# CORRECTING FOR SAMPLING BIAS IN QUANTITATIVE MEASURES OF SELECTION WHEN FITNESS IS DISCRETE

Wolf U. Blanckenhorn,[1,2] Max Reuter,[1,3] Paul I. Ward,[1] and Andrew D. Barbour[4]

[1]*Zoological Museum, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*
[2]*E-mail: wolfman@zoolmus.unizh.ch*
[4]*Applied Mathematics, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*

*Abstract.*—We show with a simulation that nonrepresentative sampling of two discrete fitness classes leads to biased estimates of selection. Systematic underestimation occurs if the selected class is overrepresented in the sample and overestimation if the unselected class is overrepresented. The bias is greater the stronger the selection intensity, the smaller the true fraction of individuals favored by selected, and the lower the sample size. We present a simple method that allows a posteriori statistical correction in cases of biased sampling given a separate estimate of the actual class representation, describe its practical implementation, and show that it works.

*Key words.*—Computer simulation, fitness, natural selection, selection differentials, selection gradients, sexual selection, viability selection.

Natural selection is one of the primary mechanisms of organismic evolution. Although selection had been studied long before (see Manly 1985), two seminal papers by Lande and Arnold (1983) and Arnold and Wade (1984a) defined a standardized statistical machinery for measuring selection. The major advance was threefold. First, these quantitative measures of selection are closely linked with formal evolutionary theory as found in standard textbooks of evolution (e.g., Maynard Smith 1992) and quantitative genetics (e.g., Falconer 1989), and hence can be interpreted in these contexts. Second, the measures are typically standardized and therefore comparable across fitness components, selection episodes, and even species. Third, they can be easily computed using standard statistical packages. These methods have since been refined and elaborated in several ways (Manly 1985; Endler 1986; Mitchell-Olds and Shaw 1987; Schluter 1988; Kingsolver and Smith 1995; Brodie and Janzen 1996; see Brodie et al. 1995), and many studies have quantified selection and furthered our understanding of evolution in natural systems.

Despite their overall success, quantitative measures of selection are not without problems. As we shall elaborate below, there are several different but related measures of selection, and often regression of fitness on trait values is employed in their estimation (Lande and Arnold 1983; Arnold and Wade 1984a,b; Brodie et al. 1995). One problem is that the major fitness components estimated may take either discrete (often binary) or continuous values. The former is frequently the case for sexual selection, where a male may be scored as having mated or not having mated with a small number of females, and for viability selection, where individuals may be scored as dead or alive (Lande and Arnold 1983; Arnold and Wade 1984b; Brodie and Janzen 1996). In contrast, fitness, in terms of number of offspring, is often a

continuous function of a trait (e.g., fecundity or fertility selection as a function of body size in bullfrogs; Howard 1979; Arnold and Wade 1984b). Despite assertions by Lande and Arnold (1983) and Arnold and Wade (1984b) that fitness need not be normally distributed, the statistical assumptions of least-squares regression used for its estimation may be violated if it is not (Mitchell-Olds and Shaw 1987), particularly when fitness is binary. This problem is largely one of significance testing as opposed to parameter estimation (Lande and Arnold 1983; but see Mitchell-Olds and Shaw 1987) and this has been extensively discussed (Lande and Arnold 1983; Manly 1985; Endler 1986; Mitchell-Olds and Shaw 1987; Kingsolver and Smith 1995; Brodie and Janzen 1996). The consensus so far was succinctly stated by Brodie et al. (1995, p. 318): "Generalized linear models (e.g., logistic regression) are more appropriate for some data sets than linear regression, but the coefficients from such models are not yet interpretable in the context of equations for evolutionary change." Therefore, researchers now often report selection coefficients derived from ordinary least-squares regression, but supply significances derived from logistic regression (e.g., Fairbairn and Preziosi 1994) or using jackknifed standard errors (e.g., Brodie and Janzen 1996).

