Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2015, **6,** 143–152 doi: 10.1111/2041-210X.12309

Meta-analysis of variation: ecological and evolutionary applications and beyond

Shinichi Nakagawa[1]*, Robert Poulin[1], Kerrie Mengersen[2], Klaus Reinhold[3], Leif Engqvist[3], Malgorzata Lagisz[1] and Alistair M. Senior[1,4,5]

[1]*Department of Zoology, University of Otago, 340 Great King Street, Dunedin 9054,New Zealand;* [2]*Science and Engineering Faculty, Mathematical Sciences, Statistical Science, Queensland University of Technology, Brisbane QLD 4001, Australia;* [3]*Evolutionary Biology, Bielefeld University, Morgenbreede 45, 33615, Bielefeld, Germany;* [4]*The Charles Perkins Centre, The University of Sydney, Sydney, NSW 2006, Australia; and* [5]*School of Biological Sciences, The University of Sydney, Sydney, NSW 2006, Australia*

Summary

1. Meta-analysis has become a standard way of summarizing empirical studies in many fields, including ecology and evolution. In ecology and evolution, meta-analyses comparing two groups (usually experimental and control groups) have almost exclusively focused on comparing the means, using standardized metrics such as Cohen's / Hedges' *d* or the response ratio.

2. However, an experimental treatment may not only affect the mean but also the variance. Investigating differences in the variance between two groups may be informative, especially when a treatment influences the variance in addition to or instead of the mean.

3. In this paper, we propose the effect size statistic lnCVR (the natural logarithm of the ratio between the coefficients of variation, CV, from two groups), which enables us to meta-analytically compare differences between the variability of two groups. We illustrate the use of lnCVR with examples from ecology and evolution.

4. Further, as an alternative approach to the use of lnCVR, we propose the combined use of ln s (the log standard deviation) and ln $\bar{x}$ (the log mean) in a hierarchical (linear mixed) model. The use of ln s with ln $\bar{x}$ overcomes potential limitations of lnCVR and it provides a more flexible, albeit more complex, way to examine variation beyond two-group comparisons. Relevantly, we also refer to the potential use of ln s and lnCV (the log CV) in the context of comparative analysis.

5. Our approaches to compare variability could be applied to already published meta-analytic data sets that compare two-group means to uncover potentially overlooked effects on the variance. Additionally, our approaches should be applied to future meta-analyses, especially when one suspects a treatment has an effect not only on the mean, but also on the variance. Notably, the application of the proposed methods extends beyond the fields of ecology and evolution.

Key-words: systematic reviews, meta-regression, effect size, variability, dispersion, parasite behaviour manipulation, sex chromosomes, coefficient of variation

Introduction

Meta-analysis has become an indispensable quantitative tool for summarizing empirical studies not only in medical and social sciences (Egger, Smith & Altman 2001; Cooper, Hedges & Valentine 2009), but also in biological sciences and especially ecology and evolution (Nakagawa & Poulin 2012; Koricheva, Gurevitch & Mengersen 2013). In almost all fields, it is common to make comparisons between the means of a certain measurement in a treatment (experimental) group and a control group, such measurements include morphological, physiological or behavioural traits.

The most common effect size statistic for comparing two means is the standardized mean difference, often referred to as Cohen's d, or its bias-corrected metric, sometimes referred to as Hedges' g or Hedges' d (referred to as d hereafter; Hedges & Olkin 1985; Nakagawa & Cuthill 2007; Borenstein *et al.* 2009). The standardized mean difference, d and its sampling variance, s_d^2 are given by:

$$d = \frac{\bar{x}_E - \bar{x}_C}{s_{\text{pooled}}} J, \qquad \text{eqn 1}$$

$$J = 1 - \frac{3}{4(n_C + n_E - 2) - 1}, \qquad \text{eqn 2}$$

$$s_{\text{pooled}} = \sqrt{\frac{(n_C - 1)s_C^2 + (n_E - 1)s_E^2}{n_C + n_E - 2}}, \qquad \text{eqn 3}$$

$$s_d^2 = \frac{n_C + n_E}{n_C n_E} + \frac{d^2}{2(n_E + n_C)}, \qquad \text{eqn 4}$$

where $\bar{x}_C$ and $\bar{x}_E$ are the sample means of the control group (C) and experimental group (E), respectively, s_C and s_E are the stan-

*Correspondence author. E-mail: shinichi.nakagawa@otago.ac.nz

© 2014 The Authors. Methods in Ecology and Evolution © 2014 British Ecological Society

dard deviations of the two groups, n_C and n_E are the sample sizes of the two groups, and J is a bias correction for small sample sizes. Note that d is referred to as Cohen's d without J correction, but when the correction is used, d stands for Hedges' d. Some authors recommend that Equation (4) be multiplied by J^2 (Borenstein *et al.* 2009) and also that one use $2(n_E + n_C - 2)$ rather than $2(n_E + n_C)$ (Nakagawa & Cuthill 2007).

Another metric that is commonly used in the fields of ecology and evolution is the response ratio, which is the natural logarithm of the ratio between the two means (lnRR). The use of lnRR in meta-analysis was first formalized in Hedges, Gurevitch & Curtis (1999); lnRR and its sampling variance, $s^2_{\ln RR}$, are given by:

$$\ln RR = \ln\left(\frac{\bar{x}_E}{\bar{x}_C}\right), \quad \text{eqn 5}$$

$$s^2_{\ln RR} = \frac{s^2_C}{n_C\bar{x}^2_C} + \frac{s^2_E}{n_E\bar{x}^2_E}. \quad \text{eqn 6}$$

One of the main reasons for the introduction of lnRR is that d, by its construction as a standardized value, is affected not only by the difference in the means of the two groups but also by s_{pooled} (Equation 3), so that the magnitudes of standard deviations of the two groups influence the comparative effect of treatments (cf. Osenberg, Sarnelle & Cooper 1997; see also Hillebrand 2008). As one can see, lnRR is free from the effects of the standard deviations, which are only present in the variance of lnRR (Equation 6). Notably, lnRR makes sense only for ratio scale data (Houle *et al.* 2011), that is, measurements truly bounded at zero (e.g. survival time or body length).

The difference in standard deviations (i.e. variation or dispersion) between two groups may also be important in itself, although meta-analyses in the past have focused almost exclusively on differences between means (but see Vehvilainen, Koricheva & Ruohomaki 2007; Leinonen *et al.* 2008). Interestingly, ecological researchers have recently urged a shift of focus from the mean to the variance, because they have noticed that dispersion (variation) of traits within species and interspecific differences in such dispersions are informative, albeit neglected (Violle *et al.* 2012; see also Nakagawa & Schielzeth 2012). What is more, experiments and treatments are likely to not only affect the mean but also the variance (Osenberg, Sarnelle & Cooper 1997). There is, however, a dearth of statistics that allow comparison of variability between two groups in a meta-analytic framework (cf. Hedges & Nowell 1995), although modelling variability has been an active area of research (e.g. Lee & Nelder 2006). Here, we derive and describe such a statistic, which is motivated by two previous papers: Raudenbush & Bryk (1987) and Hedges, Gurevitch & Curtis (1999). Then, we provide two examples from ecology and evolution to illustrate the use of this new metric. We also describe an alternative and flexible approach, which is more complex, but may overcome potential limitations of the newly proposed metric. We end by discussing the potential usage of our approaches to meta-analyse variability across studies and their implications.

