New approaches to meta-analyse differences in skewness, kurtosis, and correlation

Pietro Pollo[1,2]*, Szymon M. Drobniak[1,3], Hamed Haselimashhadi[4#], Malgorzata Lagisz[1,5#], Ayumi Mizuno[5#], Laura A. B. Wilson[6,7,8#], Daniel W. A. Noble[9‡], Shinichi Nakagawa[1,5‡*]

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

[2] School of Environmental and Life Sciences, University of Newcastle, Newcastle, NSW, 2308, Australia

[3] Institute of Environmental Sciences, Faculty of Biology, Jagiellonian University, Kraków, Poland

[4] European Bioinformatics Institute, European Molecular Biology Laboratory, Hinxton, UK

[5] Department of Biological Sciences, University of Alberta, CW 405, Biological Sciences Building, Edmonton, AB T6G 2E9, Canada

[6] School of Archaeology and Anthropology, The Australian National University, Acton, ACT 2601, Australia

[7] School of Biological, Earth and Environmental Sciences, University of New South Wales, Kensington, NSW 2052, Australia

[8] ARC Training Centre for Multiscale 3D Imaging, Modelling and Manufacturing, Research School of Physics, The Australian National University, Acton, ACT 2601, Australia

[9] Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, ACT, 2600, Australia

[#] These authors contributed equally and are listed alphabetically.

‡ These authors share senior authorship.

* Corresponding authors: pietro_pollo@hotmail.com, snakagaw@ualberta.ca

ORCID

Pietro Pollo: https://orcid.org/0000-0001-6555-5400

Szymon M. Drobniak: https://orcid.org/0000-0001-8101-6247

Hamed Haselimashhadi: https://orcid.org/0000-0001-7334-2421

Malgorzata Lagisz: https://orcid.org/0000-0002-3993-6127

Ayumi Mizuno: https://orcid.org/0000-0003-0822-5637

Laura A. B. Wilson: https://orcid.org/0000-0002-3779-8277

Daniel W. A. Noble: https://orcid.org/0000-0001-9460-8743

Shinichi Nakagawa: https://orcid.org/0000-0002-7765-5182

Abstract

Biological differences between males and females are pervasive. Researchers often focus on sex differences in the mean or, occasionally, in variation, albeit other measures can be useful for biomedical and biological research. For instance, differences in skewness (asymmetry of a distribution), kurtosis (heaviness of a distribution's tails), and correlation (relationship between two variables) might be crucial to improve medical diagnosis and to understand natural processes. Yet, there are currently no meta-analytic ways to measure differences in these metrics between two groups. We propose three effect size statistics to fill this gap: Δsk, Δku, and ΔZr, which measure differences in skewness, kurtosis, and correlation, respectively. Besides presenting the rationale for the calculation of these effect size statistics, we conducted a simulation to explore their properties and used a large dataset of mice traits to illustrate their potential. For example, in our case study, we found that females show, on average, a greater correlation between fat mass and heart weight than males. Although calculating Δsk, Δku, and ΔZr will require large sample sizes of individual data, technological advancements in data collection create increase opportunities to use these effect size statistics. Importantly, Δsk, Δku, and ΔZr can be used to compare any two groups, allowing a new generation of meta-analyses that explore such differences and potentially leading to new insights in multiple fields of study.

Key-words: covariance, individual participant meta-analysis, meta-regression, nonnormality, normal distribution, sex characteristics

Background

Sex is a biological attribute that can strongly impact organisms' traits, with differences between males and females being central to questions in the biological sciences (e.g., [1,2]). In contrast, biomedical research has primarily focused on male subjects [3], posing a danger to female health [4,5]. Aware of these issues, the US National Institutes of Health and other health agencies have demanded using multiple sexes in animal studies when possible [6]. As a consequence, the number of biological and biomedical studies using both female and male animals as research subjects has increased in the last decade [7], leading to the accumulation of data that can be used to synthesise and quantify sex differences across biological domains.

Realising the accumulation of sex-specific data, many perspective pieces have encouraged researchers to investigate sex differences more carefully (e.g., [8–10]). Yet, some of these pieces, and most of the biological literature, focus exclusively on mean differences between males and females. A fixation on mean differences has been present for a long time in science because researchers tend to focus on dimorphism in trait averages (e.g., [11]), lack sufficiently powerful data, or have limited statistical tools available (or difficulty to use them). Yet, measures such as variance, correlation, skewness, and kurtosis can be critical to understanding sex differences. For example, certain traits in mice may exhibit no disparity in average values between sexes, but substantial differences emerge in terms of variability [12,13]. These differences could be more easily assessed because of an effect size statistic that measures differences in variability between two groups (proposed by [14]), illustrating how novel statistical tools can expand possible research questions and provide new scientific insights, such as identifying sex differences in trait selection or canalisation.

Beyond variability, the relative shape of trait distributions to the normal distribution (measured by skewness and kurtosis, i.e. asymmetry of a distribution and heaviness of a distribution's tails, respectively; Fig. 1A-B) can also be crucial to understanding ecological and evolutionary processes and patterns (e.g., [15–19]), as well as improving medical diagnostics (e.g., [20,21]). For instance, skewness can bias heritability estimates because evolutionary biologists assume that phenotypic components (genetic and environmental) are normally distributed [18]. Furthermore, kurtosis can be used to understand community assembly processes (e.g., [16]). Besides the shape of trait distributions, evolutionary biologists and quantitative geneticists can quantify correlation matrices to understand trait plasticity and evolvability (e.g., [22–24]), which could then be used for group comparisons (as in [25]; Fig. 1C). Although location-scale-shape models (e.g., [26–28]) may be used to explore between-group differences (e.g., males and females) in skewness, kurtosis, or within-group correlations, there are no effect size statistics that can easily measure such differences (but see also [29]).

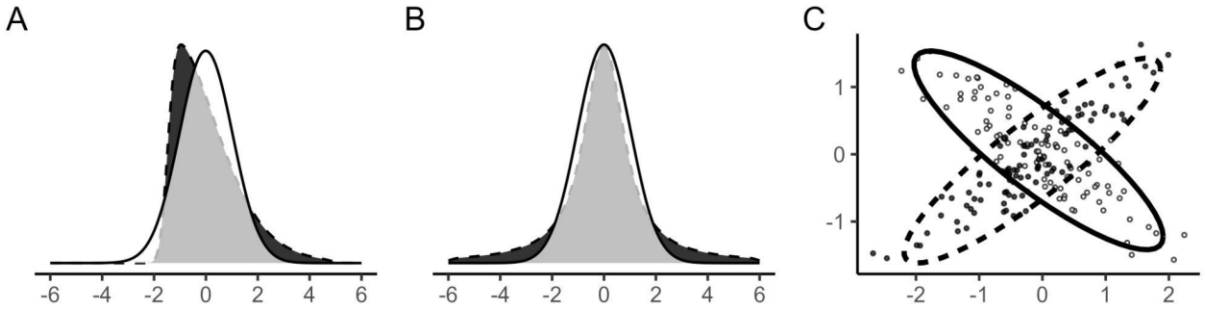


Figure 1. Simulated trait distributions for two groups with different shapes (A: distinct skewness, B: distinct kurtosis), and different correlations between two traits for two groups (C). The data and code needed to generate this Figure can be found in https://zenodo.org/records/18386956.

Here, we propose three new effect size statistics to evaluate between-group differences in skewness (Δ*sk*), kurtosis (Δ*ku*), and correlation (Δ*Zr*), key moments of a distribution that are usually unexplored. These effect size statistics will be valuable to explore sex differences but can also be applied in other fields of study and used to compare differences between any two groups of interest. Meta-analyses using these new effect sizes will create multiple avenues for novel biological enquiries. The present moment is particularly conducive for analyses using these new effect sizes because the individual-level data (e.g., individual participant data [30,31]) required for their calculation are increasingly available from new technological advances that allow faster data collection and sharing (e.g., automated phenotyping).

