

Estimating phenotypic correlations: correcting for bias due to intraindividual variability

S. C. ADOLPH*† and J. S. HARDIN‡

*Department of Biology, Harvey Mudd College, 301 Platt Blvd., Claremont, CA 91711 USA,

‡Department of Mathematics, Pomona College, 610 N. College Ave., Claremont, CA 91711 USA

Summary

1. Correlations between phenotypic traits are important in a number of contexts in physiological ecology, evolutionary physiology, and behaviour. Correlations can reflect functional connections or trade-offs among performance traits (e.g. bite force, jumping distance) and can reveal causal relationships between whole-organism traits and lower-level biochemical or morphological traits.

2. However, when one or both traits exhibit intraindividual variability (i.e. repeatability < 1), conventional estimates of Pearson product-moment correlation coefficients are biased towards zero (= attenuated). The magnitude of this bias decreases with increases in the number of measurements used to calculate the mean value of the trait for each individual. The bias varies inversely with the repeatability of each trait.

3. We present an estimator for the correlation coefficient that eliminates this bias. This estimator is based on an equation originally presented in 1904 by Spearman, and applied by researchers in psychological testing and nutritional epidemiology. The estimator is a simple function of the within- and among-individual components of variance for each of the two traits.

4. Simulations show that optimal sampling effort usually involves a small number of trials per individual and a large sample of individuals (for a fixed total sample size), although correlations between traits with low repeatabilities may be more precisely estimated with a larger number of trials per individual and a smaller number of individuals.

5. In addition to reducing the accuracy of the estimate, attenuation also reduces statistical power for detecting significant correlations. However, we do not recommend using the unbiased estimator for testing whether correlations differ from zero, because this inflates Type I error rates. Instead, the uncorrected (conventional) estimator should be used for hypothesis testing.

6. The unbiased estimator is not appropriate for correlations involving maximum or minimum values for each individual (e.g. maximum sprint speed) because sampling distributions of these extreme values typically have different properties than the sampling distributions of individual mean values.

Key-words: attenuation, experimental design, locomotion, performance, repeatability

Functional Ecology (2007) **21**, 178–184

doi: 10.1111/j.1365-2435.2006.01209.x

Introduction

Correlations between phenotypic traits are important in diverse contexts in animal behaviour, evolution and integrative biology. For example, phenotypic

correlations are used to assess functional relationships between biochemical, morphological and whole-organism traits (Garland 1984; Bennett, Garland & Else 1989; Steyermark *et al.* 2005), and to identify behavioural syndromes (Huntingford 1976; Sih, Bell & Johnson 2004a; Sih *et al.* 2004b).

Here, we alert researchers to a statistical problem that is not usually recognized in integrative biology:

†Author to whom correspondence should be addressed.
E-mail: adolph@hmc.edu

conventional estimates of Pearson product-moment correlations are biased towards zero whenever there is intraindividual variation in one or both of the traits involved (Fuller 1987). Virtually all behavioural and physiological traits show some degree of intraindividual variability (i.e. repeatability < 1.0, e.g. Bennett 1987; Huey & Dunham 1987; Boake 1989; Hayes & Jenkins 1997; Brodie & Russell 1999; Dohm 2002). Therefore, the values of phenotypic correlations reported in the biological literature are likely to underestimate the true correlations, on average.

We present an estimator that gives a theoretically unbiased estimate of the correlation coefficient between individual trait means. In addition, we use simulations to evaluate the performance of corrected vs. uncorrected estimates of correlation coefficients and to explore the allocation of sampling effort between the number of individuals vs. the number of measurements per individual.

Bias in estimating correlation coefficients

Intraindividual variability is statistically equivalent to measurement error. Statisticians have long recognized that estimates of Pearson product-moment correlation coefficients are biased when there is measurement error in either the X or the Y variable (Spearman 1904; Thouless 1939; Fuller 1987). This bias is called 'attenuation' (Spearman 1904) because on average an estimate of a correlation coefficient is biased towards zero. Attenuation of correlation coefficients is well-known in some fields, particularly in psychological and educational testing (Gulliksen 1950) and nutritional epidemiology (Willett 1998). However, researchers rarely acknowledge or correct for attenuation in many fields of biology, including physiological ecology, evolutionary physiology, and animal behaviour. A related phenomenon, the attenuation of regression slopes due to measurement error in the x variable, has been addressed by several papers in these fields (e.g. McArdle 2003).