A statistic for meta-analytic comparison of variability

Based on the results from Raudenbush & Bryk (1987), an unbiased estimator of the natural logarithm of the 'population' standard deviation ($\ln \sigma$) and its sampling variance ($s^2_{\ln \sigma}$) are expressed, respectively, as:

$$\ln \hat{\sigma} = \ln s + \frac{1}{2(n-1)}, \quad \text{eqn 7}$$

$$s^2_{\ln \hat{\sigma}} = \frac{1}{2(n-1)}. \quad \text{eqn 8}$$

where $\ln \hat{\sigma}$ is an estimate of $\ln \sigma$. It is assumed that with a large sample size and sufficiently large value of σ, $\ln \sigma$ is normally distributed with variance $s^2_{\ln \sigma}$. We note that if the standard deviation (s) is estimated from the residual variance, $(n - 1)$ in Equation (8) should be replaced by the corresponding degrees of freedom. Given Equations (7) and (8), the logarithm of the ratio of the standard deviations of the experimental and control groups (lnVR, termed 'variability ratio'; cf. Hedges & Nowell 1995) and its sampling variance ($s^2_{\ln VR}$) can be expressed as:

$$\ln\ VR = \ln\left(\frac{s_E}{s_C}\right) + \frac{1}{2(n_E-1)} - \frac{1}{2(n_C-1)}, \quad \text{eqn 9}$$

$$s^2_{\ln\ VR} = \frac{1}{2(n_C-1)} + \frac{1}{2(n_E-1)}. \quad \text{eqn 10}$$

However, lnVR may be limited in its applicability, namely because when $\bar{x}_E$ is larger than $\bar{x}_C$, it is likely that s_E is larger than s_C. This dependence between the mean and variance, known as the mean–variance relationship, is common. A natural example is when the data are counts that follow a Poisson distribution, so the mean is equal to the variance. Thus, this mean–variance relationship may make the use of lnVR as a measure of differences in variability somewhat limited, especially when one wants to know a shift in variability, which accounts for an accompanying mean change.

Therefore, we propose the natural logarithm of the ratio between the coefficients of variation from two groups (lnCVR; termed 'coefficient of variation ratio') as a more general effect size statistic than lnVR for examining variability difference between the two groups. Using Equations (5) and (9), lnCVR can be expressed as:

$$\ln CVR = \ln\left(\frac{CV_E}{CV_C}\right) + \frac{1}{2(n_E-1)} - \frac{1}{2(n_C-1)}. \quad \text{eqn 11}$$

where CV_E and CV_C are $s_E/\bar{x}_E$ and $s_C/\bar{x}_C$, respectively. An advantage of this formulation is that the use of CV removes the effects of expected changes in the standard deviation due to changes in the mean. We can derive the sampling variance for Equation (11), using Equations (6) and (10) (for the details of the derivation, see Appendix 1):

$$s^2_{\ln CVR} = \frac{s_C^2}{n_C \bar{x}_C^2} + \frac{1}{2(n_C - 1)} - 2\rho_{\ln \bar{x}_C, \ln s_C}\sqrt{\frac{s_C^2}{n_C \bar{x}_C^2}\frac{1}{2(n_C - 1)}} + \frac{s_E^2}{n_E \bar{x}_E^2} + \frac{1}{2(n_E - 1)} - 2\rho_{\ln \bar{x}_E, \ln s_E}\sqrt{\frac{s_E^2}{n_E \bar{x}_E^2}\frac{1}{2(n_E - 1)}}, \quad \text{eqn 12}$$

where $\rho \ln \bar{x}_C, \ln s_C$ and $\rho \ln \bar{x}_E, \ln s_E$ are the correlations between the means and the standard deviation in the control and experimental groups on the log scale across studies. Also, when sample size is small (i.e. in a meta-analytic context equivalent to a small number of effect sizes), it may be more practical to estimate the variance above assuming $\rho \ln \bar{x}_C, \ln s_C = \rho \ln \bar{x}_E, \ln s_E$. In this approximation, a common correlation is estimated by using all means and standard deviations in the data set jointly. We now provide two examples to illustrate the use of lnCVR (all data sets and associated files used in this study are supplied as Supporting information).

Example 1: host–parasite manipulation

Poulin (1994) investigated the effect of parasite infection on host behaviour collecting the results from the experimental studies on this topic (i.e. analysing the standardized mean differences between infected groups and control groups, using Hedges' *d*; referred to as *d* hereafter). Later, Poulin (2000) expanded this data set and tested a time-lag effect (Trikalinos & Ioannidis 2005) to examine whether the effect size for parasite infection on host behaviour had declined over time. These two meta-analytic studies investigated mean differences in behaviour, but not differences in behaviour variability between infected groups and control groups. Some parasitic species are known to manipulate host behavioural means (e.g. making hosts more active or making hosts use shelters less frequently), but it is possible that some parasites may manipulate behavioural variability (Poulin & Thomas 1999; Poulin 2013). For example, by reducing host behavioural variability (increasing predictability), hosts may become more susceptible to predators, which are the next host of the parasitic species (cf. Briffa 2013).

Here, we have updated the aforementioned data set (see Supporting information for the details of this process). We conducted meta-analyses on mean behavioural differences between infected and control groups using *d* (Equation 1) and lnRR (Equation 5) and also on differences in CV using lnCVR (Equation 11). For each meta-analytic metric, we conducted two kinds of meta-analysis: (i) a normal random-effects meta-analysis (REMA) assuming independence of all effect sizes, and (ii) a multilevel meta-analysis (MLMA) accounting for correlated structures that may arise from effect sizes originating from the same studies (see Nakagawa & Santos 2012). The rationale for running REMA is that REMA is probably the most common meta-analytic model not only in the field of ecology and evolution but also in other fields, such as medical and social sciences. REMA was conducted in the *R* package *metafor* (Viechtbauer 2010), while MLMA was undertaken using the *R* package *MCMCglmm* (Hadfield 2010). The former is a likelihood-based package, while the latter is based on Bayesian MCMC (Markov chain Monte Carlo); we note that the likelihood and Bayesian methods could produce equivalent results when we use non-informative priors, as was the case for our analyses (see Supporting information for details of statistical procedures such as the settings for Bayesian priors). Further, by extending the multilevel meta-analytic models, we constructed two multilevel meta-regression models (MLMR) with each having one of the following two moderators (predictors): (i) a three-level categorical variable, denoting which of the three parasitic phyla (Platyhelminthes, Nematoda and Acanthocephala) an effect size originated from and (ii) publication year, investigating a time-lag effect, as in Poulin (2000). Note that positive *d* and lnRR values indicate that parasites manipulated host behaviour in directions expected to increase exposure to predation, whereas positive lnCVR values mean that parasites increase behavioural variability (see Supporting information for more details).

