AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 96:83–87 (1995)

Reliability of Reliability Coefficients in the Estimation of Asymmetry

S.J. FIELDS, M. SPIERS, I. HERSHKOVITZ, AND G. LIVSHITS
Department of Anthropology, University of Michigan, Ann Arbor, Michigan 48109 (S.J.F.); Department of Anatomy and Anthropology (S.J.F., I.H., G.L.), Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel; Department of Anthropology, University of Pennsylvania, Philadelphia, Pennsylvania 19104 (M.S.)

KEY WORDS Measurement error, Anthopometric data, Odontometric data

ABSTRACT Although promising to provide insight into the interaction between genotype and environment, investigations into fluctuating asymmetry suffer from a lack of standardization in the reporting of measurement error. In the present paper we show, using both anthropometric and odontometric data, that the use of the reliability coefficient calculated for a bilateral measurement provides no indication of the reliability of the corresponding asymmetry estimate, because reliability of asymmetry depends on the relationship between measurement error and the difference between sides. Thus, we suggest that future investigations either provide reliability coefficients for asymmetry estimates specifically, or use methods that account for measurement error.

Fluctuating asymmetry (FA) is defined as the random difference between quantitative measures of a bilateral trait (Van Valen, 1962). Because the same genes control both sides of a bilateral trait, any difference between sides results from local disturbances (Mather, 1953). Thus, as a potential measure of genotype/environmental interaction FA may provide a measure of stress experienced by individuals during their development (Livshits et al., 1988; Livshits and Kobyliansky, 1991; Kieser, 1990).

Although this possibility is theoretically promising, major difficulties lie in the quantification of asymmetry because of the potentially confounding effects of measurement error, some of which have been addressed by Greene (1984) and by Palmer and Strobeck (1986). Thus, asymmetry studies often include some statistic summarizing error, either mean measurement error calculated over the different variables used in the study, or the reliability coefficient. However, measurement error varies with both the size of the trait under investigation and the definition of the landmarks used (Malina et al., 1973; Jamison and Ward, 1993).

While one can correct for the problem of size and landmark definition, the use of the reliability coefficient as an indicator of confidence in asymmetry estimates presents a different type of problem. As we shall presently show using different types of data, the reliability coefficient of a given bilateral measurement provides no indication of the reliability of the corresponding asymmetry estimate unless it is estimating the reliability of asymmetry itself.

MATERIALS AND METHODS

As part of a mixed-longitudinal growth study conducted between 1992 and 1993, one of us (S.J.F.) collected anthropometric data on 283 Israeli infants during their first

Received November 9, 1993; accepted July 25, 1994.

Address reprint requests to S.J. Fields, Department of Anatomy and Anthropology, Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel.

year of postnatal life. Ten bilateral measurements, each measured twice, provided the basis for the assessment of asymmetry, including: ear length, digit 3 length, palm length and breadth, bistyloid (wrist) and biepicondylar (elbow) breadth, foot breadth, bimalleolar (ankle) and bicondylar (knee) breadth, and tibial length.

We also used dental measurements derived from skeletal remains belonging to Epipaleolithic, Neolithic, and Chalcolithic archaeological sites in Israel. One of the authors (M.S.) measured mesiodistal and buccolingual antimeric pairs of permanent teeth twice using sliding digital calipers (0.02 mm). The four tooth types used in the analysis were: mandibular first molars, mandibular central incisors, maxillary second premolars, and maxillary canines.

Several different methods exist for calculating FA (Palmer and Strobeck, 1986). In the present study we use the simplest definition of asymmetry: the root mean square (RMS) of the variance of the difference between right and left measurements of a bilateral trait (VAR(R-L)).

Fleiss (1986) defines REL, the reliability coefficient, as the proportion of variance free of intersubject measurement error. It corresponds to the intraclass correlation coefficient and is given by:

$$\text{Rel} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2} \qquad (1)$$

where σ_T^2 equals the population variance based on the means of the first and second measurement of the same trait and σ_e^2 equals the variance of the differences between the first and second measurements of the trait. We rewrite Equation 1 to express reliability in terms of asymmetry:

$$\frac{\text{Var}\,\{\bar{X}_a\}}{\text{Var}\,\{\bar{X}_a\} + \text{Var}\,\{D_a\}} \qquad (2)$$

where X_a = the mean of the first and second asymmetry estimates and D_a = the difference between the first and second asymmetry estimates. For the sake of simplicity we will refer to the first and second measurement or estimate taken on a subject as a repeated measure.

