Ann. Zool. Fennici 34: 133–137 ISSN 0003-455X
Helsinki 27 May 1997

Commentary

Why some measures of fluctuating asymmetry are so sensitive to measurement error

Mats Björklund & Juha Merilä

Björklund, M. & Merilä, J., Department of Zoology, Uppsala University, Villavägen 9, S-752 36 Uppsala, Sweden; Present address: Merilä, J., Laboratory of Ecology and Animal Systematics, Department of Biology, Turku University, FIN-20014, Finland

Received 26 March 1997, accepted 1 April 1997

Introduction

Fluctuating asymmetry has recently become a subject of much theoretical and empirical interest as well as causing considerable discussion (Houle 1997, Leamy 1997, Markow & Clarke 1997, Møller & Thornhill 1997ab, Palmer & Strobeck 1997, Pomiankowski 1997, Swaddle 1997, Whitlock & Fowler 1997). This interest has been nurtured by results indicating that fluctuating asymmetry is caused by stress factors operating on the developmental system (Van Valen 1962, Palmer & Strobeck 1986, Parsons 1990, 1992), and thus may be a potential indicator of the amount of stress imposed upon a given population, or conversely, the ability of individuals to cope with stress during their ontogeny.

As all other estimators, measurements of fluctuating asymmetry are affected by measurement error (e.g. Palmer & Strobeck 1986, Swaddle *et al.* 1994, Merilä & Björklund 1995). This is particularly serious in the case of fluctuating asymmetry since, by definition, it is expected to take on very small values. It has previously been shown (Swaddle *et al.* 1994, Merilä & Björklund 1995) that the most commonly used measure of fluctuating asymmetry, i.e. the absolute difference between the sides ($|R–L|$), is highly sensitive to measurement error. While the fact that this measure is prone to high measurement error has been known for some time, the reason and its possible implications have not been examined. Furthermore, on the basis of our results we think that measurement error is a very important issue that has been largely overlooked in the discussion.

Theory

In this note, we will give an explanation based on basic statistical theory why both the signed and the unsigned difference measures are so sensitive to measurement error. By doing that, we will use the approach taken by Whitlock (1996) with a slight, but very important modification. Generally, fluctuating asymmetry (FA) is defined as random deviations from perfect symmetry of bilateral traits among a set of individuals (Palmer & Strobeck 1986). Random in this case means random in relation to side, either the left or the right side being larger, while the mean on average is zero. Thus, it is clear that FA is a population measure of individual asymmetry. FA can be expressed as the signed difference of the sides ($L–R$) or the unsigned (absolute) difference among the sides ($|L–R|$). If we assume that the measurements of each of the sides follow a normal distribution the variance of the signed difference is given by:

$$\sigma^2_{L-R} = \sigma^2_L + \sigma^2_R - 2\rho_{L,R}\sigma_L\sigma_R \qquad (1)$$

where $\rho_{L,\,R}$ is the correlation between sides, and σ^2_L, σ_L is the variance and the standard deviation of the left side, respectively. Note that only if the correlation is zero the variance equals $\sigma^2_L + \sigma^2_R$, as in Whitlock (1996), otherwise the variance is smaller. This can easily be seen if we set the correlation to 1.0 (eq. 1). In the case of bilateral traits, the correlation is always larger than zero, and in most cases close to 1.0, which makes the variance of the signed difference small compared with the variance of the sides.

The variance of the unsigned difference is even smaller and is given by:

$$\sigma^2_{|L-R|} = 2\sigma^2_{L-R}\frac{\pi-2}{\pi} \approx \frac{3}{4}\sigma^2_{L-R} \qquad (2)$$

(Whitlock 1996; the approximation is quite rough and should be regarded more as a rule of thumb). In the following we will make the simplifying assumption that $\sigma^2_L = \sigma^2_R = \sigma^2_S$. This simplifies eqs. (1) and (2) to:

$$\sigma^2_{L-R} = 2\sigma^2_S\left(1-\rho_{L,\,R}\right) \qquad (3)$$

$$\sigma^2_{|L-R|} = 4\sigma^2_S\left(1-\rho_{L,\,R}\right)\frac{\pi-2}{\pi} \qquad (4),$$

respectively. This makes it is even more clear that as the correlation increases, the variance decreases.