The patterns of results from the meta-analyses on mean differences were similar when quantified by both metrics (i.e. *d* and lnRR). The overall means for *d* and lnRR were positive and similar in both the random-effects and multilevel models, but the estimates for *d* and lnRR were statistically significant only in the random-effect models (i.e. 95% confidence or credible intervals not spanning across zero; see Fig. 1; REMA: meta-analytic mean for *d*, $b_{d[\text{overall mean}]} = 0{\cdot}174$, [95% confidence/credible interval], [0·050, 0·298], meta-analytic mean for lnRR, $b_{\text{lnRR[overall mean]}} = 0{\cdot}116$ [0·039, 0·193]; MLMA: $b_{d[\text{overall mean}]} = 0{\cdot}132$ [−0·121, 0·354], $b_{\text{lnRR[overall mean]}} = 0{\cdot}044$ [−0·051, 0·264]; see also Table S1 in Appendix S1). Moreover, we observed large heterogeneity (*sensu* Higgins *et al.* 2003) in the analysis of both metrics (REMA: I^2 for $d = 89{\cdot}36\%$, I^2 for lnRR = 94·64%; MLMA: I^2 for $d = 90{\cdot}78\%$, I^2 for lnRR = 96·51%; see also Table S2, for the details of I^2 for MLMA as described in Nakagawa & Santos 2012). The meta-analyses suggest that parasite-induced behavioural changes may occur, but large heterogeneity implies that such host manipulation may be species or behaviour specific. Indeed, the meta-regression models testing the effect of different phyla showed that only the species from Acanthocephala induced host behavioural changes in a direction that is expected to increase parasite transmission (MLMR: $b_{d[\text{Acanthocephala}]} = 0{\cdot}497$ [0·149, 0·891] $b_{\text{lnRR[Acanthocephala]}} = 0{\cdot}334$ [0·068, 0·590]; Fig. 1; see also Table S2). In the other meta-regression models testing a time-lag effect (originally performed in Poulin 2000), we did not find statistically significant evidence for effect sizes becoming smaller over years, but such an effect was in the expected direction (i.e. negative, MLMR: $b_{d[\text{time-lag}]} = -0{\cdot}172$ [−0·395, 0·045], $b_{\text{lnRR[time-lag]}} = -0{\cdot}138$ [−0·307, 0·008]; Fig. 1).

Unlike the results examining mean differences (i.e. *d* and lnRR), we did not find any statistically significant or notable patterns for lnCVR in meta-analytic models (REMA: $b_{\text{lnCVR[overall mean]}} = -0{\cdot}009$ [−0·095, 0·078]; MLMA: $b_{\text{lnCVR [overall mean]}} = 0{\cdot}018$ [−0·142, 0·176]) or in meta-regression models (Fig. 1; see also Table S1). However, we noted large

© 2014 The Authors. Methods in Ecology and Evolution © 2014 British Ecological Society,

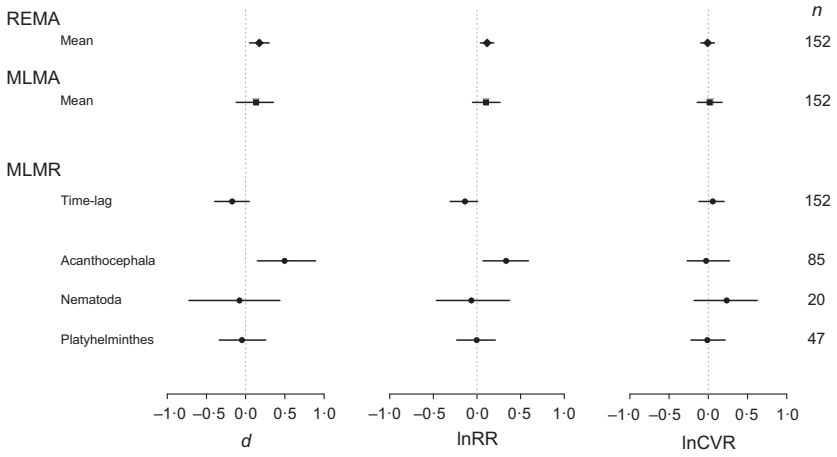

Fig. 1. Forest plots of Hedges' *d*, lnRR, lnCVR from the data set of parasitic effects on host behaviour. Point estimates and 95% confidence intervals or credible intervals (CIs) are shown. Solid diamonds are estimates from random-effects meta-analysis (REMA), while solid squares and solid circles are estimates from multilevel meta-analyses (MLMA) and multilevel meta-regression models (MLMR; note that publication year was *z*-transformed so that the regression coefficient for time-lag effect is comparable to other estimates), respectively. Sample sizes (*n*; the number of effect size values) for different estimates are listed on the right-hand side of the figure. Note that positive effect size values in *d* and lnRR indicate that parasites manipulate host behaviour to facilitate their transmission, while positive values in lnCVR mean that parasites increase variability (CV) of host behaviour.

heterogeneity in the meta-analytic models (REMA: I^2 = 90·85% MLMA: I^2 = 92·86%; Table S2). All these findings suggest that variability in behaviour may increase or decrease following parasitic infection. As far as we are aware, this is the first meta-analytic attempt to investigate changes in host behavioural variability due to parasitic infection. Clearly, we require further analysis or more studies to pinpoint exactly under what circumstances one might expect parasites to increase or decrease host behavioural variability in a manner that increases parasitic transmission.

Example 2: variability differences between the sexes

Reinhold & Engqvist (2013) tested the sex-chromosome hypothesis, which postulates that one may expect larger trait variability in the heterogametic than the homogametic sex. Indeed, the authors found strong support for this hypothesis, showing that males have a more variable body size in two male heterogametic groups (i.e. mammals and insects, excluding butterflies as this clade has homogametic males), and also that females have more variable body size in female heterogametic taxa (birds and butterflies). Interestingly, Reinhold & Engqvist (2013) used the metric termed I_{var}, defined as $\ln(CV_{male\ size}/CV_{female\ size})$. As you can see, I_{var} is conceptually identical to lnCVR above (Equation 11). However, their original analyses did not account for sampling (error) variance and phylogenetic relatedness, although some taxonomic relatedness was incorporated in the models. Here, we reanalysed the mammal and bird data sets from Reinhold & Engqvist (2013) by conducting formal meta-analyses. As in Example 1, we performed REMA and MLMA, which incorporated article and species identities (see Appendix S1 for more details). In addition, we used phylogenetic models building upon our multilevel meta-analyses, assuming a Brownian motion mode of evolution (i.e. multilevel phylogenetic meta-analyses, MLPMA; Hadfield & Nakagawa 2010; Nakagawa & Santos 2012; see also Lajeunesse, Rosenberg & Jennions 2013). To correct for phylogenetic relatedness, we used trimmed and modified versions of the mammalian tree from Bininda-Emonds *et al.* (2007) and of the avian tree from Jetz *et al.* (2012). We used the natural logarithm of the ratio between CV for the heterogametic sex (numerator) and CV for the homogametic sex (denominator) in our analysis. Therefore, positive lnCVR values indicate that the body size of the heterogametic sex is more variable than that of the homogametic sex (see Appendix S1 for more details).

In all the random-effects and multilevel meta-analyses of the mammals and birds data sets, overall means for lnCVR were positive and statistically significant (Fig. 2; REMA $b_{lnCVR[mean\ for\ mammals]}$ = 0·041 [0·012, 0·070], $b_{lnCVR[mean\ for\ birds]}$ = 0·056 [0·024, 0·089]; MLMA: $b_{lnCVR[mean\ for\ mammals]}$ = 0·046 [0·012, 0·083], $b_{lnCVR[mean\ for\ birds]}$ = 0·062 [0·026, 0·100]; Table S3). Therefore, as in the original analysis by Reinhold & Engqvist (2013), this result showed that the heterogametic sex exhibits more variability than the homogametic sex, in both mammals and birds. Our meta-analyses, however, also revealed large heterogeneity in all meta-analytic models (REMA: $I^2_{[mammals]}$ = 76·13%, $I^2_{[birds]}$ = 88·17%; MLMA: $I^2_{[mammals]}$ = 77·79%, $I^2_{[birds]}$ = 88·55%), suggesting that much of variability in effect sizes, lnCVR, is due to either differences among studies or among species (see Table S4). The incorporation of phylogenies did not change point estimates but increased the CI of the overall mean lnCVR, rendering the mean estimates statistically non-significant (MLPMA: $b_{lnCVR[mean\ for\ mammals]}$ = 0·043 [−0·032, 0·119], $b_{lnCVR[mean\ for\ birds]}$ = 0·060 [−0·002, 0·119]). We note that overall means produced by phylogenetic models could represent an ancestral value (Hadfield & Nakagawa 2010), so it is perhaps not surprising that such estimates have higher uncertainty. However, it is important to note that our point estimates from the three different meta-analytic models