Difference in skewness and kurtosis

The mean and variance represent the first and second moments of a distribution, respectively. However, the third and fourth moments of a distribution (i.e. skewness and kurtosis, respectively) can also be valuable as they characterise the distribution's shape. More specifically, skewness reflects the distribution's asymmetry around its mean. While positive skewness indicates an elongated right tail with an excess of high values, negative skewness suggests an elongated left tail with an excess of low values. This asymmetry can influence the interpretation of means and variation, as the mean tends to be larger than the median in positively skewed distributions, while the mean tends to be smaller than the median in negatively skewed distributions. Note that a perfectly normal distribution is symmetric (i.e. skewness = 0), where the mean is equal to the median. Sample skewness (*sk*) [32] can be expressed as:

$$\mathrm{sk} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left[\frac{1}{n}\sum_{i=1}^{n}(x - \bar{x})^2\right]^{\frac{3}{2}}}\frac{\sqrt{n(n-1)}}{n-2} \quad \text{(eq. 1)}$$

where x_i is a raw data value, $\bar{x}$ is the sample mean, and n is the sample size. Skewness sampling variance (s^2_{sk}) [32] can then be expressed as:

$$s^2_{sk} = \frac{6n(n-1)}{(n-2)(n+1)(n+3)} \text{ (eq. 2)}$$

On the other hand, kurtosis measures tail heaviness: high kurtosis distributions have heavier tails (i.e., proportionally more extreme values than central values), whereas low kurtosis distributions have lighter tails. For comparison, a normal distribution is expected to have kurtosis = 3. Sample excess kurtosis (*ku*) [32] can be expressed as:

$$ku = \frac{n(n+1)(n-1)}{(n-2)(n-3)} \frac{\sum_{i=1}^{n}(x_i-\bar{x})^4}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2} - \frac{3(n-1)^2}{(n-2)(n-3)} \text{ (eq. 3)}$$

with sampling variance (s^2_{ku}) [32] as:

$$s^2_{ku} = \frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)} \text{ (eq. 4)}$$

Evaluating skewness and kurtosis provides valuable insights into a variable distribution, which is crucial for interpreting means, assessing variability, and making informed decisions in statistical analyses. Although meta-analyses can use skewness (eq. 1) and kurtosis (eq. 3) to investigate single variables, effect size statistics that compare these metrics between two groups are lacking. Thus, we propose the difference between two groups in skewness (Δsk), expressed as:

$$\Delta sk = sk_1 - sk_2 \text{ (eq. 5)}$$

and its sampling variance ($s^2_{\Delta sk}$) as:

$$s^2_{\Delta sk} = s^2_{sk_1} + s^2_{sk_2} - 2\rho_{sk} s_{sk_1} s_{sk_2} \text{ (eq. 6)}$$

Where ρ_{sk} represents the sampling correlation in skewness between the two groups (zero if assumed to be independent). Similarly, we propose the difference between two groups in kurtosis (Δku), expressed as:

$$\Delta ku = ku_1 - ku_2 \text{ (eq. 7)}$$

and its sampling variance ($s^2_{\Delta ku}$) as:

$$s^2_{\Delta ku} = s^2_{ku_1} + s^2_{ku_2} - 2\rho_{ku} s_{ku_1} s_{ku_2} \text{ (eq. 8)}$$

where ρ_{ku} represents the sampling correlation in kurtosis between the two groups (zero if assumed to be independent).

However, we note that Equations 2 and 4 assume normality for sampling variances. When the underlying distributions are skewed or heavy-tailed, sampling error variances for skewness and kurtosis (Eqs. 2 and 4) and, by extension, for their between-group contrasts (Eqs. 5-8), can misestimate uncertainty. To assess robustness and to provide distribution-free alternatives, we complemented the analytic formulas with resampling-based estimators computed within each group and summed for the difference (i.e., jackknife [33]; see our simulation study below).

Difference in correlation

Numerous meta-analyses estimate the correlation between two variables (e.g., [34,35]). To do so, researchers use the effect size statistic *Zr* [36], which can be expressed as:

$$Zr = \frac{ln\left(\frac{1+r}{1-r}\right)}{2} \text{ (eq. 9)}$$

and its sampling variance (s^2_{Zr}) [36] as:

$$s^2_{Zr} = \frac{1}{n-3} \text{ (eq. 10)}$$

where *r* is Pearson's correlation coefficient between two variables and *n* is the sample size.

Although *Zr* alone remains extremely useful to test correlational hypotheses, researchers from all fields would benefit from being able to compare *Zr* values between two groups. Although Cohen [37] proposed the difference between two groups in *Zr* as *q*, he did not provide an equation to calculate its sampling variance. Consequently, this effect size statistic has not been used despite

its potential. We therefore propose the difference between two groups in *Zr* with a new name (ΔZr), as:

$$\Delta Zr = Zr_1 - Zr_2 \text{ (eq. 11)}$$

and its sampling variance ($s^2_{\Delta Zr}$) as:

$$s^2_{\Delta Zr} = s^2_{Zr_1} + s^2_{Zr_2} - 2\rho_{Zr} s_{Zr_1} s_{Zr_2} \text{ (eq. 12)}$$

where ρ_{Zr} represents the sampling correlation in Fisher's *Zr* between the two groups (zero if assumed to be independent).

Simulation study

We conducted Monte-Carlo simulations to evaluate bias and variance estimation for our new effect sizes Δsk, Δku and ΔZr. For Δsk and Δku, we simulated independent samples for two groups from Pearson distributions with known moments using the *rpearson* function from the R package *PearsonDS* v. 1.3.2 [38]. We conducted two simulations: 1) by changing skewness between groups that involved moderate departures from normality in which group-specific skewness from $sk \in \{-1, -0.5, 0, 0.5, 1\}$ and kurtosis was fixed at 3; 2) by holding skewness constant ($sk = 0$) while manipulating kurtosis from $ku \in \{2.5, 3, 4, 5, 6\}$. In all cases, we simulated scenarios where: (i) the variance between each group was the same ($\sigma^2_2 = \sigma^2_1 = 1$) or different ($2\sigma^2_2$ versus σ^2_1); (ii) the mean between the two groups was the same ($u_2 = u_1 = 0$) or different ($u_2 = 5, u_1 = 0$). For simplicity, we assumed equal sample sizes between groups with sample size varying from $n \in \{10, 20, \ldots, 100, 150, 500\}$. We created all unique combinations of the above scenarios resulting in 1,200 independent scenarios (when considering each of the 100 scenarios at each sample size). We estimated Δsk and Δku for each scenario using formulas for within-group sample skewness with small-sample correction (Eq. 1) and excess kurtosis with small-sample correction (Eq. 3) to

estimate point estimates. To estimate associated sampling variance for Δ*sk* and Δ*ku* we used the analytical variance estimators derived here (Eqs. 2 and 4) and an associated re-sampling (jackknife) approach to compute group sampling variances separately followed by pooling. Importantly, our simulations assume no correlation between groups.

For Δ*Zr* simulations, we simulated two groups each containing two variables with known correlations within each group. For Δ*Zr* we drew bivariate normal data with target within-group correlations $r \in \{-0.8, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8\}$ using the *mvnorm* function from the package *MASS* v. 7.3.61 [39]. Marginals were standard normal and group sizes varied from $n \in \{10, 20, \ldots, 100, 150, 500\}$. We created all unique combinations of scenarios resulting in 768 unique scenarios. We estimated Δ*Zr* using Fisher's Z transformation *Zr* and calculating Δ*Zr* as the difference of *Zr* across groups (Eqs. 9–11). Sampling variance for Δ*Zr* used Eq. 10 and a jackknife approach. Again, we assumed no correlation between our groups.

