POWER AND POTENTIAL BIAS IN FIELD STUDIES OF NATURAL SELECTION

Erika I. Hersch[1] and Patrick C. Phillips[2,3]
Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, Oregon 97405-5289
[1]*E-mail: ehersch@uoregon.edu*
[2]*E-mail: pphil@uoregon.edu*

Abstract.—The advent of multiple regression analyses of natural selection has facilitated estimates of both the direct and indirect effects of selection on many traits in numerous organisms. However, low power in selection studies has possibly led to a bias in our assessment of the levels of selection shaping natural populations. Using calculations and simulations based on the statistical properties of selection coefficients, we find that power to detect total selection (the selection differential) depends on sample size and the strength of selection relative to the opportunity of selection. The power of detecting direct selection (selection gradients) is more complicated and depends on the relationship between the correlation of each trait and fitness and the pattern of correlation among traits. In a review of 298 previously published selection differentials, we find that most studies have had insufficient power to detect reported levels of selection acting on traits and that, in general, the power of detecting weak levels of selection is low given current study designs. We also find that potential publication bias could explain the trend that reported levels of direct selection tend to decrease as study sizes increase, suggesting that current views of the strength of selection may be inaccurate and biased upward. We suggest that studies should be designed so that selection is analyzed on at least several hundred individuals, the total opportunity of selection be considered along with the pattern of selection on individual traits, and nonsignificant results be actively reported combined with an estimate of power.

Key words.—Experimental design, power, publication bias, regression, selection differentials, selection gradients.

Received June 27, 2003.   Accepted October 30, 2003.

A major goal in evolutionary biology is to understand how phenotypes arise and are shaped by natural selection. However, the attainment of this goal is in part impeded by the fact that selection does not operate on single traits but rather on whole organisms that are composed of many correlated traits. Since Lande and Arnold (1983) published their seminal article advocating the use of multiple regression techniques to separate the effects of direct and indirect selection on correlated traits, there has been an explosion of studies using these methods to describe selection in natural populations. Collectively these studies hold great potential for answering fundamental questions in evolutionary biology. In principle, summaries of these findings could be used to describe how selection generally shapes quantitative traits, and these estimates could then be applied to genetic models to predict the evolution of quantitative traits in nature. (Lande 1979; Lande and Arnold 1983; Hoekstra et al. 2001; Kingsolver et al. 2001). In their review of 63 studies of natural selection published between 1984 and 1997, Kingsolver et al. (2001) found that the absolute value of the median standardized selection differential (describing direct and indirect selection) was 0.13 standard deviations and that the absolute value of the median standardized selection gradient (describing direct selection only) was 0.16 standard deviations. Because these are standardized coefficients, this means that a perfectly heritable, genetically uncorrelated trait would change its mean by 13 phenotypic standard deviations in 100 generations. Although this strikes us as fairly strong selection, whether these values indicate weak or strong selection depends on one's definition of these relative terms (Conner 2001; Hoekstra et al. 2001; Kingsolver et al. 2001).

Despite the clear value of bringing together estimates of selection from diverse studies, whether the summary statistics obtained from currently published selection studies are an accurate reflection of the magnitude and type of selection in natural populations is questionable for several reasons. Because most of the studies reviewed by Kingsolver et al. (2001) had sample sizes of fewer than 135 individuals, they may have had insufficient power to detect weak levels of selection. As suggested by Kingsolver et al. (2001), this low level of statistical power could have led to a "file-drawer effect," such that small studies are less likely to be published unless they report strong and statistically significant levels of selection. The finding that magnitudes of estimated median selection gradients tend to decrease as sample sizes increase supports this possibility (see fig. 2 in Kingsolver et al. 2001). A failure to detect or to report weak levels of selection in small studies would bias reported estimates of selection, thereby skewing our view of how selection acts in natural populations. Thus, to validate current evolutionary thought on how selection shapes phenotypic distributions in natural populations, it is important to assess whether such biases exist in the literature.

Here, we formalize the suggestions of Kingsolver et al. (2001) by calculating the power of detecting selection (both indirect and direct) on correlated traits in nature and apply these calculations to published estimates of selection. It does appear that previous studies have been plagued by low power to detect selection, especially weak selection, such that current perceptions on the strength of natural selection are likely to be inaccurate. Furthermore, a consideration of power should aid in the design of future studies aimed at estimating the pattern and strength of selection.

