Correcting for Regression to the Mean
in Behavior and Ecology

Colleen Kelly[1,*] and Trevor D. Price[2]

1. Department of Mathematics and Statistics, San Diego State University, San Diego, California 92182;
2. Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637

ABSTRACT: If two successive trait measurements have a less-than-perfect correlation, individuals or populations will, on average, tend to be closer to the mean on the second measurement (the so-called regression effect). Thus, there is a negative correlation between an individual's state at time 1 and the change in state from time 1 to time 2. In addition, whenever groups differ in their initial mean values, the expected change in the mean value from time 1 to time 2 will differ among the groups. For example, birds feeding nestlings lose weight, but initially heavier birds lose more weight than lighter birds, a result expected from the regression effect. In sexual selection, males who remain unmated in the first year are, on average, less attractive than mated males. The regression effect predicts that these males will increase their attractiveness in the second year more than mated males. In well-designed experiments, changes in the experimental and control groups would be compared. In observational studies, however, no such comparison is available, and expected differential effects must be accounted for before they can be attributed to external causes. We describe methods to correct for the regression effect and assess alternative causal explanations.

Keywords: density dependence, mating success, regression, statistical artifact.

Regression toward the mean occurs in repeated-measures analyses whenever the correlation between the measurements at different times is less than perfect. At the second measurement, individuals with values above the mean will, on average, have lower values, whereas those with values below the mean will, on average, have higher ones. This so-called regression effect was first discovered by Galton (1886) through observations of human stature and in sweet pea experiments. Galton noted that if there were no regression to the mean, random error would cause the population variation to increase over time and extraordinarily small and large individuals would accumulate in the population.

Regression to the mean complicates analyses for researchers who study subsets of the population selected on the basis of their initial measurement. For example, studies that target individuals with initially large values should expect these values to decrease based on regression to the mean. The problem arises when a researcher attributes this decrease to an intervention or other causal effect. Since regression to the mean will affect both experimental and control groups, a well-designed experimental study will not be subject to this problem. However, statistical corrections must be applied to observational studies that compare groups that differ in their initial mean values. In this article we describe such tests.

The regression effect predicts that very sick patients should feel better at the next measurement even without effective treatment and thus could explain some of the placebo effect (McDonald et al. 1983). Likewise, regression to the mean predicts that recall of rare events may be improved later and could explain some of imagination inflation, an increased confidence that an imagined childhood event has occurred (Pezdek and Eddy 2001). Tversky and Kahneman (1974) noted the regression effect in training programs: students who do very well on the first trial typically do worse on the second, and students who do poorly at first typically do better later. This has led some to the erroneous conclusion that rewarding success does not work but that punishing failure does, and it leads to "a lifelong schedule in which we are [apparently] most often rewarded for punishing others and punished for rewarding" (Kahneman and Tversky 1973, p. 251). Good (1990) extended this reasoning to international relations, suggesting that the regression fallacy could be responsible for war.

* Corresponding author. Present address: School of Mathematics, Statistics and Computer Science, University of Wellington, P.O. Box 600, Wellington, New Zealand; e-mail: colleen.kelly@vuw.ac.nz.

Despite the widespread appreciation of the regression effect, recognizing its role in applications is difficult, even for statisticians well aware of the problem (Kahneman and Tversky 1973; Good 1990; Stigler 1997). For example, the statistician Secrist noticed that businesses with exceptional profits in one year tended to have smaller profits in the next and that businesses with very low profits tended to do better the next year. He used these data to conclude that all companies were converging toward mediocrity in his book *The Triumph of Mediocrity in Business* (Secrist 1933). Although the error was pointed out by Hotelling (1933) and later by Friedman (1992), other prominent economists have repeated it (Sharpe 1985; Fama and French 2000). Since regression to the mean is not likely to be recognized by sports gamblers, profitable betting opportunities may exist (Lee and Smith 2002).