Estimates of variances contain two parts, namely the true variance σ^2_{true} as well as some error variance, σ^2_{error}, due to imprecise measurements, and thus $\sigma^2_{\text{total}} = \sigma^2_{\text{true}} + \sigma^2_{\text{error}}$. To see why the two measures of FA potentially can give high levels of measurement errors, we define measurement error (ME) as:

$$\text{ME} = \frac{\sigma^2_{\text{error}}}{\sigma^2_{\text{error}} + \sigma^2_{\text{true}}} \qquad (5)$$

where both variance components are derived from a one-way ANOVA as described by Bailey and Byrnes (1990). Thus, measurement error is expressed as a proportion of total variance. The variance of the FA-estimates is given by eqs. (3) and (4), while the error variance can be derived in the same way as in eq. (1), but with one very important exception: errors are not (and should not be) correlated between the sides. Thus, the error variance is simply the sum of the error variances of the sides. Assuming equal errors on both sides, using eqs. (3) and (4), in eq. (5) we obtain:

$$\text{ME}_{L-R} = \frac{\sigma^2_{\text{error}}}{\sigma^2_{\text{error}} + \sigma^2_S\left(1-\rho_{L,\,R}\right)} \qquad (6)$$

$$\text{ME}_{|L-R|} = \frac{\sigma^2_{\text{error}}}{\sigma^2_{\text{error}} + 2\sigma^2_S\left(1-\rho_{L,\,R}\right)\frac{\pi-2}{\pi}} \qquad (7).$$

From eqs. (6) and (7) it can easily be seen that as the correlation between the sides increases, i.e. asymmetry decreases, the proportion of total variance measured that is due to measurement error rapidly approaches 1.0. When there is no asymmetry, all variation is due to measurement error. This means that in traits which are only slightly asymmetrical, the problem with measurement error becomes prominent and needs special attention.

Since the variance of the unsigned difference is smaller than the variance of the signed difference (eq. 2), the problem will be even larger with this measure, as has previously been observed (Merilä & Björklund 1995). It is of paramount importance to notice that the ME of the signed or unsigned difference is *not* the simple sum of the ME's of the sides. As is obvious from eqs. (6) and (7), the situation is more complex.

It is also important to notice that the mean (M) is related to the variance:

$$M_{|L-R|} = \sqrt{\frac{2\sigma^2_{|L-R|}}{\pi-2}} = \sigma_{|L-R|}\sqrt{\frac{2}{\pi-2}} \approx \frac{4}{3}\sigma_{|L-R|} \qquad (8)$$

(Whitlock 1996). Consequently, if the variance is inflated due to measurement error, so will be the mean value, too. Thus, if measurement error of the sides is 5%, and the correlation is 0.95 between the sides (i.e. a low degree of FA), then the measurement error of the FA-variance is in order of 50%. If the observed variance of FA is 1.0, then the true variance is only 0.67. The true mean will be 1.09, whereas the observed mean will be 1.33, i.e. a 22% overestimation. This has great importance when comparing traits with regard to amount of FA, since the mean value will increase with the amount of measurement error. Thus, a trait with an observed larger mean FA can have so as a consequence of it having larger measurement error, and not necessarily because its true value of FA is higher. In all studies comparing mean levels of FA or variances it is imperative that the influence of ME is cancelled out. As argued before, the best way of doing that is to use a mixed model two-way ANOVA, which provides an unbiased (relative to measurement error) estimate of FA (Palmer & Strobeck 1986, Merilä & Björklund 1995).

Note that if errors are somehow positively correlated, i.e. there is a systematic bias in the way the traits are measured, the error variance term decreases (compare eq. 1), and the estimate of the measurement error is deflated to some degree. This means that small values of measurement error can be obtained, which is only a result of biased measures. This, again, is a serious argument against using this measure of FA.