Theory

Regression Technique and Selection Analyses

Regression models can be used to measure the direction and magnitude of direct and indirect selection acting on quantitative traits in natural populations (Lande and Arnold 1983).

---

[3] Corresponding author.

Using this approach, a fitness variable (e.g., survival, lifetime fecundity, or mating success) is regressed onto the phenotypic traits of interest and the slope of this line describes the strength and direction of selection on these traits. From the regression approach, two groups of metrics, linear selection differentials and linear selection gradients, can be used to describe directional selection acting on correlated traits. Linear selection differentials (**s**) describe the change in the trait mean phenotype due to both direct and indirect selection acting on traits during a single episode of selection, and they equal the covariance between trait (z) and relative fitness (w). The second group of metrics, the linear selection gradients (**β**), measure the change in trait mean phenotype resulting from direct selection acting on traits, with the effects of selection on correlated traits being held constant. Mathematically, linear selection gradients are the partial regression coefficients obtained from the multiple regression of relative fitness onto the phenotypic traits of interest:

$$\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{s}, \tag{1}$$

where **P** is the phenotypic variance-covariance matrix of the measured traits (Lande and Arnold 1983). Both of these metrics are commonly reported on a standardized scale of trait standard deviations ($i = s' = s/\sigma_z = \text{cov}[z, w]/\sigma_z$ and $\beta' = \beta\sigma_z$) to facilitate comparisons of the strength of selection between traits and across different studies and species.

Power of Selection Metrics

Power is the probability of reporting a selection differential or gradient as significant when selection is in fact operating on a trait. In a statistical sense, power is the probability of not making a Type II error, which is not rejecting the null hypothesis when it is in fact false. Power depends on three factors: (1) the Type I error rate fixed by the experimenter (α-level); (2) the sample size; and (3) the size of the effect to be detected (Cohen 1992). Calculating the power of selection estimates therefore depends on the sampling distribution of the estimates themselves. Under the assumption of normality, these sampling distributions are closely related to other well-known statistical distributions.

For example, the sampling distribution of a univariate linear-regression coefficient for a bivariate normal sample is closely related to a Student's t-distribution. In particular, the quantity,

$$\frac{(b - \beta)\sigma_z}{\sigma_w\sqrt{1 - \rho^2}} \times (N - 1)^{1/2} \tag{2}$$

has a t-distribution with $(N - 1)$ degrees of freedom (Stuart and Ord 1987, cf. eqs. 16.86 and 16.87). Here N is the sample size, σ_w is the standard deviation of relative fitness, ρ is the correlation coefficient between the trait and fitness, and β and b are, respectively, the population and sample regression coefficients from the regression of relative fitness onto a single phenotypic trait of interest. If we reparameterize this equation using the fact that $\rho = \beta\sigma_z/\sigma_w = s'/\sigma_w$, the power to detect the total amount of selection on a trait can be shown to be a function of a single variable, the correlation between the trait and fitness. From equation (2), the quantity

$$\frac{r' - \rho}{\sqrt{1 - \rho^2}} \times (N - 1)^{1/2} \tag{3}$$

also has as t-distribution, with $r' = (s_w\sigma_z/s_z\sigma_w)r$, where s_z and s_w are the sample variances of the trait and fitness and r is the sample correlation coefficient. (The extra complication caused by the sample variance can easily be accounted for by using the true sampling variance for the correlation coefficient in the power calculations; Phillips 1998). Power to detect selection is thus a function of the strength of selection on the trait (s') relative to the square root of the total opportunity for selection in the population (σ_w, Crow 1958; Arnold and Wade 1984). Equations (2) and (3) show that the ability to detect selection depends not only on the strength of selection acting on the trait, but also on how much that selection explains the total amount of variation in fitness.

In the multitrait case, power of the partial regression coefficients is complicated by the pattern of correlation among the traits. For the power of total selection on a trait given above, the correlation between the trait and fitness is equivalent to the square root of the coefficient of determination of the univariate regression model ($\rho_{w,\hat{w}}^2$ and its estimator R^2; see the Appendix). To calculate the power of the partial regression coefficients, we partition the coefficient of determination as a function of the strength of selection operating on each trait and the pattern of covariance among traits,

$$\rho_{w,\hat{w}}^2 = \frac{1}{\sigma_w^2}(\mathbf{s}^T\mathbf{P}^{-1}\mathbf{s}) = \frac{1}{\sigma_w^2}(\boldsymbol{\beta}^T\mathbf{s}) \tag{4}$$

