EFFECT SIZE IN ECOLOGICAL EXPERIMENTS: THE APPLICATION OF BIOLOGICAL MODELS IN META-ANALYSIS

CRAIG W. OSENBERG,[1,*] ORLANDO SARNELLE,[2,†] AND SCOTT D. COOPER[2]

[1]Department of Zoology, University of Florida, Gainesville, Florida 32611-8525; [2]Department of Ecology, Evolution, and Marine Biology, and Marine Science Institute, University of California, Santa Barbara, California 93106

Submitted February 20, 1997; Revised June 16, 1997; Accepted June 26, 1997

Some of the most interesting and important questions in ecology require examination of the strength of different processes across environmental gradients and among organisms with different traits (Quinn and Dunham 1983; Tilman 1989; Cooper et al. 1990; Sarnelle 1992; Osenberg and Mittelbach 1996). Meta-analysis (see, e.g., Gurevitch et al. 1992; Gurevitch and Hedges 1993; Arnqvist and Wooster 1995; Curtis 1996) combines results from independent studies to examine patterns of effect across taxa or environments and, thus, may represent a powerful tool to test ecological theory. A meta-analysis requires that a common metric of effect size be extracted from each of the studies. Here, we focus on choosing a metric that best facilitates ecological inferences. We begin with a brief description of the standard definition of effect size in meta-analysis, as used in recent papers. We then discuss potential problems with this approach and suggest an alternative that is more explicitly tied to the dynamics of ecological systems. Using two examples drawn from predator-prey experiments, we then illustrate the limitations of the standard metric and the conceptual advantages of one ecologically based alternative. We conclude by discussing the link between metrics of effect size and ecological models.

EFFECT SIZE IN META-ANALYSIS

The Classic Approach

In most ecological applications of meta-analysis to date, effect size has been defined as the difference between two treatments—experimental and control—standardized by the pooled within-treatment standard deviation (Glass 1976; Hedges and Olkin 1985; see Gurevitch et al. 1992; VanderWerf 1992; Tonhasca and Byrne 1994; Wooster 1994; Arnqvist and Wooster 1995; and Curtis 1996 for ecological examples):

$$d = \frac{\bar{N}_E - \bar{N}_C}{s} J, \quad (1)$$

where

$$s = \sqrt{\frac{(n_E - 1)(s_E^2) + (n_C - 1)(s_C^2)}{n_E + n_C - 2}}, \quad (2)$$

$\bar{N}_i$ is the mean, n_i the sample size, and s_i^2 the variance within the experimental (i = E) or control (i = C) groups, and J corrects for bias due to small sample size (Hedges and Olkin 1985). Note that d is mathematically related to the t statistic and its companion meta-analytic metric, the correlation coefficient, r (e.g., Rosenthal 1994). One potential problem with d (or t or r) is that patterns may emerge that have little to do with the actual strength of a process but, instead, reflect differences in s. For example, suppose that invertebrate predators have larger effects as measured by d on stream invertebrate prey than do vertebrate predators (Wooster 1994). It is possible that vertebrates reduce prey density to the same degree (or more) as invertebrates but that studies with vertebrates were characterized by larger values of s. The latter could arise for several plausible reasons, including greater individual variation in feeding behavior, which would be more pronounced when lower densities of predators are used (as is likely with vertebrates), and the use of larger cages to enclose vertebrate predators, which might result in the cages being dispersed over a larger and more heterogeneous landscape. In these scenarios, the use of d confounds the effect of the treatments with aspects of the sampling design and the degree of spatial heterogeneity that differ among studies. Related metrics, such as r or r^2, have similar problems (Petraitis 1998).

An Alternative Approach

As an alternative to d (and related statistically derived metrics), we suggest that the choice of a metric in meta-analysis should be based on clear specification of the process being studied and the spatial and temporal scales over which the experiments were conducted. This will often require specification of a biological model or conceptual framework (see related arguments in Adler and Morris 1994; Billick and Case 1994; Wootton 1994; Laska and Wootton 1998).

For example, consider two treatments: a control, representing ambient conditions, and an experimental treatment, in which one factor has been manipulated. Assuming that the resulting dynamics of the target species can be described by an exponential model, the effect of the treatment on the target's per capita dynamics can be estimated as:

$$\Delta r = r_C - r_E = dN_C/N_C dt - dN_E/N_E dt = [\ln(N_{t,C}/N_{0,C}) - \ln(N_{t,E}/N_{0,E})]/t, \quad (3)$$

where r_C and r_E are the per capita rates of change for the target populations; $N_{t,C}$ and $N_{t,E}$ (and $N_{0,C}$ and $N_{0,E}$) are the densities of the target at the end (or start) of

* E-mail: osenberg@zoo.ufl.edu.

† Present address: Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan 48824-1222.

