently calculated the average statistical power, Type M and S error rates. The reason why adding this post-hoc analysis was that the underlying true effect sizes in some meta-analyses were likely to be so trivially small (and biologically significant) that corresponding power calculation was meaningless. In such a case, if we included those effects when estimating average power across meta-analyses in ecology and evolution, we would get a downwardly biased average power estimate. Note that relevant results were reported in Supplementary Material (Table S4).

## 3 Results

### 3.1 The pattern of small-study effects in ecology and evolutionary biology

#### 3.1.1 Within-meta-analysis level

Of the 87 ecological and evolutionary meta-analyses, 15 (17%) meta-analyses showed evidence for a (i.e., smaller studies reporting larger effect sizes) (i.e., statistically significant $\beta_{1[\mathrm{small-study}]}$; see Fig. 4A). Importantly, $\beta_{1[\mathrm{small-study}]}$ from 54 (62%) meta-analyses were in the expected direction (Fig. 4A), indicating that these meta-analyses exhibited a statistically non-significant tendency for a small-study effect.

### 3.1.2 Between-meta-analysis level

When aggregating $\beta_{1[\mathrm{small-study}]}$ obtained from the 87 meta-analyses (i.e., when conducting the second-order meta-analysis), there was a statistically significant pooled $\beta_{1[\mathrm{small-study}]}$ providing evidence for the existence of small-study effects across meta-analyses (grand mean $\beta_{1[\mathrm{small-study}]}$ = 0.084, 95% confidence intervals (CIs) = 0.034 to 0.135, *p*-value = 0.001; Fig. 5A). Moreover, the heterogeneity among the $\beta_{1[\mathrm{small-study}]}$ estimates obtained from the 87 meta-analyses was low ($\sigma^2_{among-meta-analysis}$ = 0.0050; $I^2_{among-meta-analysis}$ =10%) suggesting high generalizability of these results. Three percent of  this heterogeneity could be explained by the types of effect sizes (SMD, lnRR, *Zr*) being meta-analysed ($R^2_{marginal}$ = 0.031); the non-random pattern of the small-study effect was mainly driven by SMD (grand mean $\beta_{1[\mathrm{small-study}]}$ = 0.091, 95% CI = 0.018 to 0.165, *p*-value = 0.015; Fig. 5B) and *Zr* (grand mean $\beta_{1[\mathrm{small-study}]}$ = 0.119, 95% CI = 0.026 to 0.212, *p*-value = 0.013), but not lnRR (grand mean $\beta_{1[\mathrm{small-study}]}$ = 0.029, 95% CI = −0.072 to 0.13, *p*-value = 0.571).

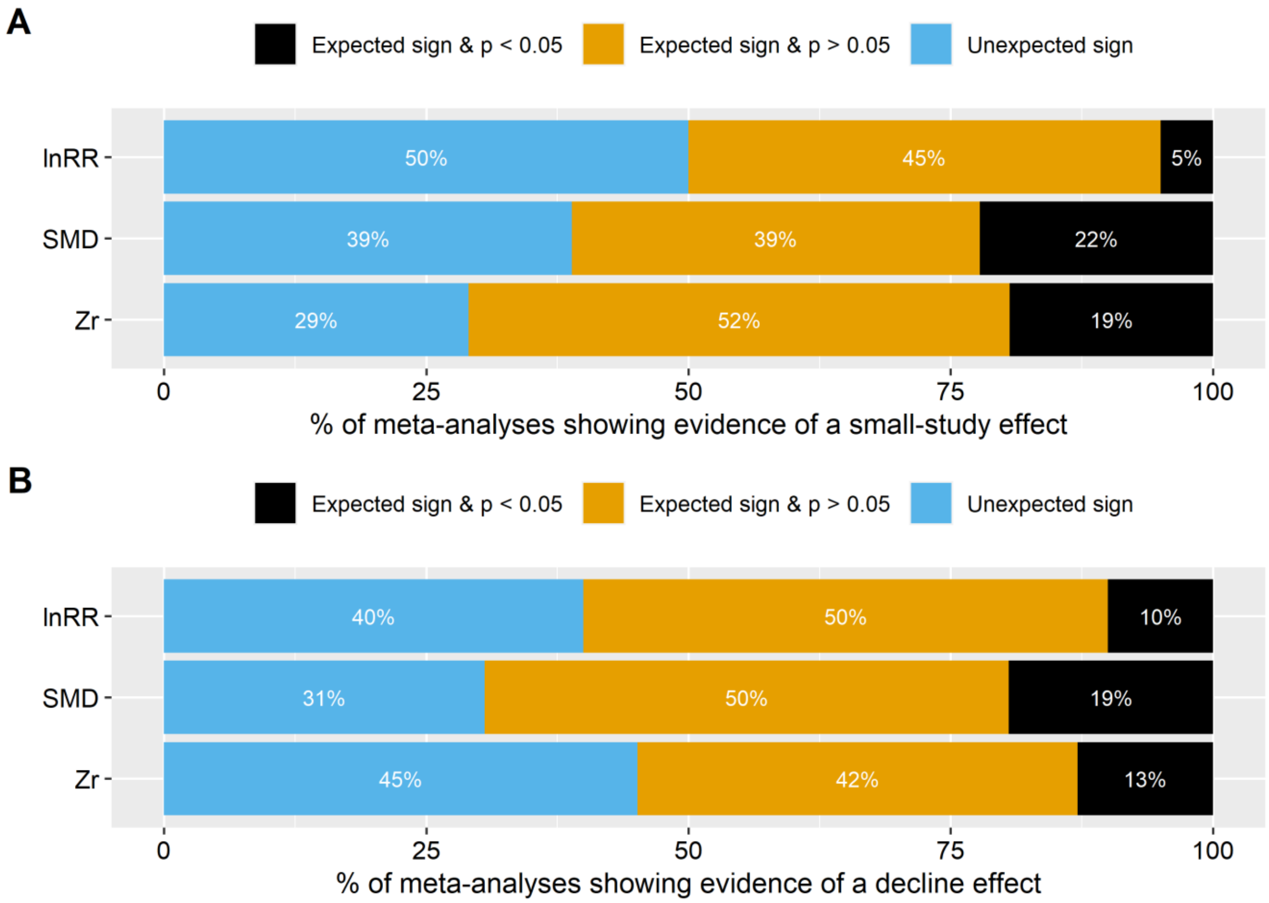

**Fig. 4**. The percentage of ecology and evolutionary meta-analyses showing evidence of publication bias. (A) a small-study effect (i.e., small non-statistically significant effects and small statistically significant negative effects are underrepresented), (B) a decline effect (; the magnitude of effect sizes changes over time). See more details in the legend of **Fig. 3**. All figures were drawn using the *geom_bar()* function in *ggplot2* R package (version 3.3.5) [58].