Expressing Equation 2 in terms of repeated right (R) and left (L) measurement, REL equals:

$$\frac{\text{Var}\,\{[(R_1 - L_1) + (R_2 - L_2)]/2\}}{\text{Var}\,\{[(R_1 - L_1) + (R_2 - L_2)]/2\} + \text{Var}\,\{[(R_1 - L_1) - (R_2 - L_2)]\}}$$

and rearranging to express reliability in terms of right and left means:

$$\frac{\text{Var}\,\{[(R_1 + R_2) - (L_1 + L_2)]/2\}}{\text{Var}\,\{[(R_1 + R_2) - (L_1 + L_2)]/2\} + \text{Var}\,\{[(R_1 - R_2) - (L_1 - L_2)]\}}.$$

Letting $\bar{X}_R$ and $\bar{X}_L$ represent the respective means of the repeated right and left measurements, D_R and D_L represent the respective differences between the repeated measurements, and Cov represent covariance, the above is equivalent to:

$$\frac{\text{Var}\,\{\bar{X}_R\} + \text{Var}\,\{\bar{X}_L\} - 2\text{Cov}\,\{\bar{X}_R,\bar{X}_L\}}{\text{Var}\,\{\bar{X}_R\} + \text{Var}\,\{\bar{X}_L\} - 2\text{Cov}\,\{\bar{X}_R,\bar{X}_L\} + \text{Var}\,\{D_R\} + \text{Var}\,\{D_L\} - 2\text{Cov}\,\{D_R,D_L\}}.$$

Assuming that Var $\{\bar{X}_R\}$ = Var $\{\bar{X}_L\}$, Var $\{D_R\}$ = Var $\{D_L\}$, and since Cov $\{\bar{X}_R,\bar{X}_L\}$ = ρVar{i}, where ρ equals the parametric correlation coefficient, the above simplifies to:

$$\frac{2\text{Var}\{i\} - 2\rho_i\text{Var}\{i\}}{2\text{Var}\{i\} - 2\rho_i\text{Var}\{i\} + 2\text{Var}\{j\} - 2\rho_j\text{Var}\{j\}} \qquad (3)$$

where Var{i} equals the variance of either the right or left means, and Var{j} equals the variance of either the right or left repeat difference. After collecting like terms, Equation 3 equals:

$$\frac{(1 - \rho_i)\text{Var}\{i\}}{(1 - \rho_i)\text{Var}\{i\} + (1 - \rho_j)\text{Var}\{j\}} \qquad (4)$$

where ρ_i is the correlation between mean right and left measurements and ρ_j is the correlation between the repeat differences of the right and left sides (i.e., D_R and D_L).

TABLE 1. Estimates of mean, variance, measurement error, and reliability for 10 right and left anthropometric measurements (100 cm) on living infants

Variable	Side	N	Mean[1]	Variance[1]	VAR(E)	REL
Ear	R	275	340.15	1,881.00	116.12	94.19
	L	275	339.67	1,806.46	115.31	94.00
Digit	R	273	249.69	843.93	89.51	90.41
	L	274	249.65	989.53	81.10	92.43
Palm length	R	267	358.93	1,639.36	149.61	91.64
	L	271	355.17	1,792.66	135.37	92.98
Palm breadth	R	269	330.09	1,631.61	87.67	94.90
	L	266	327.86	1,803.13	90.86	95.20
Wrist	R	270	225.63	862.98	43.12	95.24
	L	273	223.74	928.82	53.94	94.51
Elbow	R	273	243.50	1,578.25	79.19	95.22
	L	278	240.00	1,567.69	65.70	95.98
Foot	R	275	304.73	2,137.24	129.19	94.30
	L	275	305.49	2,008.06	129.28	93.95
Ankle	R	279	246.22	1,177.11	96.61	92.42
	L	277	247.02	1,150.25	87.02	92.97
Knee	R	283	342.49	3,386.24	76.03	97.80
	L	282	341.65	3,295.05	74.81	97.78
Tibia	R	275	1,233.91	27,910.30	257.19	99.09
	L	272	1,235.31	27,304.98	229.08	99.17

[1]Based on mean of repeated measures.