Sampling bias (nonrandom sampling) is a second problem that potentially afflicts the measurement of selection (Manly 1985). This is particularly obvious when dealing with binary fitness values in studies of viability or sexual selection. As Arnold and Wade (1984b, p. 722) put it: "to accurately estimate the sexual selection differential from cross-sectional data, it is essential to sample mated and unmated males in proportion to their actual representation in the population. In contrast, nonsystematic samples of the two fractions with greater effort generally devoted to the rarer fraction often suffice for hypothesis testing but not for parameter estimation." In other words, biased sampling leads to inaccurate estimates, but has little bearing on significance testing. This problem in measuring selection has thus far received much less attention, perhaps because sampling classes of individ-

---

[3] Present address: Institut de zoologie et d'écologie animale, Université Lausanne, Bâtiment de Biologie, CH-1015 Lausanne, Switzerland; E-mail: max.reuter@izea.unil.ch.

uals (mated/unmated or alive/dead) in proportion to their actual occurrence is possible in several species (e.g., water striders: Arnqvist 1992; Fairbairn and Preziosi 1994) and thus believed to constitute no problem. This is not the case. In many animal systems, different classes of individuals are not equally accessible, for example, nonbreeding (as opposed to breeding) birds or mammals that have no fixed site. Sampling classes in proportion to their occurrence is then simply not possible, ultimately precluding the use of quantitative selection measures. Furthermore, even when all classes are accessible, proportional sampling can be very restrictive when population sizes are low because only few individuals of the rarer class may be sampled, which impedes the quality of the estimates and their statistical significance. Such situations occur frequently when the entity relevant for reproductive competition is not the population as a whole, for example, an entire pond or stream (e.g., Howard 1979; Fairbairn and Preziosi 1994), but instead particular mating sites like food plants of hemipterans (McLain 1992; McLain et al. 1993), oviposition sites of dung flies (i.e., a dung pat; e.g., Borgia 1982), or leks of vertebrates (see Andersson 1994), all of which may feature relatively few individuals. In such cases it would be desirable to sample as many individuals of the rarer class as possible. However, it is usually easy to acquire reasonable estimates of the proportional representation of the typically two classes of individuals (mated/unmated and breeders/nonbreeders). These can be obtained just prior to sampling by simply counting the number of paired and unpaired males at a particular mating site (e.g., a host plant or a dung pat; Carroll and Salamon 1995; Blanckenhorn et al. 1999) or from separate behavioral observations or experiments in cases where these proportions are reasonably consistent over space and time (e.g., Zuk 1988). From separate, more extensive mark-recapture or mark-resight datasets it should also often be possible to acquire a more accurate estimate of the proportion of dead individuals within a large study population when estimating viability selection using a smaller subsample (cf. examples in Lande and Arnold 1983 and Manly 1985).

Here we present and evaluate a simple method that allows a posteriori statistical correction when sampling is biased, as just outlined. We first derive our correction method and then describe its practical implementation, which depends on the statistical methods used to generate selection coefficients. We then present the results of a simulation showing that the correction works. The simulation also shows that Arnold and Wade's (1984b) abovecited, intuitive, but unsubstantiated assertion is correct: biased sampling indeed leads to incorrect estimates of selection coefficients, unless corrected as described here. Our method is of importance primarily for the practical aspects of measuring selection in nature.

## Correcting Selection Coefficients for Sampling Bias

### Coefficients of Selection

There are several quantitative measures of selection (listed in Brodie et al. 1995), the most common being the opportunity for selection $I$ (which shall not be further treated here), the selection differential, and the selection gradient. These coefficients are of different historical origin and can be computed in different ways, but they are intimately related mathematically (Wade and Arnold 1980; Lande and Arnold 1983; Arnold and Wade 1984a,b; Endler 1986; Brodie et al. 1995; see below). The selection differential $S$ calculated on traits standardized to a population mean of zero and a standard deviation of one (Z-scores) is called the selection intensity $i$ (Endler 1986; Falconer 1989; Brodie et al. 1995). (Note that there is another measure with that name; Schluter 1988; Brodie et al. 1995.) In the univariate case, when selection on only one standardized trait is considered, the selection differential, the selection intensity, and the selection gradient are all identical (Endler 1986; Brodie et al. 1995). In general, these coefficients can serve to estimate linear (directional), nonlinear (disruptive or stabilizing), and correlational (in the multivariate case) components of selection (Lande and Arnold 1983; Arnold and Wade 1984a,b; Brodie et al. 1995).