Note that our simulations did not explore differences in sample size between groups. However, many groups being compared in meta-analyses have the same or very similar sample size. Additionally, simulations often show relatively small impacts of unbalanced sample sizes [40,41], which is why we originally did not vary sample size between groups in our simulations.

We resampled 2,500 times for each scenario across all simulations. Performance metrics were (a) bias of the point estimator, (b) relative bias of the sampling-variance estimator, (c) coverage (95%) and (d) Monte-Carlo standard errors (MCSEs). See supplementary material for full formulas. We also evaluated the performance of these effects for meta-analysis (see details in sections 8.4 and 9.4 of the supplementary material).

Simulation results

In all cases, we found the Monte Carlo Sampling Error (MCSEs) to be low for all our performance metrics (range of MCSEs for Δsk: 0 to 0.01; Δku: 0 to 0.624; ΔZr: 0 to 0.004). Δsk, Δku, and ΔZr point estimators exhibited small sample bias with less than 20-30 samples, except for Δku, which showed this bias below $n < 50$-60, indicating effect sizes involving kurtosis are more challenging to estimate (Fig. S1, Fig. S2). Differences in the mean and variance between groups did not differentially affect bias (Fig. S3). Regardless, small sample biases were moderate, and there was rarely a consistent over or under-estimation in point estimates across the scenarios evaluated (Fig. S1). Bias-corrected jackknife estimates reduced the small-sample bias relative to analytical bias corrected-moment estimators (mean square bias [MSB], jackknife and analytical, for Δsk: 1.109, 3.375; Δku 477.71, 891.659; ΔZr 0.029, 0.214).

In contrast to point estimators, the effectiveness of sampling variance estimators for Δsk, Δku, and ΔZr varied. Analytical sampling variance formulas for Δsk and Δku were consistently biased (Fig. S4). Jackknife resampling when combined with analytical point estimates (Fig. 2) performed the best. Under these conditions, estimators performed well when $n > 50$. In contrast, the performance of sampling variance estimators for ΔZr was best when using the analytical formulas for both the point estimator and its associated sampling variance (Fig. 2).

Coverage was close to nominal (95%) for Δsk and ΔZr across sample sizes (Fig. 2C, I). Coverage for Δku, however, was poor across many simulated scenarios (Fig. 2F). Increased sample size did not improve coverage. Poor coverage was the result of skewed sampling distributions from Jackknife approaches (Fig. S5, S6). At small sample sizes, Δku was estimated poorly when true Δku was high, leading to non-skewed distributions with good coverage. In contrast, large sample sizes improved point estimation of Δku when differences existed, but the sampling distribution

became highly skewed leading to poor coverage (Fig. S5, S6). These problems stem from the fact that the standard error formula for kurtosis assumes normality (see [42]).

Considering these simulation results, we suggest pairing the formula-based point estimators for skewness (Eq. 1) and kurtosis (Eq. 3) with jackknife standard errors for Δ*sk* and Δ*ku*. For Δ*Zr*, the standard analytic variance is recommended (Eqs. 9-12). This choice balances efficiency under normality with robustness to realistic deviations from it and aligns with our broader guidance to avoid very small group sizes for these statistics. Given the challenges in estimating Δ*ku,* and the poor properties of its sampling variance [42], we recommend weighted meta-analytic models using sample size instead of sampling variance (see supplementary material and [41]).

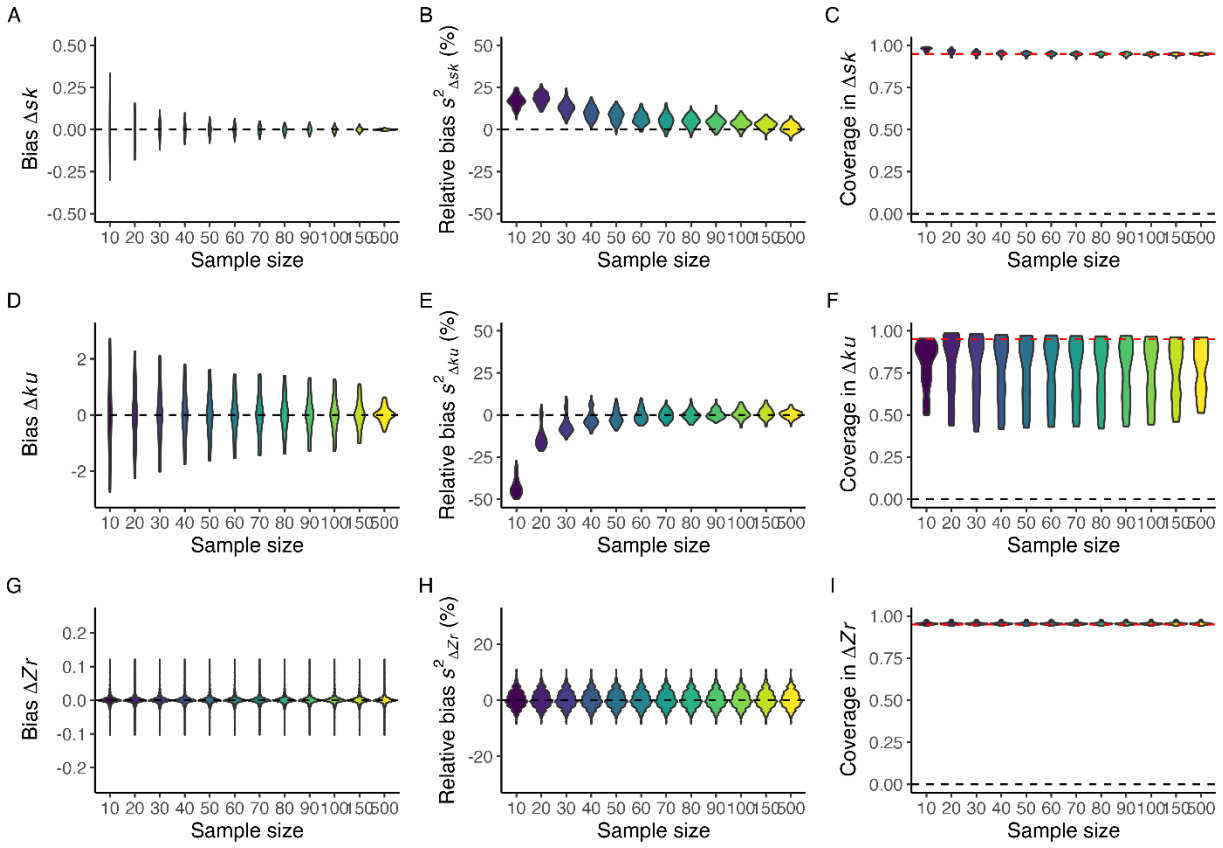


Figure 2. Bias in Δsk, Δku and ΔZr effect estimates (A, D, G), relative bias in their sampling variance using jackknife-based approximation (B, E, H), and coverage of effect estimates (C, F, I) across simulations where samples ranged in group sample sizes between $n \in \{10, 20, ..., 100, 150, 500\}$. A total of 100 simulated scenarios were assessed for Δsk and Δku whereas 64 simulated scenarios were assessed for ΔZr. We ran 2,500 simulations for each scenario. For simplicity, we only present results from our recommended point estimators and sampling variance estimators using jackknife. See supplementary material for full simulation results. The data and code needed to generate this Figure can be found in https://zenodo.org/records/18386956.

