Ann. Zool. Fennici 36: 45–52 ISSN 0003-455X
Helsinki 28 April 1999

Accuracy and power in the statistical analysis of fluctuating asymmetry: effects of between-individual heterogeneity in developmental instability

Stefan Van Dongen

Van Dongen, S., Department of Biology, University of Antwerp (UIA), Universiteitsplein 1, B-2610 Wilrijk, Belgium.

Received 7 October 1998, accepted 2 March 1999

Van Dongen, S. 1999: Accuracy and power in the statistical analysis of fluctuating asymmetry: effects of between-individual heterogeneity in developmental instability. — *Ann. Zool. Fennici* 36: 45–52.

A non-exhaustive literature search revealed that samples often show heterogeneity in the underlying developmental instability. As a consequence, the distribution of the signed asymmetry is leptocurtic. Simulations presented in this paper showed that a recently developed method to test heterogeneity in FA (likelihood ratio test of REML mixed regression models) has inflated type I error rates, whereas Levene's test suffered power reduction. The latter appeared to be the result of a lower accuracy of the estimation of population level FA. These effects became stronger with increasing leptokurtisis. In contrast, the estimation of R became more accurate and precise with increasing heterogeneity (and thus expected value of R). The estimation of small values of R is subject to extremely large sampling variation and were biased towards lower values. Implications for the analysis of FA at the individual and population level are discussed.

1. Introduction

The statistical analysis of fluctuating asymmetry (FA; small random deviations from perfect symmetry, Van Valen 1962) has received much attention since the influential paper by Palmer and Strobeck (1986) (e.g., Palmer & Strobeck 1992, Swaddle *et al.* 1994, Merilä & Björklund 1995, Whitlock 1996, Björklund & Merilä 1997, Pomory 1997, Smith *et al.* 1997, Swaddle & Cuthill 1997, Thomas & Poulin 1997, Van Dongen 1998a, Van Dongen *et al.* 1999a, Whitlock 1998, Van Dongen 1999). The use of fluctuating asymmetry as an estimate of developmental instability has four inherent difficulties. Firstly, as FA is often small it can be easily confounded with measurement error (ME) and directional asymmetry (DA). The use of within-subject repeats and mixed model analysis allows the separation of FA, ME and DA (Palmer & Strobeck 1986, Van Dongen *et al.* 1999a). Secondly, FA and antisymmetry cannot be separated by mixed model analysis and the statistical detection of antisymmetry often has low power (Palmer & Strobeck 1992). As antisymmetry may be more common than previously thought (e.g., Rowe *et al.* 1997) this poses a challenging problem for the statistical analysis of FA. Thirdly, since population-level FA is expressed as a variance, it cannot be estimated with great accuracy which has important power consequences (Palmer 1996, Björklund & Merilä 1997). Therefore, high sample sizes and possibly also high number of within-subject repeats (if ME is relatively large) are required (Van Dongen 1999). And finally, single trait individual asymmetry is only a very crude measure of individual developmental instability (Whitlock 1996), because it is an attempt to estimate a variance (developmental instability) with only two data points (left and right trait value). As a result, individual asymmetry is only loosely correlated with the presumed underlying individual developmental instability. Consequently, correlations between FA and other variables, between trait correlations in FA and heritability estimates of FA are biased downwards (Whitlock 1996, Houle 1997). The hypothetical repeatability (R, i.e. an estimate of the proportion of the total variation in the unsigned FA that can be attributed to between-individual heterogeneity in the presumed underlying developmental instability) can be used to correct for such biases (*see* Whitlock (1996) for details and Van Dongen (1998a) and Whitlock (1998) for a correction of the computational formula of R). Because of these four difficulties the use of FA as an estimate of developmental instability requires careful measurement and statistical analysis (*see* also Palmer 1996).