(Appendix). Thus, in the multitrait case it more difficult to predict when there will be sufficient power to detect selection for any given trait because power depends on three factors: the strength of selection operating on each trait, the overall opportunity for selection, and the pattern of covariance among traits. For example, in the two-trait case, power will be a function of

$$\rho_{w,\hat{w}}^2 = \frac{\rho_{z_1w}^2 + \rho_{z_2w}^2 - 2\rho_{z_1z_2}\rho_{z_1w}\rho_{z_2w}}{(1 - \rho_{z_1z_2})^2}. \tag{5}$$

Because of limitations in the correlation structure among the traits and fitness, the R^2-value necessarily always increases as additional traits are added to the model (Draper and Smith 1981). The precise fit of the model, however, depends on the balance between the direction of selection on each individual trait and the pattern of correlation between the traits. In general, power is increased when any two of the correlation coefficients are in the opposite direction as the third (see below for an analysis of the individual selection gradients).

METHODS

To assess whether summary statistics obtained from published phenotypic selection analyses, such as those reported by Kingsolver et al. (2001), reflect the actual magnitude of selection in natural populations, we first examine how sample size (N) and the correlation coefficient (ρ) jointly influence the power of detecting selection, and we then use these models to examine previous studies.

Models of power.—The power of detecting selection on single traits by regression for different combinations of sam-

ple size and correlation coefficients at a fixed α-level of 0.05 was calculated via equation (3) using the approach outlined in Phillips (1998). A simulation approach was used to estimate the power of detecting selection acting directly on a given trait in the multivariate case (the partial regression coefficients). Here, correlation matrices representing high, medium, and low correlations among traits were used to randomly generate sample multivariate datasets of specified size. Partial regression analysis was used to estimate the selection gradients for each dataset, and the power to detect selection acting upon each trait was determined as the proportion of partial regression coefficients among the 10,000 simulation runs that were significant.

Equations (2) and (3) used for calculating power are based on the assumption of multivariate normality among the traits and fitness. Many studies of selection use viability as their measure of fitness, so it is possible that any conclusions regarding power derived from the methods presented above might not be applicable to binomially distributed fitnesses (Janzen and Stern 1998). To test for this possibility, for a given theoretical correlation coefficient and sample size, sample datasets were generated under the assumption that the traits were normally distributed ($\mu = 0$, $\sigma^2 = 1$) but that fitness was binomially distributed (the probability of being alive, $p = 0.1$, 0.5, and 0.9; BIVAR, ver. 1.1; J. Miller, University of Otago, New Zealand). These sample correlation coefficients were then compared to the critical value of the correlation coefficient for a given sample size and an α-level of 0.05. Critical values were calculated from a simulated null distribution in which the correlation between trait and fitness was zero. For each sample size and fitness distribution, power was calculated as the proportion of sample correlation coefficients of 10,000 that were found to be equal to or greater than the critical value. Because the opportunity for selection is fixed by the probability of survival in the binomial case, it is impossible to separate the selection differential (s') from the correlation between trait and fitness. We therefore only present results for fixed values of ρ (although in practice changing the opportunity for selection has very little effect on the results in this case; see below).

Power of selection differentials reported in previous studies.—To resolve whether previous studies have had sufficient power to detect selection on quantitative traits in natural populations, we analyzed a subset of the articles reviewed by Kingsolver et al. (2001) that met the following criteria: (1) they had used multiple regression techniques advocated by Lande and Arnold (1983); (2) they had calculated s'; and (3) they reported the variance in relative fitness (σ_w^2) or presented data that allowed easy calculation of this value. Of the original 63 studies, we found 17 studies, comprising 298 linear selection differentials, that met the above criteria. The major limitation here was a lack of reporting of total variance for fitness. Even so, 40% of the linear selection differentials reviewed by Kingsolver et al. (2001) are represented here. Power for each of these reported selection differential was determined by using the sample size (N) and the calculated correlation coefficient ($r = s'/\sigma_w$) as outlined above. Although these estimates are frequently not independent of one another, we use the individual values here as a way of representing our current best estimate of the distribution of selection coefficients in nature (Kingsolver et al. 2001).
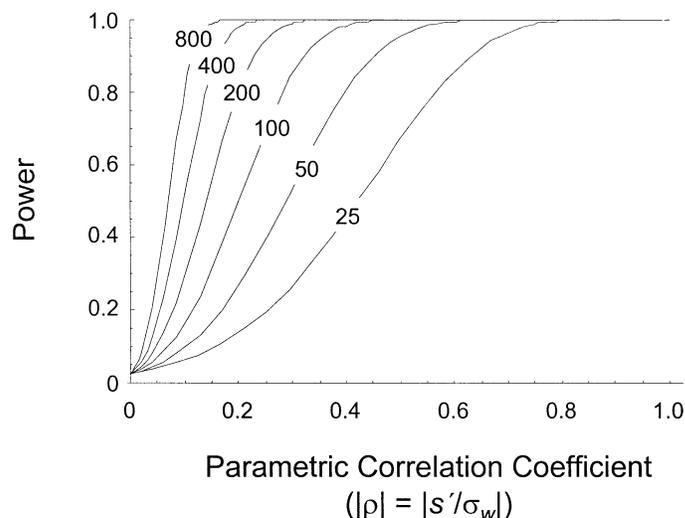