Worked examples: sex differences in mice

To illustrate the application of our proposed effect size statistics, we used data compiled by the International Mouse Phenotyping Consortium (IMPC, version 18.0; [43]; http://www.mousephenotype.org/). We examined differences between male and female mice in two pairs of traits from distinct functional domains: morphology (fat mass and heart weight) and physiology (glucose and total cholesterol). We selected these traits because they are widely understood traits, even by non-specialists, and had a large sample size (more than 10,000 individuals measured). More specifically, we assessed differences between the sexes in mean (using the natural logarithm of the response ratio [44], hereby *lnRR*), variability (using the natural logarithm of the variance ratio [14], hereby *lnVR*), skewness (using Δsk), and kurtosis (using Δku) for each trait, as well as in the difference in correlation for each trait pair (using ΔZr). The IMPC dataset contains data from multiple phenotyping centres and mice strains, so we selected the ones with the most data points for our analyses here, computing the aforementioned effect size statistics separately for each one of them.

We performed a meta-analysis for each effect size statistic to obtain a mean effect size for each trait (or pair of traits, in the case of Δ*Zr*), using 'effect size ID', 'phenotyping centre', and 'mice strain' as random factors in meta-analytical models (due to substantial heterogeneity, Table 1). In the case of Δ*ku,* we fitted a weighted meta-analytic model using sample size instead of sampling variance (see previous sections and [41]). In all these analyses, positive effect sizes denoted a greater estimate (mean, variability, skewness, kurtosis, or correlation) for males than females. We conducted all statistical analyses in the software R 4.5.1 [45]. We used the functions *moment_effects* and *cor_diff*, which have been incorporated into the package *orchaRd* v. 2.1.3 [46], to compute Δ*sk*, Δ*ku*, and Δ*Zr*. We fitted meta-analytical models using the *rma.mv* function from the package *metafor* v. 4.8-0 [47]. All methodological details and additional information can be found in our tutorial, athttps://pietropollo.github.io/new_effect_size_statistics/.

We found that males, on average, had greater fat mass and heart weight than females regardless of phenotyping centre and mice strain (Fig. 3A, B, F, G). The variability among individuals regarding these traits was also greater for males than for females, except for fat mass from one specific phenotyping centre and mice strain (Fig. 3C). By contrast, females had a similar skewness in fat mass and heart weight compared with males (Fig. 3D, I). However, Δ*sk* values for fat mass and heart weight varied across phenotyping centres and mice strains, with negative and positive values present (Fig. 3D, I). Sex differences in kurtosis for fat mass and heart weight followed a very similar pattern to the one described for skewness: Δ*ku* values overlapping zero with some variation across individual effect sizes (Fig. 3E, J). Moreover, the correlation between fat mass and heart weight was, on average, greater for females than males (Fig. 4A, B). However, this difference in correlation was absent for some phenotyping centres and mice strains (Fig. 4A, B).

We also found that male and female mice were, on average, similar in terms of blood glucose levels (Fig. 5A, B), although males had higher total cholesterol than females (Fig. 5F, G). We observed the same pattern regarding the variability of these traits: on average, the sexes were similarly variable in glucose (Fig 5C), but the variability of total cholesterol was greater in males than in females (Fig. 5H). Contrasting with morphological traits, sex differences in skewness and kurtosis were mostly absent (Fig. 5D, E, I, J). Lastly, males and females showed a similar relationship between glucose and total cholesterol, albeit this relationship was stronger for males than for females in some instances (Fig. 4C, D).

Our findings that females have, on average, lower (Fig. 3B, G), less variable (Fig. 3C, H), but similar skewness (Fig. 3D, I) and extreme values (kurtosis; Fig. 3E, I) of fat mass and heart weight compared with males may contribute to sex-related differences in the development of diseases associated with these traits and their biomarkers (e.g., QTc interval length [48]). Moreover, a stronger relationship between fat mass and heart weight in females than in males (Fig. 4B) may represent a greater risk of cardiohypertrophy arising from obesity in the former compared with the latter [49]. Meanwhile, absent or less pronounced sex differences in glucose and total cholesterol (Fig. 4) may suggest other sources of variation may contribute to sex differences in the symptomology of diseases associated with these measurements (e.g., [50–52]). Characterising sex differences in biological traits, as we have done here, can provide new perspectives on evolutionary, ecological, and medical patterns, possibly improving healthcare and environmental interventions.

Table 1. Heterogeneity estimates (I^2) for each meta-analytical model fitted in our study.

Trait(s)	**Effect size type**	I^2_{total}	$I^2_{effect\ size\ ID}$	$I^2_{phenotyping\ center}$	I^2_{strain}
Fat mass	*lnRR*	97.69	97.69	< 0.01	< 0.01

Fat mass	*lnVR*	95.71	< 0.01	33.60	62.11
Fat mass	Δ*sk*	75.61	9.82	< 0.01	65.80
Fat mass	Δ*ku*	85.81	< 0.01	< 0.01	85.81
Heart weight	*lnRR*	96.32	69.24	< 0.01	27.08
Heart weight	*lnVR*	87.15	87.15	< 0.01	< 0.01
Heart weight	Δ*sk*	68.48	38.29	30.19	< 0.01
Heart weight	Δ*ku*	97.90	< 0.01	84.60	13.30
Glucose	*lnRR*	94.76	42.67	< 0.01	52.09
Glucose	*lnVR*	70.08	< 0.01	70.08	< 0.01
Glucose	Δ*sk*	11.76	< 0.01	11.76	< 0.01
Glucose	Δ*ku*	3.60	< 0.01	1.21	2.14
Total cholesterol	*lnRR*	95.43	69.77	< 0.01	25.66
Total cholesterol	*lnVR*	94.70	84.86	< 0.01	9.84
Total cholesterol	Δ*sk*	< 0.01	< 0.01	< 0.01	< 0.01
Total cholesterol	Δ*ku*	68.94	< 0.01	68.63	0.31
Fat mass and heart weight	Δ*Zr*	64.87	< 0.01	64.87	< 0.01
Glucose and total cholesterol	Δ*Zr*	92.33	< 0.01	< 0.01	92.33

