The analysis of fluctuating asymmetry

JOHN P. SWADDLE, MARK S. WITTER & INNES C. CUTHILL
School of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, U.K.

(Received 24 March 1994; initial acceptance 3 May 1994; final acceptance 27 May 1994; MS. number: sc-1002)

Fluctuating asymmetries result from the inability of individuals to undergo identical development of a bilateral trait on both sides of the body (Van Valen 1962). It has long been realized that a negative correlation exists between asymmetry and fitness (e.g. Beardmore 1960). Recently, in behavioural ecology, particularly in studies of sexual selection, there has been a growth of interest in the study of fluctuating asymmetries, both as measures of individual quality (e.g. Möller 1990, 1992a; Swaddle & Witter 1994; Witter & Swaddle 1994) and as indicators of the strength and direction of selection (e.g. Möller & Höglund 1991; Balhoff et al. 1993; Möller & Pomiankowski 1993). However, the most appropriate method of quantifying and analysing fluctuating asymmetry has been controversial (Cuthill et al. 1993; Evans & Hatchwell 1993; Harvey et al. 1993; Möller 1993a; Sullivan et al. 1993; Watson & Thornhill 1994); in this paper, we highlight limitations in previous approaches and consider more appropriate methods of analyses.

There can be gross differences in values of fluctuating asymmetry between different populations of the same species (Valentine & Soule 1973; Picton et al. 1990; Markowski 1993; Möller 1993b, in press), between years within populations (Zhukarov 1981; Möller 1993b) and at the extremities of a population range (Downhower et al. 1990; Parsons 1993). Comparisons between populations may be of intrinsic interest (references above), but behavioural ecologists are typically interested in within-population variation as an outcome of differences in condition or as an object of mate choice, in which case, heterogeneity of source invalidates museum and field studies based on pooled samples. For example, Möller (1992b) reported data on fluctuating asymmetry from 517 male barn swallows, Hirundo rustica, from colony sites in Denmark. These data are an accumulation of measurements from 18 different colonies spanning several years, so between-year, between-site and between-individual variation are confounded. Möller, having collected all the samples himself, can in principle separate these effects statistically or analyse more homogeneous samples (e.g. Möller 1990). However, this will rarely be possible with museum collections (e.g. Wayne et al. 1986; Möller & Höglund 1991; Möller 1992a; Balhoff et al. 1993; Manning & Chamberlain 1993, 1994) or uncontrolled field sampling (e.g. Solberg & Sæther 1994). These have additional biases including the following. (1) There may be differential mortality by level of fluctuating asymmetry, ornament size, or some interaction thereof. Hence measured relationships between asymmetry and size may reflect the action of natural selection rather than developmental constraints or condition-dependent expression. (2) Humans find symmetrical and elaborate objects more aesthetically pleasing (Eisenman & Rappaport 1967; Szilagyi & Baird 1977) and collectors, hunters and museum curators seeking ‘typical’ specimens are not immune to such biases. (3) Wear and damage asymmetry may not be discriminable from fluctuating asymmetry (see Cuthill et al. 1993; Möller 1993a).

Measurement error, like (signed) fluctuating asymmetry, is normally distributed with a mean of zero. So, since fluctuating asymmetries are generally very small relative to the size of the traits being measured (typically ca 1%; Möller & Pomiankowski 1993), measurements must be replicated to distinguish true asymmetry from measurement error. Measurement error has been shown to account for up to 25% of the variation in dental asymmetry data in wild mice (Bader 1965) and humans (Greene 1984) and 75% of the variation in wing length asymmetry in nymphalid butterflies (Mason et al. 1976). This point has been made previously by Palmer & Strobeck (1986), but has gone largely unheeded in the recent behavioural literature. Several recent papers provide no repeatability analyses (e.g. Thornhill 1992; Thornhill & Sauer 1992; Radesäter & Halldorsdóttir 1993). Even where
measurements have been replicated, repeatability of the trait size has been erroneously equated with accurate measurement of asymmetry. For example, Balmford et al. (1993) estimated the repeatability ($r_1$; see Zar 1984) of tail and wing length measurements in swallows, and Manning & Chamberlain (1994) for canine length in gorillas. However, low measurement error relative to (left or right) trait length does not imply low error relative to asymmetry. The appropriate analysis is a mixed-model ANOVA with factors Individual (I), Side (S; left or right) and Replicate (R; the repeated measurements). The ratio of the I-by-S mean square to the combined I-by-S-by-R and I-by-R mean squares provides an $F$-test of whether between-individual variation in estimated asymmetry is significantly greater than can be accounted for by measurement error (see also Palmer & Strobeck 1986). We illustrate this point with our own data, where the tarsus lengths of 35 starlings, *Sturnus vulgaris*, were measured twice for each leg using Vernier calipers to 0.01 cm accuracy. Repeatabilities of the lengths of left ($r_1=0.798$) and right ($r_1=0.911$) tarsi were suitably high ($P<0.0001$; cf. other morphometric studies), yet estimated asymmetry was not significantly higher than expected from the estimated measurement error ($F_{43,480}=0.650$, $P=0.970$). However, if the number of repeats is increased to, for example, six, the accuracy of the asymmetry measurement increases dramatically relative to the measurement error ($F_{43,480}=5.44$, $P<0.001$).