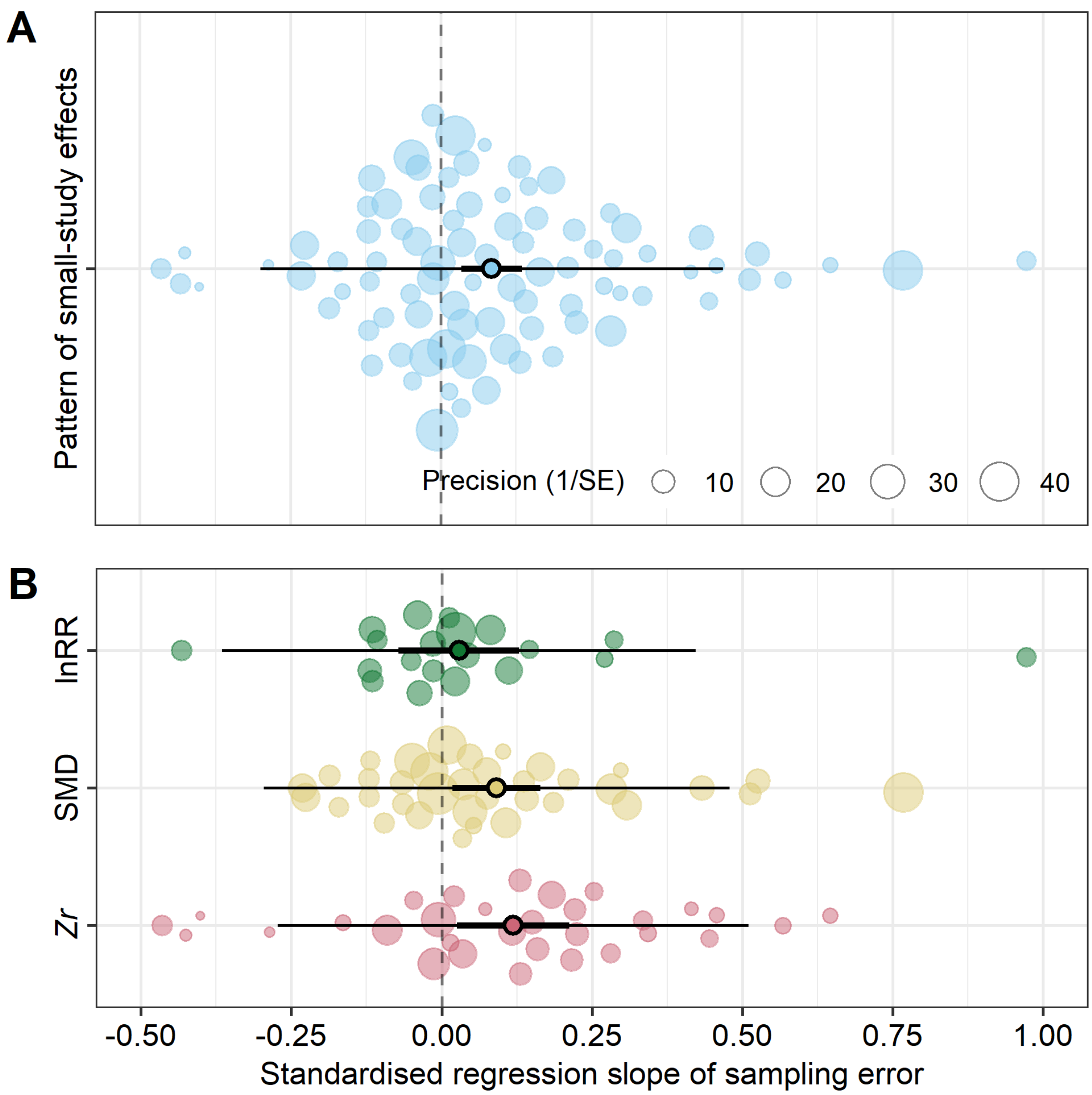

**Fig. 5.** Orchard plots showing the distribution of the indicator of small-study effect (model slope $\beta_{1[\text{small-study}]}$) for each meta-analysis and meta-analytic aggregation of $\beta_{1[\text{small-study}]}$ (pooled $\beta_{1[\text{small-study}]}$). (A) Pooled $\beta_{1[\text{small-study}]}$ across different meta-analyses and different types of effect size, indicating the pattern of small-study effects. (B) Pooled $\beta_{1[\text{small-study}]}$ for each type of effect size. Solid circles = $\beta_{1[\text{small-study}]}$ estimates obtained from each meta-analysis; the size of each solid circle is proportional to its inverse standard error (i.e., precision). Open circles = pooled $\beta_{1[\text{small-study}]}$. Thick error bars = 95% confidence intervals (CIs). Thin error bars = prediction intervals (PIs). See more details in the legend of **Fig. 2**. All panels were made using *orchard_plot()* function in *orchaRd* R package (version 2.0) [59].

3.2 The pattern of decline effects in ecology and evolutionary biology

### 3.2.1 Within-meta-analysis level

Thirteen of 87 (15%) meta-analyses showed evidence for a decline effect (i.e., effect sizes decreasing over time), with 54 (62%) meta-analyses showing a statistically non-significant decline in effect size over time (Fig. 4B).

### 3.2.2 Between-meta-analysis level

There was a statistically significant pooled $\beta_{2[\mathrm{time-lag}]}$ (grand mean $\beta_{2[\mathrm{time-lag}]}$ = −0.006, 95% CI = −0.009 to −0.002, *p*-value < 0.001; Fig. 6A) providing evidence for the existence of decline effects across meta-analyses. In addition, the estimates of $\beta_{2[\mathrm{time-lag}]}$ were homogeneous across these 87 meta-analyses (high generalizability of results) given the almost zero estimate of relative heterogeneity ($\sigma^2_{among-meta-analysis}$ = 0.0001; $I^2_{among-meta-analysis}$ < 1%). Five percent of that heterogeneity could be explained by the types of effect sizes ($R^2_{marginal}$ = 0.05); SMD and *Zr* exhibited a statistically significant pattern of decline effect (SMD: pooled $\beta_{2[\mathrm{time-lag}]}$ = -0.005, 95% CI = -0.010 to -0.001, *p*-value = 0.013; *Zr*: pooled $\beta_{2[\mathrm{time-bias}]}$ = -0.008, 95% CI = -0.015 to -0.001, *p*-value = 0.023; Fig. 6B), but lnRR did not (pooled $\beta_{2[\mathrm{time-bias}]}$ = -0.004, 95% CI = -0.01 to 0.003, *p*-value = 0.289).

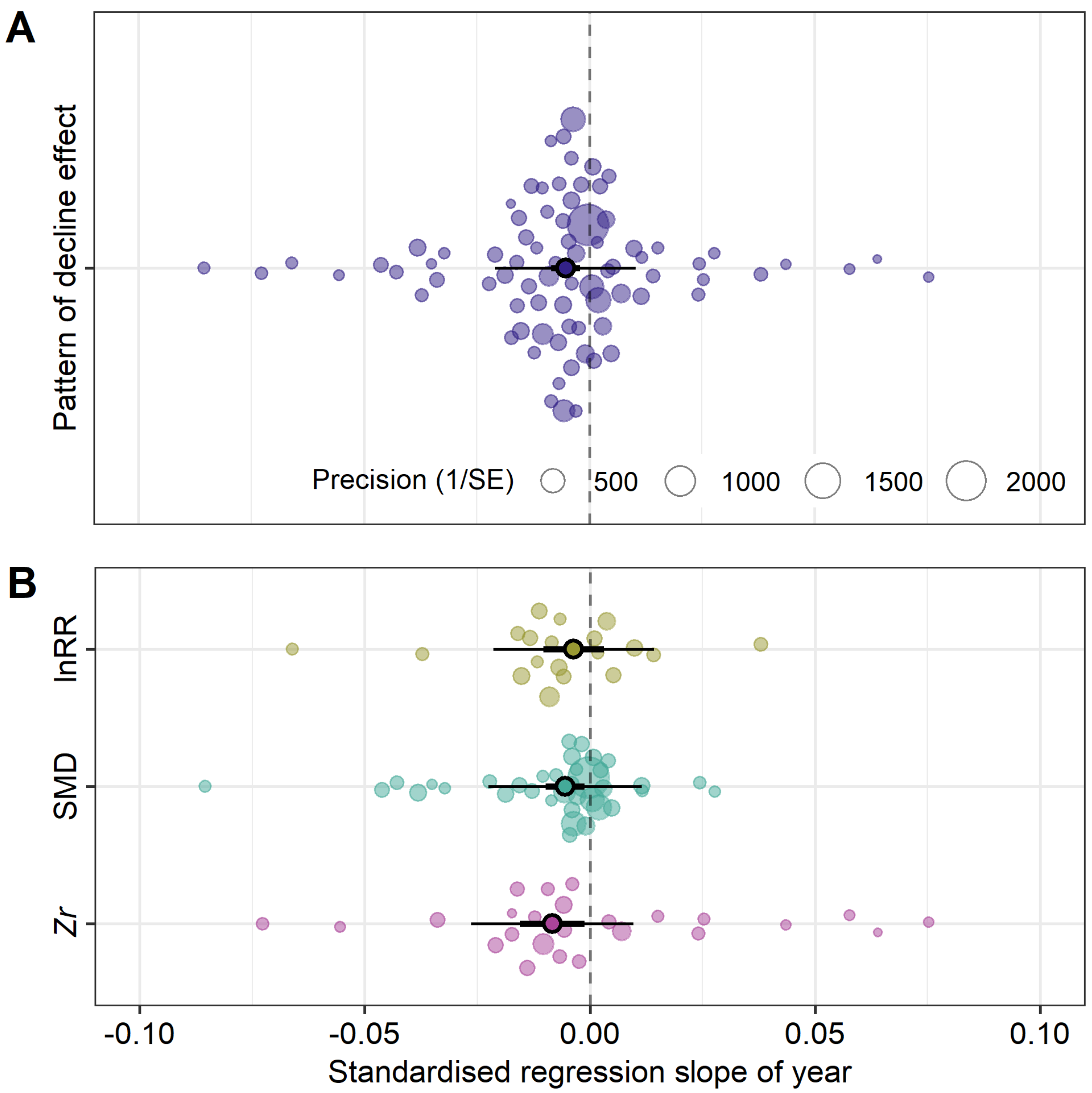

**Fig. 6.** Orchard plots showing the distribution of the indicator of decline effects (model slope $\beta_{2[\text{time-lag}]}$) for each meta-analysis and meta-analytic aggregation of $\beta_{2[\text{time-lag}]}$ (pooled $\beta_{2[\text{time-lag}]}$). (A) Pooled $\beta_{2[\text{time-lag}]}$ across different meta-analyses and different types of effect size, indicating the systematic pattern of decline effect. (B) Pooled $\beta_{2[\text{time-lag}]}$ for each type of effect size. See more details in the legend of **Figs. 2** and **3**. All panels were made using *orchard_plot()* function in *orchaRd* R package (version 2.0) [59].