311

312

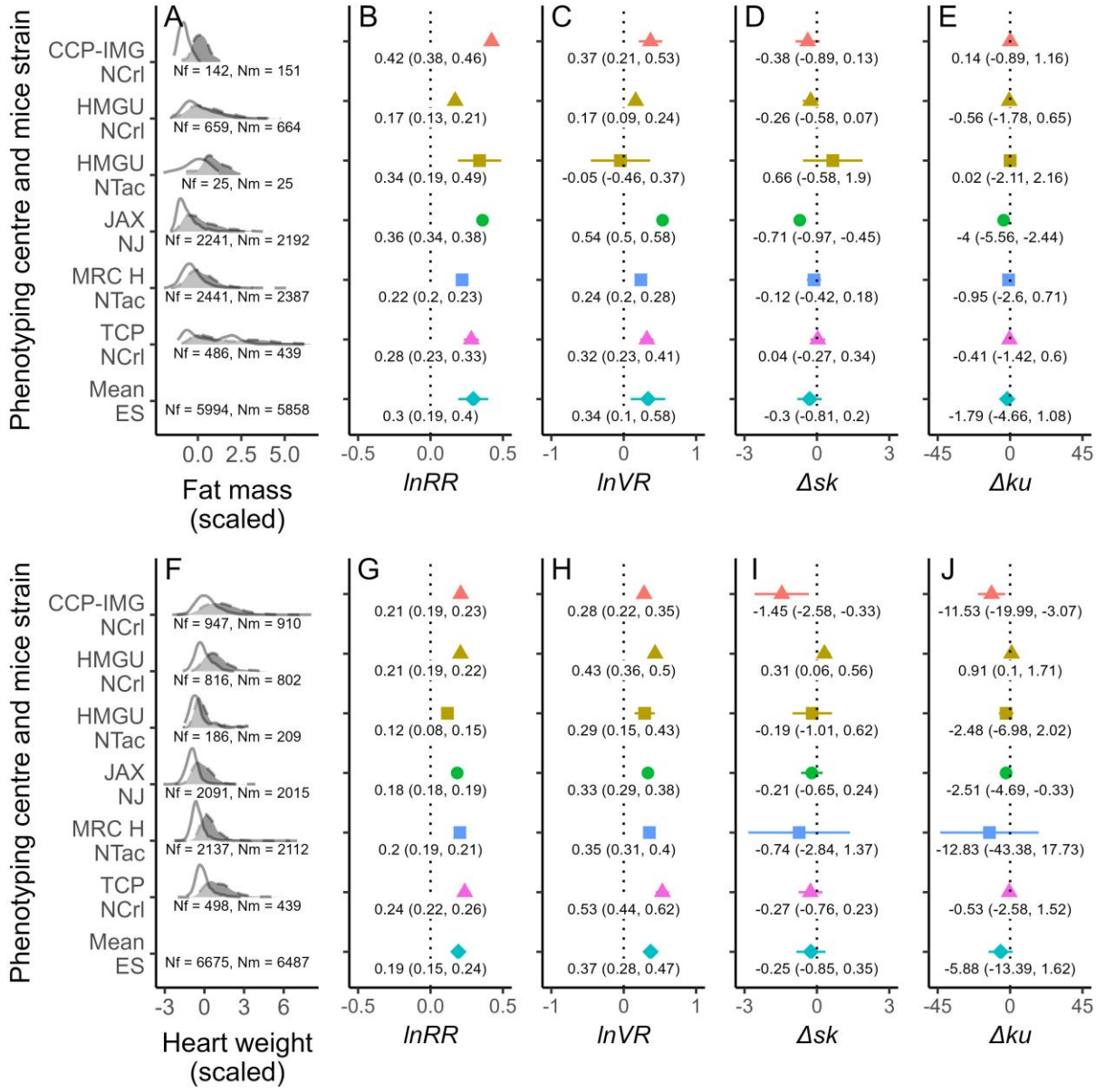
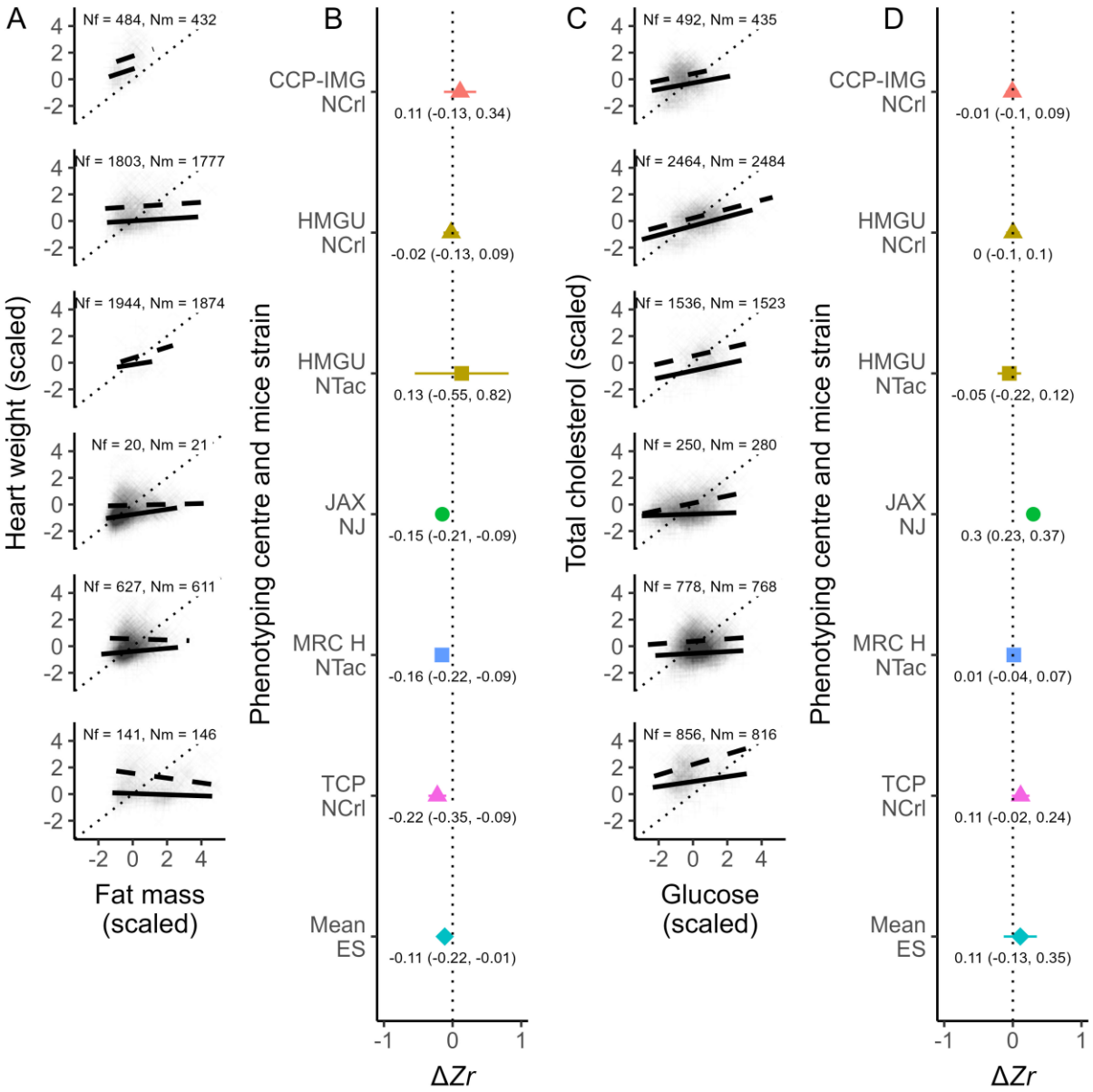


Figure 3. Examples of morphological sex differences in mice (fat mass, A-E; heart weight, F-J) for various phenotype centres (each with a different colour in panels B-E and G-J) and mice strains (each with a different shape in panels B-E and G-J), with the bottom estimate in panels B-E and G-J (turquoise diamond) representing the mean effect size. A and F show distributions of these traits (scaled by subtracting the mean from each value and then dividing the result by the standard deviation) for males (black with dashed borders) and females (white with solid borders), with the

sample size of females and males shown as Nf and Nm, respectively. Panels B-E and G-J show effect sizes (*lnRR*: natural logarithm of the response ratio; *VR*: variance ratio; Δ*sk*: difference in skewness; Δ*ku*: difference in kurtosis), with their respective point estimate and 95% confidence interval stamped. The data and code needed to generate this Figure can be found in https://zenodo.org/records/18386956.



326

Figure 4. Relationship between fat mass and heart weight (A, B) and glucose and total cholesterol (C, D) in mice. Panels A and C show these relationships (with variables scaled by subtracting the mean from each value and then dividing the result by the standard deviation) separately for males (dashed line) and females (solid line), each subpanel representing a different phenotyping centre and/or mice strain, with the sample size of females and males shown as Nf and Nm, respectively. Panels B and D then show differences in correlation (ΔZr) between males and females (point estimate and 95% confidence interval stamped), where each colour represents a distinct phenotype centre and each shape represents a distinct mice strain, with the bottom estimate in each panel (turquoise diamond) representing the mean effect size. Note that panels A and C contain individual data points, which may appear as background shading in cases with large sample sizes. The data and code needed to generate this Figure can be found in https://zenodo.org/records/18386956.