Spearman (1904) showed that the expected value of a sample correlation coefficient between two traits X and Y is given by

$$E(\hat{r}_{XY}) = \rho \sqrt{\left(\frac{\sigma_{a,X}^2}{\sigma_{a,X}^2 + \sigma_{w,X}^2}\right) \left(\frac{\sigma_{a,Y}^2}{\sigma_{a,Y}^2 + \sigma_{w,Y}^2}\right)} \quad \text{eqn 1}$$

where ρ = the true correlation coefficient between the mean values of Traits X and Y for each individual, $\sigma_{a,X}^2$ = the among-individual variance in Trait X and $\sigma_{w,X}^2$ = the within-individual variance in Trait X , with corresponding notation for Trait Y . Now suppose that we make n_X measurements per individual for Trait X , and n_Y measurements for Trait Y , and use the mean value of each individual's trials as our x,y pairs. The individual mean values have error variances $\sigma_{w,X}^2/n_X$ and $\sigma_{w,Y}^2/n_Y$, so that the expected value of a sample correlation coefficient is

$$E(\hat{r}_{XY}) = \rho \sqrt{\left(\frac{\sigma_{a,X}^2}{\sigma_{a,X}^2 + \frac{\sigma_{w,X}^2}{n_X}}\right) \left(\frac{\sigma_{a,Y}^2}{\sigma_{a,Y}^2 + \frac{\sigma_{w,Y}^2}{n_Y}}\right)} \quad \text{eqn 2}$$

This formula or its equivalent has been presented previously by several authors in other fields (Liu *et al.* 1978; Beaton *et al.* 1979; Rosner & Willett 1988). Note that the magnitude of the multiplicative bias (the square root term) is independent of ρ and of the number of individuals sampled.

Equations 1 and 2 show that if within-individual variance is present the estimated correlation coefficient will be biased towards zero. The magnitude of the bias decreases as the ratio σ_w^2/σ_a^2 decreases (i.e. as repeatability increases) and as the per-individual sample sizes n_X and n_Y increase (Fig. 1).

Equation 2 can be used to derive an unbiased estimator for the correlation coefficient:

$$\hat{r}_{\text{corrected}} = r \sqrt{\left(1 + \frac{s_{w,X}^2}{s_{a,X}^2 n_X}\right) \times \left(1 + \frac{s_{w,Y}^2}{s_{a,Y}^2 n_Y}\right)} \quad \text{eqn 3}$$

where r is the uncorrected correlation coefficient calculated using the sample mean values of each trait for each individual (\bar{x} and \bar{y}). We have replaced the parametric variance terms (σ^2) with sample variance terms (s^2) in eqn 3 because these variances will almost

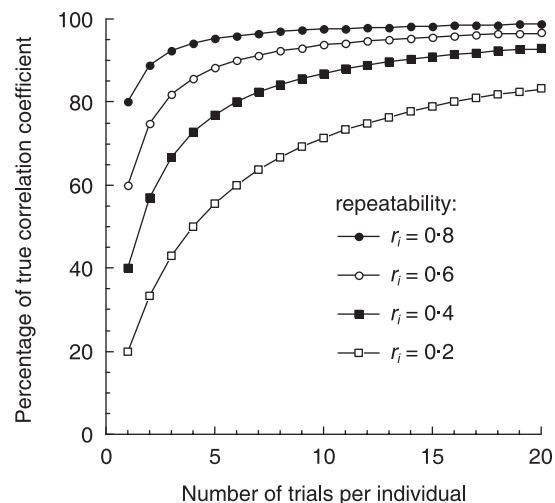


Fig. 1. Multiplicative bias of the estimated correlation coefficient between two traits when each trait exhibits intraindividual variation. Each curve shows the expected value of the correlation coefficient (eqn 2) that would be obtained from a sample, shown as a percentage of its true value, when the score for each individual is the mean of a given number of trials (= measurements). Curves are shown for four representative values of repeatability (intra-class correlation coefficient, r_i); for these examples, the repeatability is the same for traits X and Y , and the number of trials per individual are the same for both traits ($n_X = n_Y$).

always be estimated from data. It is possible for a given sample to yield a value of $\hat{r}_{corrected}$ that exceeds 1.0 (or -1.0); in this case, the estimate should be rounded to 1.0 (or -1.0).