### 3.3 The inflation of effect size estimates and distortion of meta-analytic evidence by publication bias

Across the 87 meta-analyses, the estimated absolute mean difference between the original (uncorrected) effect size ($\beta_{0[\text{overall}]}$) and its bias-corrected version ($\beta_{0[\text{bias-corrected}]}$) were statistically significant (pooled $D = 0.225$, 95% CI = 0.180 to 0.269, $p$-value < 0.001; Fig. S1A). The overestimation in SMD, lnRR, and *Zr* were 0.189, 0.195 and 0.333 standard deviation units, respectively (Fig. S1B). After back-transformation to the original scale, the publication bias led the estimates of SMD, lnRR, and *Zr* to be exaggerated by an average of 0.217, 0.116 and 0.128 (Fig. 7), respectively. Moreover, when correcting for publication bias, 33 out of 50 initially statistically significant meta-analytic mean became non-significant.

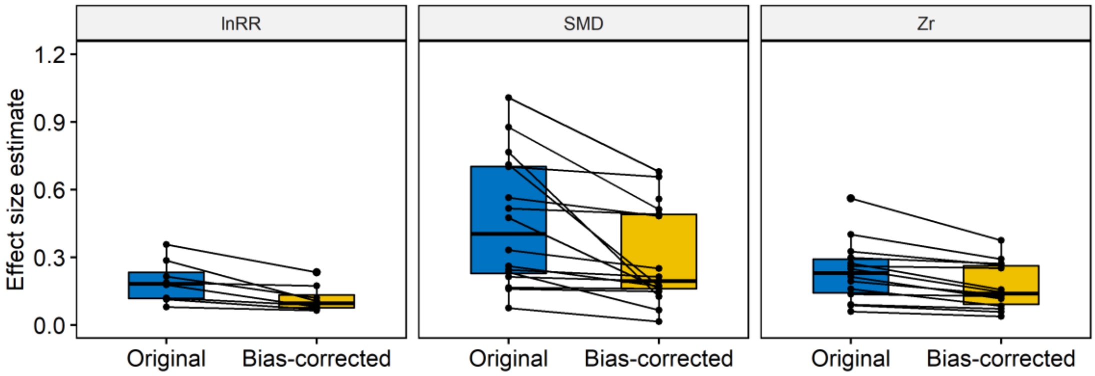

**Fig. 7.** The magnitude declines in meta-analytic estimate of effect sizes after correcting for publication bias for each meta-analysis. 9 out of 20 meta-analyses of lnRR, 17 out 36 meta-analyses of SMD, 14 out 31 meta-analyses of *Zr* had corrected directions of slope after adjusting for publication bias. The remained 11 in lnRR, 19 in SMD, and 17 in *Zr* showed wrong direction of slope because of high degree of heterogeneity that could not be controlled for. Original = uncorrected meta-analytic estimate effect sizes (i.e., $\beta_{o[\text{overall}]}$ in Equation 1). Bias-corrected = meta-analytic estimate effect size corrected for the presence of two forms of publication bias, small-study and decline effects (i.e., $\beta_{0[\text{bias-corrected}]}$ in Equation 3).

### 3.4 Statistical power, and Type S and M error rates

### 3.4.1 Sampling level (primary studies)

Overall, primary studies or single experiments (i.e., sampling level) only had a statistical power of 23% to detect the “true” effect indicated by the original (uncorrected) meta-analytic estimate of effect sizes, $\beta_{\mathrm{o[overall]}}$ (19%, 24% and 28% for sampling level of SMD, lnRR, and *Zr*, respectively; see Fig. 8 and Table S1). When bias correction was employed, the overall power to detect the “true” effect ($\beta_{0\mathrm{[bias-corrected]}}$) decreased to 15% (12%, 16%, and 18% for sampling level of SMD, lnRR, and *Zr*, respectively; see Fig. 8A and Table S1).

The primary studies infrequently showed the wrong estimation of the signs of the true effect sizes (overall Type S error = 5%; Fig. 9 and Table S2). For example, the primary study (i.e., sampling level) of lnRR and SMD had only 5% and 6% probabilities of having a direction that was opposite to the meta-analytic mean estimated as $\beta_{\mathrm{o[overall]}}$. Correcting for publication bias increased Type S error by 60% (from 5% to 8%). By contrast, the primary studies tended to exaggerate the magnitude of the meta-analytic mean estimated as $\beta_{\mathrm{o[overall]}}$ (overall Type M error = 2.7; Fig. 10 and Table S3). For example, the magnitude of lnRR, SMD and *Zr* were overestimated by an average of 2.5, 3.5 and 2 times, respectively. When correcting for the publication bias ($\beta_{0\mathrm{[bias-corrected]}}$), the Type M errors were increased to 4 (3.5 for lnRR, 6 for SMD and 3.4 for *Zr*).

### 3.4.2 Meta-analysis level

On average, at the level of individual meta-analyses lnRR and *Zr* had statistical power that was equal or over the nominal 80% level to detect the true effects estimated as $\beta_{0\mathrm{[bias-corrected]}}$: 81% for both lnRR and *Zr* (Fig. 8 and Table S1). In contrast, the estimated power of SMD was 41%, which fell short of the nominal 80% level. When detecting true effects indicated by

$\beta_{0[\text{bias-corrected}]}$, the statistical power of each meta-analysis decreased to 40%, where lnRR, SMD and *Zr* decreased to 63%, 25% and 51%, respectively.

Ecological and evolutionary meta-analyses had a relatively low probability of reporting an opposite sign to the true direction of both $\beta_{o[\text{overall}]}$ and $\beta_{0[\text{bias-corrected}]}$ (Type S = 5%–8%; Fig. 9 and Table S2). Meta-analyses considerably reduced the overestimation of the true effect size for lnRR (Type M = 1.1 for $\beta_{o[\text{overall}]}$ and 1.3 for $\beta_{0[\text{bias-corrected}]}$; Fig. 10 and Table S3), SMD (Type M = 1.9 for $\beta_{o[\text{overall}]}$ and 2.5 for $\beta_{0[\text{bias-corrected}]}$) and *Zr* (Type M = 1.1 for $\beta_{o[\text{overall}]}$ and 1.6 for $\beta_{0[\text{bias-corrected}]}$).