### The Z-Score Method

Restricting ourselves to the univariate linear case for the moment, the selection differential $S$ can be defined as the difference in trait means before and after selection, describing the total change in the mean phenotype within a generation (Brodie et al. 1995). This is the definition common in textbooks of quantitative genetics and evolution (Falconer 1989; Maynard Smith 1992). If the trait is standardized, as should be generally the case, the selection differential of a continuous trait $x$ is

$$S = i = \bar{Z}sel - \bar{Z}_{pop} = \bar{Z}_{sel} = \frac{\bar{x}_{sel} - \bar{x}_{pop}}{SD_{pop}} \tag{1}$$

because the standardized population mean $\bar{Z}_{pop}$ by definition is zero. We refer to this method of calculation as the Z-score method, as opposed to the regression method described below. This simple calculation method can only be applied if one indeed can identify a selected (and an unselected) group of individuals, for example, alive or mated individuals as opposed to dead or unmated ones. If we have such a sample of $M = 2$ discrete fitness classes with a fraction $p$ of the individuals belonging to the selected group, the population trait mean can easily be expressed as the weighted or stratified mean of the trait means of the selected and unselected groups as

$$\bar{x}_{pop} = \sum_{i=1}^{M=2} p_i \cdot \bar{x}_i = p \cdot \bar{x}_{sel} + (1 - p) \cdot \bar{x}_{unsel}, \tag{2}$$

(the $p_i$ sum to one) with the population variance being

$$Var_{pop} = \sum_{i=1}^{M=2} p_i \cdot Var_i + \sum_{i=1}^{M=2} p_i \cdot (\bar{x}_i - \bar{x}_{pop})^2$$

$$= p \cdot Var_{sel} + (1 - p) \cdot Var_{unsel} + p \cdot (\bar{x}_{sel} - \bar{x}_{pop})^2$$

$$+ (1 - p) \cdot (\bar{x}_{unsel} - \bar{x}_{pop})^2. \tag{3a}$$

Substituting equation (2) in (3) and rearranging yields

$$Var_{pop} = p \cdot Var_{sel} + (1 - p) \cdot Var_{unsel}$$

$$+ p \cdot (1 - p) \cdot (\bar{x}_{sel} - \bar{x}_{unsel})^2. \tag{3b}$$

Just as in analysis of variance, the population variance can be partitioned as the sum of the weighted variances of the individual groups plus the weighted between-group variance.

Substituting equations (2) and (3b) into (1) and rearranging yields a formula for the selection differential that is a function only of the proportion of selected individuals $p$ and the means and variances of the selected and unselected groups:

$$S = i = \frac{(1 - p) \cdot (\bar{x}_{sel} - \bar{x}_{unsel})}{SD_{pop}}$$

$$= \frac{(1 - p) \cdot (\bar{x}_{sel} - \bar{x}_{unsel})}{\sqrt{p \cdot Var_{sel} + (1 - p) \cdot Var_{unsel} + p \cdot (1 - p) \cdot (\bar{x}_{sel} - \bar{x}_{unsel})^2}}.$$

$$(4)$$

This argument and the correction generalize to cases with $M \geq 2$ fitness classes, but this is the rarer case (cf. Brodie and Janzen 1996).

Now suppose we obtained a biased sample containing a nonrepresentative fraction of individuals belonging to the selected group $p$, but we have a separate, more accurate estimate of that fraction, $p^*$. As we shall show with our simulation below, using $p$ in equation (4) to calculate $S$, which is what statistical packages with automated Z-score calculation would perform, will yield a biased estimate; but using $p^*$ instead would correct for this bias. Zuk (1988) had used a similar correction but this remained largely unnoticed. It is practical to use equation (3b) instead of (3a) for calculation of the corrected $S^*$ because equation (2) shows that the population mean changes when $p^*$ is substituted, whereas the group means remain the same.

### The Regression Method

In the majority of cases, selection coefficients are calculated using regression. This method is more general and equally easy to implement with standard statistical packages. Fitness may be discrete or continuously distributed, and both linear and nonlinear selection coefficients can be estimated for univariate and multivariate datasets (Lande and Arnold 1983; Arnold and Wade 1984a,b). Our correction should therefore also work for selection coefficients calculated using regression. Again, we treat the univariate linear case without loss of generality.