Am. Nat. 1997. Vol. 150, pp. 798–812.

the experiment in the control and experimental treatments; and t is the duration of the experiment. If we assume that the initial conditions are similar (i.e., $N_{0,C} = N_{0,E}$), equation (3) simplifies to a more familiar form (e.g., Osenberg and Mittelbach 1996):

$$\Delta r = \ln(N_{t,C}/N_{t,E})/t. \quad (4)$$

An index analogous to equation (4), but defined from a discrete time model, would yield $(N_{t,C}/N_{t,E})^{1/t}$, which gives the proportionate change in the finite growth rate of the target caused by the treatment. Equation 4 can also be expressed as a per-unit effect on the target by dividing the effect on the target's per capita growth rate by the size of the manipulation. For example, if the study focuses on the effects of predation, and the ambient (control) predator density is P whereas the experimental density is 0, then the per-predator effect is:

$$\Delta r/P = \ln(N_{t,C}/N_{t,E})/(tP). \quad (5)$$

EXAMPLES FROM STUDIES OF PREDATION

To highlight the distinction between d and Δr (eq. 4), we analyzed two sets of data exploring the effects of fish on freshwater invertebrates. The experiments are typical of many conducted in aquatic environments: cages were used to manipulate the densities of fish, and prey densities were assessed at the end of the experiment.

Bluegill Effects on Odonates: Comparisons within a Single Study

The first example is taken from Morin's (1984) classic experiment examining the effects of fish predation on a larval odonate assemblage. In this case, our analysis is concerned not with results from multiple studies but, rather, with the results from a single experiment that documented the responses of different taxa. Although this is not a meta-analysis per se, this analysis provides a clear distinction between d and Δr and illustrates that our arguments are relevant to inferences drawn from individuals studies as well as meta-analyses.

In our analysis, we focused on the proposed relationship between odonate body size and the magnitude of the fish effect on odonate prey. This hypothesis, one explicitly addressed by Morin, was motivated by the observation that bluegill selectivity increases with prey size (e.g., Mittelbach 1981, 1988; Olson et al. 1995). We calculated effect sizes, d and Δr (using eqq. [1] and [4]), for the seven most abundant odonate species, whose densities in cages with and without fish (primarily bluegill) were reported in table 2 of Morin (1984). We then asked whether effect sizes varied significantly with odonate body size (Morin 1984, table 1).

Although d was not significantly correlated with odonate size (fig. 1*A*: $r = .06$, $n = 8$, $P = .88$), Δr was ($r = -.84$, $n = 7$, $P = .017$). The analysis using Δr was somewhat conservative because, although the largest species incurred the greatest effect, this datum was dropped from the analysis because its density was 0 in the presence of fish (i.e., $\Delta r = -\infty$). This analysis with Δr suggests

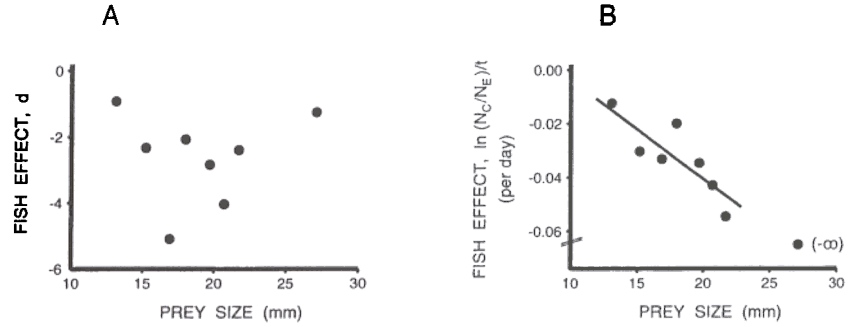

FIG. 1.—Relationship between fish effects and prey size using two metrics of effect size: (*A*) d (eq. [1]) and (*B*) Δr (eq. [4]). Data are from Morin's (1984) study of fish predation on larval odonates. Each point represents the response of one odonate species to the manipulation of fish density (presence/absence). Final odonate densities (number/cage) were calculated as the sum of the number of metamorphs and overwintering larvae reported in Morin's table 2. One very rare species (*Epicordulia regina*) was excluded from the analysis because its average density was <1 animal cage^{-1}. Standard deviations of final density (needed to estimate d) were estimated as the square root of the sum of the variance in metamorphs and overwintering larval numbers. Body size is the mean final instar body length and was taken from Morin's table 1. The regression in *B* excludes the point for the largest species (*Tramea lacerata*), which was not recorded in any cage with fish (thus, the effect size was undefined and noted here with a value of $-\infty$).

that larger prey incurred greater per capita mortality rates from bluegill predation than did smaller prey (fig. 1*B*). Not only did Δr yield a clear pattern of response (fig. 1*B* vs. *A*), it also provided a quantitative assessment of the relationship between fish effects on prey dynamics and prey size. For example, based on the fitted regression, the effect of fish on the per capita mortality rate of a large odonate (say, 25-mm body length) is over five times greater than that for a species of half the size (12.5-mm body length).

Effects of Specialist and Generalist Molluscivores: Comparisons across Studies

The second example is a meta-analysis in which we synthesized data from multiple studies that examined the effects of fish on freshwater gastropods. The experiments in these studies varied in their duration, which can influence the choice of a metric as well as the set of studies that can be appropriately combined. For example, under conditions of exponential growth, many metrics—such as $\ln(N_{t,C}/N_{t,E})$ or $N_{t,C} - N_{t,E}$ (or in many cases, d)—will assume values close to zero at the start of the experiment and increase in magnitude as the experiment progresses. Thus, two studies, with exactly the same dynamics and response to predation, would yield different effect sizes using any of these metrics if the studies had different durations. However, in this scenario Δr would be time invariant because it focuses on differences in rates and would thus be suitable for a meta-analysis. This would not be the case, however, if studies lasted

sufficiently long to reequilibrate. In these cases, Δr would vary with duration, while other metrics (such as $\ln(N_{t,C}/N_{t,E})$ or $N_{t,C} - N_{t,E}$) would be time invariant. The challenge then is to select a metric that is appropriate to the question and the timescale of the experiments being compared. For studies that are brief, relative to the speed of feedbacks in the system, a metric that focuses on rates (such as Δr) seems most appropriate (Billick and Case 1994). For long-term studies that measure equilibrial responses, other metrics, based on state variables (such as $\ln(N_{t,C}/N_{t,E})$ or $N_{t,C} - N_{t,E}$), will probably be best.