Palmer and Strobeck (1992) argued that distributions of the signed FA (left-right) deviating from ideal FA (i.e. mean zero and normal distribution) are unsuitable as descriptors of developmental instability because a fraction of the asymmetry variation may have a genetic basis. However, deviations from normality, and leptokurtisis in particular, are very common in studies of FA (Whitlock 1996, Björklund & Merilä 1997, Gangestad & Thornhill 1999; Table 1). These deviations may have different origins. The sample may contain a minority of antisymmetrical individuals or may consist of a mixture of individuals exhibiting different levels of developmental instability (Palmer & Strobeck 1992, Van Dongen 1998b). Since antisymmetry may have a genetic basis it does not necessarily reflect developmental instability (Palmer & Strobeck 1992, but *see* Graham *et al.* 1993). Developmental instability on the other hand may be heritable as well (Møller & Thornhill 1997, but *see* e.g. Leamy 1997, Markow & Clarke 1997, Whitlock & Fowler 1997), and environmental conditions are often heterogeneous at a small spatial scale (e.g., Van Dongen *et al.* 1997). Therefore, it can be expected that samples will often contain individuals with different degrees of developmental instability and thus leptokurtically distributed signed FA (as indicated in Table 1). The estimation of individual developmental instability via individual asymme-

Table 1. Hypothetical repeatabilities (R) of individual single trait fluctuating asymmetry for different species and traits. The hypothetical repeatability equals zero for samples exhibiting normally distributed signed FA, and increases with the degree of leptokurtisis (Van Dongen 1998a,1998b). The maximal value of R can be shown to equal 0.637. Values of R were calculated following Van Dongen (1998a) or Whitlock (1998).

Species	Trait	R	Reference
Agelaius phoeniceus	various	0.31–0.54	Dufour & Weatherhead 1996
Various	various	0–0.51	Van Dongen 1998b
Gorilla gorilla gorilla	canines	0.38	Manning & Chamberlain 1994
Various birds	wing/tail	0.42–0.63	Møller & Höglund 1991
Various	various	0–0.63	Whitlock 1996
Forficula auricularia	forceps	0.59	Radesäter & Halldòrdòttir 1993
Operophtera brumata	tibia/wing	0–0.52	Van Dongen *et al.* 1999b

try requires heterogeneity in the underlying developmental instability because otherwise all between-individual variation in the unsigned FA is due to sampling variation. This lack of heterogeneity results in a hypothetical repeatability of zero whereas leptokurtically distributed signed FA yields positive values of *R* (Palmer 1996, Van Dongen 1998a, Whitlock 1998). Heterogeneity in developmental instability within a sample is, therefore, indispensable for the analysis of patterns in individual FA (Van Dongen 1998a), whereas it may hamper the analysis of population level FA (Palmer & Strobeck 1992). As leptokurtisis is common in FA studies, it is very useful to evaluate its effects on accuracy, precision and power in the statistical analysis of FA.

In this paper, I present simulation data investigating the effect of between-individual variation in developmental instability within a sample, on the performance of a newly developed technique to test between sample heterogeneity based on REML estimation (Van Dongen *et al.* 1999a). I compare the performance of this test with Levene's method, which has previously been shown to be quite robust against deviations from normality and to have relatively high statistical power (Palmer & Strobeck 1992). I also compare the accuracy of the estimation of the hypothetical repeatability and of population level FA under various degrees of between-individual heterogeneity in developmental instability. I did not consider antisymmetry as this form of asymmetry requires different statistical methods. A SAS program was used to perform accuracy and power estimation.

2. Simulations

Four different degrees of between-individual heterogeneity in developmental instability were analysed, ranging from no heterogeneity and consequently a repeatability of zero, to relative high heterogeneity and a repeatability of 0.48 (Table 2). As a consequence of this heterogeneity, the distributions became leptokurtic and the coefficient of variation of the unsigned FA became larger than 76% (i.e. the value expected under normality, Björklund & Merilä 1997) (Table 2). For each of these 4 degrees of heterogeneity (further called no, low, moderate and high), 5 different levels of overall developmental instability were modelled. The lowest level of overall developmental instability was multiplied by 1.5, 2, 3 and 4 (Table 2). In all simulations two within-subject repeats were obtained and a low degree of measurement error was modelled (Variance = 0.0625). Note that within the four heterogeneity groups the value of *R* slightly increases with increasing level of overall developmental instability. This is because ME remains constant and thus becomes relatively less important for higher developmental instability.