Because the repeatability (= intraclass correlation coefficient; Sokal & Rohlf 1994) for Trait X is given by $r_{i,X} = \sigma_{a,X}^2 / (\sigma_{a,X}^2 + \sigma_{w,X}^2)$, eqn 3 can be rewritten in terms of the repeatabilities of the two traits:

$$\hat{r}_{corrected} = r \sqrt{\left(1 + \frac{1 - r_{i,X}}{r_{i,X} n_X}\right) \times \left(1 + \frac{1 - r_{i,Y}}{r_{i,Y} n_Y}\right)} \quad \text{eqn 4}$$

Equation 4 can be used to obtain unbiased estimates of correlations from previously published work if the repeatabilities are known, and also demonstrates that using the mean of several measurements is comparable with increasing repeatability (Falconer 1981; Arnold 1994; Hayes & Jenkins 1997).

We ran computer simulations to evaluate the performance of the unbiased estimator (eqn 3). We drew samples from a bivariate normal distribution with a known true correlation coefficient ($\rho = 0.4$ and 0.8) and examined the effects of number of individuals ($N_{subjects} = 20, 40$ and 80), the number of measurements per individual ($n_{trials} = 2, 5$ and 10), and repeatability ($r_i = 0.2, 0.5$ and 0.8). We used the same values of n_{trials} and r_i for Traits X and Y . We added to each variate a random, normally distributed error term adjusted to yield the desired repeatability. For each parameter combination we drew 5000 independent samples, and from each sample we calculated four different estimates of the correlation coefficient.

- 1 Uncorrected for bias, but without adding error terms (i.e. $\sigma_{w,X}^2 = 0$), as a check on the simulation procedure. This should yield an unbiased estimate, on average.
- 2 Uncorrected for bias. On average, this should yield a biased estimate (eqn 2).
- 3 Corrected for bias using eqn 3, using the known parametric variance components. This should yield an unbiased estimate.
- 4 Corrected for bias using eqn 3, with variance components estimated from the sample. This should yield an unbiased estimate.

Simulations were run in the statistical computing language R (R Development Core Team 2005) and yielded several noteworthy results. First, the uncorrected (conventional) correlation coefficients were biased towards zero as predicted by theory, whereas the corrected correlation coefficients on average were unbiased (Figs 2 and 3). In each case the mean correlations were essentially identical to the analytical predictions from eqns 2 and 3. Second, the sampling distributions differed in variance (Fig. 2). The variance of the distribution with $\sigma_{w,X}^2 = 0$ was entirely due to sampling of individuals from the population. The

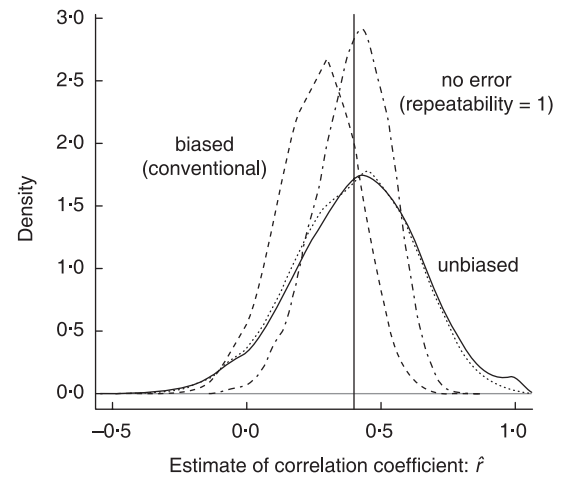


Fig. 2. Sampling distributions of Pearson product-moment correlation coefficients obtained via simulation. Ten thousand independent samples of $N = 40$ observations were drawn from a bivariate normal distribution with $\rho = 0.4$ (indicated by the vertical line) and repeatabilities $r_{i,X} = r_{i,Y} = 0.5$. Correlations were calculated using the mean values of two trials per individual for each trait ($n_{trials,X} = n_{trials,Y} = 2$). Dashed line: conventional correlation coefficient, uncorrected for bias. Dotted line: unbiased correlation coefficient (eqn 4), calculated using the known parametric values of the variance components. Solid line: unbiased correlation coefficient (eqn 4), with variance components estimated from the sample. Dashes and dots: correlation coefficients calculated for observations with zero intraindividual variability ($r_{i,X} = r_{i,Y} = 1.0$); the variance of this distribution reflects only the sampling variation due to the finite number of individuals.

other three distributions exhibited variance due to both sampling of individuals and sampling of values within individuals. The greater sampling variance of the unbiased estimators compared with the conventional estimator is due to the fact that the unbiased estimate involves multiplication by a scalar (the correction factor); recall that $\text{Var}(aX) = a^2 \text{Var}(X)$ for any random variable X and scalar a . The distributions of the unbiased estimators were virtually identical, indicating that estimating the variance components in eqn 3 added little variance to the total sampling variance.