FIG. 1. Effect of the magnitude of the correlation between the trait and fitness and sample size on power to detect overall selection acting on a trait. Each line gives the power of detecting a given correlation coefficient (the standardized selection differential divided by the square root of the opportunity for selection) for an experiment with sample size *N*. Lines represent successive doubling of sample size from $N = 25$ to $N = 800$. The *x*-axis is the true correlation in the population and the *y*-axis is the probability of detecting this correlation for a given sample size. Curves for positive and negative correlations of the same magnitude are identical.

Publication bias.—We define publication bias as the tendency to only report the results of a study if at least one of the selection gradients is found to be significant. To assess the potential for publication bias to explain the finding that the magnitude of selection tends to decrease as study sample sizes increase (Kingsolver et al. 2001), we simulated selection studies conducted at various sample sizes. The studies reviewed by Kingsolver et al. (2001) examined on average 4.6 correlated traits per study and the values of the corresponding standardized selection gradients were approximately normally distributed ($\mu = 0.058$, $\sigma = 0.194$). To simulate these experiments, five parametric selection gradients were drawn at random from this distribution and data were simulated under three different coefficients of determination ($R^2 = 0.1$, 0.4, and 0.7) following equations (1) and (4). The absolute values of the estimated selection gradients were tested for significance ($\alpha = 0.05$), and if any were found to be significant, then all were retained; otherwise they were all discarded. Estimates for each combination of parameters were simulated 1000 times for each sample size. Because publication bias can be caused both by variance in the actual strength of selection across studies and by variance in the estimated strength of selection (sampling error), we tested the latter effect by choosing a fixed combination of selection gradients for input into the simulation procedure outlined above.

RESULTS AND DISCUSSION
Models of Power

Power to detect selection is increased when either the sample size or the correlation coefficient is large (Fig. 1; Phillips
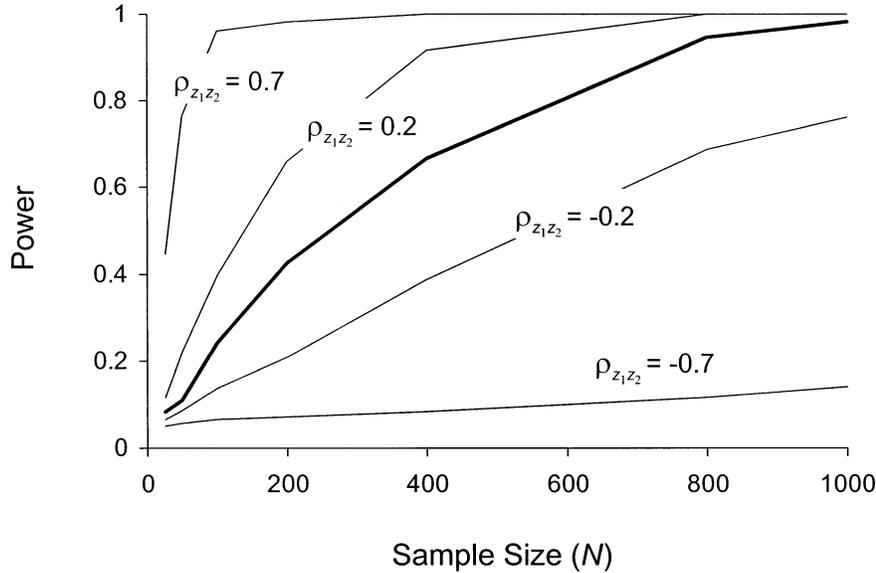
FIG. 2.   Effect of the inclusion of single versus multiple traits and sample size on power to detect direct selection. The bold line gives the power of detecting direct selection when only the first trait is measured in an experiment (correlation between the trait and fitness, $\rho_{z_1w} = -0.12$). Thin lines give the power of detecting direct selection (the partial regression coefficient) on the first trait when selection is also estimated on a second trait ($\rho_{z_2w} = 0.20$). Each line represents a different value of the correlation between two traits.