TABLE 2. Estimates of mean, variance, measurement error, and reliability for right and left mesiodistal (MD) and buccolingual (BL) measurements on four teeth (100 mm) from archaeological samples

Variable	Side	N	Mean[1]	Variance[1]	VAR(E)	REL
Incisor						
MD	L	43	543.09	1,364.11	25.65	98.15
MD	R	44	536.40	1,795.20	26.67	98.54
BL	L	36	619.26	1,369.32	61.86	95.68
BL	R	38	613.59	1,495.92	42.64	97.23
Canine						
MD	L	42	787.29	2,142.23	29.66	98.63
MD	R	41	784.44	3,092.35	47.61	98.48
BL	L	41	880.30	3,032.45	51.45	98.33
BL	R	41	881.66	4,161.09	39.35	99.06
Premoler						
MD	L	53	665.92	2,111.09	33.39	98.44
MD	R	54	663.39	1,610.51	27.20	98.34
BL	L	54	962.85	3,249.17	43.19	98.69
BL	R	54	965.09	3,618.12	20.93	99.42
Molar						
MD	L	44	1,144.32	4,765.77	60.31	98.75
MD	R	45	1,142.77	5,237.75	48.45	99.08
BL	L	43	1,073.01	2,812.29	72.24	97.50
BL	R	45	1,074.93	2,638.93	74.28	97.26

[1]Based on mean of repeated measures.

RESULTS

Table 1 provides the means, variance, error variance [VAR(E)], and reliability (REL), for both sides of the 10 bilateral anthro- pometric traits estimated from the 283 infants. Reliability ranges from 90.41 to 99.17, implying that measurement error accounts for approximately 1 to 10 percent of the observed variation of the respective measures.

The odontometric data also display high reliabilities, ranging from 95.68 to 99.54% (Table 2). With a reliability averaging 98.30%, measurement error constitutes a relatively small percent of the total odontometric variance.

In Table 3 we provide VAR(R-L), defined as the variances of the mean side differences, VAR(E), defined as the error variance or the variance of the difference between re-

TABLE 3. Variance, measurement error, and reliability of asymmetry estimates for the 10 anthropometric and 8 odontometric variables

Variable	N	VAR(R-L)	VAR(E)	REL
Anthropometrics (100 cm)				
Ear	273	128.75	168.19	43.36
Digit	262	41.49	121.55	25.45
Palm length	256	48.27	148.14	24.58
Palm breadth	255	38.20	117.04	24.61
Wrist	258	30.81	54.78	36.00
Elbow	267	38.04	94.81	28.63
Foot	270	44.75	109.55	29.00
Ankle	275	78.31	133.24	37.02
Knee	282	37.82	98.91	27.66
Tibia	272	131.30	267.52	32.92
Odontometrics (100 mm)				
Incisor				
MD	43	439.06	47.55	90.23
BL	33	293.45	82.04	78.15
Canine				
MD	41	926.22	82.81	91.79
BL	40	767.69	53.47	93.49
Premolar				
MD	53	788.18	44.95	94.60
BL	54	595.50	80.02	88.15
Molar				
MD	44	582.73	174.33	76.97
BL	43	192.38	169.68	53.14

peated asymmetry estimates, and REL of the asymmetry estimates for both the anthropometric and odontometric traits. Subtracting VAR(E) from VAR(R-L) estimates asymmetry that is free of measurement error (Greene, 1984). If VAR(E) exceeds VAR(R-L), asymmetry will be indistinguishable from measurement error and reliability will be small. This is the case with the anthropometric data, which show asymmetry reliabilities ranging from 24.45 to 43.36%. By contrast, VAR(R-L) exceeds VAR(E) values by as much as 17-fold (premolar MD) in the odontometric data. Thus, reliabilities of the asymmetry estimates for the odontometric data are higher than those seen in the anthropometric data, ranging from 53.14 to 94.60%.

DISCUSSION

Comparing the reliabilities of the bilateral measurements given in Tables 1 and 2 with the reliabilities of the corresponding asymmetries in Table 3 clearly shows the lack of congruity between the two metrics. Although both anthropometric and odontometric reliabilities both exceed 90%, the reliabilities of the respective asymmetries varies from 24 to 94%. Thus, the reliability of a particular measure provides no indication of the reliability of the corresponding estimate of asymmetry. This is because reliability of asymmetry depends on the relationship between measurement error and the size of the difference between sides and not on measurement error alone. Examination of Equation 4 clearly demonstrates why this is so.