Overall, the simulations confirmed that the uncorrected estimator is biased, and that the corrected estimator eliminated this bias (Fig. 3). The simulations also illustrated the effects of sample size and repeatability on bias: the bias was greatest for the smallest number of samples per individual ($n_{trials} = 2$) and for the lowest repeatabilities ($r_i = 0.2$). In these cases, almost the entire distribution of sample correlation coefficients fell below the true value of ρ , clearly an undesirable property for an estimator. Another key result is that using a larger sample size of individuals ($N_{subjects}$) did not reduce bias, although it did reduce the sampling variances of both the unbiased and biased estimators; recall that $N_{subjects}$ does not appear in eqn 2. The sampling variance of the unbiased estimator was

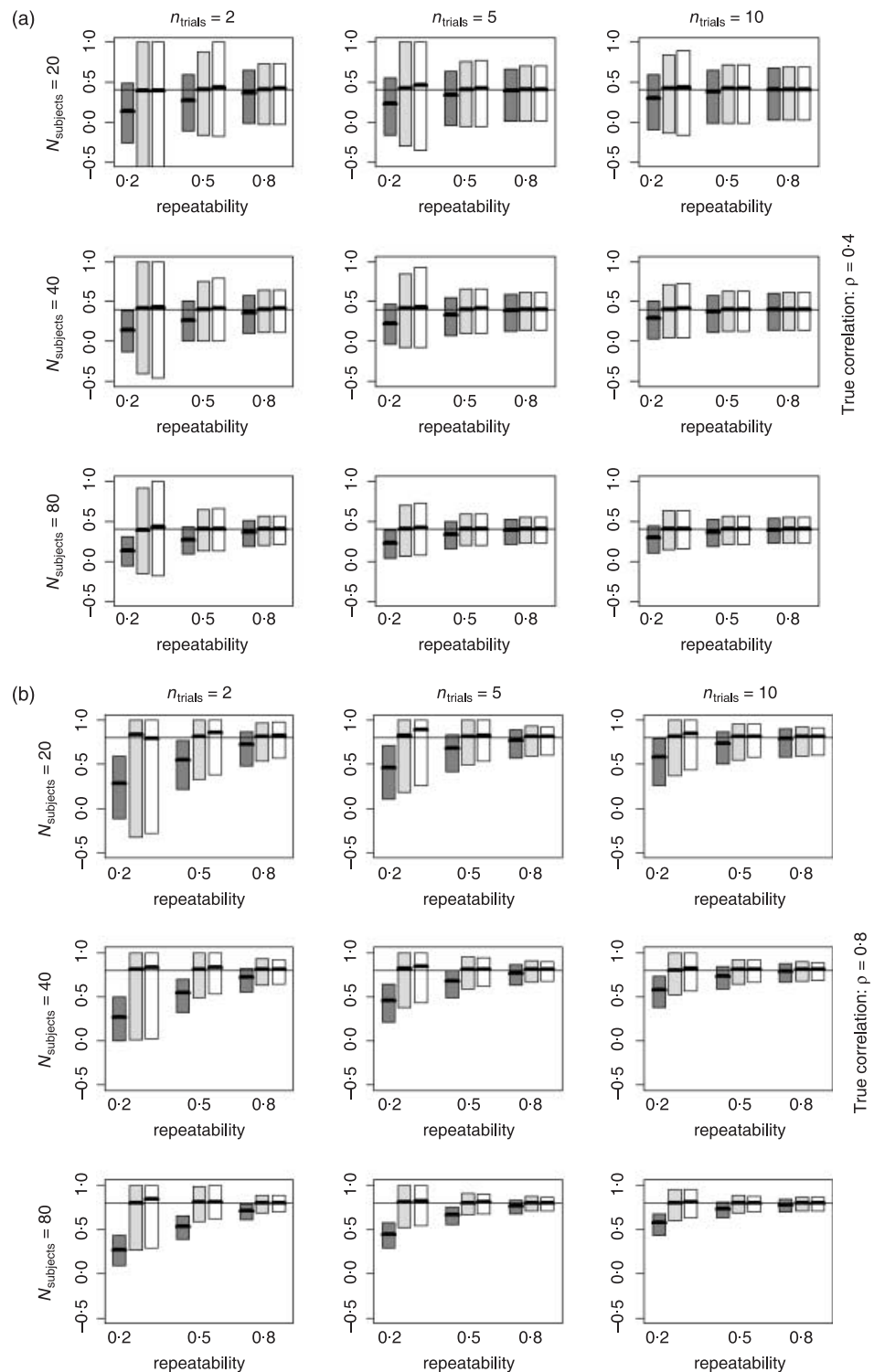


Fig. 3. Sampling distributions of Pearson product-moment correlation coefficients obtained via computer simulation, for selected sample sizes and repeatabilities. Each panel shows the median (heavy line) and central 90% of sample correlation coefficients from 5000 independent samples, using three different estimators: dark grey is the conventional estimator; light grey is the unbiased estimator with variance components estimated from the sample, and white is the unbiased estimator using known variance components. Each panel includes results for three different values of repeatability ($r_{i,X} = r_{i,Y} = 0.2, 0.5$ or 0.8). The horizontal line in each panel indicates the true value of ρ (either 0.4 or 0.8). Three different sample sizes of individuals are used ($N_{\text{subjects}} = 20, 40$ or 80), and three different numbers of trials for each individual ($n_{\text{trials}} = 2, 5$ or 10).

greater than that of the biased (uncorrected) estimator, as discussed above for Fig. 2. However, this difference decreased with increases in repeatability and the number of trials per individual. Our results are consistent

with the conclusions of Fan (2003), who conducted a similar study to address the issue of measurement reliability and its effect on correlations between test scores in educational and psychological assessment.