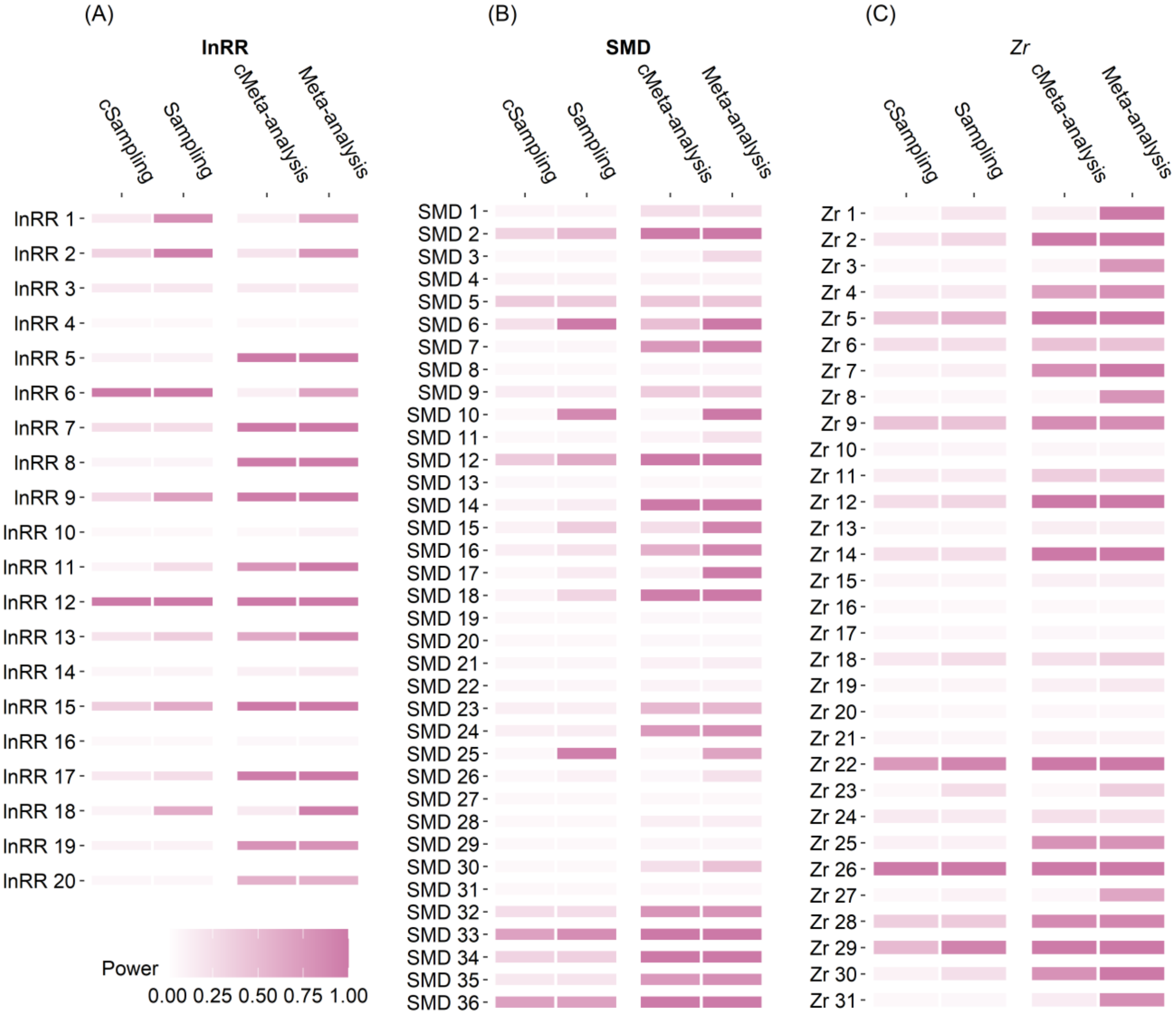

**Fig. 8.** Ecological and evolutionary studies' median statistical power to detect "true" effects that were approximated by meta-analytic mean effect size estimates (labels: Meta-analysis, Sampling) and their bias-corrected versions (labels: cMeta-analysis, cSampling). On the y-axis, effect size metrics with different subscripts represent different individual meta-analyses (see **Fig. 2**). Sampling = statistical power at sampling level (primary studies). cSampling = statistical power at sampling level after correcting for publication bias. Meta-analysis = statistical power at meta-analysis level. cMeta-analysis = statistical power at meta-analysis level after correcting for publication bias. See more details in the legend of **Fig. 3**. All figures were drawn via *geom_tile()* function in *ggplot2* R package (version 2.0) [58].

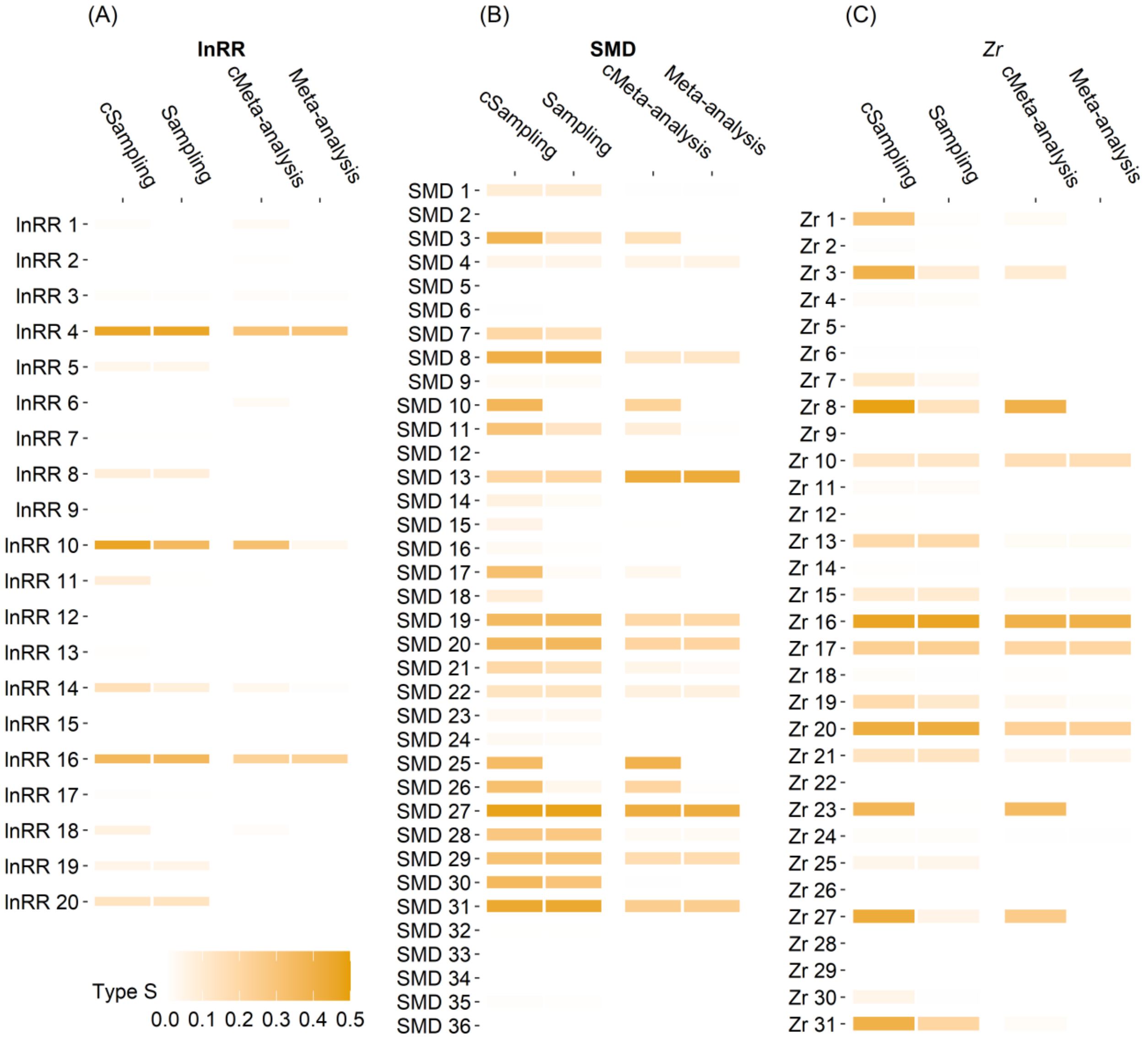

**Fig. 9.** Ecological and evolutionary studies' median Type S error rates (sign error) in detecting "true" effects that were approximated by meta-analytic mean effect size estimates (labels: Meta-analysis, Sampling) and their bias-corrected versions (labels: cMeta-analysis, cSampling). On the y-axis, effect size metrics with different subscripts represent different individual meta-analyses (see **Fig. 2**). Sampling = statistical power at sampling level (primary studies). See more details in the legend of **Figs. 3 and 8**. All figures were drawn via *geom_tile()* function in *ggplot2* R package (version 2.0) [58].