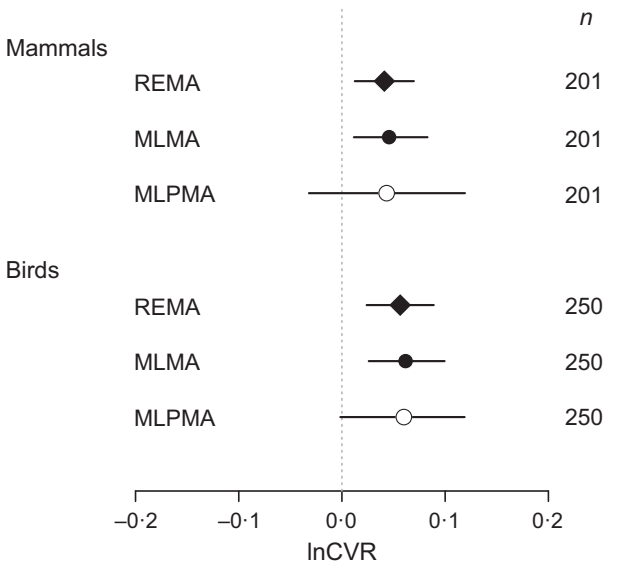

Fig. 2. A forest plot of meta-analytical comparisons of variability (CV) in the body size of the two sexes performed using lnCVR metric on two data sets: mammals and birds. Point estimates and 95% confidence intervals or credible intervals (CIs) are shown. Solid diamonds are estimates from random-effects meta-analysis (REMA), solid circles are estimates from multilevel meta-analyses (MLMA), and empty circles are estimates from multilevel phylogenetic meta-analyses (MLPMA). Sample sizes (*n*; the number of effect size values) for the two data sets are shown on the right side. Note that positive values in lnCVR mean that the heterogametic sex has higher variability (CV) in body size than the homogametic sex, as predicted by the sex-chromosome hypothesis.

are very similar, and these point estimate values are in accordance with the sex-chromosome hypothesis. Nonetheless, the results from our MLPMA indicate the importance of considering phylogeny in meta-analysis, as recently demonstrated by Chamberlain *et al.* (2012).

Limitations of lnCVR and an alternative approach

We see two potential limitations for lnCVR. First, as for the log response ratio, lnRR, the use of lnCVR is only limited to ratio scale data (note that this is not the case for $\ln \hat{\sigma}$ or lnVR). Secondly, more importantly, the use of CV in lnCVR assumes that the standard deviation is proportional to the mean. In many ecological and evolutionary data sets, such an assumption may not be supported. This is especially so given Taylor's law, also known as the power law. Taylor's law is an empirically derived relationship, which states that the variance is a power function of the mean in many biological and physical systems (Taylor 1961; see also Kilpatrick & Ives 2003). Taylor's law can be defined as:

$$s^2 = a\bar{x}^b, \qquad \text{eqn 13}$$

where a and b are some constants. As one can see, when Equation (13) holds, the standard deviation is not proportional to the mean under most circumstances. However, on the log scale, the mean and standard deviation (or variance) have a linear relationship:

$$2\ln s = \ln a + b\ln \bar{x}. \qquad \text{eqn 14}$$

Therefore, when the Taylor's law appears to hold in data, we recommend an alternative approach that is equivalent to meta-analysis using lnCVR. This alternative approach uses a hierarchical model with $\ln \hat{\sigma}$ as the response (Equation 7), and $\ln \bar{x}$ and groupings (i.e. control and experiment) as predictors. We describe this model below, but first, we feel it is also important to describe a comparable meta-analytic model for clarity.

The REMA using lnCVR such as those used in Examples 1 and 2, can be written as:

$$\ln \mathrm{CVR}_i = \mu + \tau_i + m_i, \qquad \text{eqn 15}$$

$$\tau_i \sim N(0, \sigma^2_\tau), \qquad \text{eqn 16}$$

$$m_i \sim N(0, \sigma^2_{\ln \mathrm{CVR}_i}), \qquad \text{eqn 17}$$

where $\ln\mathrm{CVR}_i$ is the effect statistic, as in Equation (11), for the ith study ($i = 1, 2, \ldots, k$; k is the number of studies or papers), μ is the overall mean (i.e. meta-analytic mean), τ_i is a random-effects term describing the deviation from μ for the ith study (i.e. a study-specific effect), which is assumed to be normally distributed around 0 with a variance of σ^2_τ (i.e. the between-study variance), m_i is a random-effects term describing the sampling variation for the ith study, which is assumed to be normally distributed with $\sigma^2_{\ln \mathrm{CVR}_i}$ (i.e. the within-study variance for the ith study). Typically, $\sigma^2_{\ln \mathrm{CVR}_i}$ is substituted by the plug-in value $s^2_{\ln \mathrm{CVR}_i}$ (Equation 12), although the variation in this estimate can also be included in the analysis (Schmid & Mengersen 2013). We can use a hierarchical/multilevel (2-level) model, or a random-slope linear mixed-effects model (Raudenbush & Bryk 2002), to compare variability between two groups, using $\ln \hat{\sigma}$ and $\ln \bar{x}$. An equivalent model to Equations (15–17) can be written as (cf. Schielzeth & Forstmeier 2009):

$$\ln \hat{\sigma}_j = (\beta_0 + \tau_i) + (\beta_1 + \varphi_i)\mathit{Group}_j + \beta_2 \ln \bar{x}_j + \varepsilon_j + m_j, \qquad \text{eqn 18}$$

$$\begin{pmatrix} \tau_i \\ \varphi_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_\tau & \rho\sigma_\tau\sigma_\varphi \\ \rho\sigma_\tau\sigma_\varphi & \sigma^2_\varphi \end{pmatrix} \right), \qquad \text{eqn 19}$$

$$\varepsilon_j \sim N(0, \sigma^2_\varepsilon), \qquad \text{eqn 20}$$

$$m_j \sim N(0, \sigma^2_{\ln \sigma_j}), \qquad \text{eqn 21}$$

where $\ln \hat{\sigma}_j$ is the jth effect size as in Equation (7) ($j = 1, 2, \ldots, n$; n is the number of effect sizes), $\ln \bar{x}_j$ is the mean estimate for the jth effect size, *Group* is a (binary) dummy variable (e.g. the control group = 0 and the treatment group = 1), β_0 is the grand intercept (the overall mean for control groups), β_1 is the grand slope or regression coefficient for *Group* (it is perhaps most intuitive to think β_1 as the difference between control and treatment groups), β_2 is the slope or regression coefficient for $\ln \bar{x}$, τ_i (random intercept) is the deviation from β_0 for the ith study ($i = 1, 2, \ldots, k$), φ_i (random slope) is the deviation from β_1 for the ith study, τ_i and φ_i have a multivariate normal