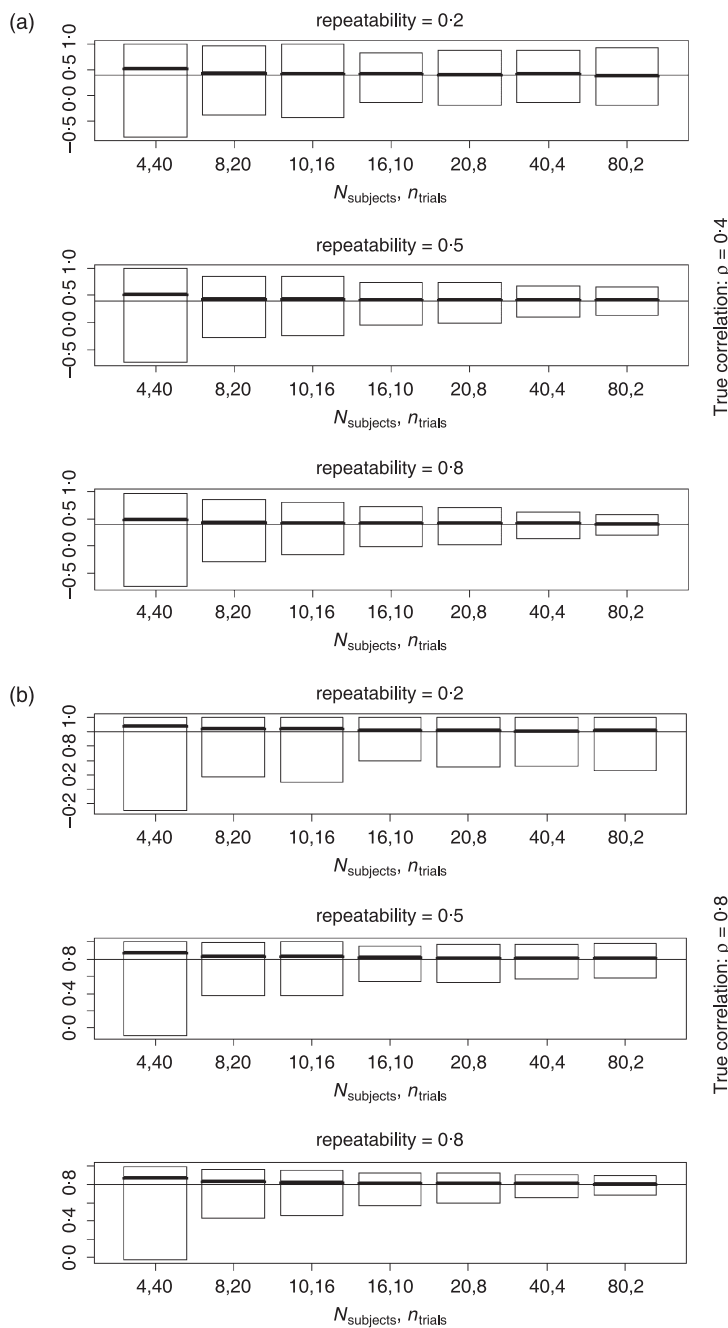


Fig. 4. Effect of trading off the number of trials per individual (n_{trials}) vs. number of individuals (N_{subjects}) on the sampling variance of unbiased correlation coefficients obtained via simulation. Each bar shows the median and central 90% of sample correlation coefficients (eqn 2) from 5000 samples drawn from a bivariate normal distribution. The total number of observations ($n_{\text{trials}} \times N_{\text{subjects}}$) was fixed at 160; the combination of n_{trials} and N_{subjects} corresponding to each bar is indicated. Top three panels: $\rho = 0.4$; bottom three panels: $\rho = 0.8$; horizontal line indicates true value of ρ .

Hypothesis testing: which estimator to use?

At first glance it might seem preferable to use the more accurate unbiased estimator (eqn 3) for testing the null hypothesis $\rho = 0$. Instead, it is actually more appropriate to use the uncorrected (conventional) estimate for this hypothesis test, for the following reason: if $\rho = 0$, then there is no bias, and therefore no need to correct the

estimate. The tables of critical values for the correlation coefficient were established for the uncorrected estimator; different tables would need to be constructed for the unbiased estimator. Using an unbiased estimate of ρ for hypothesis testing with the conventional critical values would inflate the Type I error rate beyond the desired rate, because the estimate would exceed the critical value more frequently. We verified this effect by running simulations in which $\rho = 0$ (results not shown). This suggests that researchers should first use the conventional estimator for testing for whether $\rho = 0$. Then, if H_0 is rejected (i.e. $\rho \neq 0$), eqn 3 should be used for an unbiased estimate of ρ .

Much less frequently, researchers wish to test null hypotheses involving nonzero values of ρ . In this case it seems reasonable to use the unbiased estimator. Rosner & Willet (1988) and Charles (2005) give formulas for the standard error and confidence intervals of $\hat{\rho}_{\text{corrected}}$, which could be adapted for hypothesis testing.

Allocation of research effort: number of trials vs. number of individuals