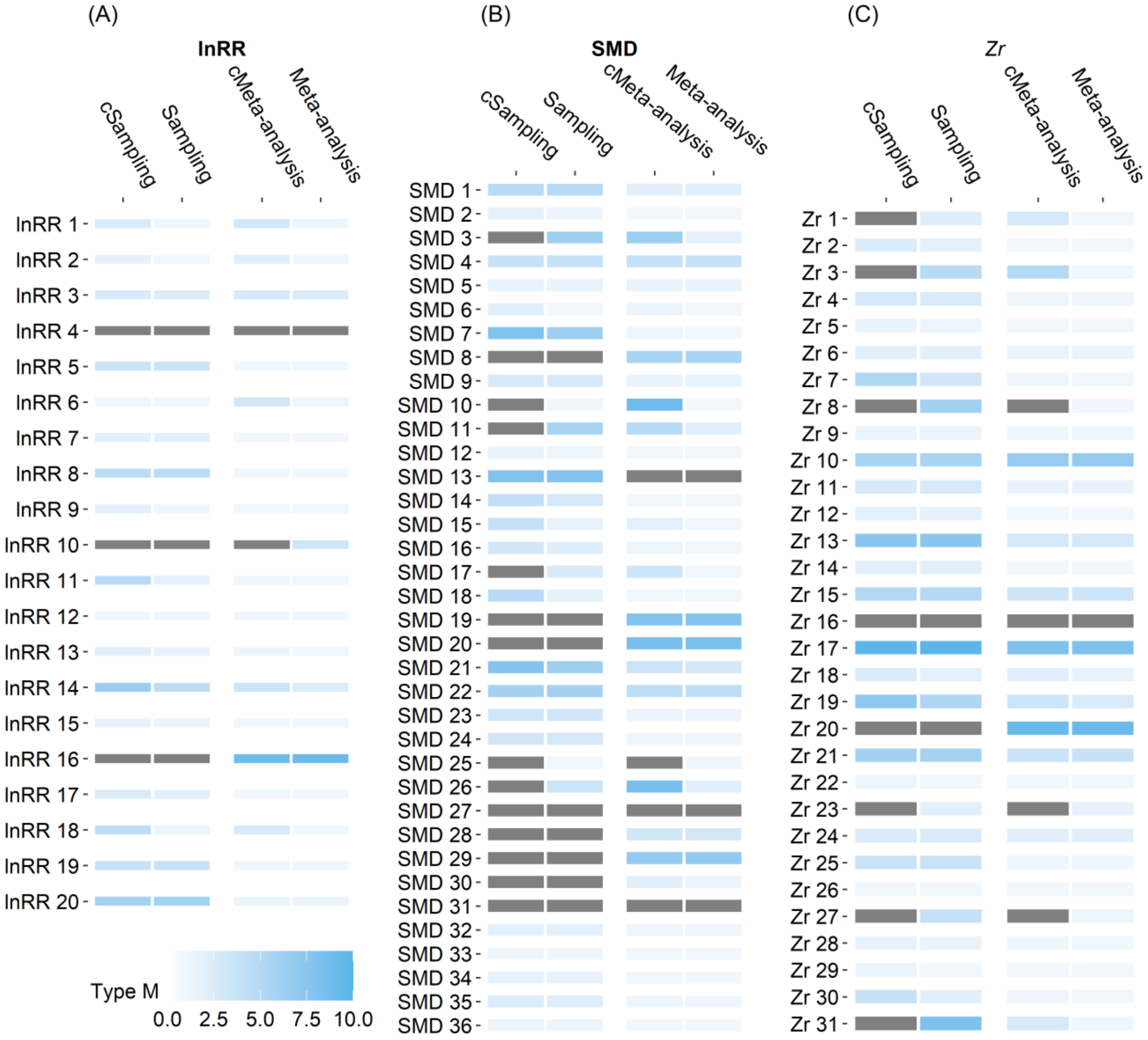

**Fig. 10.** Ecological and evolutionary studies' median Type M error rates (magnitude error) in detecting "true" effects that were approximated by meta-analytic mean effect size estimates (labels: Meta-analysis, Sampling) and their bias-corrected versions (labels: cMeta-analysis, cSampling). On the y-axis, effect size metrics with different subscripts represent different individual meta-analyses (see **Fig. 2**). Gray cells indicate that Type M errors are greater than 10. See more details in the legend of **Figs. 3 and 8**. All figures were drawn via *geom_tile()* function in *ggplot2* R package (version 2.0) [58].

## 4 Discussion

We have conducted the first comprehensive investigation of the prevalence and severity of two common forms of publication bias (i.e., small-study and decline effects) in the fields of ecology and evolutionary biology. Overall, we found strong support for small-study and decline effects (time-lag bias) with little heterogeneity across studies. The prevalence of such publication bias resulted in overestimating meta-analytic mean effect size estimates by at least 0.12 standard deviations. and substantially distorted the ecological and evolutionary evidence. When estimating power using bias-corrected effect size estimates, ecological and evolutionary studies and experiments consistently showed low statistical power (15%). Primary studies had a 4-fold overestimation of effects (Type M error = 4.4) and the signs of the effects were small yet not trivial (Type S error = 8%; error in the direction can lead to a completely opposite conclusion). To place these in perspective with the replication crisis [5, 6], we conclude that prior published findings in ecology and evolutionary biology, at least for the current dataset (87 meta-analyses, 4,250 primary studies, 17,638 effect sizes) are likely to have low replicability.

### 4.1 The persistent and non-negligible publication bias in ecological and evolutionary meta-analyses

#### 4.1.1 Small-study and decline effects are general phenomena

We have found that 17% of ecological and evolutionary meta-analyses show evidence for small-study effects (i.e., smaller studies reporting larger effect sizes). Medical researchers found a similar percentage of meta-analyses showing small-study effects (7% – 18%) in a survey of 6,873 meta-analyses (which was able to obtain a large sample because medical research has a large pool of meta-analyses to draw from and because that study extracted a much narrower scope of data from each meta-analysis than did our study; [7, 60]. Similarly, 13% – 25% of psychological meta-analyses presented evidence for small-study effects [61, 62]. These values may seem relatively small, but this is in part because, for a given meta-analysis, bias detection methods often lack statistical power to identify a small-study effect [42, 60, 63]. Indeed, simulations have shown that the power to detect a moderate small-study effect in a medical meta-analysis with 10 studies was as low as 21% [14].

Given the limited power to detect a small-study effect [14], it seems reasonable to focus on the sign and magnitude of the relationship between effect size and sampling error rather than decisions based on *p*-values (i.e., null-hypothesis significance testing). By doing so, we found that more than 60% of meta-analyses had a positive statistically non-significant relationship between the effect size and its sampling error, indicating that small studies (i.e., with large sampling error or small precision) tend to report larger effects. We confirmed these results by employing a more powerful approach, i.e., a second-order meta-analysis or meta-meta-analysis, which showed a statistically significant positive relationship between effect size and sampling error. This result is in line with recent investigations revealing an negative mean association of effect size and sample size in psychology and psychiatry meta-analyses [48, 64]. Moreover, our analysis also showed a small amount of heterogeneity among these 87 slopes. This positive and homogenous effect, therefore, implies that small-study effects are commonplace in ecology and evolutionary biology. Similar conclusions were reached in investigations of economic and psychological meta-analyses: small-study effects are widespread phenomena [65-67].

We conclude that decline effects are also widespread in the field. More than 50% of ecological and evolutionary meta-analyses showed a negative relationship between effect size and their year of publication, indicating that effect sizes decrease over time. As mentioned above, the principal reason for failing to detect a decline effect in a single meta-analysis root in the low statistical power of the available detection methods [13, 42, 68]. The observed power to determine a decline effect in the current set of 87 meta-analyses was considerably low (median = 13%). This low power was similar to that observed in another much larger survey of 464 ecological meta-analyses (median = 17%; [68, 69]). Importantly, our second-order meta-analysis have found a statistically significant and homogeneous effect (Figure 6A), corroborating that decline effects are common in both sub-fields previously explored (status signalling: [70], plant and insect biodiversity: [17, 71] and ocean acidification: [72]) and general fields of ecology and evolutionary biology [12, 68]. Evidence from other disciplines also reveals the pervasiveness of decline effects (medical and social sciences: [48, 73, 74]).

### 4.1.2 The distorted meta-analytic estimate of effect sizes and evidence by publication bias

By combining the observed bias from both small-study and decline effects, we found evidence that magnitudes of effect sizes might have been overestimated by 0.217, 0.116 and 0.128 their original units for lnRR, SMD and *Zr*, respectively). A recent investigation of 433 psychological meta-analyses also showed a statistically significant, albeit small, downward change in meta-analytic estimates after correcting for publication bias [75]. A comparison of meta-analyses that were published without pre-registration versus registered reports (which are less prone to publication bias) has also shown that unregistered meta-analyses substantially overestimated effect sizes although bias-correction like the one used in this study can eliminate the difference in results between registered meta-analyses and registered reports [76]. Accordingly, in our dataset, correcting for publication bias led to 33 of 50 initially statistically significant meta-analytic estimates becoming non-significant, suggesting that 66% of published ecological and evolutionary meta-analyses might have been overconfident in mean effect size estimates, and in some cases, potentially concluding that biologically meaningful effects exist when in fact they do not. Recent psychological investigations revealed a similar percentage (60%) of erroneous conclusions of meta-analytic evidence because of publication bias [77].