A second definition of the linear selection differential $S$ (and $i$) is the covariance between relative fitness and the standardized trait,

$$S = i = COV(w, z) = COV\left(\frac{W}{\bar{W}}, \frac{x - \bar{x}}{SD(x)}\right)$$

$$= \frac{COV(W, x)}{\bar{W} \cdot SD(x)},$$

$$(5)$$

where relative fitness $w$ is absolute fitness $W$ divided by mean fitness $\bar{W}$, and $z$ is the standardized trait $x$ (Lande and Arnold 1983; Arnold and Wade 1984a,b; Brodie et al. 1995). This directly relates to the regression coefficient $b$, the general formula for which is

$$b(y, x) = \frac{COV(y, x)}{SD^2(x)} = \frac{COV(y, x)}{Var(x)}.$$

$$(6)$$

For standardized trait values $z$, the regression coefficient equals the covariance because the variance $Var(z) = SD^2(z)$ is one.

Returning to our case of two groups of individuals of which $p$ are in the selected group, we in the simplest case (and typically) assign the absolute fitnesses $W_{sel} = 1$ and $W_{unsel} = 0$; mean fitness $\bar{W}$ then equals $p$ and the corresponding relative fitnesses are $1/p$ and zero (Brodie and Janzen 1996). Using equation (5) and the familiar theorem for the covariance (e.g., DeGroot 1986, p. 215), the covariance definition of $S$ indeed turns out to be identical to that given by equation (1):

$$COV(w, z) = \frac{COV(W, x)}{p \cdot SD(x)} = \frac{E(W \cdot x) - E(W) \cdot E(x)}{p \cdot SD(x)}$$

$$= \frac{p \cdot \bar{x}_{sel} - p \cdot \bar{x}_{pop}}{p \cdot SD(x)} = \frac{\bar{x}_{sel} - \bar{x}_{pop}}{SD(x)},$$

$$(7)$$

where $SD(x) = SD_{pop}$. If, as above, we have a sample containing a nonrepresentative fraction of individuals belonging to the selected group $p$ and a more accurate estimate of that fraction $p^*$, the regression (and selection) coefficient is biased. When regressing relative fitness $w$ on the standardized trait $z$ based on the actual data, we obtain from equation (4)

$$b(w, z) = S = \frac{(1 - p) \cdot (\bar{x}_{sel} - \bar{x}_{unsel})}{SD_{pop}},$$

$$(8a)$$

whereas we are interested in

$$b^*(w, z) = S^* = \frac{(1 - p^*) \cdot (\bar{x}_{sel} - \bar{x}_{unsel})}{SD_{pop}^*},$$

$$(8b)$$

where the asterisks denote the population mean and variance corrected according to equations (2) and (3). Solving equation (8a) for $(\bar{x}_{sel} - \bar{x}_{unsel})$, which is the stable entity, and substituting into equation (8b) and rearranging yields the corrected regression coefficient $b^*$ as a function of the biased regression coefficient $b = S$, the sample and corrected population standard deviations, and the actual and estimated fractions $p$ and $p^*$:

$$b^* = S^* = \frac{(1 - p^*)}{(1 - p)} \cdot \frac{SD_{pop}}{SD_{pop}^*} \cdot b.$$

$$(8c)$$

Consequently, after calculating the regression coefficient with a statistical package using the actual, nonrepresentative dataset, the corrected value $b^* = S^*$ can be easily obtained using equation (8c). In the univariate case $b^*$ is identical to the corrected linear selection gradient $\beta^*$ and the selection intensity $i^*$. Following Lande and Arnold (1983), this correction directly extends to multivariate studies of selection, where the partial regression coefficients from multiple regression are the linear selection gradients $\beta$.