To illustrate this point and assess the timescale over which Δr is likely to be time invariant in the molluscivore analysis, we used data from Martin et al.'s (1992) study in which gastropod densities were sampled over 2 yr following the manipulation of sunfish. Martin et al. (1992) had four treatments: zero fish, five small fish m^{-2}, 0.5 large fish m^{-2}, and five small + 0.5 large fish m^{-2}. We estimated Δr by comparing each of the three "fish" treatments with the "no fish" treatment and examined the behavior of Δr through time. Effect size, Δr, showed no consistent trends through time for approximately the first 100 d of the experiment (fig. 2). After that time, and as expected, Δr decayed toward 0. Other metrics more suitable to long-term responses (e.g., $\ln(N_{t,C}/N_{t,E})$, or $N_{t,C} - N_{t,E}$) showed a different pattern; for example, $\ln(N_{t,C}/N_{t,E})$ decreased linearly from 0 to ~ -4 over the first 100 d but then appeared to stabilize after longer time periods.

We found 10 experiments that reported snail density responses to manipulations of fish density. Some experiments were only sampled once, after 43–93 d. For experiments that were sampled on more than one date, we restricted our analysis to a single date per study, chosen at a timescale (i.e., 43–86 d) comparable to the studies lacking multiple sampling dates. Given the duration of these studies, and our analysis of Martin et al.'s (1992) data (fig. 2), we used Δr as our ecologically derived metric of effect size.

We categorized the studies based on the functional morphology of the fish species that were manipulated. Fish were labeled as "specialists" if they possessed a specialized feeding morphology that allowed them to crush snails or "generalists" if they were incapable of crushing snails. Specialists consisted of tench, redear, and large pumpkinseed, whereas generalists consisted of bluegill, largemouth bass, Eurasian perch, and small pumpkinseed. Each study yielded more than one estimate of effect size because the studies included multiple levels of fish density or additional factors that were crossed with the fish manipulations.

In all cases, we defined effect size based on comparisons of one treatment with fish to another treatment (often without fish) that best isolated the effects of generalist or specialist predators. In most studies snail biomass in plots (cages) with versus without fish were compared; however, in some cases, the comparisons were more complex (e.g., five bluegill vs. five bluegill + 10 pumpkinseed yielded the effect of 10 pumpkinseed). In several cases effects of specialists and generalists could not be isolated from one another; we refer to these cases as "mixtures." For each study, we also quantified the strength of the ex-

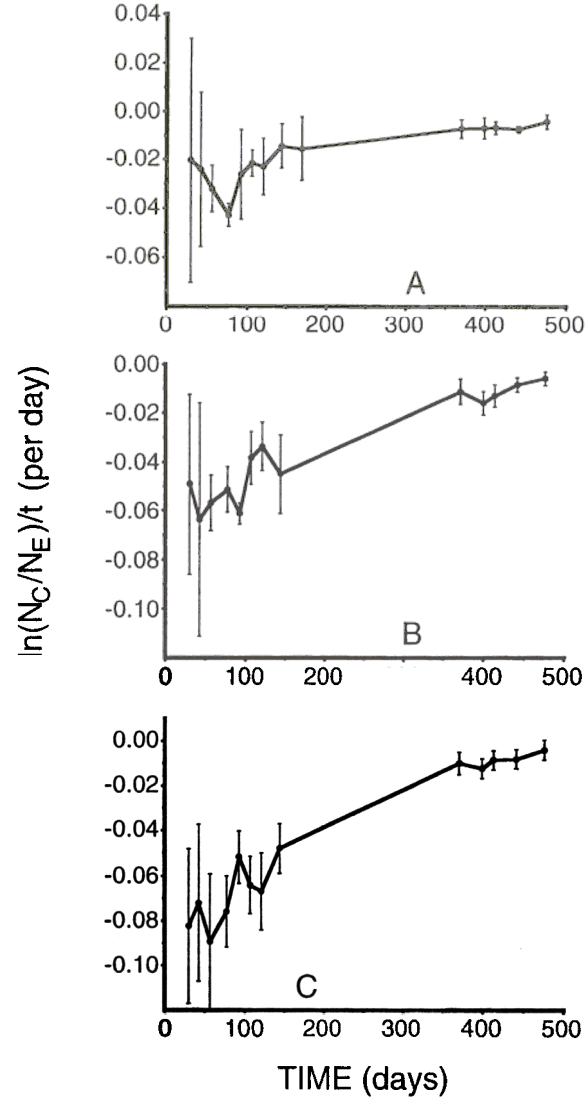

FIG. 2.—Effect size (Δr) as a function of duration using data from Martin et al.'s (1992) study of fish effects on snail density. *A*, Small fish (5 fish m^{-2}); *B*, large fish (0.5 fish m^{-2}); *C*, small and large fish together (5 + 0.5 fish m^{-2}). Confidence intervals for Δr were approximated as $(1.96/t)\sqrt{(SE_C^2/N_C^2 + SE_E^2/N_E^2)}$, where SE is the standard error of the mean density (N) of snails in the control or experimental treatments, and t is the time that had elapsed from the start of the experiment to the chosen sampling date (A. Stewart-Oaten, personal communication). Results for the first posttreatment survey (t = 16 d) gave confidence intervals ~3 times greater than for any other date and are not shown. Data from t = 170 d is not shown in *B* and *C* because Δr was undefined (no snails were collected in the cages with fish).

perimental perturbation as the difference in fish density (individuals m^{-2}) between the two treatments.