## 4.2 Low statistical power and high Type M error in ecological and evolutionary studies

### 4.2.1 Ecological and evolutionary studies lack power and are prone to Type M error

Primary studies in ecology and evolutionary biology included in our sample of meta-analyses, on average, only had a power of 15% to detect the effect size identified in the meta-analysis, which is consistent with earlier findings in the sub-fields of global change biology [53, 78] and animal behaviour: [10, 20]. When excluding the effects that are not statistically significant, the corresponding average power of primary studies was still very low (17%; Table S4). As a result, only studies with largely exaggerated effect sizes (4-fold) have reached statistical significance. Contrastingly, Type S error was small yet not trivial (8%); note that making an error in the direction can result in a completely opposite conclusion. The lack of statistical power seems to be a general phenomenon in scientific research, low power has been identified in many disciplines (medical sciences = 20% [79], neuroscience = 21% [22], psychological sciences = 36% [24], economics = 18% [80]). Given this widespread bias in individual studies, meta-analysis with appropriate bias correction is an important part of generating reliable estimates of effect sizes[27]. Statistically speaking, meta-analysis is an effective way to approximate population-level estimates by combining sampling level estimates, despite its shortcomings, some of which were shown above. Science is a process of evidence accumulation in which primary studies are the basis that can be used to produce high-order and high-quality evidence (e.g., via systematic review and meta-analysis).

### 4.2.1 Publication bias aggravates the low power and high Type M error

Publication bias is expected to aggravate low power and type M error rates because it creates a non-random sample of effect size evidence. We show that correcting for publication bias resulted in a

65% decrease in statistical power (23% *vs.* 15%), a 60% increase in Type S error rates (5% *vs.* 8%) and a 63% increase Type M error rates (2.7 *vs.* 4.4). Psychological and economic research also confirm that meta-analyses without bias adjustments overestimate the estimate of statistical power [24, 80]. Therefore, the exaggeration of power and effect size might have been even more severe in ecological and evolutionary studies if no bias corrections were to be made [6], providing further support to recent concerns about low replicability ("the replication crisis") in the fields of ecology and evolutionary biology [5, 10].

### 4.3 Limitations

There are three limitations in the present registered report. First, when calculating statistical power to detect true effects in ecology and evolutionary studies, we used the meta-analytic mean effect size (and corresponding bias-corrected version) as the true effect for each meta-analysis. We assumed that the multiple primary studies included in the same meta-analyses share a common true effect. However, the high heterogeneity in ecology and evolutionary meta-analyses indicates that each primary study may have a true effect size that is dependent on the research context (e.g., population and species; [44]). Therefore, using such context-dependent effects as the proxies of true effect is probably more reasonable [78]. Second, in the post-hoc analysis, we used the statistical significance ($p$-value < 0.05) of the meta-analytic mean effect size as the threshold to decide whether the true effect in a meta-analysis is so tiny that can be biologically neglected and subsequently excluded in the calculation of average power. We acknowledge that this categorization is arbitrary because the statistical significance does not represent biological significance [4]. However, it is difficult to exactly know the magnitude of the minimal effect that has biological importance. Third, the meta-analytic effect size estimates after correcting for publication bias may still be overestimated. This is caused by the incomplete reporting of important moderators in meta-analyses prevented us from accurately correcting for publication bias using our regression-based method [39, 43].

## 4.4 Implications

### 4.4.1 How to properly test for publication bias and correct for its impacts?

Given the strong and widespread evidence of publication bias found in this study (and others), publication bias tests should be a standard part of meta-analyses. A recent survey showed that publication bias tests have become more widespread in ecology and evolution in recent years [42]; however, inappropriate bias detection methods still dominate the literature [42]. Generally, regression-based methods are more powerful than other methods such as correlation-based methods [14, 60]. The regression-based method in the multilevel model framework used in the current study can further handle non-independence and high heterogeneity, which are common in the field, to bring down the rate of false positives [42-44]. Importantly, the method used here provides an intuitive quantification of the severity of publication bias. For example, the magnitude of pooled $\beta_{1[\text{small-study}]}$ (0.119) of *Zr* was larger than that of SMD (0.091), indicating publication bias in *Zr* is more severe than in SMD. Regression-based methods have been shown to produce effect size estimates similar to those of registered reports [76]. We strongly recommend that meta-analysts employ the regression-based method, which was used by current paper, to routinely test the presence of publication bias, correct for its impact and report the corrected effect sizes, allowing stakeholders to better judge how robust the reported effects are.

### 4.4.2 How to increase power and mitigate overestimation of effect for primary studies and meta-analyses?

For primary studies, a fundamental solution to increase statistical power and mitigate effect size overestimation is to increase sample sizes by building up more big-team science [81] or global-scale collaborative scientific networks such as Nutrient Network [82], US Long-Term Ecological Research network [83], and Zostera Experimental Network [84]. Our results confirm that lnRR is a more powerful effect size metric than SMD [78]. Power of meta-analyses using lnRR was almost twice as large as SMD (lnRR *vs.* SMD: 81% *vs.* 41%). Moreover, lnRR was less prone to exaggeration (lnRR *vs.* SMD: 1 *vs.* 2). Practically, we recommend using lnRR as the main effect size when conducting meta-analyses if the biological questions focus on mean differences (but see [85]). Yet it is often best to use both in a meta-analysis with one as the main analysis and the other as sensitivity analysis (see [78, 86] for comparisons of the pros and cons of lnRR and SMD).

### 4.5 Conclusions

We indirectly yet empirically examined the extent of the replication crisis in ecology and evolutionary biology using two inter-related indicators: publication bias and statistical power. Our results demonstrate indicate that both forms of publication bias, small-study and decline effects, are persistent and non-negligible in the field. Primary studies in ecology and evolutionary biology are often underpowered and prone to overestimation of the magnitude of the effect (i.e., Type M error). The pervasive publication bias leads to overestimated effect sizes, meta-analytic evidence and statistical power, and to an underestimated Type M error rate, undermining the reliability of previous findings. Although no single indicator can capture the true extent or all relevant evidence of the replication crisis [87], our research provides clear evidence that, as in many other disciplines [1, 2, 4], previously published findings are likely to have low replicability in ecology and evolutionary biology. The likely replication crisis in these fields highlights the importance of (i) designing high-power primary studies by building up big-team science [7, 81] where possible, (ii) adopting appropriate publication bias detection and correction methods for meta-analyses [42], (iii) embracing publication-bias-robust publication forms (e.g., Registered Reports – like the current article) for both empirical studies and meta-analyses alike. More generally, researchers need to adhere more closely to open and transparent research practices [88], such as (pre-)registration [89], data and code sharing [90, 91], and transparent reporting [6], to achieve credible, reliable and reproducible ecology and evolutionary biology.

## Author’s contributions

Yefeng Yang: conceptualising the paper, collecting the data, analysing the data, and drafting the manuscript

Alfredo Sánchez-Tójar: collecting the data, commenting, and editing the manuscript

Rose E. O’Dea: collecting the data, analysing the data, commenting, and editing the manuscript

Daniel W.A. Noble: collecting the data, commenting, and editing the manuscript

Julia Koricheva: collecting the data, commenting, and editing the manuscript

Michael D. Jennions: collecting the data, commenting, and editing the manuscript

Timothy H. Parker: collecting the data, commenting, and editing the manuscript

Malgorzata Lagisz: visualising, collecting the data, commenting, editing the manuscript, and supervising the project

Shinichi Nakagawa: conceptualising the paper, collecting the data, analysing the data, commenting, editing the manuscript, and supervising the project

## Availability of data and materials

The relevant data and code that reproduce the results of this registered report are available for review in https://github.com/Yefeng0920/EcoEvo_PB and will be archived in Zenodo upon formal acceptance of stage 2.

## Funding

We thank the Faculty of Science and the office of Deputy Vice-chancellor of Research, UNSW, Sydney for the support to YY and SN. YY was funded by the National Natural Science Foundation of China (NO. 32102597). SN and ML were funded by the Australian Research Council Discovery Grant (DP210100812). DN was supported by an ARC Discovery Grant (DP210101152).

## Declarations

The authors declare that they have no competing interests.