The accuracy of the estimation of FA and *R* was simulated for three different sample sizes (20, 40 and 80). Accuracy was expressed as the coefficient of variation (CV) to allow comparison between estimates with different mean values (Sokal & Rohlf 1995). Bias was calculated as the expected value minus the observed average. The variance and mean squared error (MSE) (two other commonly used accuracy measures) can be easily calculated from the CV, the observed average and the bias (e.g., Sokal & Rohlf 1995) and were not reported to promote readability of the table. Effects of the degree of heterogeneity in developmental instability on the power and type I error rate of both Levene's test (i.e. a one-way ANOVA on the unsigned FA, e.g. Palmer & Strobeck 1992) and the REML approach (i.e. likelihood ratio test, Van Dongen *et al.* 1999a) were analysed for the same three different sample sizes. The overall level of developmental instability was compared between two samples. The lowest degree of overall developmental instability was compared to all different levels. In this way the ratio of the two levels of overall developmental instability ranged between 1 and 4. The proportion of significant tests ($p < 0.05$) was used as an estimate of power. In the simulations where the level of developmental instability was the same for the two samples, this proportion is an estimate of the type I error rate, which ideally should approach the nominal level of 5%. All simulations were performed in SAS (version. 6.12) and were of size 1000.

3. Results

3.1. The accuracy and precision of FA and *R*

The CV and bias of single sample FA values and of the hypothetical repeatability are summarised in Table 3. As expected, both FA and *R* are estimated more accurately when sample sizes are increased. With increasing heterogeneity in between-individual variation in developmental instability, the accuracy of the estimation of FA decreased as judged from the 70%–80% increase of CV. There was no consistent bias. The estimation of *R* became more accurate with increasing heterogeneity. For small sample sizes and low heterogeneity in the underlying individual developmental instability, *R* was underestimated. This bias decreased with increasing sample size and expected value of *R* (Table 3). In Fig. 1, the distribution of *R* is given for the three sample sizes and the low level of between-individual heterogeneity in developmental instability (i.e. expected value of $R = 0.08$, *see* also Table 2). The mean of the distribution approached the expected value of *R* with increasing sample size. The distribution of *R* for the smaller sample sizes (20 and 40) showed a long tail towards lower values. Obviously, negative values of *R* are meaningless. However, the median and mode of the distributions of *R* were lower than the expected value as well. This effect was again stronger for small samples and relative low expected values of *R* (*see* also Fig. 1).

3.2. Power and type I error rate of Levene's and the likelihood ratio test

The likelihood ratio test had higher power relative to Levene's test when all individuals within a sample had the same level of developmental stabil-

Table 2. Distribution details of the different parameter combinations. For each parameter combination a sample of 10 000 individuals was generated in SAS (Version. 6.12). Within population heterogeneity in developmental instability (4 levels: no, low, moderate and high) was generated by up to three different degrees (FA1, FA2 and FA3 with different frequencies (prop.)). The simulations are ordered in blocks with different levels of within-sample heterogeneity. Within blocks, variation in overall developmental instability levels was generated by multiplying FA1, FA2 and FA3 of each heterogeneity class with 1, 1.5, 2, 3, and 4. Measurement error (ME) was held constant (= 0.0625). For each simulation, the expected variance (V_{exp}= weighed average of FA1, FA2 and FA3 + ME) and observed variance (V_{obs}) as well as the kurtosis (K) of the signed FA are reported. For the unsigned FA, the variance (V_{lfal}) and the coefficient of variation (CV) are given. The hypothetical repeatabilities were estimated following Van Dongen (1998a).