If asymmetry measures are repeatable, it is then necessary to distinguish fluctuating asymmetry from directional asymmetry or antisymmetry. The Kolmogorov–Smirnov or, worse still, the sign test, which considers only the direction of deviations from zero (see Siegel & Castellan 1988), are commonly employed to identify fluctuating asymmetries (e.g. all papers on fluctuating asymmetry cited above), but are relatively weak at detecting departures from normality. Normal probability plots are better; the (Filliben) correlation coefficient derived from these is powerful at detecting skew (Aitken et al. 1989). Having ascertained that the distribution is normal, the data should now be investigated to identify whether they are centred about zero using a one-sample $t$-test.

Often, if is of interest to compare the size of the asymmetry, relative to the size of the trait in question. It is important to realize that the use of relative measures of fluctuating asymmetry (i.e. $[(L - R)/0.5 \times (L + R)]$) 'control' for trait size only if the relationship between fluctuating asymmetry and trait size is isometric and intercepts the origin (see Cuthill et al. 1993). Where there is clearly a linear relationship between trait length and asymmetry, it may be more appropriate to control for trait size using analysis of covariance (cf. Packard & Boardman 1987; but see below). However, since one would predict a U-shaped relationship between asymmetry and trait size in traits under stabilizing selection, linear regression and correlation analyses would be inappropriate in such circumstances, although they are still applied (e.g. Radesäter & Halldörsdóttir 1993; Wakefield et al. 1993).

Absolute (unsigned) asymmetry ($|L - R|$) has a characteristic 'half-normal' distribution (Van Valen 1962). Although the assumptions of normality and homogeneity of variances for parametric statistics are likely to be violated, $t$-tests, ANOVA and linear regression are frequently used (e.g. Möller 1990; Thornhill & Sauer 1992; Radesäter & Halldörsdóttir 1993; Wakefield et al. 1993; Wilber et al. 1993; Manning & Chamberlain 1994; Solberg & Sæther 1994). The residuals from regression analyses may turn out to be normally distributed (e.g. Möller 1990), but more often they will not (e.g. Thornhill 1992; Radesäter & Halldörsdóttir 1993; Solberg & Sæther 1994); residuals in simple ANOVA or $t$-tests will never be. Two-parameter Box-Cox transformations of the form $(Y + \lambda x)^{\lambda}$ are often suitable for normalizing skewed positive data containing zeroes (Palmer & Strobeck 1986; Aitken et al. 1989) and we have explored their utility using GLIM (Numerical Algorithms Group 1985; Aitken et al. 1989). For a variety of feather (Swaddle & Witter 1994), tarsus (see above) and randomly generated data, values of $\lambda_1$ around 0.3 and $\lambda_2$ set to be somewhat smaller than the smallest non-zero asymmetry work well. The limitation of this approach is that small differences in asymmetry near zero have a large influence and, if zero asymmetry is common, then so too does the choice of $\lambda_2$ (see Aitken et al. 1989). However, for sexually selected characters, where average asymmetry is large (Möller & Pomiankowski 1993), this is less likely to be a problem. Otherwise, non-parametric techniques (see Siegel & Castellan 1988) should be routinely employed.

To summarize, it is imperative to ensure that: (1) groups of individuals are not excluded
from measurement because of sampling bias; (2) measures of asymmetry are repeatable, tested via mixed-model ANOVA; (3) asymmetries are correctly identified as fluctuating asymmetries using normal probability plots and one-sample $t$-tests; (4) due consideration is given to the statistical properties of the data set and the underlying assumptions about the developmental process (see Evans & Hatchwell 1993) when quantifying measures of fluctuating asymmetry; (5) the assumptions of parametric analyses are fulfilled, or the data are suitably transformed (using, for example, Box-Cox transformations), before applying parametric analyses on absolute (unsigned) fluctuating asymmetries.

M.S.W. and J.P.S. were supported by a NERC Research Fellowship and a SERC Research Studentship, respectively.

REFERENCES