## References

1. Collaboration OS: **Estimating the reproducibility of psychological science**. *Science* 2015, **349**(6251).
2. Camerer CF, Dreber A, Forsell E, Ho T-H, Huber J, Johannesson M, Kirchler M, Almenberg J, Altmejd A, Chan T: **Evaluating replicability of laboratory experiments in economics**. *Science* 2016, **351**(6280):1433-1436.
3. Ebersole CR, Mathur MB, Baranski E, Bart-Plange D-J, Buttrick NR, Chartier CR, Corker KS, Corley M, Hartshorne JK, IJzerman H: **Many Labs 5: Testing pre-data-collection peer review as an intervention to increase replicability**. *Advances in Methods and Practices in Psychological Science* 2020, **3**(3):309-331.
4. Baker M: **Reproducibility crisis**. *Nature* 2016, **533**(26):353-366.
5. Kelly CD: **Rate and success of study replication in ecology and evolution**. *PeerJ* 2019, **7**:e7654.
6. Parker TH, Forstmeier W, Koricheva J, Fidler F, Hadfield JD, Chee YE, Kelly CD, Gurevitch J, Nakagawa S: **Transparency in ecology and evolution: real problems, real solutions**. *Trends in Ecology & Evolution* 2016, **31**(9):711-719.
7. O'Dea RE, Parker TH, Chee YE, Culina A, Drobniak SM, Duncan DH, Fidler F, Gould E, Ihle M, Kelly CD: **Towards open, reliable, and transparent ecology and evolutionary biology**. *BMC biology* 2021, **19**(1):1-5.
8. Fraser H, Barnett A, Parker TH, Fidler F: **The role of replication studies in ecology**. *Ecology and Evolution* 2020, **10**(12):5197-5207.
9. Nakagawa S, Parker TH: **Replicating research in ecology and evolution: feasibility, incentives, and the cost-benefit conundrum**. *BMC biology* 2015, **13**(1):1-6.
10. Feng X, Park DS, Walker C, Peterson AT, Merow C, Papeş M: **A checklist for maximizing reproducibility of ecological niche models**. *Nature Ecology & Evolution* 2019, **3**(10):1382-1395.
11. Rosenthal R: **The file drawer problem and tolerance for null results**. *Psychological bulletin* 1979, **86**(3):638.
12. Jennions MD, Møller AP: **Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution**. *Proceedings of the Royal Society of London Series B: Biological Sciences* 2002, **269**(1486):43-48.
13. Koricheva J, Kulinskaya E: **Temporal instability of evidence base: a threat to policy making?** *Trends in ecology & evolution* 2019, **34**(10):895-902.
14. Sterne JA, Gavaghan D, Egger M: **Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature**. *Journal of clinical epidemiology* 2000, **53**(11):1119-1129.
15. McShane BB, Böckenholt U, Hansen KT: **Average power: A cautionary note**. *Advances in Methods and Practices in Psychological Science* 2020, **3**(2):185-199.
16. Barto EK, Rillig MC: **Dissemination biases in ecology: effect sizes matter more than quality**. *Oikos* 2012, **121**(2):228-235.
17. Crystal-Ornelas R, Lockwood JL: **Cumulative meta-analysis identifies declining but negative impacts of invasive species on richness after 20 yr**. *Ecology* 2020, **101**(8):e03082.
18. Cohen J: **Statistical power analysis for the behavioral sciences**, 2nd edn. Hillsdale, New Jersey: Lawrence Erlbaum.; 1988.
19. Jennions MD, Møller AP: **A survey of the statistical power of research in behavioral ecology and animal behavior**. *Behavioral Ecology* 2003, **14**(3):438-445.

20. Smith DR, Hardy IC, Gammell MP: **Power rangers: no improvement in the statistical power of analyses published in Animal Behaviour**. *Animal Behaviour* 2011, **1**(81):347-352.

21. Correll J, Mellinger C, McClelland GH, Judd CM: **Avoid Cohen's 'small','medium', and 'large'for power analysis**. *Trends in cognitive sciences* 2020, **24**(3):200-207.

22. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR: **Power failure: why small sample size undermines the reliability of neuroscience**. *Nature reviews neuroscience* 2013, **14**(5):365-376.

23. Ioannidis JP, Stanley TD, Doucouliagos H: **The power of bias in economics research**. *Economic Journal* 2017, **127**:F236–F265.

24. Stanley T, Carter EC, Doucouliagos H: **What meta-analyses reveal about the replicability of psychological research**. *Psychological bulletin* 2018, **144**(12):1325.

25. Gelman A, Tuerlinckx F: **Type S error rates for classical and Bayesian single and multiple comparison procedures**. *Computational Statistics* 2000, **15**(3):373-390.

26. Gelman A, Carlin J: **Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors**. *Perspectives on Psychological Science* 2014, **9**(6):641-651.

27. Gurevitch J, Koricheva J, Nakagawa S, Stewart G: **Meta-analysis and the science of research synthesis**. *Nature* 2018, **555**(7695):175-182.

28. O'Dea RE, Lagisz M, Jennions MD, Koricheva J, Noble DW, Parker TH, Gurevitch J, Page MJ, Stewart G, Moher D: **Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension**. *Biological Reviews* 2021.

29. Hedges LV: **Estimation of effect size from a series of independent experiments**. *Psychological bulletin* 1982, **92**(2):490.

30. Hedges LV, Gurevitch J, Curtis PS: **The meta-analysis of response ratios in experimental ecology**. *Ecology* 1999, **80**(4):1150-1156.

31. Nakagawa S, Cuthill IC: **Effect size, confidence interval and statistical significance: a practical guide for biologists**. *Biological reviews* 2007, **82**(4):591-605.

32. Wood JL, Yates MC, Fraser DJ: **Are heritability and selection related to population size in nature? Meta-analysis and conservation implications**. *Evolutionary applications* 2016, **9**(5):640-657.

33. Murren CJ, Maclean HJ, Diamond SE, Steiner UK, Heskel MA, Handelsman CA, Ghalambor CK, Auld JR, Callahan HS, Pfennig DW: **Evolutionary change in continuous reaction norms**. *The American Naturalist* 2014, **183**(4):453-467.

34. Caruso CM, Eisen KE, Martin RA, Sletvold N: **A meta-analysis of the agents of selection on floral traits**. *Evolution* 2019, **73**(1):4-14.

35. Yates MC, Fraser DJ: **Does source population size affect performance in new environments?** *Evolutionary Applications* 2014, **7**(8):871-882.

36. Barrientos R: **Adult sex-ratio distortion in the native European polecat is related to the expansion of the invasive American mink**. *Biological Conservation* 2015, **186**:28-34.

37. Wehi P, Nakagawa S, Trewick S, Morgan-Richards M: **Does predation result in adult sex ratio skew in a sexually dimorphic insect genus?** *Journal of evolutionary biology* 2011, **24**(11):2321-2328.

38. Koricheva J, Gurevitch J: **Uses and misuses of meta-analysis in plant ecology**. *Journal of Ecology* 2014, **102**(4):828-844.

39. Nakagawa S, Santos ES: **Methodological issues and advances in biological meta-analysis**. *Evolutionary Ecology* 2012, **26**(5):1253-1274.

40. Viechtbauer W: **Bias and efficiency of meta-analytic variance estimators in the random-effects model**. *Journal of Educational and Behavioral Statistics* 2005, **30**(3):261-293.

41. Viechtbauer W: **Conducting meta-analyses in R with the metafor package**. *Journal of statistical software* 2010, **36**(3):1-48.