A number of articles have pointed out problems due to regression effect in several disciplines, including epidemiology and clinical studies (Davis 1976; Curnow 1987; Bland and Altman 1994; Yudkin and Stratton 1996; Barnett et al. 2005), exercise and sports science (Shephard 2003; Nevill et al. 2004), psychiatry (Streiner 2001), chronobiology (Atkinson et al. 2001), communication (Zhang and Tomblin 2003), and hierarchical linear modeling (Marsh and Hau 2002). However, there has been no general survey in behavior and ecology. Here, we discuss several recent biological and ecological examples where causal explanations have been proposed for the regression effect. We describe a method that can be used to correct for the regression effect and illustrate its use in the examples.

Regression to the mean will be present whenever individuals or populations are measured at two different times. We note four general manifestations of regression to the mean that may be mistakenly attributed to causal factors. First, there is a negative correlation between an individual's first value and changes in that value between the first and second measurements (Cichoń et al. 1999). Second, there will be change in the mean of a single group whenever the mean of that group differs from the population mean. Third, there will be different changes in two groups whenever these two groups differ in their mean values at the first measurement. This applies even if both groups lie above the population mean or both groups lie below the population mean. Fourth, the most subtle artifact occurs when all individuals show an increase but those below the mean increase more than those above the mean (or all decrease, but those above the mean show a greater decrease than those below the mean; Gebhardt-Henrich et al. 1998; Griffith and Sheldon 2001).

We describe examples of both observational and experimental studies that illustrate these problems. Our examples include changes in mass, sexually selected traits, costs of cooperative breeding, and density dependence in

forests. We have not been provided with data sets for all these examples and are not able to reanalyze all of them. We start by describing some of the examples.

Examples

Mass Loss

Our primary example considers the debate on the correct statistical method to test whether mass loss depends on initial mass (Cichoń et al. 1999; Gebhardt-Henrich 2000; Ruf 2000). Contrary to other examples, change is defined here as the initial state minus the final state (corresponding to mass loss), and regression to the mean implies a positive correlation between the initial weight and the mass lost. Indeed, a positive correlation between incubation mass and subsequent mass loss in birds has been noted in several observational studies (Norberg 1981; Nur 1988; Hillström 1995; Merilä and Wiggins 1997; Gebhardt-Henrich et al. 1998). An example is in figure 1 (*top*). This result has often been interpreted as "parents in initially better condition can 'afford' to lose more mass than those in poorer condition, and that this energy can be allocated to their offspring" (Cichoń et al. 1999, p. 191), but it is expected from the regression effect (Cichoń et al. 1999). Cichoń et al. (1999) developed a method to correct for the regression effect based on resampling. Their method assumed that the initial and final values were uncorrelated, which is equivalent to complete regression to the mean. As noted by Ruf (2000), this is a very unreasonable assumption. Gebhardt-Henrich (2000) suggested an alternative statistical analysis that also assumes independence of the initial and final measurements.

Sexual Selection

In a noncontrolled experiment, Witte and Curio (1999) measured the attractiveness of male Javanese mannikins (*Lonchura leucogastroides*) to females. They then attached red feathers to the males' crowns and found that previously unattractive males gained attractiveness, whereas previously attractive males lost attractiveness. There was a significant negative relationship ($r = -0.76$, $n = 11$, $P = .003$) between initial attractiveness and change in attractiveness. Witte and Curio suggested that this is because attractive males have their phenotype disrupted by addition of what is otherwise an inherently attractive trait. However, as a result of regression to the mean, such a relationship is expected when there is no effect of the experiment.