For each study we obtained estimates of mean snail biomass (grams/area) and standard deviations for each treatment, as well as experimental duration and magnitude of the manipulation. From these data we estimated effect sizes, defined as d or Δr, and examined how effect size differed among fish groups (generalists vs. specialists vs. mixtures) using magnitude of the manipulation (i.e., fish density differential) as a covariate. If the per-predator effect of fish on prey is constant, then the effect of the predator population (Δr) should be zero when fish density has not been manipulated and should decrease linearly as the fish density differential increases. Deviations from this expectation might suggest important biological mechanisms, such as predator interference or shifts in prey vulnerability with fish density.

In the analysis using d, one study was an extreme outlier ($d = -32.7$) because of unusually low among-replicate variance. The study was not an outlier in the analysis using Δr (fig. 3*B*), indicating that the study was exceptional in its variance, not the effect of predators on prey. To facilitate subsequent comparisons, we excluded this study from all analyses involving d (its inclusion only worsens the case for d).

Both metrics of effect size showed similar qualitative results (fig. 3), owing in part to the relatively high correlation between the two metrics ($\Delta r = .71$, $n = 25$, $P = .0001$). On average, specialists had greater (more negative) effects on snails than did generalists based on both d (specialists: -1.91 ± 0.96 [mean $\pm$ 95% confidence interval (CI)]; generalists: -0.13 ± 0.31) and Δr (eq. [4]; -0.031 ± 0.010 d^{-1} vs. -0.0024 ± 0.0039 d^{-1}), and the few points for the "mixtures" fell between the other two groups. Despite this crude similarity, there were critical differences between the two analyses concerning the relationships between effect size and fish density (fig. 3). Using ANOVA—with dietary group as the class variable, fish density differential as the covariate, and an interaction term to reflect the expected difference in the impact of generalists and specialists across predator densities—we found that d was unable to detect the expected relationships. There was neither a significant interaction between dietary group and fish density ($F = 0.75$, df = 2, 19, $P = .49$) nor a significant effect of fish density after dropping the nonsignificant interaction term ($F = .68$, df = 2, 21, $P = .42$). The analysis using Δr did, however, detect a significant interaction between dietary group and fish density ($F = 5.57$, df = 2, 20, $P = .012$). The slope for specialists was steeper than for generalists, and the effects of specialists (at any given fish density) were stronger (i.e., more negative) than those of generalists (i.e., a specialist imposed a greater effect on snails than did a generalist).

The conceptual advantages of Δr, in this context, are further highlighted by the ease with which this metric facilitates additional analyses and estimation. For example, we fit linear relationships (with zero intercepts) to the generalist and specialist data shown in figure 3*B*. This yielded estimates of the slope of the relationships (i.e., the per-fish effect on snail dynamics): $b_{\Delta r, P(\text{specialists})} = -.0595 \pm .0184$ d^{-1} $fish^{-1}$; $b_{\Delta r, P(\text{generalists})} = -.00127 \pm .00123$ d^{-1} $fish^{-1}$ (95% CI). Alternatively, these estimates can be obtained using equation (5) (i.e., by dividing each estimate in fig. 3*B* by the associated fish density differential and then averaging within a dietary group: $\Delta r/P_{\text{specialists}} = -.0798 \pm .0240$ d^{-1} $fish^{-1}$; $\Delta r/P_{\text{generalists}} = -.0014 \pm .0036$ d^{-1} $fish^{-1}$; means $\pm$ 95% CI). Assuming most of these effects arise from the direct consumption of snails by fish, these two procedures indicate that a specialist imposes a mortality rate ~50 times greater than does a generalist. These results provide an expeditious way to use meta-analytic results to generalize to new situations or to summarize the entire data set. For example, say that in a particular lake, generalists outnumber specialists by 10:1, then these results indicate that the specialist population (despite their numerical inferiority) imposes five times the mortality on snails than do the generalists ($10 \times 1{:}1 \times 50 = 1{:}5$).

To further illustrate the advantages of an ecologically based metric and to explore the experimental dataset, we defined a new variable, "effective predator density," based on a linear combination of the specialist and generalist density differentials in each experiment. Specialists were given a weight of 1, whereas generalists were given a weight of 1/50 (see analyses above). Weighting fish by their relative ecological impacts enabled us to put all studies (including those