Repeatability and coefficient of variation (CV)

Another way of looking at this is to examine the repeatability (R) of FA-measures, defined as:

$$R = \frac{1}{1 + \left(1 + \frac{1}{\mathrm{CV}^2}\right)\frac{\pi - 2}{2} + \frac{\sigma^2_{\text{error}}}{\sigma^2_{|L-R|}}} \tag{9}$$

where $\mathrm{CV} = \sqrt{\sigma^2_{|L-R|} / M^2_{|L-R|}}$ (Whitlock 1996). Note that the mean is just a constant times the variance (eq. 8), and thus CV is also a constant, CV = $\sqrt{(\pi - 2)/2}$ (also noted by Houle (1997) and Pomiankowski (1997)). This reduces (9) to:

$$R = \frac{1}{1 + \frac{\pi}{2} + \frac{\sigma^2_{\text{error}}}{\sigma^2_{|L-R|}}} \approx \frac{1}{2.57 + \frac{\sigma^2_{\text{error}}}{\sigma^2_{|L-R|}}} \tag{10}$$

This means that the maximum repeatability that can be found is 1/2.57 ≈ 0.39. This is only about 60% of the value reported by Whitlock (1996), but he assumes CV to take all values when it is really a constant. If measurement error on the FA estimate is 50%, which it very well can be, then repeatability is 1/3.57 ≈ 0.28.

The fact that CV is a constant and approximately equal to 0.76 allows us to evaluate figures recently published. In table 1 in Whitlock (1996) we find values ranging from 0.56–6.3, and while those reported by Møller and Höglund (1992) ranged from 1.3 to 9.5. In relation to the expected value these values correspond to a slight underestimation (– 25%) to huge overestimations (+ 1 250%)! In fact, 28/30 of the values presented in Whitlock's table 1 and all of the values presented by Møller and Höglund (1992) are overestimates, in the former case on average by 233%, and in the latter on average by 526%. It can be argued that since both the mean and the variance are inflated by measurement error, this will cancel out when using CV. However, since the variance increases more with increasing measurement error than the mean, then CV increases with increasing measurement error. This clearly illustrates that the variance of FA as reported in these studies is seriously biased by measurement error. Unfortunately, data on ME is not given in the papers cited. Whitlock's (1996) own data on *Felis concolor vancouverensis* is about 20% too high, and measurement error said to be 'essentially zero'. However, even if measurement error on the sides is very low, the measurement error on the FA-estimate can be substantial, if the correlation between the sides is large enough, i.e. if the asymmetry is slight. For Whitlock's data on *Canis lupus,* measurement error is said to be corrected for, but still CV is overestimated by approximately a factor two. The reason for this may be in the way in which measurement error was corrected for.

The finding that CV is a constant has important implications for comparisons of CV for different traits, or different populations. A finding that some traits have a higher CV (as in Møller & Höglund 1992) only means that these traits are more prone to measurement error. Thus, consistent differences in CV means consistent differences in susceptability to measurement error.

Naturally ME will affect the estimation of the correlation of FA from different traits, as already noted by Whitlock (1996). The covariance of the measures of FA is unbiased, but the variance of FA has to be devaluated by the repeatability ($r_{\text{obs}} = r_{\text{true}}\sqrt{R_1 R_2}$; Whitlock 1996). If we assume that the true variance is 1.0 for both traits and the covariance is 0.5, i.e. a correlation between the two measures of $r = 0.4$, then the correlation given maximum repeatability is only $r = 0.16$. With a 50% measurement error of FA, this then reduces to $r = 0.11$. Thus, in both cases the estimated correlation will drop to a level which in most studies will be found not to be significantly different from zero (critical levels for $P = 0.05$ indicate that sample sizes need to be in the order of 300 individuals). It is worth remembering in this context that a finding of a low correlation does not mean that the true one is really higher but masked by measurement error, it may very well be that the two measures are uncorrelated. The conclusion reached here is that even if there is a high correlation, it is likely to go undetected.