In another example from sexual selection, Blows (1998) examined the mating success of hybrid *Drosophila serrata* × *Drosophila birchii* when male or female hybrids
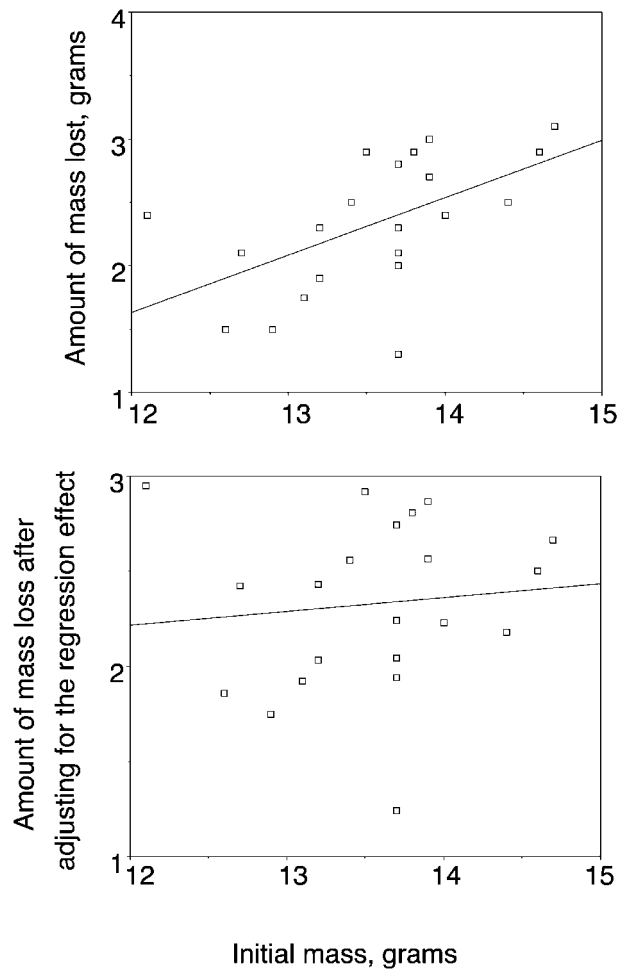
Figure 1: Loss of mass of female blue tits during the raising of their brood. *Top*, original data (from Gebhardt-Henrich et al. 1998). *Bottom*, mass loss adjusted for the regression effect using equation (6).

were placed with *D. serrata* in 29 replicated lines. Blows compared hybrid mating success five generations after initiating the hybrid lines with hybrid mating success 19 generations later. Among females, he found a tendency for hybrid lines with high mating success at generation 5 to have lower success at generation 24 and for those with low mating success at generation 5 to have higher success at generation 24. In males, there was a general increase in mating success between the two generations (fig. 2), but males with low mating success increased their mating success to a greater extent. Blows (1998) concluded that all lines were evolving toward a single regression line that describes a linear association between mean male mating success and mean female mating success. Blows considered the possibility that the regression effect may have influenced the results. He tested whether each line had signif-

icantly changed its mating success between generation 5 and generation 24 with a two-sample *t*-test. The results of these *t*-tests were combined into a single *P* value. The combined *P* value suggested that the mating success of the lines had indeed changed during the course of the experiment, but this does not in itself control for the regression effect, which also predicts a change.

Badyaev and Duckworth (2003) measured area of red on the breast of male house finches (*Carpodacus mexicanus*) to investigate how prior mating status affects an individual's investment in sexual ornaments. The red area increased from one year to the next, but it increased more for unpaired males than for paired males. Badyaev and Duckworth (2003) concluded that prior mating status had a strong influence on future development of sexually selected traits. However, red area is an attractive trait to females (Hill 1994), so paired males presumably initially had more red area than unpaired males. A greater increase among the unpaired group is expected from the regression effect, and this must be accounted for before differences can be assigned to mating status. Before considering these and other examples in more depth, we now consider methods to correct for the regression effect.

Theory

Most methods to adjust for the regression effect deal with the situation in which study subjects are selected on the basis of a large (or small) initial measurement (James 1973;
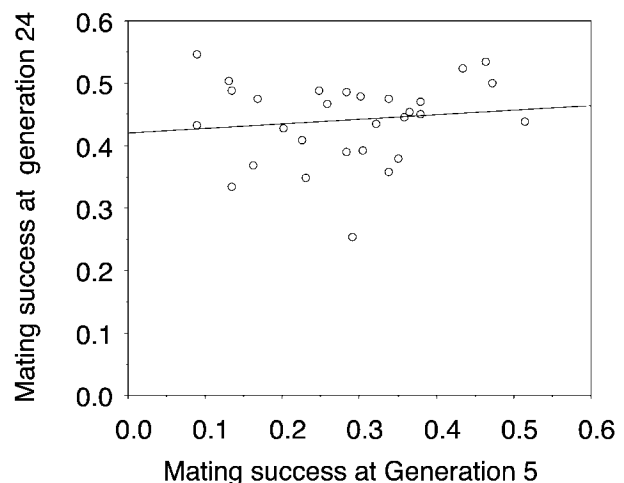

Figure 2: Hybrid male (*Drosophila serrata* × *Drosophila birchii*) mating success measured as the proportion of female *D. serrata* inseminated in 29 lines, with each line measured at generation 5 and again at generation 24 (Blows 1998). The data here have been arcsine–square root transformed. Note the much larger variance at generation 5 ($s_1^2 = 0.013$) than at generation 24 ($s_2^2 = 0.004$).