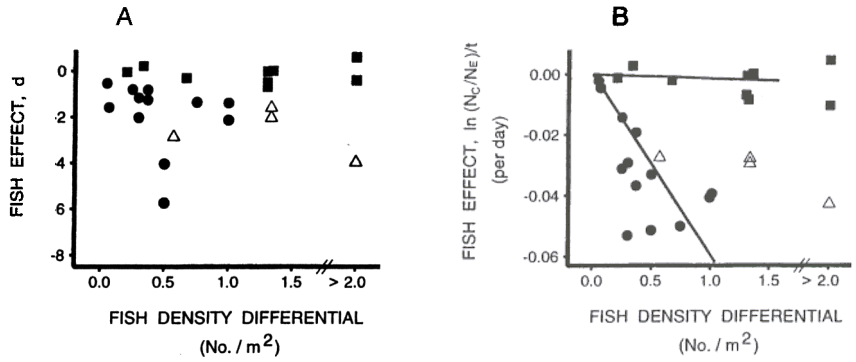

FIG. 3.—Meta-analyses of the effects of different groups of fish on snails using (*A*) d (eq. [1]) and (*B*) Δr (eq. [4]). The *X*-axis gives the difference in fish density between the two treatments. Each estimate was categorized by the type of fish whose density was manipulated: specialists (*circles*), generalists (*squares*), and mixtures of the two (*triangles*). A total of 10 experiments reported in nine publications (Crowder and Cooper 1982; Mittelbach 1988; Osenberg 1988; Bronmark et al. 1992; Martin et al. 1992; Osenberg et al. 1992; Bronmark 1994; Olson et al. 1995; Huckins 1996) were used. Most experiments yielded two estimates of effect size either because multiple densities of fish were used (thus, more than one density was compared to the "no fish" treatment) or an additional factor was crossed with the fish manipulation (thus, effect size was calculated for each level of the other factor). With one exception, data were provided by the primary investigators. For Crowder and Cooper (1982) snail densities were estimated from their figures 1 and 6; standard deviations, which were not available, were assumed equal to the standard deviation in total invertebrate density taken from their figure 1. Snail biomass per unit area was used in all calculations.

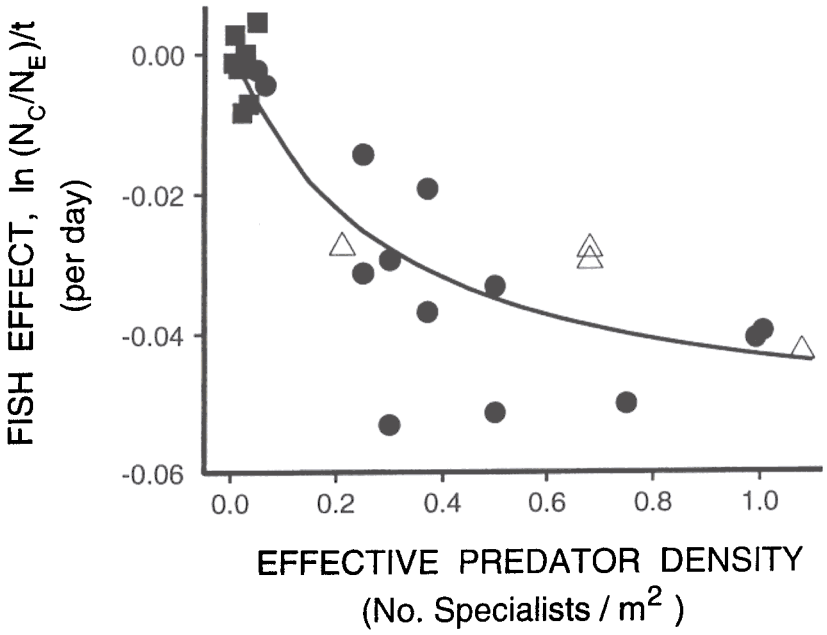

FIG. 4.—The relationship between the effect of fish on snails (Δr) and the effective predator density (i.e., total fish density expressed as the equivalent density of "specialists," which was obtained by a weighted sum of generalists [weight = 1/50] and specialists [weight = 1]). *Squares* = generalists; *circles* = specialists; *triangles* = mixtures of the two.

with mixtures of specialists and generalists) onto a common scale so that all results could be combined into a single relationship. The relationship between effect size (Δr) and effective predator density appeared nonlinear (fig. 4); therefore, we fit the data with a nonlinear function that had a zero intercept and reached an asymptote at high fish densities: $Y = b_1X/(1 + b_2X)$; $n = 26$, $r^2 = 0.78$, $b_1 = -0.186 \pm 0.120$, $b_2 = 3.32 \pm 3.28$ (estimates ± 95% CI). This represents a significant improvement over the linear model (with a zero intercept), which accounted for only 54% of the variance. The nonlinearity in this relationship suggested that the per capita effect of the predators decreased as predator density increased, indicating that our previous linear models may have underestimated fish effects at low densities (especially for specialists). This nonlinearity might arise because of interference among the predators or because of shifts in the size structure, species composition, or behavior of snails leading to a decrease in the vulnerability of survivors as fish density increases (e.g., Osenberg and Mittelbach 1996; Turner 1996, 1997). These ideas could be tested with additional data gleaned from the surveyed studies or by performing new experiments.

LINKING EFFECT SIZE TO BIOLOGICAL MODELS

Population-Level Responses

Under some conditions, we can more explicitly relate Δr (or $\Delta r/P$, or other selected metrics) to parameters or terms in ecological models and then relate the patterns observed across studies to changes in these parameters. For example, consider the previous examples of predator-prey interactions and assume that the effects of predators on prey dynamics are driven by mortality imposed by the predator and that these effects are measured before any indirect effects occur. In this situation, $\Delta r/P$ provides an estimate of the interaction strength of the predator on the prey (e.g., $\partial(dN/Ndt)/\partial P$, which is also the attack coefficient [a_{NP}] in the Lotka-Volterra predator-prey equations). In this situation, interaction strength is a subset of what we have termed *effect size* (see Laska and Wootton 1998). Quantifying interaction strength and its variation across systems and species is critical to the development and tests of many ecological hypotheses (Paine 1992; Laska and Wootton 1998). Meta-analysis, with an appropriate definition of effect size, will help facilitate these goals, but only if we are explicit about how our metrics relate to models and how (or if) such models are appropriate for the particular types of experiments that are being synthesized.