The unbiased estimator for the correlation coefficient (eqns 3 and 4) requires at least two measurements per individual so that σ_{within}^2 and σ_{among}^2 can be estimated. Clearly, measuring more than two values per individual has several benefits: it decreases the bias (see eqn 3), thereby increasing statistical power, and it also should give a better estimate of the variance components σ_{within}^2 and σ_{among}^2 . However, for a fixed total sample size ($N_{\text{subjects}} \times n_{\text{trials}}$) there is a trade-off between N_{subjects} and n_{trials} . This trade-off would be particularly important to consider when nontrivial effort is required for each phenotypic measurement (e.g. in studies of exercise metabolism).

We ran simulations to explore how this trade-off between N_{subjects} and n_{trials} influences the precision of estimating ρ . We assumed that a researcher could make 160 total measurements, and chose combinations of N_{subjects} (from 4 to 80) and n_{trials} (from 2 to 40) that would yield this total. We then obtained 5000 random samples for each allocation and for each possible combination of repeatabilities ($r_r = 0.2, 0.5, 0.8$) and true correlations ($\rho = 0.4, 0.8$), using the same repeatabilities and sample sizes for Traits X and Y .

In the intermediate- and high-repeatability cases, the sampling variance of the correlation coefficient decreased with higher values of N_{subjects} and lower values of n_{trials} (Fig. 4). This decrease in sampling variance was most consistent in the high-repeatability case. In the low-repeatability simulation, the sampling variance was lowest at intermediate values of N_{subjects} and n_{trials} . These findings were essentially the same for high ($\rho = 0.8$) and low ($\rho = 0.4$) true correlations. These results suggest that the optimal allocation of sampling effort, using the criterion of minimizing the sampling variance of the correlation coefficient (i.e.

increasing the precision of the estimate), depends on repeatability. For traits with high repeatability, the best design appears to involve making two measurements on as many individuals as possible. For traits with intermediate repeatabilities, sampling variance was relatively insensitive to allocation as long as $N_{\text{subjects}} > n_{\text{trials}}$. For traits with low repeatabilities, choosing a higher n_{trials} at the expense of a lower N_{subjects} appears to yield slightly more precise estimates of ρ . When repeatabilities are unknown prior to beginning an experiment, Fig. 4 suggests that choosing a small value of n_{trials} (i.e. 2, 3 or 4), along with the highest feasible value of N_{subjects} , would probably yield the most precise estimate of ρ .