Senn and Brown 1985; Curnow 1987; Mee and Chua 1991; Chen and Cox 1992; Chuang-Stein 1993; Chen et al. 1998; Naranjo and McKean 2001). Here we consider studies in which there is no initial selection bias but where individuals with different initial values are separated and compared with respect to their subsequent values. The question of interest in these studies is, are there differential effects on the groups with initially low and high values beyond that expected from the regression effect? As pointed out by Galton (1886) and Hotelling (1933), a differential effect would change the variance of the population. Thus, if heavy individuals lose more weight than expected by regression to the mean and/or lighter individuals gain more, we expect convergence to the mean and a corresponding reduction in variance. If heavy individuals lose less weight than expected and/or lighter individuals gain less, we expect an increased variance in the population. Thus, a suitable test for a differential effect is a test of the equality of variance in the two time groups. We outline the test following Berry et al. (1984) and Chuang-Stein (1993).

Let the measurements at times 1 and 2 be X_1 and X_2 and the change in these measurements between time intervals be $D = X_1 - X_2$. We assume that the measurements have a bivariate normal distribution with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation ρ. We define an additive effect as one that affects all subjects equally; that is, it is the mean difference between the two sets of measurements, $\Delta = \mu_1 - \mu_2$. When measuring body mass at two time points, this is the average mass loss. We define a differential effect as one that affects individuals differently based on their initial values, X_1. If there is a differential effect that pushes values toward the mean (above and beyond the regression effect), then the variance at time 2 will be smaller than that at time 1: $\sigma_1^2 > \sigma_2^2$. If there is no differential effect, then the variances will be the same: $\sigma_1^2 = \sigma_2^2$. When there is only an additive effect and no differential effect, the expected change from time 1 to time 2 of the individual values, D, is given by the regression function of D on X_1,

$$\mathrm{E}[D|X_1] = (1 - \rho)(X_1 - \mu_1) + \Delta, \tag{1}$$

and the expected percentage change is

$$\mathrm{E}\left[\frac{D}{X_1}\Big|X_1\right] = (1 - \rho)\left(1 - \frac{\mu_1}{X_1}\right) + \frac{\Delta}{X_1}. \tag{2}$$

These equations show that the expected change (and the expected percentage change) is positive for values above the mean and negative for values below the mean and that the magnitude of the expected change increases for mea-

surements X_1 farther from the mean. The correlation between the initial value and the change is

$$\sqrt{\frac{1 - \rho}{2}}. \tag{3}$$

The magnitude of the differential effect is measured as the ratio of the standard deviations of the two sets of observations, $\theta = \sigma_2/\sigma_1$. Assuming only that the measurements are bivariate normal, we obtain the following relationship between the initial state and the change in state:

$$\mathrm{E}[D|X_1] = (1 - \rho\theta)(X_1 - \mu_1) + \Delta. \tag{4}$$

Note that $\rho\theta$ is the regression of X_2 on X_1. If $\theta = 1$, there is no differential treatment effect: individuals above and below the mean decrease or increase in a manner that is expected from the regression effect.

Following Berry et al. (1984), we test for a differential effect by testing the null hypothesis H_0 ($\sigma_1^2 = \sigma_2^2$) against the alternative H_a ($\sigma_1^2 \neq \sigma_2^2$) with Pitman's (1939) test for the equality of variances in paired samples,

$$T = \frac{\sqrt{n - 2}[(s_1/s_2) - (s_2/s_1)]}{2\sqrt{1 - r^2}}, \tag{5}$$

where T has a Student's t distribution with $n - 2$ degrees of freedom, n is the sample size, and s_1, s_2, and r are the usual estimates of σ_1, σ_2, and ρ, respectively.