The correction formulas for the nonlinear and the correlational selection differentials can be derived analogously from the formulas given in Brodie et al. (1995). The univariate nonlinear (quadratic) selection coefficient is

$$C = Var(z_{sel}) - Var(z_{pop}) + S^2 = \frac{Var(x_{sel})}{Var_{pop}} - 1 + S^2,$$

$$(9a)$$

and, because $Var(z_{pop}) = 1$, the corrected coefficient becomes

$$C^* = \frac{Var_{pop}}{Var_{pop}^*}(C + 1 - S^2) - 1 + S^{*2}.$$

$$(9b)$$

The bivariate nonlinear (correlational) selection coefficient for two variables $x$ and $y$ is

$$C_{x,y} = COV[z(x)_{sel}, z(y)_{sel}] - COV[z(x)_{pop}, z(y)_{pop}] + S_x \cdot S_y$$

$$= \frac{COV(x_{sel}, y_{sel})}{SD_x \cdot SD_y} - r_{x,y} + S_x \cdot S_y, \quad (10a)$$

and, because $COV[z(x)_{pop}, z(y)_{pop}] = r_{x,y}$ (the Pearson product-moment correlation coefficient), the corrected coefficient becomes

$$C_{x,y}^* = \frac{SD_x \cdot SD_y}{SD_x^* \cdot SD_y^*}(C_{x,y} + r_{x,y} - S_x \cdot S_y) - r_{x,y}$$

$$+ S_x^* \cdot S_y^*. \quad (10b)$$

The asterisks again denote the corrected parameters. This correction also extends to the nonlinear selection gradients $\beta$ derived using multiple regression (Lande and Arnold 1983; Brodie et al. 1995).

Having established an a posteriori correction for quantitative measures of selection in case of biased sampling, we now use a simulation to show that this correction works.

## SIMULATION METHODS

The simulation mimicked a population of individuals under directional selection for one continuous trait that influences an individual's chance of being mated. The program sampled from a total population of 400 individuals whose trait values were normally distributed with a mean of zero and a standard deviation of one. Of those, either 200 (50%), 100 (25%), or 40 (10%) individuals were randomly sampled to study the effect of sample size $N$ on the selection coefficients obtained. Each replicate run generated an independent population of $N$ individuals consisting of two subpopulations of selected and unselected individuals. For this purpose, all individuals were ranked in descending order of their trait values and then assigned to the selected (as opposed to the unselected) group with a probability decreasing linearly with rank. The slope of this probability function could be changed to simulate different expected selection intensities at any given number of individuals belonging to the selected and unselected groups. We simulated a range of selection differentials $S$ (= intensities $i$) commonly measured in nature (Endler 1986): 0.05, 0.2, and 0.5.

The parameter crucial to evaluating the correction method is the fraction $p$ of individuals that belong to the selected group (see previous section). We simulated three such fractions spanning a range that should be common in nature: $\frac{1}{2}$, $\frac{1}{4}$, and $\frac{1}{6}$. For each replicate run within a treatment combination, we took two independent samples of the same total sample size $N$ to compare the representative (unbiased according to Arnold and Wade 1984b) and the nonrepresentative (biased) sampling methods. For the representative method, the numbers of individuals belonging to the selected and unselected groups were sampled in proportion to the fraction $p$ simulated. For the nonrepresentative method, we took samples of about equal numbers of individuals belonging to the selected and unselected groups. So as to not always sample the exact same number of individuals in the selected

group for a given total sample size $N$, the fraction of $N$ assigned to the selected group was actually a random variable with a mean of $0.5 \pm 0.125$ (range).

Our design was thus a three $S$ by three $N$ by three $p$ factorial design with two sets of 100 (independent) replicate simulation runs for each treatment combination that served to compare the two different sampling methods. This sample size per treatment combination was chosen to simulate a typical sample size obtainable in nature. For each simulation run $S$ was calculated using the Z-score method, that is, equation (4) above. In the case of representative sampling, $p$ was the "true" fraction sampled, that is, $\frac{1}{2}$, $\frac{1}{4}$, or $\frac{1}{6}$. In the case of nonrepresentative sampling, we used the nonrepresentative fraction $p$ actually sampled to calculate (a biased) $S$, but we also calculated a corrected $S^*$ by substituting the "true" fraction $p^*$ of $\frac{1}{2}$, $\frac{1}{4}$, or $\frac{1}{6}$. The program was written in $C$ and performed on a UNIX system. The code is available upon request from MR.

## SIMULATION RESULTS

The $2 \times 2700$ values of $S$ and $S^*$ obtained by simulation ranged from $-1.18$ to $+1.38$. This spans the bulk of the values typically measured in nature (Endler 1986). When using representative sampling, the selection differential estimates obtained did not differ from the "true" $S$ we attempted to simulate (Fig. 1A; Table 1A). This shows that our simulation worked.