We have cautioned that the selection of a metric is a difficult task and that no single metric will be appropriate for all analyses. Timescale will be a critical element to consider in defining and interpreting effect-size metrics (e.g., fig. 2). Here we provide a more explicit consideration of the issues that we empirically addressed in figure 2. Consider a simple example in which the dynamics of a focal species (e.g., the prey) can be modeled as a linear combination of its own density and that of its predator:

$$dN/Ndt = r + a_{NN}N + a_{NP}P. \quad (6)$$

Following the removal of the predator, the prey population density will initially diverge (relative to the control) at a rate given by a_{NP}. If prey densities are sampled early in this trajectory, $\Delta r/P$ provides an estimate of a_{NP}, the direct effect of the predator (i.e., $\partial(dN/Ndt)/\partial P$). However, as time goes by, prey density will increase, thus increasing the strength of intraspecific competition relative to the control. Under these conditions, $\Delta r/P$ no longer estimates a_{NP}, and this metric can no longer be interpreted clearly (i.e., it confounds direct effects with changes in the intensity of intraspecific competition in the prey population). If the experiment's duration is sufficiently long that the prey species has reequilibrated, then we can take an alternate approach that focuses on a balance between opposing processes of predation and intraspecific competition rather than on the short-term responses themselves. Following reequilibration, the direct effect of the predator (i.e., $a_{NP}P$) must be completely compensated by increased intraspecific competition (i.e., $a_{NN}[N_E - N_C]$):

$$a_{NP}P = a_{NN}(N_E - N_C), \quad (7)$$

where N_E and N_C are equilibrial prey densities in the absence and presence of the predator, respectively. Thus, we can estimate:

$$\alpha_{NP} = a_{NP}/a_{NN} = (N_E - N_C)/P. \quad (8)$$

This metric is conceptually identical to the standardized competition coefficients of the Lotka-Volterra equations (but defined for a predator rather than a competitor). In other words, it provides a measure of the number of predators that are required to impose the same effect on the prey's per capita growth rate as one prey.

If this system had consisted of k additional species (effects of which could be modeled with linear terms), all of which went to a new equilibrium following the predator removal, then the direct effect of the predator on the prey would eventually be balanced by changes in the direct effect of the focal species on itself plus changes in the direct effects of the other k species on the focal species (Levine 1976; Lawlor 1979; Bender et al. 1984). In this case, $(N_E - N_C)/P$ (eq. 8) still estimates the per capita effect of the predator on the equilibrium prey density, but it now includes a multitude of indirect effects mediated through the k other species.

Appropriately chosen metrics should also help link studies of individual attributes with their consequences for population or community dynamics. For example, there have been numerous studies of the feeding behavior and prey preferences of predators. Because preference is a measure of the relative mortality rates imposed by a predator (Chesson 1978; Osenberg and Mittelbach 1989), these studies of individual behavior can be used to help interpret the results of field experiments that measure the effect of predators on the rate of change of prey populations or variation in the impacts of the predator on different prey species (Peckarsky et al. 1997). To illustrate this, we used Mittelbach's (1981) model of encounter rates between bluegill and invertebrates to generate a quantitative prediction for the impact of bluegill on odonates in Morin's (1984) experiment (see fig. 1). Bluegill were the most abundant fish in Morin's study; other fish species in the pond probably had similar effects on odonates (Olson et al. 1995). Based on Mittelbach's model, and assuming that encounter rates primarily determine feeding rates (Mittelbach and Osenberg 1994), per capita mortality rates should be proportional to odonate size raised to the 1.045 power (Mittelbach 1981). The observed effect (using Δr) was highly correlated with the predicted effect (using Mittelbach's model) ($r = 0.84$, $n = 7$, $P = .017$). However, the scaling of the observed effect to prey size (exponent of power function = 2.31 ± 1.64 [95% CI]) was twice as large as that predicted based on Mittelbach's laboratory studies (i.e., 1.045). Although the large confidence interval on the scaling parameter includes the estimate from Mittelbach's model, the results suggest that additional size-specific mechanisms might further accentuate fish effects on large odonate species (e.g., large prey might incur even greater mortality rates because of greater attack probabilities as well as greater encounter rates; Werner et al. 1983). Such insights could not be gained by use of d or any other metric lacking a biological basis.