We can also adjust each value by subtracting the change that is expected as a result of the regression effect, $D^* = D - \mathrm{E}[X_1 - X_2|X_1]$. From equation (1), we estimate D^* with

$$\hat{D}^* = \hat{\rho}(X_1 - \bar{X}_1) - (X_2 - \bar{X}_2), \tag{6}$$

where $\hat{\rho} = r$ if the null hypothesis of equal variances is rejected and

$$\hat{\rho} = \frac{2rs_1s_2}{s_1^2 + s_2^2}$$

when it is not rejected. The adjusted differences, $\hat{D}^*$, can be regressed against X_1 and graphically viewed in a scatter plot, or they can be used to test for the influence of a measured factor that is correlated with the initial measurements. This is illustrated further in the next section.

Analysis of Examples

In this section, we reanalyze examples for which we have been able to obtain data from the original article or have been provided the data by the authors.

Mass Loss

Gebhardt-Henrich et al. (1998) presented data on loss of mass of female blue tits *Parus caeruleus* during the feeding of the brood. Their figure 2, reproduced here as figure 1 (*top*), shows that all blue tits lost mass but that females that were initially heavier lost more mass than those that were initially lighter (correlation between mass and loss of mass is $r = 0.55$, $P = .0096$). We analyzed these data using equation (5) and could find no evidence for a differential effect ($T = 0.74$, df $= 19$, $P = .46$). We constructed adjusted values by using equation (6) and added these values to the change in the mean (fig. 1, *bottom*). This shows that there is a lack of association between change in mass and initial mass after the regression effect has been accounted for. Note that these adjusted values could also be used in other tests. For example, some females may be in poor habitat and initially of less weight than others in good habitat. If one wished to determine if habitat influenced weight loss, a simple *t*-test comparing females in the two habitats would be confounded with the regression effect. One approach to deal with this would be to compare the adjusted values (fig. 1, *bottom*) between females in poor and good habitats using a *t*-test.

Cichoń et al. (1999) presented initial mass and mass loss from three other studies (on the blue tit, the collared flycatcher *Ficedula albicollis*, and the pied flycatcher *Ficedula hypoleuca*). They concluded that although the correlation between initial mass and mass loss is positive, for all three species the observed empirical correlations were significantly weaker than those expected under null expectations derived by bootstrapping (under the assumption of independent initial and final masses). The result led them to suggest that lighter birds actually lost more mass than expected and that heavy birds lost less mass than expected. Their correction thus led to a conclusion that was opposite to that reached by the original analysis, but this conclusion depends on an obviously falsifiable null model that initial and final weights are uncorrelated (Ruf 2000). We applied the tests outlined in this article and found evidence of differential mass loss only in the collared flycatcher ($T = 2.22$, df $= 273$, $P = .04$). In this species, the correlation between initial and final mass was $r = 0.33$. Heavier birds lost more mass than expected (variance of values at the second measurement was 77% that of values at the first measurement), opposite to the conclusions of Cichoń et al. (1999).

Sexual Selection

We found no evidence that male mating success is affected by addition of a red feather in Witte and Curio's (1999) study of Javanese mannikins ($T = 0.403$, df $= 9$, two-tailed $P = .696$). We emphasize that other results in the article stand, and the article should be consulted to evaluate this result in context.

In the study by Blows (1998), the correlation between (arcsine–square root transformed) hybrid female mating success in generation 5 and change in hybrid female mating success between generation 5 and generation 24 was $r = -0.78$ ($n = 29$ replicate lines). It is not possible to exclude the regression effect as the underlying reason for this correlation ($T = 1.081$, df $= 27$, two-tailed $P = 0.289$). Indeed, the low correlation between hybrid female mating success at generation 5 and generation 24 ($r = -0.06$) results in a large regression effect (see eq. [1]).

In males, there was a general increase in mating success between the two generations (fig. 2), but males with low mating success increased their mating success to a greater extent. The correlation between (arcsine transformed) mating success at generation 5 and change in mating success is $r = -0.85$. In this example, the male lines have converged. The variance decreased by 67% (from 0.013 to 0.004 on the transformed scale; see fig. 2), and this is significant ($T = 3.051$, df $= 27$, $P = .005$). This implies that males with particularly low values of mating success at generation 5 are indeed increasing their mating success more than those with high mating success. Thus, there is support for Blows's conclusion that all lines were converging toward a common trajectory, at least through male mating success.