1998). Interestingly, power does not depend on the strength of selection per se (the size of the effect to be detected, s'), but rather on the strength of selection relative to how much fitness variation is present in the population (the correlation coefficient; eqs. 2, 3). This suggests that it is easier to detect selection for traits that contribute the most to the variance in fitness within a population. Future studies should be de-
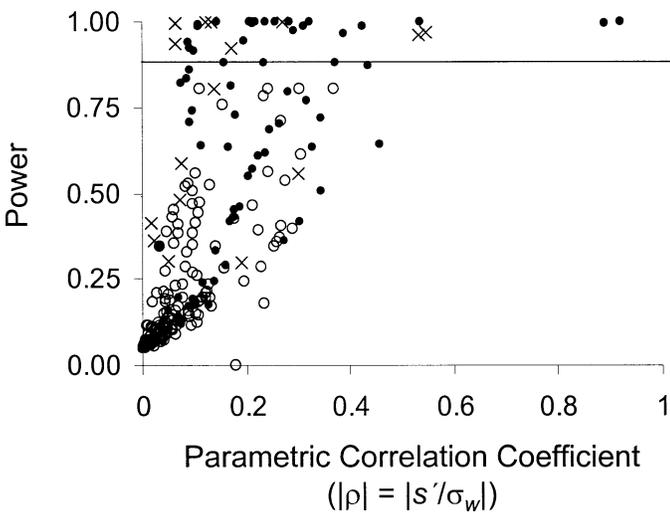

FIG. 3.   Power as a function of the correlation coefficient between a trait and fitness from reported studies (298 estimates from 17 studies). Correlation coefficients (r_{zw}) are equal to the standardized selection differential divided by the standard deviation in relative fitness. The statistical significance (at the α-level of 0.05) of each estimate is given: filled circles indicate significant estimates; open circles indicate not significant; crosses indicate the significance level of the estimate was not reported. The solid line indicates a 90% probability of detecting selection when it is operating on a trait (power = 0.9).

signed with the consideration that different sample sizes are needed to achieve equivalent levels of power, even for the same strength of selection, depending on the particulars of the population under study.

Although the analysis of multiple traits increases the overall power of the model to detect selection (eq. 4), what is more interesting is whether the addition of multiple traits alters the power of detecting selection acting directly on a given trait (the partial regression coefficient). As shown in Figure 2, the power to detect direct selection on an individual trait can be either increased or decreased relative to the case when only a single trait is included in the model (see also eq. 5). As above for the overall power for the two-trait case (eq. 5), power of an individual selection gradient is increased when any two correlation coefficients ($\rho_{z_1z_2}$, ρ_{z_1w}, ρ_{z_2w}) are in the opposite direction as the third. These results are robust to variations in the magnitude and direction of the correlation coefficients between each trait and fitness and between the traits.