distribution with the variance–covariance structure specified in Equation 19 (ρ is the correlation between τ_i and φ_i), and ε_j is the jth residual value which is normally distributed with σ^2_{ε}, and m_j is a sampling error effect for the jth effect size, normally distributed with $\sigma^2_{\ln \sigma_j}$, which is the sample variance for the jth effect size (in practice, $s^2_{\ln \hat{\sigma}_i}$ as in Equation (8) is usually used). In this formulation, we are assuming that each study has only one pair of control and treatment groups, but this does not need to be the case (e.g. more than one pairs of such groups can come from one study, see Example 3 and Appendix S1). Importantly, the overall mean (μ) in Equation (15) and the regression coefficient (β_1) in Equation (18) are the equivalent parameters of interest, because they both represent the difference in variability between the two groups. We can flexibly change Equation (18) to make this model suitable for different types of data. For example, if we use $\bar{x}$ and possibly the square term of $\bar{x}$, instead of $\ln \bar{x}$, then there is no requirement for ratio scale data. That is, sets of interval scale data, which can span around 0 (e.g. the degree Celsius, °C), can also be used to compare variability. Additionally, the term relating to the mean ($\bar{x}$) is not necessary if the mean and standard deviation are independent (i.e. no mean–variance relationship). We note that one possible shortcoming of this hierarchical model formulation is that we assume that $\bar{x}$ is estimated without error. Clearly, this is usually not the case. However, often the variation induced by this estimate is negligible compared with the total variation in the data, so can be justifiably ignored (cf. Raudenbush & Bryk 1987, 2002). This is analogous to the common practice of ignoring the variability associated with the estimation of the within-study variance. Alternatively, we could accommodate these sources of variation in the model by including measurement error, often called measurement error models (i.e. explicitly modelling sample variance in predictors; see Buonaccorsi 2010). We could also employ a bivariate-response meta-analytic model by using the first response as $\ln \hat{\sigma}$ with sampling (error) variance, $1/2\,(n-1)$ (as in Equation 8) and the second response as $\ln \bar{x}$ with the sampling variance, $s^2/n\bar{x}^2$ (cf. Equation 6), thus simultaneously modelling both effects with variation (see Raudenbush & Bryk 2002; Nam, Mengersen & Garthwaite 2003). A difficulty of such a bivariate meta-analytic model is that we do not know sampling (error) covariance (Riley 2009), although we note some solutions to this issue have been proposed (Riley, Thompson & Abrams 2008). Obviously, the implementation of bivariate meta-analytic models, as well as measurement error models, is more complex than the random-slope mixed models described above (Equation 18) and beyond the scope of this study.

Example 3: reanalysis of variability differences between the sexes

We used models based on Equation (18) to reanalyse data sets from Example 2. We fitted two kinds of random-slope mixed-effects models (i.e. without phylogeny, RSMM and with phylogeny, PRSMM) for both mammal and bird data sets (see Appendix S1 for more details). As seen in Fig. 3, our main assumption that $\ln s$ and $\ln \bar{x}$ have a linear relationship seemed to be well supported (Fig. 3b,d). In contrast (and as anticipated given the previous observation), the linearity between s and $\bar{x}$ seemed to be less well supported with variation in s increasing with increasing $\bar{x}$ (i.e. heteroscedasticity; Fig. 3a,c). Nonetheless, our results from the random-slope mixed models are largely consistent with those from models using lnCVR (Example 2). The *Group* effect (as described in Equation 18) was in the expected direction (Fig. 4). In other words, the heterogametic sex showed more variation in body size in both mammals and birds regardless of phylogenetic corrections in the models (RSMM: $b_{\ln \hat{\sigma}[Group\ \text{for mammals}]} = 0{\cdot}036$ [0·006–0·064], $b_{\ln \hat{\sigma}[Group\ \text{for birds}]} = 0{\cdot}053$ [0·026–0·090]; PRSMM: $b_{\ln \hat{\sigma}[Group\ \text{for mammals}]} = 0{\cdot}039$ [0·010–0·067], $b_{\ln \hat{\sigma}[Group\ \text{for birds}]} = 0{\cdot}053$ [0·026 = 1–0·088]; details of results in Table S5). These results are in line with the original conclusions of Reinhold & Engqvist (2013), which find support for the sex-chromosome hypothesis.

An extension: phylogenetic comparative analysis

The hierarchical/multilevel model described above (Equation 18) can be used in the context of comparative biology (Hadfield & Nakagawa 2010; Nakagawa & Santos 2012). For example, we may be interested in evaluating interspecific differences in variability of a trait, after accounting for sampling (error) variability (e.g. leaf size; cf. Violle *et al.* 2012). We may also want to test a hypothesis regarding how aspects of species' distribution (e.g. altitude) explain such variability. In such instances, a model can be written as:

$$\ln \hat{\sigma}_l = \beta_0 + \beta_1 w_1 + \beta_2 \ln \bar{x}_l + a_l + u_l + m_l, \quad \text{eqn 22}$$

$$a_l \sim N(0, \sigma^2_a \mathbf{A}), \quad \text{eqn 23}$$

$$u_l \sim N(0, \sigma^2_u), \quad \text{eqn 24}$$

$$m_l \sim N(0, s^2_{\ln \hat{\sigma}_l}), \quad \text{eqn 25}$$

where $\ln \hat{\sigma}_l$ is the estimated log standard deviation of a trait of interest (e.g. leaf size) for the lth species ($l = 1, 2, \ldots, t$; t is the number of species), $\ln \bar{x}_l$ is the log trait mean for the lth species, w_l is a value of a predictor for the lth species (e.g. altitude), β_0 is the intercept, β_1 is the slope or regression coefficient for w, β_2 is the slope or regression coefficient for $\ln \bar{x}$, a_l is the phylogenetic effect for the lth species, which is normally distributed with $\sigma^2_a \mathbf{A}$ where σ^2_a is the variance due to phylogeny and $\mathbf{A}$ is a l by l matrix of distances between species derived from a phylogenetic tree (see below), u_l is the species-specific effect (including a residual) for the lth species, which is normally distributed with a variance of σ^2_u, and m_l is sampling variance for the lth species and is normally distributed with a variance of $s^2_{\ln \hat{\sigma}_l}$ (as in Equation 8). A group of models, which incorporate phylogenetic relatedness (e.g. using $\mathbf{A}$ in Equation 23), are often referred to as phylogenetic comparative methods (Paradis 2012). Thus, this model (Equation 22) can be seen as another example of just such a method. The correlation matrix $\mathbf{A}$ can be changed according

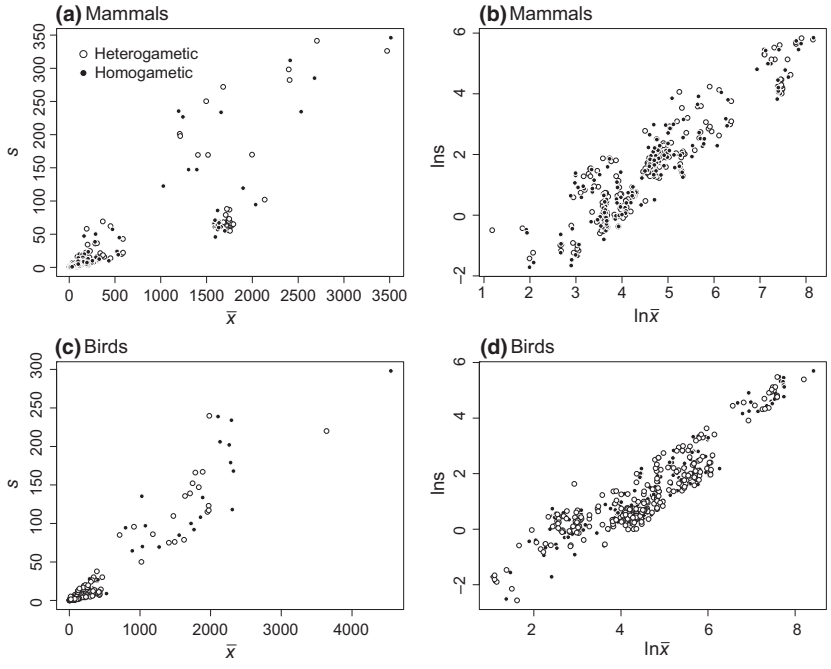

Fig. 3. Scatter plots of the standard deviation and the mean for the mammalian data set (a) and for the avian data set (c), and of the log standard deviation and the log mean for the mammalian data set (b) and for the avian data set (d); empty data points are for the heterogametic sex, while solid data points are for the homogametic sex.