FA1	Prop.	FA2	Prop.	FA3	Prop.	V_{exp}	V_{obs}	K	V_{lfal}	CV	R
No											
–	–	0.25	(100)	–	–	0.31	0.31	0.04	0.12	0.76	0
–	–	0.375	(100)	–	–	0.44	0.44	0.04	0.16	0.75	–0.01
–	–	0.5	(100)	–	–	0.56	0.55	–0.03	0.20	0.75	–0.01
–	–	0.75	(100)	–	–	0.81	0.83	0.02	0.31	0.76	0
–	–	1	(100)	–	–	1.06	1.07	–0.01	0.39	0.76	0
Low											
0.0625	(75)	0.25	(25)	–	–	0.17	0.17	0.72	0.07	0.81	0.08
0.0938	(75)	0.375	(25)	–	–	0.22	0.22	0.77	0.09	0.82	0.10
0.125	(75)	0.5	(25)	–	–	0.28	0.28	1.10	0.12	0.83	0.11
0.1875	(75)	0.75	(25)	–	–	0.39	0.39	1.20	0.16	0.84	0.11
0.25	(75)	1	(25)	–	–	0.50	0.49	1.26	0.21	0.85	0.13
Moderate											
0.0625	(50)	0.25	(25)	1	(25)	0.41	0.39	2.80	0.19	0.97	0.25
0.0938	(50)	0.375	(25)	1.5	(25)	0.58	0.58	2.93	0.29	1.01	0.28
0.125	(50)	0.5	(25)	2	(25)	0.75	0.75	3.34	0.38	1.02	0.29
0.1875	(50)	0.75	(25)	3	(25)	1.09	1.09	3.75	0.57	1.04	0.30
0.25	(50)	1	(25)	4	(25)	1.44	1.42	3.75	0.74	1.04	0.30
High											
0.0625	(75)	–	–	4	(25)	1.11	1.08	7.42	0.71	1.38	0.45
0.0938	(75)	–	–	6	(25)	1.63	1.61	7.14	1.07	1.42	0.46
0.125	(75)	–	–	8	(25)	2.16	2.19	7.19	1.50	1.47	0.47
0.375	(75)	–	–	12	(25)	3.20	3.21	7.17	2.22	1.50	0.48
0.25	(75)	–	–	16	(25)	4.25	4.38	7.53	3.04	1.51	0.48

ity. However, the type I error rate was inflated when the distribution of the signed FA was leptocurtic. This effect became stronger with increasing values of R and thus with increasing between-individual heterogeneity in developmental instability (Fig. 2). Higher values of R resulted in a power reduction in Levene's test (Fig. 2).

4. Discussion

The simulations presented here reveal four important patterns. Firstly, the recently developed method for testing differences in FA based on a REML mixed regression model (Van Dongen *et al.* 1999a) is sensitive to departures from normality as the type I error rate is inflated. Thus, this method is inappropriate to test for heterogeneity in FA between samples if individuals differ strongly in their developmental instability within samples. In the mixed regression model, individual signed asymmetry is modelled as a random slope (Van Dongen *et al.* 1999a), hereby making the assumption that the slope follows a normal distribution (e.g., Verbeke 1997). The inflated type I error rate is thus likely to be directly attributable to the violation of this assumption. Secondly, the power of Levene's test is strongly reduced for leptokurtic distributions, whereas the type I error rate remains largely unaffected (*see* also Palmer & Strobeck 1992 for similar results). Thirdly, and related to the previous point, the accuracy of FA estimation at the population level decreases with increasing within-sample heterogeneity in developmental instability. And fourthly, the estimation of the hypothetical repeatability is biased towards lower values for small samples and relative low degree of heterogeneity in individual developmental instability within a sample. In addition, the accuracy of the estimation of R decreases with decreasing degree of heterogeneity as well. Thus, if heterogeneity in individual developmental instability is low, FA can be estimated with relatively high accuracy and differences can be tested with

Table 3. Accuracy and precision of the estimation of population level FA and the hypothetical repeatability (R). Simulations of size 1 000 were performed for the four levels of between-individual heterogeneity in developmental instability (no, low, moderate and high) and for 3 different sample sizes (20, 40 and 80). For each simulation FA was estimated from a mixed regression model (Van Dongen *et al.* 1999a) and R was estimated following Van Dongen (1998a). Mean (±SD) values of FA and R are given for all simulation conditions. Accuracy is expressed by the coefficient of variation (CV) and the precision as the bias (expected value-observed mean). For the hypothetical repeatability there was a consistent bias towards lower values, whereas for FA no consistent bias was found. To illustrate the magnitude of the bias I also present it as a percentage. Note that for the simulations where the expected value of R ($R_{exp.}$) equalled zero, the CV and proportional bias were not calculated as this would involve division by values close to zero.