Individual-Level Responses

In many cases ecological experiments focus specifically on individual-based responses, such as habitat use, diet composition, and individual size or growth rate. This is particularly common in studies of interspecific competition (Gurevitch et al. 1992) and physiological responses to environmental factors (e.g., Curtis 1996). In these cases, the above metrics (derived by consideration of population phenomenon) must be replaced by metrics that are tied to models describing the response of individuals. For example, consider the common situation where the response variable reported in a set of experimental studies is body size (e.g., in studies manipulating competitor densities; Gurevitch et al. 1992). A standard option might be to compare body size in the treatments; however, this approach focuses on a state variable, rather than a rate, and thus does not isolate a parameter in a dynamic model (see also Billick and Case 1994; Wootton 1994). Furthermore, the effect size based on the difference in body size between a treatment and control group would not be time invariant because the difference would, in most cases, increase through time. Instead of analyzing body size directly, the data might instead be used to estimate individual growth rates or growth curves. If a specific growth model were applicable to the studies, then treatment effects could be defined by changes in the parameters of the model. For example, assume that an organism grows according to the VonBertalanfy growth equation:

$$L_t = L_\infty(1 - e^{-g(t-t_0)}), \quad (9)$$

where L_t is the length of the organism at time t, L_∞ is the asymptotic maximum length, g is the growth constant, and t_0 a fitted constant that defines the time at which the organism had zero length. The parameters L_∞ and g might both be expected to vary as a function of treatments and environmental features of a study. If measurements of size were made at several times over the experiment, such that these parameters could be estimated, then a meta-analysis could be based on the effect of the treatment on these parameters. For example, possible metrics of effect size include $L_{\infty,E}/L_{\infty,C}$, which quantifies the effect of the treatment on asymptotic size relative to the control or $g_E - g_C$, which quantifies the absolute difference in the observed growth constants. Note that the raw data (i.e., individual sizes) are not directly analyzed in the meta-analysis; instead, the raw data must be used indirectly to estimate the parameter of interest (e.g., g or L_∞), which is then used to calculate the effect size (e.g., $g_E - g_C$). The link between the raw data and the framework for defining effect size emerged only after a model was specified. The development of such models, their link to experimental designs, and their relationship to biologically interpretable metrics of effect size should be a goal of future research.

CONCLUSIONS

Meta-analysis holds much promise for providing a rigorous, quantitative approach to the synthesis of experimental results. Its future applications and its successes, however, will depend on our ability to extract a suitable metric, common to a collection of studies. Although we are optimistic, extracting suitable metrics poses a difficult problem, one that we emphasize cannot be solved by

reference to generic metrics such as d or by reliance on a single biologically derived metric (e.g., Δr). Not all studies that address a particular process (such as effects of predation or competition) should be combined because they often address different questions, apply different approaches, or operate on vastly different temporal (or spatial) scales. Defining and justifying a particular metric is an important and difficult task that will often require specification of a biological model, which will (and should) lead to the exclusion of many potential studies from a meta-analysis. The future success of meta-analysis will likely be decided not by the number of studies that can be amassed but by the specificity with which metrics can be defined, studies can be selected, and questions can be resolved.

ACKNOWLEDGMENTS

Our ideas and presentation have benefited from the comments of many people, including S. Diehl, N. Gotelli, J. Gurevitch, K. Kratz, M. Laska, D. Peart, P. Petraitis, C. St. Mary, A. Underwood, members of the University of Florida Research Review in Ecology and Evolution evening seminar, and participants in the National Center for Ecological Analysis and Synthesis (NCEAS) workshop (see below). We are particularly grateful to C. Huckins, G. Mittelbach, and M. Olson for providing unpublished data from their studies; C. Bronmark and T. Martin for providing us with raw data from figures reported in their papers; and A. Stewart-Oaten for the derivation of the confidence interval for Δr (fig. 2). Support was provided through grants from the National Science Foundation (NSF): DEB-9596204/9528445 to C.W.O., DEB-9308375/9629473 to O.S., and DEB-9407591 to S.D.C. This work was also facilitated by a working group ("Meta-analysis, Interaction Strength and Effect Size: Application of Biological Models to the Synthesis of Experimental Data") supported by the NCEAS, which is funded by the NSF (DEB-94-21535), the University of California at Santa Barbara, and the state of California. Additional support was also provided for O.S. as an NCEAS Postdoctoral Associate.

LITERATURE CITED

Adler, F. R., and W. F. Morris. 1994. A general test for interaction modification. Ecology 75: 1552–1559.

Arnqvist, G., and D. Wooster. 1995. Meta-analysis—synthesizing research findings in ecology and evolution. Trends in Ecology & Evolution 10:236–240.

Bender, E. A., T. J. Case, and M. E. Gilpin. 1984. Perturbation experiments in community ecology: theory and practice. Ecology 65:1–13.

Billick, I., and T. J. Case. 1994. Higher order interactions in ecological communities: what are they and how can they be detected? Ecology 75:1529–1543.

Bronmark, C. 1994. Effects of two benthivorous fishes, tench and perch, on interactions in a freshwater, benthic food chain. Ecology 75:1818–1828.

Bronmark, C., S. P. Klosiewski, and R. A. Stein. 1992. Indirect effects of predation in a freshwater benthic food chain. Ecology 73:1662–1674.

Chesson, J. 1978. Measuring preference in selective predation. Ecology 59:211–215.

Cooper, S. D., S. J. Walde, and B. L. Peckarsky. 1990. Prey exchange rates and the impact of predators on prey populations in streams. Ecology 71:1503–1514.

Crowder, L. B., and W. E. Cooper. 1982. Habitat structural complexity and the interaction between bluegills and their prey. Ecology 63:1802–1813.

Curtis, P. S. 1996. A meta-analysis of leaf gas exchange and nitrogen in trees grown under elevated carbon dioxide. Plant, Cell and Environment 19:127–137.

Glass, G. V. 1976. Primary, secondary, and meta-analysis of research. Educational Researcher 5:3–8.