When sampling was nonrepresentative, that is, when the fraction of the rarer, in this case the selected group, was overrepresented, the "true" $S$ was underestimated in a way that seems intuitive: The bias was stronger the higher the "true" selection differential, the smaller the fraction of selected individuals, and the lower the sample size (Fig. 1B), which resulted in significant effects of nearly all factor combinations tested in an analysis of variance (Table 1B). This proves Arnold and Wade's (1984b) abovecited assertion correct. However, when the values obtained in Figure 1B were corrected as described above using our accurate estimate of $p^*$, this bias disappeared: The corrected values of $S$ were not different from those simulated and therefore statistically indistinguishable from those obtained when sampling was representative (Fig. 1C; Table 1C). The correction appears to work equally well for all three sample sizes $N$ tested, so there is little reason to suppose it would not also work for sample sizes less than 40, though we did not specifically test this.

## DISCUSSION

As pointed out by Arnold and Wade (1984b) and demonstrated by simulation here, sampling bias leads to biased estimates of selection coefficients. This is particularly true when fitness classes are binary or, more generally, few and discrete, as is often the case when studying sexual or viability selection. This bias is predictable (i.e., directional). When the individuals favored by selection are the rare class, as is more typically the case in sexual selection, and greater sampling effort is invested in obtaining them so that this class is overrepresented in the sample, selection coefficients will generally be underestimated. The underestimate is greater the stronger the selection intensity, the smaller the true fraction
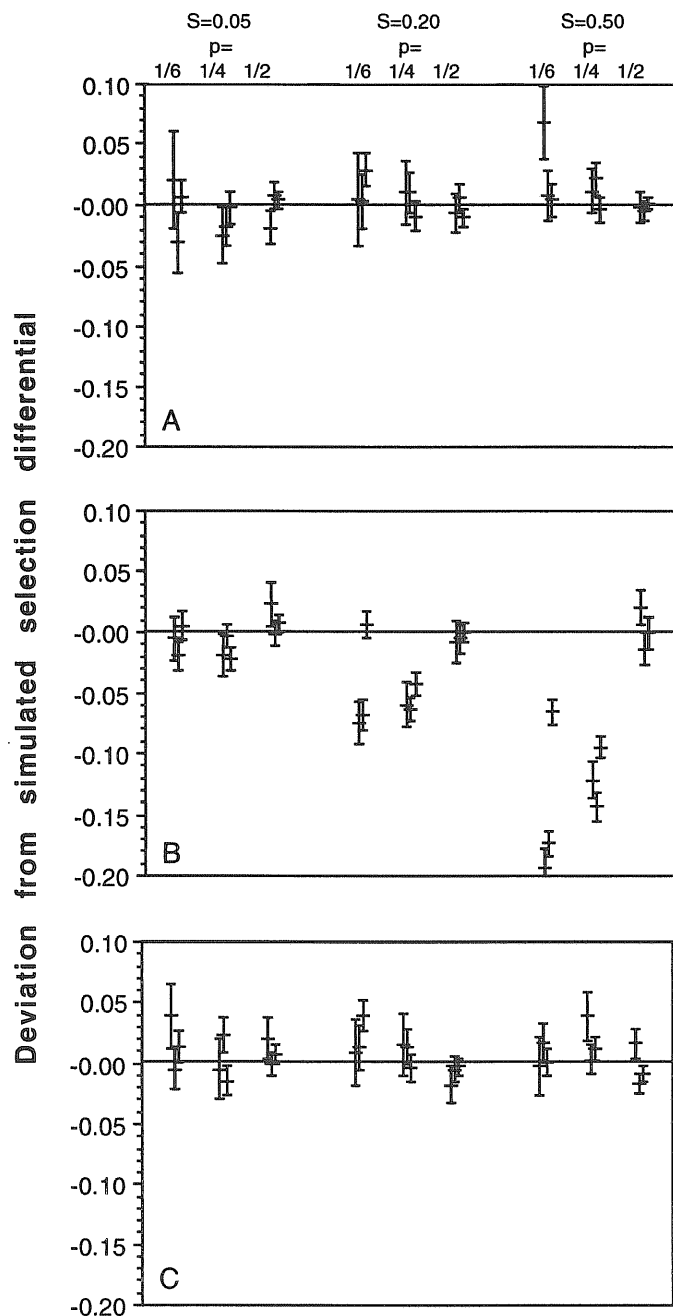
FIG. 1. Mean (± SE) deviation of the selection differential $S$ obtained from that simulated for three different simulated $S$, three fractions $p$ of individuals in the selected group, and three sample sizes $N$ of 40, 100, and 200 (from left to right in every group of three): (A) representative sample; (B) nonrepresentative sample, not corrected; and (C) nonrepresentative sample, corrected.