Power of Selection Differentials Reported in Previous Studies

Consistent with the findings of Kingsolver et al. (2001), most of the studies reviewed here have insufficient power (power below 0.8–0.9) to detect the reported levels of selection (Fig. 3). As would be expected, selection differentials that were reported as significant tended to have higher power than nonsignificant selection differentials. Note that a finding of low power does not indicate an error in reported levels of selection, because significant selection differentials are significant regardless of power. However, a finding of low power may indicate that our perception regarding the magnitude of directional selection shaping quantitative traits in natural populations may be inflated because the design of previous
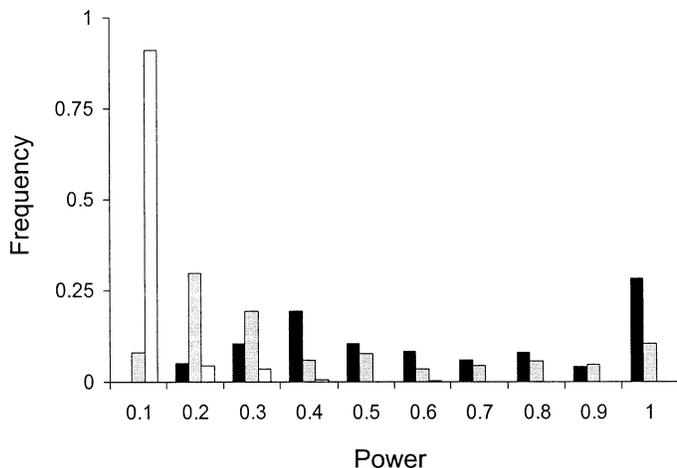
FIG. 4. Frequency histograms displaying the power that studies of current sample sizes would have for a given specified level of relative selection ($|\rho| = |s'/\sigma_w|$). Bars represent the frequency of studies that would have a given power to detect a specified level of selection: black bars, relatively strong selection $|\rho| = 0.20$; gray bars, average selection $|\rho| = 0.12$; white bars, relatively weak selection $|\rho| = 0.02$.

FIG. 5. Comparison of the power to detect selection ($\rho_{z,w} = 0.12$) for normally distributed versus binomially distributed fitnesses. The dark bar shows power for the normal distribution case, whereas the dashed lines show power for three different binomial distributions (probability of living, $p = 0.1$, 0.5, and 0.9). The latter curves are nearly indistinguishable on the scale shown.

studies precluded the detection of weaker levels of selection. To examine whether the low power of reported selection differentials, in general, could be attributed to small sample sizes or to the detection of weak levels of selection relative to the variance in fitness acting on traits (small correlations between the trait and fitness), we explored the power that studies with the reported sample sizes would have had to detect strong ($r = 0.20$), average ($r = 0.12$), and weak ($r = 0.02$) levels of relative selection, where the strength of selection used here is relative to that actually found in these studies. We found that current sample sizes could frequently detect strong levels of selection acting on traits, but that as relative selection intensity decreased, most studies would not have adequate samples sizes to detect selection acting on correlated traits in nature (Fig. 4).

The above findings indicate that summary statistics, such as those reported by Kingsolver et al. (2001), may not accurately reflect the intensity of selection acting in natural populations. There is some risk of overstatement here because the model used to assess power assumed that fitness measures were normally distributed, whereas most empirical studies employ fitness measures that are binomially distributed (i.e., individuals are dead or alive). However, when fitness measures have a binomial distribution, power to detect selection is actually decreased relative to the case when fitness measures have a normal distribution (Fig. 5). This suggests that our finding that most studies have had insufficient power to detect selection on correlated traits is most likely conservative.

Publication Bias

Our simulations of publication bias match the general trends reported in previous studies (see fig. 2 in Kingsolver et al. 2001); the median value of reported standardized selection gradients tend to decrease as sample size increases (Fig. 6). These results are robust to variation in both the
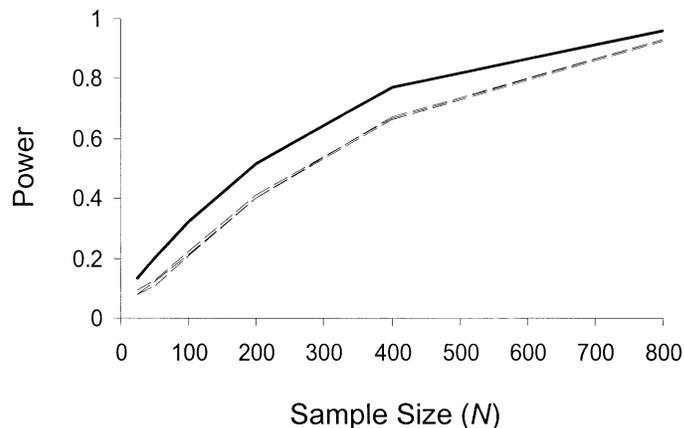
coefficient of determination and in the structure of the specified correlation matrix (**P**). Two factors could generate this pattern. First, the strength of selection in nature is variable. Studies of populations in which selection is strong are more likely to be published (and perhaps investigated in the first place) than studies of populations in which selection is weak (Kingsolver et al. 2001). Failure to publish results of weak selection will tend to generate an upward bias in the average reported strength of selection. Second, any estimate of selection is subject to sampling error. If the power of a study is low, then studies are more likely to be published when the apparent strength of selection is strong because this leads to a significant result, even if the actual strength of selection is likely to be smaller than this particular estimate (Beavis 1994). The influence of sampling effect was separated from the overall pattern of bias by repeatedly estimating the pattern of selection for a single parametric set of selection gradients. The sampling effect can explain a surprisingly large fraction of the overall pattern of bias (Fig. 6B). In general, publication bias in the reporting of selection gradients could distort our view of how selection shapes the distribution of correlated traits in nature; in general, conclusions on the magnitude of selection will depend on the size of studies used to estimate that selection.