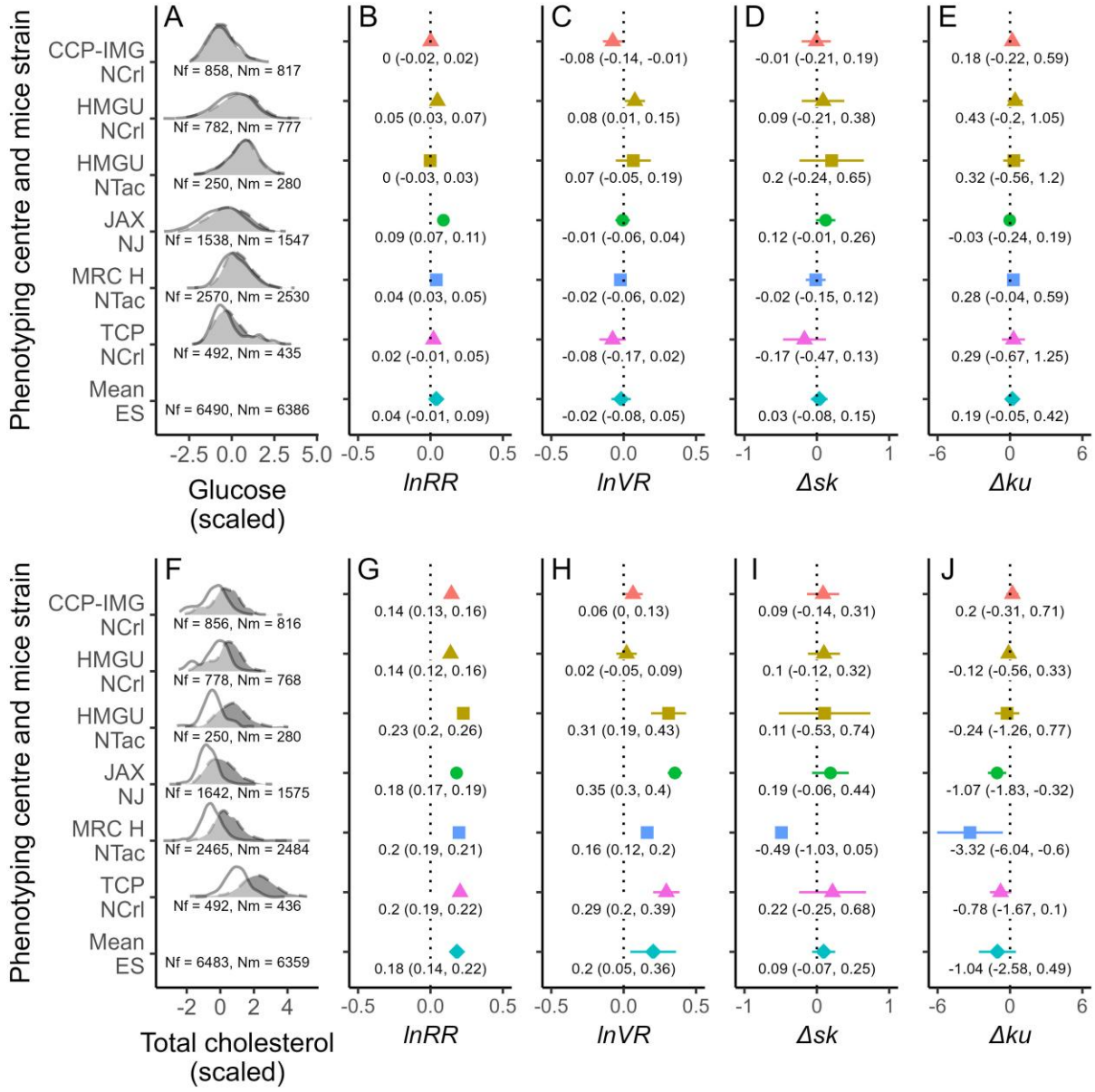


Figure 5. Examples of physiological sex differences in mice (glucose, A-E; total cholesterol, F-J) for various phenotype centres (each with a different colour in panels B-E and G-J) and mice strains (each with a different shape in panels B-E and G-J), with the bottom estimate in panels B-E and G-J (turquoise diamond) representing the mean effect size. A and F show distributions of these traits (scaled by subtracting the mean from each value and then dividing the result by the standard deviation) for males (black with dashed borders) and females (white with solid borders), with the

sample size of females and males shown as Nf and Nm, respectively. Panels B-E and G-J show effect sizes (*lnRR*: natural logarithm of the response ratio; *VR*: variance ratio; Δ*sk*: difference in skewness; Δ*ku*: difference in kurtosis), with their respective point estimate and 95% confidence interval stamped. The data and code needed to generate this Figure can be found in https://zenodo.org/records/18386956.

Limitations

Despite the enormous potential of the effect size statistics we proposed here, they are not free of limitations. For instance, skewness and kurtosis (and therefore the difference in these estimates between two groups; i.e., Δ*sk* and Δ*ku*, respectively) are more likely to become extreme with small sample sizes and with variables with few unique values, either because the variable is discrete or because it is naturally constant (e.g., number of vertebrae in mice). We thus recommend that researchers only compute Δ*sk* and Δ*ku* for continuous variables with a minimum sample size of 50 for each group (as shown in our simulations). Importantly, we found that Δ*ku* variance estimates can be biased in many situations, highlighting that exploring *Δku* should be a priority for future work. Because of this issue, meta-analysing Δ*ku* requires sample size-based weights instead of the standard sampling variance (see supplementary material and [41]). Lastly, although Δ*sk*, Δ*ku*, and Δ*Zr* can be calculated, respectively, from reported skewness, kurtosis, or within-group correlations for different samples, empirical studies rarely report these estimates. Therefore, calculating these effect sizes will probably require raw data, which, fortunately, are now becoming more readily available.

Future opportunities

The effect size statistics proposed in the present study can be useful across the life sciences, social sciences, and medicine. This is because skewness and kurtosis, and consequently differences between any two or more groups in these estimates (i.e., Δsk and Δku), may help researchers to understand epidemiological trends [53], genetic patterns relevant to medical diagnosis [20,21], disruptive selection on quantitative traits [54], body size patterns across individuals [55] and species [56], reproductive patterns [57], regime shifts in ecosystems [58], heritability [18], community assembly processes [16], and possibly many other topics. Meanwhile, comparisons regarding correlations have been used to explore memory processing during sleep [59], physiological patterns in patients with certain medical conditions [60], and selection patterns [22–24], to name a few. Because ΔZr can be used in virtually any comparison between two groups of correlational data, the opportunities for its use are endless. Most importantly, Δsk, Δku, and ΔZr are unitless measures, so they can be meta-analysed to uncover patterns between two groups (e.g., males and females). Moreover, the growing availability of raw data and big data approaches, facilitated by technological advances, makes these effect size statistics particularly valuable for modern research.

Acknowledgements

We thank Yefeng Yang for his contribution in the early stage of this study.

Data and code availability

All data and code used in this study are available at https://github.com/pietropollo/new_effect_size_statistics and https://zenodo.org/records/18386956.

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.

Author contributions

Conceptualisation: PP, SN; data curation: PP; formal analysis: PP, SMD, DWAN, SN; funding acquisition: SN; methodology: PP, SN; project administration: PP, SN; software: PP, DWAN; supervision: DWAN, SN; visualisation: PP, DWAN; writing – original draft: PP, SN; writing – review & editing: all authors.

Competing interests

We declare no competing interests.