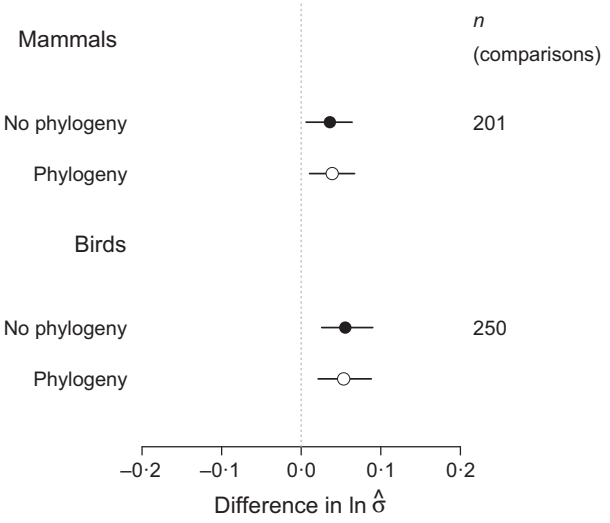

Fig. 4. A forest plot of comparisons of variability (ln $\hat{\sigma}$) in the body size of the two sexes performed using random-slope mixed-effects models, RSMM (with and without phylogeny) on two data sets: mammals and birds. Point estimates and 95% credible intervals (CIs) are shown. Solid circles are estimates from random-slope mixed-effects models without phylogeny (RSMM), and empty circles are estimates from random-slope mixed-effects models with phylogeny (PRSMM), Sample sizes (*n*; the number of group pairs or comparisons) for the two data sets are shown on the right side. Note that positive values in the difference in ln $\hat{\sigma}$, that is $\ln(\hat{\sigma}_E/\hat{\sigma}_C)$, means that the heterogametic sex has higher variability (ln $\hat{\sigma}$) in body size than the homogametic sex, as predicted by the sex-chromosome hypothesis.

to different models of evolution (Nakagawa & Santos 2012; Paradis 2012); note that in the above analyses, we used the Brownian motion model of evolution, which assumes that trait differences are proportional to phylogenetic distances. As discussed earlier, we could employ a bivariate-response model with $\ln \bar{x}$ and $\ln \hat{\sigma}$. In such a model, one could investigate what are termed, 'phylogenetically heritable', 'additive phylogenetic' or just 'phylogenetic' correlations (*sensu* Lynch 1991; Housworth, Martins & Lynch 2004); this correlation is the phylogenetic equivalent of a genetic correlation in quantitative genetics. Interestingly, the phylogenetic correlation could represent the degree of co-evolution or independent evolution between the mean and variance of the trait.

We also note that if the mean and standard deviation are proportional, the use of the log CV (lnCV) could be a useful simplification in the context of comparative analysis. The effect statistic, lnCV and its sampling variance can be given by (see Appendix 1):

$$\ln \mathrm{CV} = \ln s - \ln \bar{x} + \frac{1}{2(n-1)}, \qquad \text{eqn 26}$$

$$s^2_{\ln \mathrm{CV}} = \frac{s^2}{n\bar{x}^2} + \frac{1}{2(n-1)} - 2\rho_{\ln \bar{x}, \ln s}\sqrt{\frac{s^2}{n\bar{x}^2}\frac{1}{2(n-1)}}. \qquad \text{eqn 27}$$

Then, the model equivalent to Equation (22) using lnCV is:

$$\ln \mathrm{CV}_l = \beta_0 + \beta_1 w_l + a_l + s_l + m_l. \qquad \text{eqn 28}$$

In comparative analysis, ecologists and evolutionary biologists have already used CV as a statistic (e.g. Shine & Seigel 1996; Garcia-Gonzalez *et al*. 2012). However, CV is bounded at zero so that in many cases, CV may not conform to the normality assumption, when used as the response variable in a statistical model. The use of lnCV as well as ln $\hat{\sigma}$, which are both unbounded, may prove a useful solution to this shortcoming of CV. Finally, we can easily extend the relatively simple comparative models above to more complex models, by adding any arbitrary number of fixed and random factors as required (such models can be implemented, for example, in

MCMCglmm; see Hadfield & Nakagawa 2010; de Villemereuil & Nakagawa 2014).

Discussion

In this paper, we have proposed the meta-analytic metric, lnCVR (and lnVR), to compare differences in variability between two groups. As one can see from Examples 2 and 3, the groups do not necessarily need to pertain to a treatment (experimental) group and a control group, as long as the pairwise comparison is biologically or physically meaningful. An advantage of the proposed meta-analytic metric, lnCVR (Equation 11) is that it can be used along with all previously developed statistical tools for meta-analysis and meta-regression. For instance, although not implemented in our examples, there are numerous methods to examine publication bias in meta-analytic data such as funnels plots, Egger's tests and trim-and-fill methods (reviewed in Rothstein, Sutton & Borenstein 2005; for examples of funnels plots, see Figs S1 and S2 in Appendix S1), all of which can be conducted on meta-analysis of variance using lnCVR. Importantly, lnCVR can be used in conjunction with any standard meta-analytic package currently available (reviewed in Schmid *et al.* 2013).

However, we have also outlined the limitations of using lnCVR. We recommend that one check for linearity between the mean and the standard deviation before the use of lnCVR in one's meta-analysis. As shown above, our proposed alternative method using $\ln s$ and $\ln \bar{x}$, in conjunction with RSMM, is likely to overcome the limitations of lnCVR. Further, this approach provides a very flexible platform for not only comparing variability between two groups, but also for modelling variability meta-analytically or comparatively under numerous circumstances. A disadvantage of the random-slope mixed-effects method is its relative complexity compared to the use of lnCVR; unlike lnCVR, we cannot use many of the statistical tools developed for meta-analysis (e.g. quantification of publication bias). Therefore, both approaches should probably be used in a complementary way, bearing in mind the assumption associated with the use of lnCVR.

Another issue may be how one interprets magnitudes of effect size in the proposed metric lnCVR and when using $\ln s$ (with $\ln \bar{x}$). Famously, Cohen (1988) established benchmarks for traditional metrics such as d and r (correlation coefficient) to help practical, clinical or biological interpretation of effect sizes. For example, d values of 0·3, 0·5 and 0·8 are considered to be small, moderate and large, respectively. Unfortunately, Cohen's benchmarks cannot be obtained for metrics such as lnRR, lnVR and lnCVR. However, these three metrics can be interpreted in a very intuitive way because they represent ratios between two values once back-transformed to the original scale. For example, the value of 0·05 in lnCVR is *c.* 1·051 on the original scale (i.e. the exponentiation of 0·05), and it can be interpreted as the CV of the experimental group (or the group in the numerator) being 5·1% higher than that of the control group (or the group in the denominator). In a similar manner, the *Group* effect (Equation 18) represents the difference in $\ln s$ between the two groups after correction for differences in the mean; in other words, the ratio of standard deviations between the two groups on the log scale, that is $\ln(\hat{\sigma}_E/\hat{\sigma}_C)$. Therefore, the *Group* effect of 0·05 can be interpreted as the standard deviation of the experimental group (coded as 1 in the dummy variable) being 5·1% higher than that of the control group (coded as 0 in the dummy variable) after controlling for $\ln \bar{x}$ (assuming the model includes $\ln \bar{x}$ as for Equation 18).