TABLE 1. Analysis of variance results for the deviation of the obtained selection differential from that simulated, as depicted in Figure 1.

| Variable | df | MS | $F$ | $P$ |
|---|---|---|---|---|
| A. Representative sample | | | | |
| $S$ | 2 | 0.070 | 1.955 | 0.142 |
| $p$ | 2 | 0.060 | 1.659 | 0.191 |
| $N$ | 2 | 0.010 | 0.290 | 0.749 |
| $S$ by $p$ | 4 | 0.019 | 0.523 | 0.719 |
| $S$ by $N$ | 4 | 0.031 | 0.860 | 0.487 |
| $p$ by $N$ | 4 | 0.056 | 1.559 | 0.183 |
| $S$ by $p$ by $N$ | 8 | 0.026 | 0.719 | 0.675 |
| Residual | 2673 | 0.036 | | |
| B. Nonrepresentative sample: not corrected | | | | |
| $S$ | 2 | 1.605 | 92.667 | <0.001 |
| $p$ | 2 | 1.347 | 77.811 | <0.001 |
| $N$ | 2 | 0.254 | 14.684 | <0.001 |
| $S$ by $p$ | 4 | 0.383 | 22.143 | <0.001 |
| $S$ by $N$ | 4 | 0.068 | 3.919 | 0.004 |
| $p$ by $N$ | 4 | 0.151 | 8.735 | <0.001 |
| $S$ by $p$ by $N$ | 8 | 0.032 | 1.835 | 0.066 |
| Residual | 2673 | 0.017 | | |
| C. Nonrepresentative sample: corrected | | | | |
| $S$ | 2 | 0.001 | 0.049 | 0.952 |
| $p$ | 2 | 0.056 | 2.124 | 0.120 |
| $N$ | 2 | 0.030 | 1.146 | 0.318 |
| $S$ by $p$ | 4 | 0.042 | 1.607 | 0.170 |
| $S$ by $N$ | 4 | 0.025 | 0.952 | 0.433 |
| $p$ by $N$ | 4 | 0.019 | 0.737 | 0.567 |
| $S$ by $p$ by $N$ | 8 | 0.038 | 1.460 | 0.167 |
| Residual | 2673 | 0.026 | | |

of individuals favored by selection, and the lower the sample size. By extension, the converse case, where the individuals favored by selection are the common class and may be underrepresented in a sample relative to those not favored, leads to overestimates of selection coefficients. This case may be rarer in nature but may occur in studies of viability selection when most or all individuals that died but only a subsample of those alive are sampled (cf. examples in Lande and Arnold 1983).

In any case, we have offered a method that permits a posteriori statistical correction of such sampling bias. All that it requires is an additional good estimate of the true representation of the different fitness classes in nature. Such an estimate may be obtained from separate studies or from the same study using different methods (Zuk 1988; Carroll and Salamon 1995; Blanckenhorn et al. 1999); sampling may thereafter proceed freely and unconstrained. As pointed out in the introduction, the major advantages of our correction are that sampling without restriction is easier to carry out in the field, the problem of accessibility of different classes is circumvented, and it maximizes sample size in situations of low population densities. The correction is easy to implement after having calculated the selection coefficients with either method. To do this, the somewhat cumbersome equations are best entered in a spreadsheet.