Power

We have earlier argued that the ANOVA approach is superior (Merilä & Björklund 1995) because it provides a means of decoupling measurement error and FA, which the traditional estimate fails to do. However, the power of the ANOVA approach has to be assessed. This is fairly straightforward, and we have chosen the same approach as is common in quantitative genetic studies in estimating variance components. In particular, we will follow the approach used by Kearsey & Pooni (1996: 342–346) for the estimation of power and precision of a NCII breeding design. Here, we will only outline the method briefly, while for a full treatment the reader is referred to the original source.

The estimator of FA in the ANOVA approach is the interaction variance component which has the expected mean square $\mathrm{MS}_{\mathrm{FA}} = \sigma^2_{\mathrm{err}} + r\sigma^2_{\mathrm{FA}}$, where r is the number of measures on the same variable, and the error MS is an unbiased estimator of the true error variance. Thus, the FA estimate is defined as:

$$\sigma^2_{\mathrm{FA}} = \frac{\left(\mathrm{MS}_{\mathrm{FA}} - \mathrm{MS}_{\mathrm{err}}\right)}{r} \qquad (11)$$

Testing the significance of FA, tests the interaction mean square against the error mean square with $(n_1 - 1)(n_2 - 1)$ and $n_1 n_2 (r - 1)$ df, where n_1 is the number of individuals, and n_2 is the number of sides. Of course, when analysing FA, the number of sides is 2. Thus, the dfs are in the case of FA $n_1 - 1$ and $2n_1(r - 1)$, respectively. The expected value is then $E[F] = 1 + r\sigma^2_{\mathrm{FA}} / \sigma^2_{\mathrm{err}}$. If the critical value (say 5%) is F, let $F' = F/E[F]$ and then power is the corresponding probability for the F'-distribution and the appropriate degrees of freedom. Thus, for $n_1 = 20$, $r = 2$, and $\sigma^2_{\mathrm{FA}} = \sigma^2_{\mathrm{err}} = 2.0$, $E[F]$ is 3.0, $F_{0.05[19,40]} = 1.85$, $F' = 0.62$, and power is thus $P = 0.87$.

The σ^2_{FA}-estimator has great power even for low sample sizes (Fig. 1). If the σ^2_{FA} is twice the error, then even with a sample size of 10 individuals we are highly likely to achieve a significant result. On the other hand, if the error is twice σ^2_{FA}, then about 40 individuals are needed to get a significant result in more than 80% of the cases. It should be noted that power in this case is determined by the relationship between FA and ME and not the absolute values. Thus, even in cases of relatively high ME, a significant FA can be found provided FA is large enough. For example, even if ME = 20%, it is highly likely that we get a significant FA even for moderate sample sizes provided that FA is about 10% of the total variation.

Finally, the problems reported here also apply to measures of individual asymmetry. If asymmetry is slight even very small levels of measurement error can seriously distort the individual's values, and any correlation between asymmetry and, for example, any fitness measure will be uncertain. The only way to handle this, as we can see it, is to measure the individuals a number of times and take the mean values. Since the variance of the estimate decreases with the number of measurements, this seems to be rewarding.

Conclusion

In conclusion, measuring FA by taking the absolute difference of the sides should be avoided since the influence of even small amounts of measurement error on the original variables can lead to drastically higher levels of measurement error of the FA-estimate. The fact that CV of FA is a constant (≈ 0.76) can be used as a quick way of checking the influence of measurement error. If CV is substantially higher than 0.76 than the measure-