Other Examples

Griffith and Sheldon (2001) found a negative correlation between the size of an unpigmented plumage patch in male collared flycatchers and change in the size of the patch across a season ($r = -0.4$, $n = 80$). Although they did not attach much significance to this finding, they have kindly provided us with the data, and we find that the result can be explained by the regression effect ($T = 0.51$, df $= 78$, $P = .61$).

Discussion

Regression to the mean results whenever there is a less-than-perfect correlation between successive measurements. In fact, the magnitude of the regression to the mean is proportional to $(1 - \rho)$. Because measurement error lowers the correlation, regression to the mean can be reduced by using more accurate measurement methods or

the mean of replicate measurements (Gardner and Heady 1973; Blomqvist 1987; Griffith and Sheldon 2001). However, even when this is done, the regression effect will be present because the correlation will be less than perfect, and the best approach to avoiding the regression fallacy is to use an appropriate control and compare changes in the experimental and control groups. Thus, in the study by Witte and Curio (1999), it would have been possible to examine changes in a group of nonmanipulated males and see if they differed from those in the manipulated males. Russell et al. (2003) found that in cooperative meerkats *Suricata suricatta,* individuals investing heavily in one breeding event significantly reduce their contribution in the following event, whereas those previously investing little significantly increase their contribution. They recognized that this could be due to the regression effect and used a supplemental feeding experiment (as well as other correlative evidence) to infer causality.

In observational studies, a measured causal factor can be investigated by examining its relationship with values adjusted for the regression effect. This would be possible in the house finch study by Badyaev and Duckworth (2003), for example. In this study, unpaired males increased the area of red in their plumage 85% more than the paired males did. Assuming that they initially had less red in their plumage, a differential increase is expected. The causal factor (mating status) is known, so the adjusted values could be calculated according to equation (6) and then compared using a *t*-test. Unfortunately, we have not been provided with the original data for this study, and the information needed to correct for the regression effect is not extractable from the article. In other studies (e.g., those of mass loss, where larger individuals are postulated to be in better "condition"), the causal factor is not measured, and only the change-in-variance test can be used.

Wills et al. (1997) and Wills and Condit (1999) suggested that strong density-dependent effects were occurring in tropical trees. They found that quadrats in experimental plots with few individuals of one species tended to recruit more individuals than quadrats with many individuals. They interpreted this as a result of density dependence, perhaps acting through parasites and pathogens; again the causal factor is postulated but not measured. There have been many other studies purportedly demonstrating a role of density dependence in enabling coexistence of tree species (Lambers et al. 2002). Indeed, Lambers et al. (2002) suggested several reasons why density dependence may be underestimated. In general, the supporting evidence relies on the finding of a negative correlation between recruitment and prior density of trees in a quadrat, which would be predicted based on the regression effect. The possibility of regression effects can be quite subtle. Lambers et al. (2002) measured the propor-

tion of seeds germinated in 1 × 1-m plots and showed that this was negatively correlated with density of seeds in the plot. However, if high density of seed is partly a result of high germination at a previous time period, then a negative correlation is expected from the regression effect.

We have not been able to analyze results on trees. However, in this case it is worth noting that at equilibrium, the variance in tree distributions across quadrats remains the same from one generation to the next (at least when measured at the same life-history stage). This means that it is impossible to refute the regression effect as a cause of a negative association between change and initial value and hence impossible to detect density dependence by these methods. Variance decreases due to high-density quadrats decreasing in number and low-density quadrats increasing in number must be compensated for by intermediate quadrats both increasing and decreasing. Unless one is inclined to invoke special factors causing all these changes, it is more parsimonious to assign all increases and decreases to random factors. More sophisticated methods using time series data that explicitly incorporate an error term to account for regression toward the mean are required if density dependence is to be detected (Lande et al. 2002).

Whenever two sets of measurements are not perfectly correlated, there will be regression toward the mean. Thus, ascribing biological significance to regression to the mean is equivalent to ascribing significance to a correlation of less than unity. Many traits are influenced by multiple factors, so correlations are rarely unity, and regression toward the mean is inevitable.