N	Degree of heterogeneity			
	No	Low	Moderate	High
$R_{exp.}$:	0	0.08	0.25	0.45
$FA_{exp.}$:	0.25	0.109	0.344	1.047
20	FA = 0.255 (0.10), CV = 40 bias = 0.005	FA = 0.115 (0.07), CV = 59 bias = 0.006	FA = 0.345 (0.21), CV = 60 bias = –0.001	FA = 1.077 (0.77), CV = 73 bias = 0.030
	R = –0.061 (1.38) bias = –0.061	R = 0.017 (0.23), CV = 1360 bias = –0.063 (78.8%)	R = 0.192 (0.19), CV = 98 bias = –0.058 (1.5%)	R = 0.395 (0.15), CV = 37 bias = –0.055 (2.5%)
40	FA = 0.249 (0.08), CV = 30 bias = –0.001	FA = 0.111 (0.05) , CV = 45 bias = 0.002	FA = 0.355 (0.15), CV = 42 bias = 0.011	FA = 1.038 (0.54), CV = 52 bias = –0.009
	R = –0.018 (0.44) bias = –0.018	R = 0.047 (0.15), CV = 328 bias = –0.033 (41.3%)	R = 0.205 (0.13), CV = 64 bias = –0.045 (1.1%)	R = 0.431 (0.08), CV = 18 bias = –0.019 (0.8%)
80	FA = 0.245 (0.05), CV = 21 bias = –0.005	FA = 0.110 (0.03), CV = 30 bias = 0.001	FA = 0.346 (0.10), CV = 28 bias = –0.002	FA = 1.050 (0.37), CV = 35 bias = 0.003
	R = –0.008 (0.12) bias = –0.0008	R = 0.059 (0.113), CV = 191 bias = –0.021 (26.3%)	R = 0.224 (0.08), CV = 36 bias = –0.026 (0.6%)	R = 0.439 (0.04), CV = 10 bias = –0.011 (0.3%)

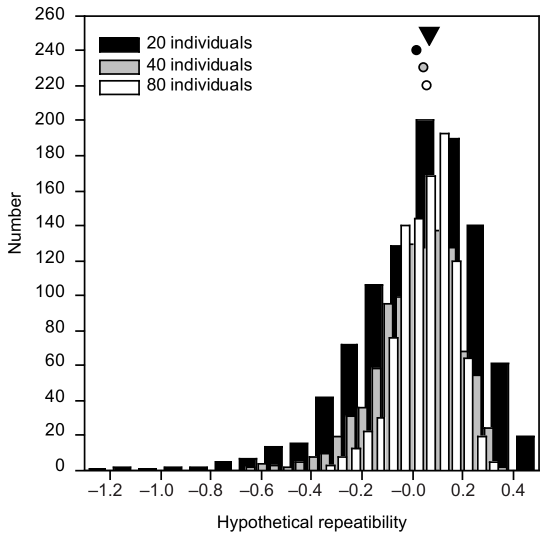

Fig. 1. Sampling distribution of the hypothetical repeatability. The distribution is based on 1 000 simulations for 3 different sample sizes (20, 40 and 80) with a low between-individual heterogeneity in developmental instability. The inverse triangle indicates the expected value of R (i.e. = 0.08) whereas the different circles represent the means of the three distributions. Fill colours of the circles correspond to the different sample sizes.

relatively high power. Yet, the estimation of the hypothetical repeatability then requires very high sample sizes.

Palmer and Strobeck (1992) have argued that only normally distributed FA should be used as an estimate of population-level developmental stability, because other forms, that cause deviations from normality, may have a genetic basis (but *see* Graham *et al.* 1993). Leung and Forbes (1997) have argued that leptokurtic distributions may also express ideal FA and therefore should not be a priori discarded (*see* also Van Dongen 1998a). However, besides the statistical difficulties shown the present paper (*see* also Palmer & Strobeck 1992) and the problem that antisymmetry in a small fraction of the population may result in leptokurtisis as well, there are also difficulties in interpreting patterns in population level FA for leptocurtic distributed signed FAs. The simulations presented here are relatively simplistic in the sense that between population heterogeneity in developmental instability is generated by a proportional difference in all levels of developmental instability within a population. However, the frequencies of these different levels as well as their