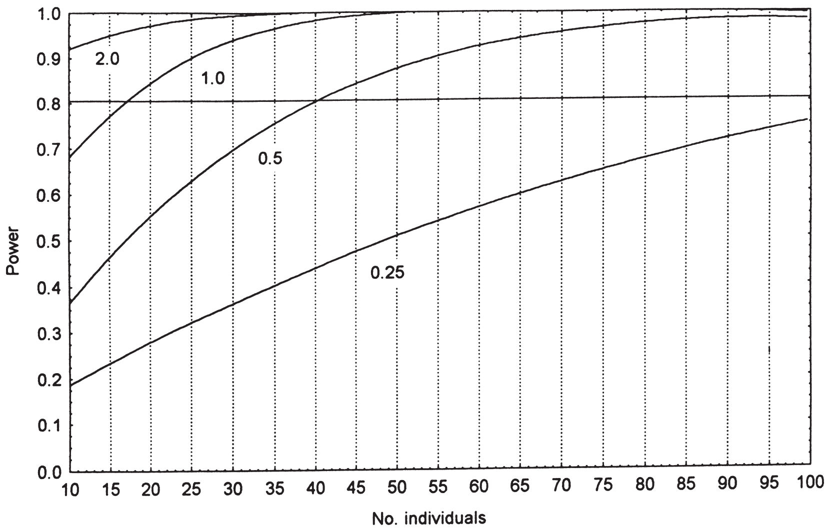

Fig. 1. Power analysis of the interaction variance component as a measure of FA and levels of measurement error. Lines corresponding to ratio FA:ME, i.e. 1.0 means equal proportions of total variance is explained by FA and ME.

ments should be avoided. Luckily, there is an alternative, namely a two-way mixed model ANOVA which provides independent estimates of both measurement error and FA, and in addition has considerable power, even for small sample sizes.

Acknowledgments. We thank Esa Ranta and Staffan Ulfstrand for valuable comments. MB was supported by the Swedish Natural Science Research Council.

References

Bailey, R. C. & Byrnes, J. 1990: A new, old method of estimating measurement error in both univariate and multivariate studies. — Syst. Zool. 39: 124–130.

Houle, D. 1997: Comment on “A meta-analysis of the heritability of developmental stability” by Møller and Thornhill. — J. Evol. Biol. 10: 17–20.

Kearsey, M. J. & Pooni, H. S. 1996: The genetical analysis of quantitative traits. — Chapman & Hall, London.

Leamy, L. 1997: Is developmental stability heritable? — J. Evol. Biol. 10: 21–29.

Markow, T. A. & Clarke, G. M. 1997: Meta-analysis of the heritability of developmental stability: a giant step backward. — J. Evol. Biol. 10: 31–37.

Merilä, J. & Björklund, M. 1995: Fluctuating asymmetry and measurement error. — Syst. Biol. 44: 97–101.

Møller, A. P. & Höglund, J. 1992: Patterns of fluctuating asymmetry in avian feather ornaments: implications for models of sexual selection. — Proc. R. Soc. Lond. B. 245: 1–5.

Møller, A. P. & Thornhill, R. 1997a: A meta-analysis of the heritability of developmental stability. — J. Evol. Biol. 10: 1–16.

Møller, A. P. & Thornhill, R. 1997b: Developmental instability is heritable. — J. Evol. Biol. 10: 69–76.

Palmer, A. R. & Strobeck, C. 1986: Fluctuating asymmetry: measurement, analysis, pattern. — Ann. Rev. Ecol. Syst. 17: 391–421.

Palmer, A. R. & Strobeck, C. 1997: Fluctuating asymmetry and developmental stability: heritability of observable variation vs. heritability of inferred cause. — J. Evol. Biol. 10: 39–49.

Parsons, P. A. 1990: Fluctuating asymmetry: an epigenetic measure of stress. — Biol. Rev. 65: 131–145.

Parsons, P. A. 1992: Fluctuating asymmetry: a biological monitor of environmental and genomic stress. — Heredity 68: 361–364.

Pomiankowski, A. 1997: Genetic variation in fluctuating asymmetry. — J. Evol. Biol. 10: 51–55.

Swaddle, J. P. 1997: On the heritability of developmental stability. — J. Evol. Biol. 10: 57–61.

Swaddle, J. P., Witter, M. S. & Cuthill, I. C. 1994: The analysis of fluctuating asymmetry. — Animal Behaviour 48: 986–989.

Van Valen, L. 1962: A study on fluctuating asymmetry. — Evolution 16: 125–142.

Whitlock, M. C. 1996: The heritability of fluctuating asymmetry and the genetic control of developmental stability. — Proc. R. Soc. Lond. B. 263: 849–854.

Whitlock, M. C. & Fowler, K. 1997: The instability of studies of instability. — J. Evol. Biol. 10: 63–67.