As mentioned earlier, our proposed methods are most useful when one expects an effect to induce changes not only in the mean but also in the variability of a measure. A familiar example of this may be that climate change could induce not only changes in the average temperature but also changes in temperature variability (e.g. Schar *et al.* 2004). Although we provided examples from ecology and evolution, we see applications of the proposed methods in a number of fields. The first ever meta-analysis using the standardized mean difference (Glass's Δ, which is a special case of Cohen's/Hedges' d) examined the effectiveness of psychotherapy and showed its considerable efficacy (Smith & Glass 1977). Another finding of this pioneering study was that there were little differences (in means) among various types of psychotherapy (e.g. behavioural and non-behavioural therapies). It remains to be tested whether outcomes of different types of therapy differ in variability and thus in their reliability. Additionally, in medical fields, it may be desirable to reduce variance rather than means using a treatment or a drug (e.g. for heart rate or sugar levels in the blood).

Finally, there have been hundreds, if not thousands, of meta-analyses comparing the means of two groups (Nakagawa & Poulin 2012; Koricheva, Gurevitch & Mengersen 2013). We point out that our methods of comparing variability between two groups can be retrospectively applied to these past meta-analyses to gain additional insights. Notably, studies making use of d and lnRR have already gathered all the data necessary to quantify lnCVR or the other relevant metrics outlined here. Furthermore, in future meta-analyses, we recommend examining not only mean differences but also variability differences when comparing two groups. In particular, paying more attention to changes in variance will facilitate a better understanding of phenomena not only in biological fields but also medical and social sciences.

Acknowledgements

We are thankful to Holger Schielzeth, David Fletcher, two anonymous reviewers and the Editor, Bob O'Hara for comments, which improved our manuscript. Further, S.N. greatly benefited from the insightful discussions on the method with the members of the Society for Research Synthesis Methodology (SRSM), especially, Larry Hedges and Elena Kulinskaya. S.N. gratefully acknowledges the Rutherford Discovery Fellowship (New Zealand). L.E. is funded by German Research Foundation (DFG) (project number FOR 1232).

Data accessibility

All data files and R scripts, which were used in this work, are available in the Supporting information.

References

Bininda-Emonds, O.R.P., Cardillo, M., Jones, K.E., MacPhee, R.D.E., Beck, R.M.D., Grenyer, R. *et al.* (2007) The delayed rise of present-day mammals. *Nature*, **446**, 507–512.

Borenstein, M., Hedges, L.V., HigginS, J.P.I. & Rothstein, H.R. (2009) *Introduction to Meta-Analysis*. Wiley, Chichester.

Briffa, M. (2013) Plastic proteans: reduced predictability in the face of predation risk in hermit crabs. *Biology Letters*, **9**, 20130592.

Buonaccorsi, J.P. (2010) *Measurement Error: Models, Methods, and Applications*. CRC Press, Boca Raton.

Chamberlain, S.A., Hovick, S.M., Dibble, C.J., Rasmussen, N.L., Van Allen, B.G., Maitner, B.S. *et al.* (2012) Does phylogeny matter? Assessing the impact of phylogenetic information in ecological meta-analysis. *Ecology Letters*, **15**, 627–636.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Lawrence Erlbaum, Hillsdale, New Jersey.

Cooper, H., Hedges, L.V. & Valentine, J.C. (2009) *The Handbook of Research Synthesis and Meta-Analysis*, 2nd edn. Russell Sage Foundation, New York.

Egger, M., Smith, G.D. & Altman, D.G. (2001) *Systematic Reviews in Health Care: Meta-Analysis in Context*, 2nd edn. BMJ, London.

Garcia-Gonzalez, F., Simmons, L.W., Tomkins, J.L., Kotiaho, J.S. & Evans, J.P. (2012) Comparing evolvabilities: common errors surrounding the calculation and use of coefficients of additive genetic variation. *Evolution*, **66**, 2341–2349.

Hadfield, J. (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, **33**, 1–22.

Hadfield, J. & Nakagawa, S. (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology*, **23**, 494–508.

Hedges, L.V., Gurevitch, J. & Curtis, P.S. (1999) The meta-analysis of response ratios in experimental ecology. *Ecology*, **80**, 1150–1156.

Hedges, L.V. & Nowell, A. (1995) Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, **269**, 41–45.

Hedges, L. & Olkin, I. (1985) *Statistical Methods for Meta-Analysis*. Academic Press, New York.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J. & Altman, D.G. (2003) Measuring inconsistency in meta-analyses. *British Medical Journal*, **327**, 557–560.

Hillebrand, H. (2008) Meta-analysis in ecology. *Encyclopedia of Life Sciences*, doi: 10.1002/9780470015902.a9780470003272. Wiley, Chichester.

Houle, D., Pelabon, C., Wagner, G.P. & Hansen, T.F. (2011) Measurement and meaning in biology. *Quarterly Review of Biology*, **86**, 3–34.

Housworth, E.A., Martins, E.P. & Lynch, M. (2004) The phylogenetic mixed model. *American Naturalist*, **163**, 84–96.

Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K. & Mooers, A.O. (2012) The global diversity of birds in space and time. *Nature*, **491**, 444–448.

Kilpatrick, A.M. & Ives, A.R. (2003) Species interactions can explain Taylor's power law for ecological time series. *Nature*, **422**, 65–68.

Koricheva, J., Gurevitch, J. & Mengersen, K. (2013) *The Handbook of Meta-Analysis in Ecology and Evolution*. Princeton University Press, Princeton.

Lajeunesse, M.J., Rosenberg, M.S. & Jennions, M.D. (2013) Phylogenetic nonindependence and meta-analysis. *The Handbook of Meta-Analysis in Ecology and Evolution* (eds J. Koricheva, J. Gurevitch & K. Mengersen), pp. 284–299. Princeton University Press, Princeton.

Lee, Y. & Nelder, J.A. (2006) Double hierarchical generalized linear models. *Journal of the Royal Statistical Society Series C-Applied Statistics*, **55**, 139–167.

Leinonen, T., O'Hara, R.B., Cano, J.M. & Merila, J. (2008) Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *Journal of Evolutionary Biology*, **21**, 1–17.

Lynch, M. (1991) Methods for the analysis of comparative data in evolutionary biology. *Evolution*, **45**, 1065–1080.

Nakagawa, S. & Cuthill, I.C. (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, **82**, 591–605.

Nakagawa, S. & Poulin, R. (2012) Meta-analytic insights into evolutionary ecology: an introduction and synthesis. *Evolutionary Ecology*, **26**, 1085–1099.

Nakagawa, S. & Santos, E.S.A. (2012) Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*, **26**, 1253–1274.

Nakagawa, S. & Schielzeth, H. (2012) The mean strikes back: mean-variance relationships and heteroscedasticity. *Trends in Ecology & Evolution*, **27**, 474–475.

Nam, I.S., Mengersen, K. & Garthwaite, P. (2003) Multivariate meta-analysis. *Statistics in Medicine*, **22**, 2309–2333.

Osenberg, C.W., Sarnelle, O. & Cooper, S.D. (1997) Effect size in ecological experiments: the application of biological models in meta-analysis. *American Naturalist*, **150**, 798–812.

Paradis, E. (2012) *Analysis of Phylogenetics and Evolution with R*, 2nd edn. Springer, New York.