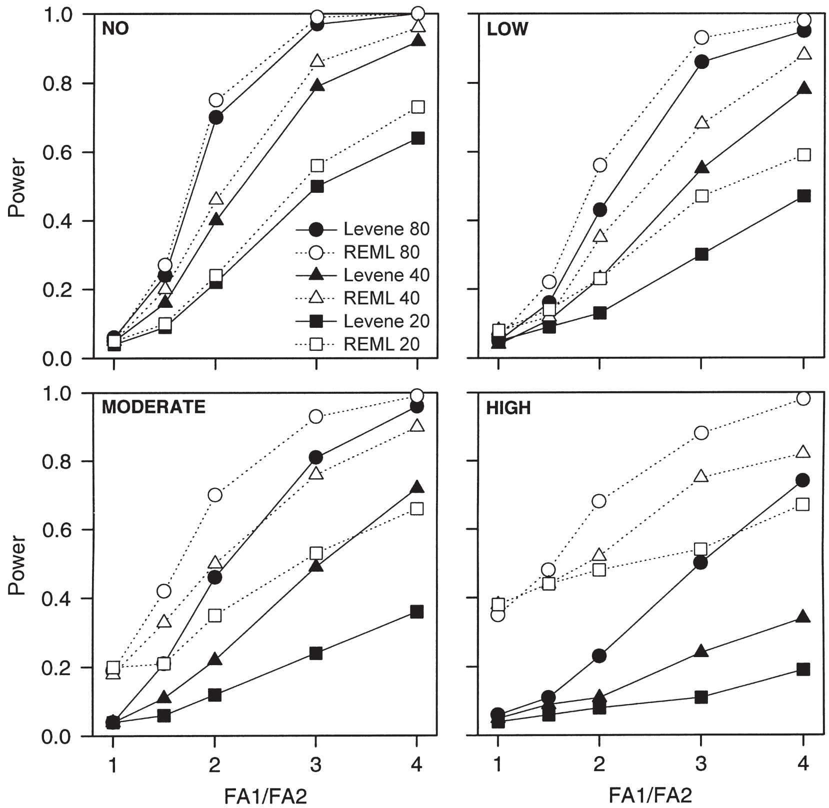

Fig. 2. Power curves for Levene's test (black symbols) and the likelihood ratio test based on a REML mixed regression model (open symbols) testing heterogeneity in developmental instability between two samples (FA1 and FA2). Power curves were estimated for the four levels of within-sample heterogeneity in developmental instability (no, low, moderate and high)

magnitudes may vary with stress, but also with several other factors. Suppose for example two hypothetical populations each of which contain two genotypes exhibiting different levels of FA (or different levels of susceptibility to some forms of stress). Due to random genetic drift, the allele frequencies may change in time and differ between the two populations such that observed differences in FA may have various origins. Thus, heterogeneity within samples may complicate the interpretation of patterns in FA. I therefore suggest that population level studies should attempt to sample as homogeneously as possible. Studies analysing individual asymmetry on the other hand, require heterogeneous sampling.

5. Conclusions

The results presented here have important consequences for the planning of field work and experiments. Depending on the degree of between-individual heterogeneity in developmental instability, as well as on the type of analysis (population vs. individual level analyses) that are planned, sample size requirements are different. Population-level analyses should attempt to sample as homogeneously as possible in order to increase power and accuracy. Individual-level analyses, on the other hand, should try to sample populations with high between-individual heterogeneity in developmental instability in order to increase the power of finding a relationship between FA and other measures. In addition, the bias as well as the extremely low accuracy of the estimation of R when its expected value is low can seriously bias transformation of patterns in FA into patterns in developmental instability. In particular, as these translations involve a division by R (Whitlock 1996, 1998), an underestimation of R will result in an overestimate of the patterns in developmental instability. Thus the application of R to translate patterns in FA should be done with great caution. This bias should, off course, not result in correlations between individual asymmetry and, for example, fitness.

Acknowledgements: The author is research assistant with the Science Fund Flanders (FWO Vlaanderen). I thank Luc Lens and an anonymous referee for their comments on an earlier version of the manuscript.

References

Björklund, M. & Merilä, J. 1997: Why some measures of fluctuating asymmetry are so sensitive to measurement error. — *Ann. Zool. Fennici* 34: 133–137.