Gurevitch, J., and L. V. Hedges. 1993. Meta-analysis: combining the results of independent experiments. Pages 378–398 *in* S. M. Scheiner and J. Gurevitch, eds. Design and analysis of ecological experiments. Chapman & Hall, New York.

Gurevitch, J., L. L. Morrow, A. Wallace, and J. S. Walsh. 1992. A meta-analysis of competition in field experiments. American Naturalist 140:539–572.

Hedges, L. V., and I. Olkin. 1985. Statistical methods for meta-analysis. Academic Press, Orlando, Fla.

Huckins, C. J. F. 1996. Functional relationships between morphology, feeding performance, diet, and competitive ability in molluscivorous sunfish. Ph.D. diss. Michigan State University, East Lansing.

Laska, M. S., and J. T. Wootton. 1998. Theoretical concepts and empirical approaches to measuring interaction strength. Ecology 79 (in press).

Lawlor, L. R. 1979. Direct and indirect effects of n-species competition. Oecologia (Berlin) 43:355–364.

Levine, S. H. 1976. Competitive interactions in ecosystems. American Naturalist 110:903–910.

Martin, T. H., L. B. Crowder, C. F. Dumas, and J. M. Burkholder. 1992. Indirect effects of fish on macrophytes in Bays Mountain Lake: evidence for littoral trophic cascade. Oecologia (Berlin) 89:476–481.

Mittelbach, G. G. 1981. Foraging efficiency and body size: a study of optimal diet and habitat use by bluegills. Ecology 62:1370–1386.

———. 1988. Competition among refuging sunfishes and effects of fish density on littoral zone invertebrates. Ecology 69:614–623.

Mittelbach, G. G., and C. W. Osenberg. 1994. Using foraging theory to study trophic interactions. Pages 45–59 *in* D. J. Stouder, K. L. Fresh, and R. J. Feller, eds. Theory and application in fish feeding ecology. Belle W. Baruch Library in Marine Sciences, no. 18. University of South Carolina Press, Columbia.

Morin, P. J. 1984. The impact of fish exclusion on the abundance and species composition of larval odonates: results of short-term experiments in a North Carolina farm pond. Ecology 65: 53–60.

Olson, M. H., G. G. Mittelbach and C. W. Osenberg. 1995. Competition between predator and prey: resource-based mechanisms and implications for stage-structured interactions. Ecology 76: 1758–1771.

Osenberg, C. W. 1988. Body size and the interaction of fish predation and food limitation in a freshwater snail community. Ph.D. diss. Michigan State University, East Lansing.

Osenberg, C. W., and G. G. Mittelbach. 1989. The effects of body size on the predator-prey interaction between pumpkinseed sunfish and gastropods. Ecological Monographs 59:405–432.

———. 1996. The relative importance of resource limitation and predator limitation in food chains. Pages 134–148 *in* G. A. Polis and K. O. Winemiller, eds. Food webs: integration of patterns and dynamics. Chapman & Hall, New York.

Osenberg, C. W., G. G. Mittelbach, and P. C. Wainwright. 1992. Two-stage life histories in fish: the interaction between juvenile competition and adult performance. Ecology 73:255–267.

Paine, R. T. 1992. Food-web analysis through field measurement of per capita interaction strength. Nature (London) 355:73–75.

Peckarsky, B. L., S. D. Cooper, and A. R. McIntosh. 1997. Extrapolating from individuals to populations and communities in streams. Journal of the North American Benthological Society 16: 375–390.

Petraitis, P. S. 1998. How can we compare the importance of ecological processes if we never ask, "compared to what?" *In* W. J. Resetarits, Jr., and J. Bernardo, eds. Experimental ecology: issues and perpsectives. Oxford University Press, New York.

Quinn, J. F., and A. E. Dunham. 1983. On hypothesis testing in ecology and evolution. American Naturalist 122:602–617.

Rosenthal, R. 1994. Parametric measures of effect size. Pages 231–244 *in* H. Cooper and L. V. Hedges, eds. The handbook of research synthesis. Sage, New York.

Sarnelle, O. 1992. Nutrient enrichment and grazer effects on phytoplankton in lakes. Ecology 74: 551–560.

Tilman, D. 1989. Discussion: population dynamics and species interactions. Pages 89–100 *in* J. Roughgarden, ed. Perspectives in ecological theory. Princeton University Press, Princeton, N.J.

Tonhasca, A., and D. N. Byrne. 1994. The effects of crop diversification on herbivorous insects: a meta-analysis approach. Ecological Entomology 19:239–244.

Turner, A. M. 1996. Fresh-water snails alter habitat use in response to predation. Animal Behavior 51: 747–756.

———. 1997. Contrasting short-term and long-term effects of predation risk on consumer habitat use and resource dynamics. Behavioral Ecology 8:120–125.

VanderWerf, E. 1992. Lack's clutch size hypothesis: an examination of the evidence using meta-analysis. Ecology 73:1699–1705.

Werner, E. E., G. G. Mittelbach, D. J. Hall, and J. F. Gilliam. 1983. Experimental test of optimal habitat use in fish: the role of relative habitat profitability. Ecology 64:1525–1539.

Wooster, D. 1994. Predator impacts on stream benthic prey. Oecologia (Berlin) 99:7–15.

Wootton, J. T. 1994. Putting the pieces together: testing the independence of interactions among organisms. Ecology 75:1544–1551.

Associate Editor: Nicholas J. Gotelli