Funding

SMD was supported by a National Science Centre (Poland; https://www.ncn.gov.pl/) grant (UMO-2020/39/B/NZ8/01274). ML was supported by an ARC (Australian Research Council; https://www.arc.gov.au/) Discovery Project grant (DP230101248). LABW was supported by an

ARC Future Fellowship grant (FT200100822). DWAN was supported by an ARC Future Fellowship (FT220100276). SN was supported by an ARC Discovery Project grant (DP210100812) and the Canada Excellence Research Chair Program (CERC-2022-00074; https://www.cerc.gc.ca/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supporting information

S1 Supplementary information. An HTML file containing all steps to reproduce simulations and meta-analyses presented in our study, as well as supplementary figures. **Fig. S1**. Bias in Δ*sk*, Δ*ku*, and Δ*Zr* effect estimates across simulations where samples ranged in group sample sizes between n ∈ {10, 20, …, 100, 150, 500}. A total of 100 simulated scenarios were assessed for Δ*sk* and Δ*ku* whereas 64 simulated scenarios were assessed for Δ*Zr*. We ran 2,500 simulations for each scenario. The data and code needed to generate this Figure can be found in https://zenodo.org/records/18386956. **Fig. S2.** Bias of analytical point estimators in relation to the absolute difference in skewness and kurtosis between groups. A) skewness and B) kurtosis. Colour of points correspond to the sample size and each point is a single simulated scenario. The dotted line is the zero bias line. The data and code needed to generate this Figure can be found in https://zenodo.org/records/18386956. **Fig. S3.** Bias for Δ*sk* and Δ*ku* for simulated scenarios was not related to group means or variances being different. We ran 2,500 simulations for each scenario. The data and code needed to generate this Figure can be found in https://zenodo.org/records/18386956. **Fig. S4.** Relative bias in Δ*sk*, Δ*ku* and Δ*Zr* effect estimates across simulations where samples ranged in group sample sizes between n ∈ {10, 20, …, 100, 150, 500}. A total of 100 simulated scenarios were assessed for Δ*sk* and Δ*ku* whereas 64 simulated

scenarios were assessed for Δ*Zr*. Note that for relative bias different combinations of point estimates and sampling variance estimates were used in their calculation as indicated in their titles which show the calculation. Notation is as follows ku and sk are the skewness and kurtosis calculated using original formulas. sk_sv and ku_sv are the sampling variance estimates using the original formulas. jack_skew_sv and jack_ku_sv are the sampling variance estimates for skewness and kurtosis using jackknife. jack_skew_bc and jack_ku_bc are the bias corrected point estimates from the jackknife. We ran 2,500 simulations for each scenario. **Fig. S5.** Coverage of 95% confidence intervals for Δ*sk*, Δ*ku* and Δ*Zr* effect estimates across simulations where samples ranged in group sample sizes between n ∈ {10, 20, …, 100, 150, 500}. A total of 100 simulated scenarios were assessed for Δ*sk* and Δ*ku* whereas 64 simulated scenarios were assessed for Δ*Zr*. We ran 2,500 simulations for each scenario. The data and code needed to generate this Figure can be found in https://zenodo.org/records/18386956. **Fig. S6.** Example sampling distributions of three different scenarios (Δ*ku* = 0, 1, or 2.5) for n = 10 and n = 500 samples for each group. We ran 2,500 simulations for each scenario. The data and code needed to generate this Figure can be found in https://zenodo.org/records/18386956. (HTML)

References

1. Maklakov AA, Lummaa V. Evolution of sex differences in lifespan and aging: causes and constraints. BioEssays. 2013;35: 717–724. doi:10.1002/bies.201300021
2. Harrison LM, Noble DWA, Jennions MD. A meta-analysis of sex differences in animal personality: no evidence for the greater male variability hypothesis. Biol Rev. 2022;97: 679–707. doi:10.1111/brv.12818

3. Zucker I, Beery AK. Males still dominate animal studies. Nature. 2010;465: 690. doi:10.1038/465690a

4. Karp NA, Mason J, Beaudet AL, Benjamini Y, Bower L, Braun RE, et al. Prevalence of sexual dimorphism in mammalian phenotypic traits. Nat Commun. 2017;8. doi:10.1038/ncomms15475

5. Zucker I, Prendergast BJ, Beery AK. Pervasive neglect of sex differences in biomedical research. Cold Spring Harb Perspect Biol. 2021;14: a039156. doi:10.1101/cshperspect.a039156

6. Clayton JA, Collins FS. Policy: NIH to balance sex in cell and animal studies. Nature. 2014;509: 282–283. doi:10.1038/509282a

7. Woitowich NC, Beery AK, Woodruff TK. A 10-year follow-up study of sex inclusion in the biological sciences. eLife. 2020;9: 1–8. doi:10.7554/eLife.56344

8. Tannenbaum C, Ellis RP, Eyssel F, Zou J, Schiebinger L. Sex and gender analysis improves science and engineering. Nature. 2019;575: 137–146. doi:10.1038/s41586-019-1657-6

9. Phillips B, Haschler TN, Karp NA. Statistical simulations show that scientists need not increase overall sample size by default when including both sexes in in vivo studies. Munafò M, editor. PLoS Biol. 2023;21: e3002129. doi:10.1371/journal.pbio.3002129

10. Drobniak SM, Lagisz M, Yang Y, Nakagawa S. Realism and robustness require increased sample size when studying both sexes. PLoS Biol. 2024;22: e3002456. doi:10.1371/journal.pbio.3002456

11. Fairbairn DJ, Blanckenhorn WU, Székely T. Sex, size and gender roles: evolutionary studies of sexual size dimorphism. Oxford, UK: Oxford University Press; 2007. doi:10.1093/acprof:oso/9780199208784.001.0001

12. Zajitschek SRK, Zajitschek F, Bonduriansky R, Brooks RC, Cornwell W, Falster DS, et al. Sexual dimorphism in trait variability and its eco-evolutionary and statistical implications. eLife. 2020;9: 1–17. doi:10.7554/eLife.63170
13. Wilson LAB, Zajitschek SRK, Lagisz M, Mason J, Haselimashhadi H, Nakagawa S. Sex differences in allometry for phenotypic traits in mice indicate that females are not scaled males. Nat Commun. 2022;13: 7502. doi:10.1038/s41467-022-35266-6
14. Nakagawa S, Poulin R, Mengersen K, Reinhold K, Engqvist L, Lagisz M, et al. Meta-analysis of variation: ecological and evolutionary applications and beyond. Methods Ecol Evol. 2015;6: 143–152. doi:10.1111/2041-210X.12309
15. McGuigan K, Van Homrigh A, Blows MW. Genetic analysis of female preference functions as function-valued traits. Am Nat. 2008;172: 194–202. doi:10.1086/588075
16. Cornwell WK, Ackerly DD. Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. Ecol Monogr. 2009;79: 109–126. doi:10.1890/07-1134.1
17. Reid JM, Arcese P, Nietlisbach P, Wolak ME, Muff S, Dickel L, et al. Immigration counter-acts local micro-evolution of a major fitness component: migration-selection balance in free-living song sparrows. Evol Lett. 2021;5: 48–60. doi:10.1002/evl3.214
18. Pick JL, Lemon HE, Thomson CE, Hadfield JD. Decomposing phenotypic skew and its effects on the predicted response to strong selection. Nat Ecol Evol. 2022;6: 774–785. doi:10.1038/s41559-022-01694-2
19. Stemkovski M, Dickson RG, Griffin SR, Inouye BD, Inouye DW, Pardee GL, et al. Skewness in bee and flower phenological distributions. Ecology. 2023;104: 1–9. doi:10.1002/ecy.3890

20. Church B V., Williams HT, Mar JC. Investigating skewness to understand gene expression heterogeneity in large patient cohorts. BMC Bioinformatics. 2019;20: 1–14. doi:10.1186/s12859-019-3252-0

21. Kulminski AM, Philipp I, Loika Y, He L, Culminskaya I. Haplotype architecture of the Alzheimer's risk in the APOE region via co-skewness. Alzheimer's Dement Diagnosis, Assess Dis Monit. 2020;12: 1–10. doi:10.1002/dad2.12129