Dufour, K. & Weatherhead, P. J. 1996: Estimation of organism-wide asymmetry in red-winged blackbirds and its relation to studies of mate selection. — *Proc. R. Soc. London B* 263: 769–775.

Graham, J. H., Freeman, D. C. & Emlen, J. M. 1993: Antisymmetry, directional asymmetry, and chaotic morphogenesis. — *Genetica* 89: 121–137.

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

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

Leung, B. & Forbes, M. R. 1997: Modelling fluctuating asymmetry in relation to stress and fitness. — *Oikos* 78: 397–405.

Manning, J. T. & Chamberlain, A. T. 1994: Fluctuating asymmetry, sexual selection and canine teeth in primates. — *Proc. R. Soc. London B* 251: 83–87.

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. 1991: Patterns of fluctuating asymmetry in avian feather ornaments: implications for models of sexual selection. — *Proc. R. Soc. London B* 245: 1–5.

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

Palmer, A. R. 1996: Waltzing with asymmetry. — *Bioscience* 46: 518–532.

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

Palmer, A. R. & Strobeck, C. 1992: Fluctuating asymmetry as a measure of developmental stability: Implications of non-normal distributions and power of statistical tests. — *Acta Zool. Fennica* 191: 57–72.

Pomory, C. M. 1997: Fluctuating asymmetry: biological relevance or statistical noise? — *Anim. Behav.* 53: 225–227.

Radesäter, T. & Halldòrdòttir, H. 1993: Fluctuating asymmetry and forceps size in earwigs, *Forficula auricularia*. — *Anim. Behav*. 45: 626–628.

Rowe, L., Repasky, R. R. & Palmer, R. 1997: Size-dependent asymmetry: fluctuating asymmetry versus antisymmetry and its relevance to condition-dependent signalling. — *Evolution* 51: 1401–1408.

Smith, D. R., Crespi, B. J. & Bookstein, F. L. 1997: Fluctuating asymmetry in the honey bee, *Apis mellifera*: effects of ploidy and hybridization. — *J. Evol. Biol.* 10: 551–574.

Sokal, R. R. & Rohlf, F. J. 1995: *Biometry*. 3rd ed. Freeman & Co, New York.

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

Swaddle, J. P. & Cuthill, I. C. 1997: The biological relevance of testing for perfect symmetry. — *Anim. Behav.* 54: 475–476.

Thomas, F. & Poulin, R. 1997: Using randomisation techniques to analyse fluctuating asymmetry data. — *Anim. Behav*. 54: 1027–1029.

Van Dongen, S., Backeljau, T., Matthysen, E. & Dhondt A. A., 1997: Synchronization of hatching date with budburst of individual host trees (*Quercus robur*) in the winter moth (*Operophtera brumata*) and its fitness consequences. — *J. Anim. Ecol*. 66: 113–121.

Van Dongen, S. 1998a: How repeatable is the estimation of individual fluctuating asymmetry. — *Proc. R. Soc. London B* 263: 1423–1427.

Van Dongen, S. 1998b: The distribution of individual fluctuating asymmetry: why are the coefficients of variation of the unsigned FA so high? — *Ann. Zool. Fennici* 35: 79–85.

Van Dongen, S. 1999: Accuracy and power in fluctuating asymmetry studies: effects of sample size and number of within-subject repeats. — *J. Evol. Biol*. 12. [In Press].

Van Dongen, S., Molenberghs, G. & Matthysen, E. 1999a: Statistical analysis of fluctuating asymmetry: REML estimation of a mixed regression model. — *J. Evol. Biol*. 12: 94–102.

Van Dongen, S., Sprengers, E., Löfstedt, C. & Matthysen, E. 1999b: Heritability of tibia fluctuating asymmetry and developmental instability in the winter moth (*Operophtera brumata* L.) (Lepidoptera, Geometridae). — *Heredity* 82. [In Press].

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

Verbeke, G. 1997: Linear mixed models for longitudinal data. — In: Verbeke, G. & Molenberghs, G. (eds.), *Linear mixed models in practice: A SAS oriented approach*. Lecture notes in Statistics. Springer-Verlag, New York.

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

Whitlock, M. 1998: The repeatability of fluctuating asymmetry: A revision and extension. — *Proc. R. Soc. London B* 263: 1429–1431.

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