As the correlation between the right and left repeated measurements, ρ_j, increases, reliability approaches unity because $(1 - \rho_j)\text{Var}\{j\}$ approaches zero. Similarly, as the variance of the repeated measures, Var{j}, decreases relative to Var{i}, reliability increases. However, as ρ_j decreases, it inflates the relative contribution of Var{j} so that if Var{j} is large relative to Var{i}, reliability will decrease. Similarly, as the correlation between sides, ρ_i, approaches unity, the term $(1 - \rho_i)\text{Var}\{i\}$ approaches zero, asymmetry decreases, and reliability becomes extremely sensitive to measurement error. In such a case, Var{j} will have to be very small for there to be any meaningful measure of asymmetry.

Two important points arise from these results. First, the reliability of a measure, such as left tibial length, gives no indication of the reliability of asymmetry estimates because the reliability of asymmetry depends on the relationship between measurement error and the size of the difference between sides and not solely on measurement error. Second, the accuracy and repeatability of the asymmetry estimate depends on the variable under investigation. In this study we compared measurements taken from archaeological specimens with measurements from living infants under 1 year of age. While comparisons of asymmetry estimates from archaeological data have their own set of problems (Smith et al., 1982), they do not present the same set of difficulties inherent in measuring squirming babies.

Given the difficulties with the reliability of FA measures, it is interesting to note a relatively large number of studies showing significant levels of FA even though Smith et al. (1982) have shown that small sample size decreases the level of statistical power. While the possibility exists that these studies reflect true biological phenomena, the lack of proper control of measurement error with respect to right/left differences blurs the difference between observer error and real genotype/environmental interaction.

We thus recommend that investigators either directly estimate reliability coefficients for their asymmetry measures or use statistical methods that account for measurement error such as the two-way analysis of variance discussed by Palmer and Strobeck (1986). Without such information, the interpretation of the results of asymmetry studies holds little promise of clarifying the genotype/environmental interaction its users claim to be investigating.

ACKNOWLEDGMENTS

We wish to express our thanks for helpful discussion to Professors F. Bookstein, S.M. Garn, and A.R. Frisancho, and to Mr. G. Goldsands. We are also grateful to the two anonymous reviews for their thoughtful and helpful suggestions. This work was supported in part by N.S.F. Doctoral Dissertation Improvement grant BNS-9115973 to S.F.

LITERATURE CITED

Fleiss JL (1986) The Design and Analysis of Clinical Experiments. New York: John Wiley and Sons.

Greene DL (1984) Fluctuating dental asymmetry and measurement error. Am. J. Phys. Anthropol. *65:*283–289.

Jamison PL, and Ward RE (1993) Measurement size, precision, and reliability in craniofacial anthropometry: Bigger is better. Am. J. Phys. Anthropol. *90:*495–500.

Kieser JA (1990) Human Adult Odontometrics. The Study of Variation in Adult Tooth Size. Cambridge: Cambridge University Press.

Livshits G, and Kobyliansky E (1991) Fluctuating asymmetry as a possible measure of developmental homeostasis in humans. Hum. Biol. *63:*441–466.

Livshits G, Davidi L, Kobyliansky E, Ben-Amitai D, Levi Y, and Merlob P (1988) Decreased developmental stability as assessed by fluctuating asymmetry of morphometric traits in preterm infants. Am. J. Med. Genet. *29:*793–805.

Malina RM, Hamill PVV, and Lemeshow S (1973) Body dimensions and proportions, White and Negro children 6-11 years, United States. *Vital and Health Statistics Series 11, No 143.* Washington, DC: U.S. Department of Health and Human Services.

Mather K (1953) Genetic control of stability in development. Heredity *7:*397–436.

Palmer AR, and Strobeck C (1986) Fluctuating asymmetry: Measurement, analysis, patterns. Ann. Rev. Ecol. Syst. *17:*391–421.

Smith BH, Garn SM, and Cole PE (1982) Problems of sampling and inference in the study of fluctuating dental asymmetry. Am. J. Phys. Anthropol. *58:*281–290.

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