Acknowledgments

Literature Cited

Atkinson, G., J. Waterhouse, T. Reilly, and B. Edwards. 2001. How to show that unicorn milk is a chronobiotic: the regression to the mean statistical artifact. Chronobiology International 18:1041–1053.

Badyaev, A. V., and R. A. Duckworth. 2003. Context-dependent sexual advertisement: plasticity in development of sexual ornamentation throughout the lifetime of a passerine bird. Journal of Evolutionary Biology 16:1065–1076.

Barnett, A. G., J. C. van der Pols, and A. J. Dobson. 2005. Regression to the mean: what it is and how to deal with it. International Journal of Epidemiology 34:215–220.

Berry, D. A., M. L. Eaton, B. P. Ekholm, and T. L. Fox. 1984. Assessing differential drug effect. Biometrics 40:1109–1115.

Bland, J. M., and D. G. Altman. 1994. Some effects of regression towards the mean. British Medical Journal 309:780.

Blomqvist, N. 1987. On the bias caused by regression toward the mean in studying the relation between change and initial value. Journal of Clinical Periodontology 14:34–37.

Blows, M. W. 1998. Evolution of a mate recognition system after hybridization between two *Drosophila* species. American Naturalist 151:538–544.

Chen, S., and C. Cox. 1992. Use of baseline data for estimation of treatment effects in the presence of regression to the mean. Biometrics 48:593–598.

Chen, S. D., C. Cox, and L. Cui. 1998. A more flexible regression-to-the-mean model with possible stratification. Biometrics 54:939–947.

Chuang-Stein, C. 1993. The regression fallacy. Drug Information Journal 27:1213–1220.

Cichoń, M., J. Merilä, L. Hillström, and D. Wiggins. 1999. Mass-dependent mass loss in breeding birds: getting the null hypothesis right. Oikos 87:191–194.

Curnow, R. N. 1987. Correcting for regression in assessing the response to treatment in a selected population. Statistics in Medicine 6:113–117.

Davis, C. E. 1976. The effect of regression to the mean in epidemiologic and clinical studies. American Journal of Epidemiology 104:493–498.

Fama, E. F., and K. R. French. 2000. Forecasting profitability and earnings. Journal of Business 73:161–175.

Friedman, M. 1992. Do old fallacies ever die? Journal of Economic Literature 30:2129–2132.

Galton, F. 1886. Regression toward mediocrity in hereditary stature. Journal of the Anthropological Institute 15:246–263.

Gardner, M. J., and J. A. Heady. 1973. Some effects of within-person variability in epidemiological studies. Journal of Chronic Diseases 26:781–795.

Gebhardt-Henrich, S. G. 2000. When heavier birds lose more mass during breeding: statistical artifact or biologically meaningful? Journal of Avian Biology 31:245–246.

Gebhardt-Henrich, S. G., P. Heeb, H. Richner, and F. Tripet. 1998. Does loss of mass during breeding correlate with reproductive success? a study on blue tits *Parus caeruleus*. Ibis 140:210–213.

Good, I. J. 1990. The regression effect as a cause of war. Journal of Statistical Planning and Inference 26:123–124.

Griffith, S. C., and B. C. Sheldon. 2001. Phenotypic plasticity in the expression of sexually selected traits: neglected components of variation. Animal Behaviour 61:987–993.

Hill, G. E. 1994. Geographic variation in male ornamentation and female mate preference in the house finch: a comparative test of models of sexual selection. Behavioral Ecology 5:64–73.

Hillström, L. 1995. Body mass reduction during reproduction in the pied flycatcher *Ficedula hypoleuca*: physiological stress or adaptation for lowered costs of locomotion? Functional Ecology 9:807–817.

Hotelling, H. 1933. Review of *The triumph of mediocrity in business*. Journal of the American Statistical Association 28:463–465.

James, K. E. 1973. Regression toward the mean in uncontrolled clinical studies. Biometrics 29:121–130.

Kahneman, D., and A. Tversky. 1973. Psychology of prediction. Psychological Review 80:237–251.