22. Rausher MD. The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. Evolution. 1992;46: 616–626. doi:10.1111/j.1558-5646.1992.tb02070.x

23. Blows MW. Complexity for complexity's sake? J Evol Biol. 2007;20: 39–44. doi:10.1111/j.1420-9101.2006.01241.x

24. Hansen TF, Houle D. Measuring and comparing evolvability and constraint in multivariate characters. J Evol Biol. 2008;21: 1201–1219. doi:10.1111/j.1420-9101.2008.01573.x

25. Noble DWA, Radersma R, Uller T. Plastic responses to novel environments are biased towards phenotype dimensions with high additive genetic variation. Proc Natl Acad Sci U S A. 2019;116: 13452–13461. doi:10.1073/pnas.1821066116

26. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. J R Stat Soc Ser C Appl Stat. 2005;54: 507–554. doi:10.1111/j.1467-9876.2005.00510.x

27. Stasinopoulos DM, Rigby RA. Generalized additive models for location scale and shape (GAMLSS) in R. J Stat Softw. 2007;23. doi:10.18637/jss.v023.i07

28. Umlauf N, Klein N, Zeileis A. BAMLSS: Bayesian additive models for location, scale, and shape (and beyond). J Comput Graph Stat. 2018;27: 612–627. doi:10.1080/10618600.2017.1407325

29. Malgady RG. How skewed are psychological data? A standardized index of effect size. J Gen Psychol. 2007;134: 355–359. doi:10.3200/GENP.134.3.355-360

30. Riley RD, Tierney JF, Stewart LA. Individual participant data meta-analysis: a handbook for healthcare research. Hoboken, NJ: John Wiley & Sons; 2021.

31. Tierney JF, Stewart LA, Clarke M. Individual participant data. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane handbook for systematic reviews of interventions. Hoboken, NJ: John Wiley & Sons; 2024. pp. 643–658.

32. Joanes DN, Gill CA. Comparing measures of sample skewness and kurtosis. J R Stat Soc Ser D (The Statistician). 1998;47: 183–189. doi:10.1111/1467-9884.00122

33. Efron B. The jackknife, the bootstrap and other resampling plans. Philadelphia, PA: Society for Industrial and Applied Mathematics; 1982. doi:10.1137/1.9781611970319

34. Pollo P, Lagisz M, Macedo-Rego R, Mizuno A, Yang Y, Nakagawa S. Synthesis of nature's extravaganza: an augmented meta-meta-analysis on (putative) sexual signals. Ecol Lett. 2025;28: e70215.

35. Machado G, Macedo-Rego RC. Benefits and costs of female and male care in amphibians: a meta-analytical approach. Proc R Soc B Biol Sci. 2023;290: 1–12. doi:10.1098/rspb.2023.1759

36. Hedges L V., Olkin I. Statistical methods for meta-analysis. Amsterdam, Netherlands: Elsevier; 1985. doi:10.1016/C2009-0-03396-0

37. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

38. Becker M, Klößner S. PearsonDS: Pearson distribution system. 2025. Available: https://cran.r-project.org/package=PearsonDS

549 39. Venables WN, Ripley BD. Modern applied statistics with S. New York, NY: Springer New
550 York; 2002. doi:10.1007/978-0-387-21706-2
551 40. Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, White JK, et al. High-throughput
552 discovery of novel developmental phenotypes. Nature. 2016;537: 508–514.
553 doi:10.1038/nature19356
554 41. Lajeunesse MJ. Bias and correction for the log response ratio in ecological meta-analysis.
555 Ecology. 2015;96: 2056–2063. doi:10.1890/14-2402.1
556 42. Nakagawa S, Noble DWA, Lagisz M, Spake R, Viechtbauer W, Senior AM. A robust and
557 readily implementable method for the meta-analysis of response ratios with and without
558 missing standard deviations. Ecol Lett. 2023;26: 232–244. doi:10.1111/ele.14144
559 43. Wright DB, Herrington JA. Problematic standard errors and confidence intervals for
560 skewness and kurtosis. Behav Res Methods. 2011;43: 8–17. doi:10.3758/s13428-010-0044-
561 x
562 44. Hedges L V., Gurevitch J, Curtis PS. The meta-analysis of response ratios in experimental
563 ecology. Ecology. 1999;80: 1150–1156. doi:10.1890/0012-
564 9658(1999)080[1150:TMAORR]2.0.CO;2
565 45. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R
566 foundation for statistical computing; 2025. Available: https://www.r-project.org/
567 46. Nakagawa S, Lagisz M, O'Dea RE, Pottier P, Rutkowska J, Senior AM, et al. orchaRd 2.0:
568 an R package for visualising meta-analyses with orchard plots. Methods Ecol Evol.
569 2023;2023: 1–21. doi:10.1111/2041-210X.14152
570 47. Viechtbauer W. Conducting meta-analyses in R with the metafor. J Stat Softw. 2010;36: 1–
571 48. doi:10.18637/jss.v036.i03

48. Yazdanpanah MH, Bahramali E, Naghizadeh MM, Farjam M, Mobasheri M, Dadvand S. Different body parts' fat mass and corrected QT interval on the electrocardiogram: the Fasa PERSIAN Cohort Study. BMC Cardiovasc Disord. 2021;21: 1–11. doi:10.1186/s12872-021-02095-2

49. Cuspidi C, Rescaldani M, Sala C, Grassi G. Left-ventricular hypertrophy and obesity. J Hypertens. 2014;32: 16–25. doi:10.1097/HJH.0b013e328364fb58

50. Regitz-Zagrosek V. Sex and gender differences in health. EMBO Rep. 2012;13: 596–603. doi:10.1038/embor.2012.87

51. Regitz-Zagrosek V, Gebhard C. Gender medicine: effects of sex and gender on cardiovascular disease manifestation and outcomes. Nat Rev Cardiol. 2023;20: 236–247. doi:10.1038/s41569-022-00797-4

52. Kautzky-Willer A, Leutner M, Harreiter J. Sex differences in type 2 diabetes. Diabetologia. 2023;66: 986–1002. doi:10.1007/s00125-023-05891-x

53. Guharay S. A data-driven approach to study temporal characteristics of COVID-19 infection and death Time Series for twelve countries across six continents. BMC Med Res Methodol. 2025;25: 1. doi:10.1186/s12874-024-02423-y

54. Débarre F, Yeaman S, Guillaume F. Evolution of quantitative traits under a migration-selection balance: when does skew matter? Am Nat. 2015;186: S37–S47. doi:10.1086/681717

55. Poulin R, Morand S. Parasite body size distributions: interpreting patterns of skewness. Int J Parasitol. 1997;27: 959–964. doi:10.1016/S0020-7519(97)00055-6

56. Kozłowski J, Gawelczyk AT. Why are species' body size distributions usually skewed to the right? Funct Ecol. 2002;16: 419–432. doi:10.1046/j.1365-2435.2002.00646.x

595 57. Olivier LA, Higginson AD. Tests of reproductive skew theory: a review and prospectus.
596 Evol Ecol. 2023;37: 871–892. doi:10.1007/s10682-023-10263-3
597 58. Guttal V, Jayaprakash C. Changing skewness: an early warning signal of regime shifts in
598 ecosystems. Ecol Lett. 2008;11: 450–460. doi:10.1111/j.1461-0248.2008.01160.x
599 59. Verma K, Pandey K, Kashyap N. Relation between sleep spindles and semantically induced
600 false memory. Sleep Breath. 2025;29: 26. doi:10.1007/s11325-024-03186-y
601 60. Fedulovs A, Janevica J, Kruzmane L, Sokolovska J. Glucose control and variability
602 assessed by continuous glucose monitoring in patients with type 1 diabetes and diabetic
603 kidney disease. Biomed Rep. 2024;22: 23. doi:10.3892/br.2024.1901
604