The trade-off between n_{trials} and N_{subjects} has implications for statistical power as well as for the precision of the estimate. Increasing either n_{trials} or N_{subjects} will increase statistical power for detecting nonzero correlation coefficients, but for different reasons. Using a larger N_{subjects} increases the degrees of freedom for the hypothesis test, and also decreases the sampling variance of the estimated correlation coefficient (see Fig. 3). Using a larger n_{trials} decreases bias (eqn 2), which could be a substantial benefit when repeatability is low. Given a limited total sample size, increasing N_{subjects} is probably a more effective way to increase power than increasing n_{trials} .

Correlations involving maximum rather than mean values

In many studies of organismal performance, such as burst speed (Bennett 1980; Wilson 2005), jumping distance (Watkins 1997), and bite force (Huyghe *et al.* 2005), researchers use maximum rather than mean values because maxima are likely to be more relevant measures of performance in ecological and evolutionary contexts. Like mean values, maximum values are estimated with error due to intraindividual variability; individuals do not always perform at their physiological maximum. Therefore, when a correlation is calculated between individual maximum values for two traits, it will be closer to zero on average than the true correlation.

Unfortunately, the unbiased estimator presented in eqn 3 is not appropriate for correlations between maximum values, because the sampling distribution of a maximum is different than the sampling distribution of a mean value. Sampling distributions of maximum values are more complicated than those of mean values (Gumbel 1958); for example, they depend on the underlying distribution of individual performances, whereas the distribution of sample means is approximately normal for a variety of underlying distributions (hence the Central Limit Theorem). Also, the expected value of a maximum increases with n_{trials} (Losos, Creer & Schulte 2002; S.C. Adolph & T. Pickering, unpublished MS), whereas the expected value of a mean is independent of n_{trials} . Because of these complications, new statistical procedures need to be devised to correct

for bias in correlations between maximum performances. We suggest that empirical studies report correlations between performance traits using individual mean values in addition to correlations using individual maxima, particularly when the goal is to identify physiological and morphological traits that affect whole-organism performance.

Acknowledgements

We thank J.-F. Le Galliard and three anonymous reviewers for helpful comments on the manuscript, T. Pickering and H. Groves for their laboratory and simulation work that led to this project, T. Garland and V. Malcarne for suggesting key references, and R. Carroll for thoughtful discussions about correlation bias. S.C.A. thanks Monica Geber and the Department of Ecology and Evolutionary Biology at Cornell University for providing space and assistance during his sabbatical. Funding for this research was provided by a grant from the W.M. Keck Foundation to the Center for Quantitative Life Sciences at Harvey Mudd College.

References

- Arnold, S.J. (1994) Multivariate inheritance and evolution: a review of concepts. *Quantitative Genetic Studies of Behavioral Evolution* (ed. C.R.B. Boake), pp. 17–48. University of Chicago Press, Chicago, IL.
- Beaton, G.H., Milner, J., Corey, P., McGuire, V., Cousins, M., Stewart, E., de Ramos, M., Hewitt, D., Grambsch, P.V., Kassim, N. & Little, J.A. (1979) Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. *American Journal of Clinical Nutrition* **32**, 2456–2559.
- Bennett, A.F. (1980) The thermal dependence of lizard behaviour. *Animal Behaviour* **28**, 752–762.
- Bennett, A.F. (1987) Interindividual variability: an underutilized resource. *New Directions in Ecological Physiology* (eds M. Feder, A.F. Bennett, R.B. Huey & W. Burggren), pp. 147–169. Cambridge University Press, Cambridge.
- Bennett, A.F., Garland, T. Jr & Else, P.L. (1989) Individual correlation of morphology, muscle mechanics and locomotion in a salamander. *American Journal of Physiology* **256**, R1200–R1208.
- Boake, C.R.B. (1989) Repeatability: its role in evolutionary studies of mating behavior. *Evolutionary Ecology* **3**, 173–182.
- Brodie, E.D. III & Russell, N.H. (1999) The consistency of individual differences in behaviour: temperature effects on antipredator behaviour in garter snakes. *Animal Behaviour* **57**, 445–451.
- Charles, E.P. (2005) The correction for attenuation due to measurement error: clarifying concepts and creating confidence sets. *Psychological Methods* **10**, 206–226.
- Dohm, M.R. (2002) Repeatability estimates do not always set an upper limit to heritability. *Functional Ecology* **16**, 273–280.
- Falconer, D.S. (1981) *Introduction to Quantitative Genetics*, 2nd edn. Longman, London.
- Fan, X. (2003) Two approaches for correcting correlation attenuation caused by measurement error: implications for research practice. *Educational and Psychological Measurement* **63**, 915–930.