Conclusion

Multiple regression models used to tease apart direct from indirect selection acting on correlated traits are a useful tool for evolutionary biologists interested in natural selection. However, we suggest that caution is warranted both in using these methods and in synthesizing their results. Caution is urged because we found that most studies have had insufficient power to detect moderate to weak levels of selection (Fig. 4). Although the power required to justify conducting any particular study is an individual decision, the field as a whole should encourage studies with more power to detect varying levels of selection. If there has been a historical tendency to only report selection gradients when at least one
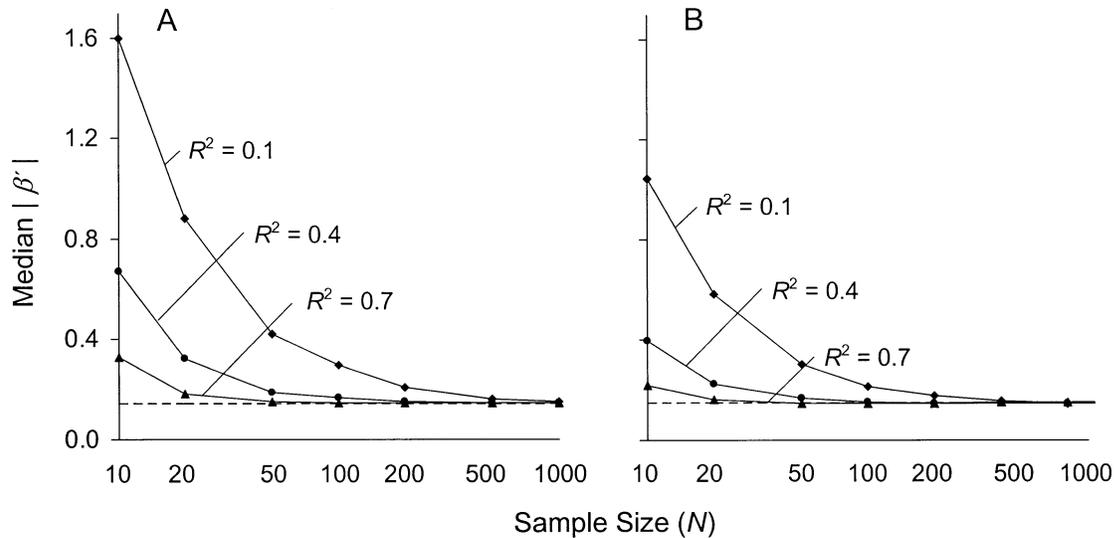
FIG. 6. Publication bias can lead to elevated selection estimates. (A) Median values of simulated standardized linear selection gradient estimates ($|\beta'|$) when five true parametric selection gradients are drawn from the empirical distribution of selection estimates, but only samples with at least one significant estimate are retained. (B) Estimates based on a single set of five selection gradients drawn from (A). Panel (A) reveals the bias generated from both real differences among selection experiments and errors in statistical estimation caused by sampling, whereas panel (B) displays the fraction of (A) that is generated by statistical sampling alone. Lines represent different input values of the model coefficient of determination R^2) for the total selection model. The dashed line gives the unbiased median estimate. The potential for bias increases as fraction of the total opportunity for selection explained by the traits under study decreases. Small samples with low predictive power can lead to upward of a 10-fold overestimate of the median strength of selection.

gradient is found to be significant (compare Fig. 6 with fig. 2 in Kingsolver et al. 2001), then statistical summaries may not accurately reflect the average strength of selection acting on quantitative traits in natural populations. In fact, these potential biases suggest that selection may be weaker than is currently assumed, although this cannot be confirmed with current data.