Poulin, R. (1994) Meta-analysis of parasite-induced behavioral changes. *Animal Behaviour*, **48**, 137–146.

Poulin, R. (2000) Manipulation of host behaviour by parasites: a weakening paradigm? *Proceedings of the Royal Society of London Series B-Biological Sciences*, **267**, 787–792.

Poulin, R. (2013) Parasite manipulation of host personality and behavioural syndromes. *Journal of Experimental Biology*, **216**, 18–26.

Poulin, R. & Thomas, F. (1999) Phenotypic variability induced by parasites: extent and evolutionary implications. *Parasitology Today*, **15**, 28–32.

Raudenbush, S.W. & Bryk, A.S. (1987) Examining correlates of diversity. *Journal of Educational and Behavioral Statistics*, **12**, 241–269.

Raudenbush, S.W. & Bryk, A.S. (2002) *Hierarchical Linear Models: Application and Data Analysis Methods*. Sage, Thousand Oaks.

Reinhold, K. & Engqvist, L. (2013) The variability is in the sex chromosomes. *Evolution*, **67**, 3662–3668.

Riley, R.D. (2009) Multivariate meta-analysis: the effect of ignoring within-study correlation. *Journal of the Royal Statistical Society Series A-Statistics in Society*, **172**, 789–811.

Riley, R.D., Thompson, J.R. & Abrams, K.R. (2008) An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics*, **9**, 172–186.

Rothstein, H., Sutton, A.J. & Borenstein, M. (2005) *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. John Wiley, Chichester.

Schar, C., Vidale, P.L., Luthi, D., Frei, C., Haberli, C., Liniger, M.A. & Appenzeller, C. (2004) The role of increasing temperature variability in European summer heatwaves. *Nature*, **427**, 332–336.

Schielzeth, H. & Forstmeier, W. (2009) Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology*, **20**, 416–420.

Schmid, C.H. & Mengersen, K. (2013) Bayesian meta-analysis. *The Handbook of Meta-Analysis in Ecology and Evolution* (eds J. Koricheva, J. Gurevitch & K. Mengersen), pp. 145–173. Princeton University Press, Princeton.

Schmid, C.H., Stewart, G.B., Rothstein, H.R., Lajeunesse, M.J. & Gurevitch, J. (2013) Software for statistical meta-analysis. *The Handbook of Meta-Analysis in Ecology and Evolution* (eds J. Koricheva, J. Gurevitch & K. Mengersen), pp. 174–191. Princeton University Press, Princeton.

Shine, R. & Seigel, R.A. (1996) A neglected life-history trait: clutch-size variance in snakes. *Journal of Zoology*, **239**, 209–223.

Smith, M.L. & Glass, G.V. (1977) Meta-analysis of psychotherapy outcome studies. *American Psychologist*, **32**, 752–760.

Taylor, L.R. (1961) Aggregation, variance and the mean. *Nature*, **189**, 65–68.

Trikalinos, T.A. & Ioannidis, J.P. (2005) Assessing the evolution of effect sizes over time. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments* (eds H. Rothstein, A.J. Sutton & M. Borenstein), pp. 241–259. John Wiley, Chichester.

Vehvilainen, H., Koricheva, J. & Ruohomaki, K. (2007) Tree species diversity influences herbivore abundance and damage: meta-analysis of long-term forest experiments. *Oecologia*, **152**, 287–298.

Viechtbauer, W. (2010) Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, **36**, 1–48.

de Villemereuil, P. & Nakagawa, S. (2014) General quantitative genetic methods for comparative biology. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology* (ed. L.Z. Garamszegi), pp. 287–303. Springer, Berlin, Heidelberg.

Violle, C., Enquist, B.J., McGill, B.J., Jiang, L., Albert, C.H., Hulshof, C., Jung, V. & Messier, J. (2012) The return of the variance: intraspecific variability in community ecology. *Trends in Ecology & Evolution*, **27**, 244–252.

Received 5 March 2014; accepted 5 November 2014
Handling Editor: Robert B. O'Hara

Appendix 1: Derivation of the variance for lnCVR (the logarithm of the ratio between two coefficients of variation)

We use two basic properties of variance. When we have two random variables A and B and a constant is c, then:

$$\mathrm{var}(A - B) = \mathrm{var}(A) + \mathrm{var}(B) - 2r_{A,B}\sqrt{\mathrm{var}(A)\mathrm{var}(B)} \tag{A1}$$

$$\mathrm{var}(A + c) = \mathrm{var}(A), \tag{A2}$$

where 'var' indicates variance, and $r_{A,\ B}$ is the correlation between A and B. Equation (11) (the main text) can be rewritten as:

$$\ln \mathrm{CVR} = \ln s_E - \ln \bar{x}_E + \frac{1}{2(n_E - 1)} - \left(\ln s_C - \ln \bar{x}_C + \frac{1}{2(n_C - 1)}\right), \tag{A3}$$

where the symbols are as in eqns 5–12. Because both $1/2(n_E - 1)$ and $1/2(n_C - 1)$ are bias correction factors, these can be regarded as scalars (i.e. c in Equation A2). As in Hedges, Gurevitch & Curtis (1999) and Raudenbush & Bryk (1987), we assume that with large sample sizes, all $\ln \bar{x}_C$, $\ln \bar{x}_E$, $\ln s_C$ and $\ln s_E$ are normally distributed. The means and the variances ($\ln \bar{x}_C$ and $\ln s_C$, and $\ln \bar{x}_E$ and $\ln s_E$, respectively) covary, whereas the means for the two groups ($\ln \bar{x}_C$ and $\ln \bar{x}_E$) and the variances ($\ln s_C$ and $\ln s_E$) are independent from each other. Thus, using Equations (6) and (10) and Equations (A1) and (A2), the variance of lnCVR can be expressed as:

$$s^2_{\ln \mathrm{CVR}} = \frac{s_C^2}{n_C \bar{x}_C^2} + \frac{1}{2(n_C - 1)} + \frac{s_E^2}{n_E \bar{x}_E^2} + \frac{1}{2(n_E - 1)} - 2\rho_{\ln \bar{x}_E, \ln s_E}\sqrt{\frac{s_E^2}{n_E \bar{x}_E^2}\frac{1}{2(n_E - 1)}}. \tag{A4}$$

Equation A4 is the same as Equation (12).

Alternatively, CV can be seen as the inverse of z value. In this case, the equation for the variance of lnCVR can be simplified to Equation (6).

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Data S1. R code for calculating d (dEffectSizes.R).

Data S2. R code for Example 1 (ParasiteAnalysis.R).

Data S3. R code for calculating ln$s$ (lnVarEffectSizes.R).

Data S4. R code for calculating lnRR (RREffectSizes.R).

Data S5. R code for Example 2 (ChromosomeAnalysis1.R).

Data S6. R code for Example 3 (ChromosomeAnalysis2.R).

Data S7. R code for calculating lnCVR (VarianceEffectSizes.R).

Data S8. Bird data file for Example 3 (birdslongformat.csv).

Data S9. Tree file for birds (108spsubtreeHackett.tre).

Data S10. Bird data file for Example 2 (birdsMLfinaltrait.csv).

Data S11. Mammal data file for Example 2 (mammalsMLfinaltrait.csv).

Data S12. Data file for Example 1 (ParasiteData.csv).

Data S13. Mammal data file for Example 3 (mammalslongformat.csv).

Data S14. Tree file for mammals (mammals81spsubtree.tre).

Appendix S1. Detailed methods for Examples 1–3 (supinfo.doc).