- Fuller, W.A. (1987) *Measurement Error Models*. John Wiley and Sons, New York.
- Garland, T. Jr (1984) Physiological correlates of locomotor performance in a lizard: an allometric approach. *American Journal of Physiology* **247**, R806–R815.
- Gulliksen, H. (1950) *Theory of Mental Tests*. John Wiley and Sons, New York.
- Gumbel, E.J. (1958) *Statistics of Extremes*. Columbia University Press, New York.
- Hayes, J.P. & Jenkins, S.H. (1997) Individual variation in mammals. *Journal of Mammalogy* **78**, 274–293.
- Huey, R.B. & Dunham, A.E. (1987) Repeatability of locomotor performance in natural populations of the lizard *Sceloporus merriami*. *Evolution* **41**, 1116–1120.
- Huntingford, F.A. (1976) The relationship between anti-predator behaviour and aggression among conspecifics in the three-spined stickleback, *Gasterosteus aculeatus*. *Animal Behaviour* **24**, 245–260.
- Huyghe, K., Vanhooydonck, B., Scheers, H., Molina-Borja, M. & Van Damme, R. (2005) Morphology, performance and fighting capacity in male lizards, *Gallotia galloti*. *Functional Ecology* **19**, 800–807.
- Liu, K., Stamler, J., Dyer, A., McKeever, J. & McKeever, P. (1978) Statistical methods to assess and minimize the role of intraindividual variability in obscuring the relationship between dietary lipids and serum cholesterol. *Journal of Chronic Diseases* **31**, 399–418.
- Losos, J.B., Creer, D.A. & Schulte, J.A. I.I. (2002) Cautionary comments on the measurement of maximum locomotor capabilities. *Journal of Zoology, London* **258**, 57–61.
- McArdle, B.H. (2003) Lines, models, and errors: regression in the field. *Limnology and Oceanography* **48**, 1363–1366.
- R Development Core Team. (2005) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Rosner, B. & Willett, W.C. (1988) Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. *American Journal of Epidemiology* **127**, 377–386.
- Sih, A., Bell, A. & Johnson, J.C. (2004a) Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology and Evolution* **19**, 372–378.
- Sih, A., Bell, A.M., Johnson, J.C. & Ziemba, R.E. (2004b) Behavioral syndromes: an integrative overview. *Quarterly Review of Biology* **79**, 241–277.
- Sokal, R.R. & Rohlf, F.J. (1994) *Biometry*, 3rd edn. W.H. Freeman, San Francisco.
- Spearman, C. (1904) The proof and measurement of association between two things. *American Journal of Psychology* **15**, 72–101.
- Steyermark, A.C., Miamen, A.G., Feghahati, H.S. & Lewno, A.W. (2005) Physiological and morphological correlates of among-individual variation in standard metabolic rate in the leopard frog *Rana pipiens*. *Journal of Experimental Biology* **208**, 1201–1208.
- Thouless, R.H. (1939) The effects of errors of measurement on correlation coefficients. *British Journal of Psychology* **29**, 383–403.
- Watkins, T.B. (1997) The effect of metamorphosis on the repeatability of maximal locomotor performance in the Pacific tree frog *Hyla regilla*. *Journal of Experimental Biology* **200**, 2663–2668.
- Willett, W. (1998) *Nutritional Epidemiology*, 2nd edn. Oxford University Press, Oxford.
- Wilson, R.S. (2005) Consequences of metamorphosis for the locomotor performance and thermal physiology of the newt *Triturus cristatus*. *Physiological and Biochemical Zoology* **78**, 967–975.

Received 19 March 2006; revised 5 July 2006; accepted 8 August 2006