Once again, we stress that our correction does not affect the significance level of the selection coefficient (Arnold and Wade 1984b), whether it be determined with one-sample $t$-tests (in case of the Z-score method) or regression. Our correction also does not bear upon the issue of whether least-squares regression is the appropriate statistical technique to generate selection coefficients and to test for their significance (Lande and Arnold 1983; Manly 1985; Endler 1986; Mitchell-Olds and Shaw 1987; Fairbairn and Preziosi 1994; Brodie et al. 1995; Kingsolver and Smith 1995; Brodie and Janzen 1996) or on the issue of which of several possible

numerical fitness values are best assigned if fitness classes are discrete (Brodie and Janzen 1996). However, our correction does accommodate cases with more than two discrete fitness classes (Brodie and Janzen 1996) and can therefore be useful for a wide range of conditions likely to be encountered in the field.

## LITERATURE CITED

ANDERSSON, M. 1994. Sexual selection. Princeton Univ. Press, Princeton, NJ.

ARNOLD, S. J., AND M. J. WADE. 1984a. On the measurement of natural and sexual selection: theory. Evolution 38:709–719.

———. 1984b. On the measurement of natural and sexual selection: applications. Evolution 38:720–734.

ARNQVIST, G. 1992. Spatial variation in selective regimes: sexual selection in the water strider, *Gerris odontogaster.* Evolution 46:914–929.

BLANCKENHORN, W. U., C. MORF, C. MÜHLHÄUSER, AND T. REUSCH. 1999. Spatio-temporal variation in selection on body size in the dung fly *Sepsis cynipsea.* J. Evol. Biol. 12:*In press.*

BORGIA, G. 1982. Experimental changes in resource structure and male density: size-related differences in mating success among male *Scathophaga stercoraria.* Evolution 36:307–315.

BRODIE, E. D., AND F. J. JANZEN. 1996. On the assignment of fitness values in statistical analyses of selection. Evolution 50:437–442.

BRODIE, E. D., A. J. MOORE, AND F. J. JANZEN. 1995. Visualizing and quantifying natural selection. Trends Ecol. Evol. 10:313–318.

CARROLL, S. P., AND M. H. SALAMON. 1995. Variation in sexual selection on male body size within and between populations of the soapberry bug. Anim. Behav. 50:1463–1474.

DEGROOT, M. H. 1986. Probability and statistics. Addison Wesley, Reading, MA.

ENDLER, J. A. 1986. Natural selection in the wild. Princeton Univ. Press, Princeton, NJ.

FAIRBAIRN, D. J., AND R. F. PREZIOSI. 1994. Sexual selection and the evolution of allometry for sexual size dimorphism. Am. Nat. 144:101–118.

FALCONER, D. S. 1989. Introduction to quantitative genetics. Longman Scientific and Technical, Harlow, U.K.

HOWARD, R. D. 1979. Estimating reproductive success in natural populations. Am. Nat. 114:221–231.

KINGSOLVER, J. G., AND S. G. SMITH. 1995. Estimating selection on quantitative traits using capture-recapture data. Evolution 49:384–388.

LANDE, R., AND S. J. ARNOLD. 1983. The measurement of selection on correlated characters. Evolution 37:1210–1226.

MANLY, B. F. J. 1985. The statistics of natural selection. Chapman and Hall, London.

MAYNARD SMITH, J. 1992. Evolutionary genetics. Oxford Univ. Press, Oxford.

MCLAIN, D. K. 1992. Population density and the intensity of sexual selection on body length in spatially and temporally restricted natural populations of a seed bug. Behav. Ecol. Sociobiol. 30:347–356.

MCLAIN, D. K., L. B. BURNETTE, AND D. A. DEEDS. 1993. Within season variation in the intensity of sexual selection on body size in the bug *Margus obscurator.* Ethol. Ecol. Evol. 5:75–86.

MITCHELL-OLDS, T., AND R. G. SHAW. 1987. Regression analysis of natural selection: statistical inference and biological interpretation. Evolution 41:1149–1161.

SCHLUTER, D. 1988. Estimating the form of natural selection on quantitative traits. Evolution 42:849–861.

WADE, M. J., AND S. J. ARNOLD. 1980. The intensity of sexual selection in relation to male sexual behaviour, female choice, and sperm precedence. Anim. Behav. 28:446–461.

ZUK, M. 1988. Parasite load, body size, and age of wild-caught male field crickets (Orthoptera: Grillidae): effects on sexual selection. Evolution 42:969–976.

Corresponding Editor: E. Brodie III