Lambers, J. H. R., J. S. Clark, and B. Beckage. 2002. Density-dependent mortality and the latitudinal effect in species diversity. Nature 417:732–735.

Lande, R., S. Engen, B.-E. Sæther, F. Filli, E. Matthysen, and H. Weimerskirch. 2002. Estimating density dependence from population time series using demographic theory and life-history data. American Naturalist 159:321–337.

Lee, M., and G. Smith. 2002. Regression to the mean and football wagers. Journal of Behavioral Decision Making 15:329–342.

Marsh, H. W., and K. T. Hau. 2002. Multilevel modeling of longitudinal growth and change: substantive effects or regression toward the mean artifacts? Multivariate Behavioral Research 37:245–282.

McDonald, C. J., S. A. Mazzuca, and G. P. McCabe Jr. 1983. How much of the placebo "effect" is really statistical regression? Statistics in Medicine 2:417–427.

Mee, R. W., and T. C. Chua. 1991. Regression toward the mean and the paired sample *t*-test. American Statistician 45:39–42.

Merilä, J., and D. A. Wiggins. 1997. Mass loss in breeding blue tits: the role of energetic stress. Journal of Animal Ecology 66:452–460.

Naranjo, J. D., and J. W. McKean. 2001. Adjusting for regression effect in uncontrolled studies. Biometrics 57:178–181.

Nevill, A., R. Holder, G. Atkinson, and J. Copas. 2004. The dangers of reporting spurious regression to the mean. Journal of Sports Sciences 22:800–802.

Norberg, A. R. 1981. Temporary weight decrease in breeding birds may result in more fledged young. American Naturalist 118:838–850.

Nur, N. 1988. The cost of reproduction in birds: an examination of the evidence. Ardea 476:155–168.

Pezdek, K., and R. M. Eddy. 2001. Imagination inflation: a statistical artifact of regression toward the mean. Memory and Cognition 29:707–718.

Pitman, E. J. G. 1939. A note on normal correlation. Biometrika 31:9–12.

Ruf, T. 2000. Mass-dependent mass loss: how to get the null hypothesis wrong. Oikos 90:413–416.

Russell, A. F., L. L. Sharpe, P. N. M. Brotherton, and T. H. Clutton-Brock. 2003. Cost minimization by helpers in cooperative vertebrates. Proceedings of the National Academy of Sciences of the USA 100:3333–3338.

Secrist, H. 1933. The triumph of mediocrity in business. Bureau of Business Research, Northwestern University, Evanston, IL.

Senn, S. J., and R. A. Brown. 1985. Estimating treatment effects in clinical trials subject to regression to the mean. Biometrics 41:555–560.

Sharpe, W. F. 1985. Investments. 3rd ed. Prentice Hall, Englewood Cliffs, NJ.

Shephard, R. J. 2003. Regression to the mean: a threat to exercise science? Sports Medicine 33:575–584.

Stigler, S. M. 1997. Regression toward the mean, historically considered. Statistical Methods in Medical Research 6:103–114.

Streiner, D. L. 2001. Regression toward the mean: its etiology, diagnosis, and treatment. Canadian Journal of Psychiatry 46:72–76.

Tversky, A., and D. Kahneman. 1974. Judgement under uncertainty: heuristics and biases. Science 185:1124–1131.

Wills, C., and R. Condit. 1999. Similar non-random processes maintain diversity in two tropical rainforests. Proceedings of the Royal Society of London B 266:1445–1452.

Wills, C., R. Condit, R. B. Foster, and S. P. Hubbell. 1997. Strong density- and diversity-related effects help to maintain tree species

diversity in a Neotropical forest. Proceedings of the National Academy of Sciences of the USA 94:1252–1257.

Witte, K., and E. Curio. 1999. Sexes of a monomorphic species differ in preference for mates with a novel trait. Behavioural Ecology 10: 15–21.

Yudkin, P. L., and I. M. Stratton. 1996. How to deal with regression to the mean in intervention studies. Lancet 347:241–243.

Zhang, X. Y., and J. B. Tomblin. 2003. Explaining and controlling regression to the mean in longitudinal research designs. Journal of Speech Language and Hearing Research 46:1340–1351.