We have only considered power and bias in estimates of linear (directional) selection. These problems are likely to be even worse with estimates of quadratic selection (Kingsolver et al. 2001). Therefore, discussions of whether selection tends to be disruptive or stabilizing will need to be prefaced by the likely low power of such studies (Conner 2001), as well as a consideration of the full multivariate pattern of selection (Blows and Brooks 2003).

We hope that our findings will be an incentive for researchers to design more powerful studies when using multiple regression techniques to examine directional selection on correlated traits in nature. Power curves, which relate power to both sample size and the correlation between trait and relative fitness, should be useful in designing experiments (Fig. 1). However, even studies with large sample sizes will have insufficient power to detect weak levels of selection. Indeed, it may be impossible to design a study of sufficient size to detect small, but biologically important, strengths of selection (Fig. 4). In addition, too few studies have considered the strength of selection they have estimated relative to the total opportunity for selection in the population. Therefore, if understanding how selection acts on traits in natural populations remains a central a goal in evolutionary biology, we suggest that future studies should report nonsignificant selection metrics in addition to a measure of the power of detecting observed magnitudes of selection.

LITERATURE CITED

Arnold, S. J., and M. J. Wade. 1984. On the measurement of natural and sexual selection: theory. Evolution 38:709–719.
Beavis, W. 1994. The power and deceit of QTL experiments: lessons from comparative QTL studies. Pp. 252–268 in Proceedings of the 49th Annual Corn and Sorghum Research Conference. American Seed Trade Association, Washington, D.C.
Blows, M. W., and R. Brooks. 2003. Measuring non-linear selection. Am. Nat. 162:815–820.
Cohen, J. 1992. A power primer. Psychol. Bull. 112:155–159.
Conner, J. K. 2001. How strong is natural selection? Trends Ecol. Evol. 16:215–217.
Crow, J. F. 1958. Some possibilities for measuring selection intensities in man. Hum. Biol. 30:1–13.
Draper, N. R., and H. Smith. 1981. Applied regression analysis. Wiley, New York.
Hoekstra, H. E., J. M. Hoekstra, D. Berrigan, S. N. Vignieri, A. Hoang, C. E. Hill, P. Beerli, and J. G. Kingsolver. 2001. Strength and tempo of directional selection in the wild. Proc. Natl. Acad. Sci. USA 98:9157–9160.
Janzen, F. J., and H. S. Stern. 1998. Logistic regression for empirical studies of multivariate selection. Evolution 52:1564–1571.
Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gilbert, and P. Beerli. 2001.

The strength of phenotypic selection in natural populations. Am. Nat. 157:245–261.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body allometry. Evolution 33:402–416.

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution 37:1210–1226.

Phillips, P. C. 1998. Designing experiments to maximize the power of detecting correlations. Evolution 52:251–255.

Stuart, A., and J. K. Ord. 1987. Kendall's advanced theory of statistics. Vol. 1, Distribution theory. Oxford Univ. Press, New York.

Wolfram, S. 1996. The Mathematica book. Wolfram Media, Champaign, IL.

Corresponding Editor: J. Fry

Identity of the squared correlation between a trait and relative fitness ($\rho^2_{z,w}$) and the coefficient of determination of the linear univariate regression model ($\rho^2_{w,\hat{w}}$) can be shown by noting that

$$\rho^2_{w,\hat{w}} = \text{cov}(w, \hat{w})^2/\sigma^2_w \, \sigma^2_{\hat{w}}, \quad (A1)$$

where $\hat{w} = \alpha + \beta z$. Since $\text{cov}(w, \alpha + \beta z) = \text{cov}(z, w)^2/\sigma^2_z$ and $\sigma^2_{\hat{w}} = \text{var}(\alpha + \beta z) = \text{cov}(z\,w)^2\sigma^2_z$, $\rho^2_{w,\hat{w}} = \text{cov}(z, w)^2/\sigma^2_z \, \sigma^2_w = \rho^2_{z,w}$. The coefficient of determination in the multitrait case can be determined using the previous relationships by noting that $\hat{w} = \alpha + \beta_1 z_1 + \beta_2 z_2 + \ldots = \alpha + \boldsymbol{\beta}^T \mathbf{z}$. Because both $\text{cov}(w, \alpha + \boldsymbol{\beta}^T \mathbf{z}) = \boldsymbol{\beta}^T \mathbf{s}$ and $\sigma^2_{\hat{w}} = \boldsymbol{\beta}^T \mathbf{s}$, combining equations (1) and (A1) yields equation (4).