42. Nakagawa S, Lagisz M, Jennions MD, Koricheva J, Noble D, Parker TH, Sánchez-Tójar A, Yang Y, O'Dea RE: **Methods for testing publication bias in ecological and evolutionary meta-analyses**. *Methods in Ecology and Evolution* 2021.
43. Noble DW, Lagisz M, O'Dea RE, Nakagawa S: **Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses**. *Molecular Ecology* 2017, **26**(9):2410-2425.
44. Senior AM, Grueber CE, Kamiya T, Lagisz M, O'dwyer K, Santos ES, Nakagawa S: **Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications**. *Ecology* 2016, **97**(12):3293-3299.
45. Senior AM, Viechtbauer W, Nakagawa S: **Revisiting and expanding the meta-analysis of variation: The log coefficient of variation ratio, lnCVR**. *Research Synthesis Methods* 2020:e176.
46. Stanley TD, Doucouliagos H, Ioannidis JP: **Finding the power to reduce publication bias**. *Statistics in medicine* 2017, **36**(10):1580-1598.
47. Schielzeth H: **Simple means to improve the interpretability of regression coefficients**. *Methods in Ecology and Evolution* 2010, **1**(2):103-113.
48. Fanelli D, Costas R, Ioannidis JP: **Meta-assessment of bias in science**. *Proceedings of the National Academy of Sciences* 2017, **114**(14):3714-3719.
49. Nakagawa S, Samarasinghe G, Haddaway NR, Westgate MJ, O’Dea RE, Noble DW, Lagisz M: **Research weaving: visualizing the future of research synthesis**. *Trends in ecology & evolution* 2019, **34**(3):224-238.
50. Leone F, Nelson L, Nottingham R: **The folded normal distribution**. *Technometrics* 1961, **3**(4):543-550.
51. Nakagawa S, Lagisz M: **Visualizing unbiased and biased unweighted meta-analyses**. *Journal of evolutionary biology* 2016, **29**(10):1914-1916.
52. Morrissey MB: **Meta-analysis of magnitudes, differences and variation in evolutionary parameters**. *Journal of Evolutionary Biology* 2016, **29**(10):1882-1904.
53. Lemoine NP, Hoffman A, Felton AJ, Baur L, Chaves F, Gray J, Yu Q, Smith MD: **Underappreciated problems of low replication in ecological field studies**. *Ecology* 2016, **97**(10):2554-2561.
54. Cohn LD, Becker BJ: **How meta-analysis increases statistical power**. *Psychological methods* 2003, **8**(3):243-253.
55. Bates D, Mächler M, Bolker B, Walker S: **Fitting linear mixed-effects models using lme4**. *Journal of statistical software* 2014, **67**(1):1-46.
56. Stanley TD, Doucouliagos H: **Meta-regression approximations to reduce publication selection bias**. *Research Synthesis Methods* 2014, **5**(1):60-78.
57. Stanley TD: **Limitations of PET-PEESE and other meta-analysis methods**. *Social Psychological and Personality Science* 2017, **8**(5):581-591.
58. Wickham H, Chang W, Wickham MH: **ggplot2: Elegant graphics for data Analysis**. New York: Springer; 2016.
59. Nakagawa S, Lagisz M, O'Dea RE, Rutkowska J, Yang Y, Noble DW, Senior AM: **The orchard plot: Cultivating a forest plot for use in ecology, evolution, and beyond**. *Research Synthesis Methods* 2021, **12**(1):4-12.
60. Ioannidis JP, Trikalinos TA: **The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey**. *Cmaj* 2007, **176**(8):1091-1096.
61. Van Aert RC, Wicherts JM, Van Assen MA: **Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis**. *PloS one* 2019, **14**(4):e0215052.
62. Ferguson CJ, Brannick MT: **Publication bias in psychological science: prevalence, methods for identifying and controlling, and implications for the use of meta-analyses**. *Psychological methods* 2012, **17**(1):120.
63. Ioannidis JP: **Why most discovered true associations are inflated**. *Epidemiology* 2008:640-648.

64. Kühberger A, Fritz A, Scherndl T: **Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size**. *PloS one* 2014, **9**(9):e105825.
65. Doucouliagos C, Stanley TD: **Are all economic facts greatly exaggerated? Theory competition and selectivity**. *Journal of Economic Surveys* 2013, **27**(2):316-339.
66. Franco A, Malhotra N, Simonovits G: **Underreporting in psychology experiments: Evidence from a study registry**. *Social Psychological and Personality Science* 2016, **7**(1):8-12.
67. Bartoš F, Maier M, Wagenmakers E-J, Nippold F, Doucouliagos H, Ioannidis J, Otte WM, Sladekova M, Fanelli D, Stanley T: **Footprint of publication selection bias on meta-analyses in medicine, economics, and psychology**. *arXiv preprint arXiv:220812334* 2022.
68. Yang Y, Nakagawa S, Lagisz M: **Decline effects are rare in ecology: Comment**. *EcoEvoRxiv:1032942/osfio/qc7bx* 2022.
69. Costello L, Fox JW: **Decline effects are rare in ecology**. *Ecology* 2022:e3680.
70. Sanchez-Tojar A, Nakagawa S, Sanchez-Fortun M, Martin DA, Ramani S, Girndt A, Bokony V, Kempenaers B, Liker A, Westneat DF: **Meta-analysis challenges a textbook example of status signalling and demonstrates publication bias**. *Elife* 2018, **7**:e37385.
71. Van Klink R, Bowler DE, Gongalsky KB, Swengel AB, Gentile A, Chase JM: **Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances**. *Science* 2020, **368**(6489):417-420.
72. Clements JC, Sundin J, Clark TD, Jutfelt F: **Meta-analysis reveals an extreme "decline effect" in the impacts of ocean acidification on fish behavior**. *PLoS biology* 2022, **20**(2):e3001511.
73. Fanshawe TR, Shaw LF, Spence GT: **A large-scale assessment of temporal trends in meta-analyses using systematic review reports from the Cochrane Library**. *Research Synthesis Methods* 2017, **8**(4):404-415.
74. Pietschnig J, Siegel M, Eder JSN, Gittler G: **Effect declines are systematic, strong, and ubiquitous: A meta-meta-analysis of the decline effect in intelligence research**. *Frontiers in psychology* 2019:2874.
75. Sladekova M, Webb LE, Field AP: **Estimating the change in meta-analytic effect size estimates after the application of publication bias adjustment methods**. *Psychological Methods* 2022.
76. Kvarven A, Strømland E, Johannesson M: **Comparing meta-analyses and preregistered multiple-laboratory replication projects**. *Nature Human Behaviour* 2020, **4**(4):423-434.
77. Bartoš F, Maier M, Shanks D, Stanley T, Sladekova M, Wagenmakers E-J: **Meta-Analyses in Psychology Often Overestimate Evidence for and Size of Effects**. 2022.
78. Yang Y, Hillebrand H, Lagisz M, Cleasby I, Nakagawa S: **Low statistical power and overestimated anthropogenic impacts, exacerbated by publication bias, dominate field studies in global change biology**. *Global Change Biology* 2022, **28**(3):969-989.
79. Lamberink HJ, Otte WM, Sinke MR, Lakens D, Glasziou PP, Tijdink JK, Vinkers CH: **Statistical power of clinical trials increased while effect size remained stable: an empirical analysis of 136,212 clinical trials between 1975 and 2014**. *Journal of Clinical Epidemiology* 2018, **102**:123-128.
80. Ioannidis JP, Stanley TD, Doucouliagos H: **The power of bias in economics research**. *The Economic Journal* 2017, **127**(605):F236–F265.
81. Coles NA, Hamlin JK, Sullivan LL, Parker TH, Altschul D: **Build up big-team science**. In.: Nature Publishing Group; 2022.
82. Harpole WS, Sullivan LL, Lind EM, Firn J, Adler PB, Borer ET, Chase J, Fay PA, Hautier Y, Hillebrand H: **Addition of multiple limiting resources reduces grassland diversity**. *Nature* 2016, **537**(7618):93-96.
83. Crossley MS, Meier AR, Baldwin EM, Berry LL, Crenshaw LC, Hartman GL, Lagos-Kutz D, Nichols DH, Patel K, Varriano S: **No net insect abundance and diversity declines**

**across US Long Term Ecological Research sites**. *Nature Ecology & Evolution* 2020, **4**(10):1368-1376.

84. Wu PP-Y, Mengersen K, McMahon K, Kendrick GA, Chartrand K, York PH, Rasheed MA, Caley MJ: **Timing anthropogenic stressors to mitigate their impact on marine ecosystem resilience**. *Nature communications* 2017, **8**(1):1-11.
85. Bakbergenuly I, Hoaglin DC, Kulinskaya E: **Estimation in meta-analyses of response ratios**. *BMC medical research methodology* 2020, **20**(1):1-24.
86. Nakagawa S, Poulin R, Mengersen K, Reinhold K, Engqvist L, Lagisz M, Senior AM: **Meta-analysis of variation: ecological and evolutionary applications and beyond**. *Methods in Ecology and Evolution* 2015, **6**(2):143-152.
87. Fidler F, Chee YE, Wintle BC, Burgman MA, McCarthy MA, Gordon A: **Metaresearch for evaluating reproducibility in ecology and evolution**. *BioScience* 2017, **67**(3):282-289.
88. Gallagher RV, Falster DS, Maitner BS, Salguero-Gómez R, Vandvik V, Pearse WD, Schneider FD, Kattge J, Poelen JH, Madin JS: **Open Science principles for accelerating trait-based science across the Tree of Life**. *Nature ecology & evolution* 2020, **4**(3):294-303.
89. Parker T, Fraser H, Nakagawa S: **Making conservation science more reliable with preregistration and registered reports**. *Conservation Biology* 2019, **33**(4):747-750.
90. Parr CS, Cummings MP: **Data sharing in ecology and evolution**. *Trends in Ecology & Evolution* 2005, **20**(7):362-363.
91. Culina A, van den Berg I, Evans S, Sánchez-Tójar A: **Low availability of code in ecology: A call for urgent action**. *PLoS Biology* 2020, **18**(7):e3000763.