

Re-analysis of meta-analysis: support for the stress-gradient hypothesis

CHRISTOPHER J. LORTIE and RAGAN M. CALLAWAY*

Department of Biology, York University, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3, and *Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA

Summary

1 Using meta-analysis, Maestre *et al.* (2005, *Journal of Ecology*, **93**, 748–757) recently rejected the predictions of the stress-gradient hypothesis for arid and semi-arid environments in entirety. It was concluded that neither positive nor negative effects of neighbours increased with abiotic stress and that different theoretical models are now needed.

2 In light of this sweeping conclusion, we re-examined the analytical approach and explored the general synthetic power of meta-analysis.

3 Detailed statistical re-analyses demonstrated that some of the meta-analyses of Maestre *et al.* were robust. However, more rigorous data selection criteria, changing gradient lengths between studies and covariance in response effects did not support their original conclusions. Additionally, application of more rigorous data selection criteria did allow us to detect a significant and consistent positive effect of neighbours, which suggests that facilitation is important at many points along stress gradients.

4 Careful evaluation of the studies used by Maestre *et al.* also revealed serious limitations. Many studies included in the meta-analyses were not conducted along stress gradients, did not identify a stress gradient within the study, focused on invasive species or were not peer reviewed. Most importantly, however, gradient lengths were not quantified and appeared to differ dramatically among studies. This crucial source of variation was not accounted for statistically nor in the interpretations.

5 Meta-analyses are useful tools for synthesis and description but are inherently limited by the appropriateness of the data selected. Unfortunately, in this particular instance, the data available to and selected by Maestre *et al.* did not adequately test the stress-gradient hypothesis and cannot thus reject its value for understanding the organization of plant communities in arid systems.

6 The ecological implication of our synthesis is that meta-analytical summary statistics may not always tell the whole story. Alternative interpretations of differences in effect sizes (or lack thereof) are possible because studies will vary in their ability to test specific predictions of a hypothesis, and furthermore, a certain level of judgement is required to infer the relative importance of certain ideas to synthetic progress within a discipline.

Key-words: abiotic stress, competition, criteria for data selection, facilitation, gradients, meta-analysis, plant–plant interactions, progress, statistics, synthesis

Journal of Ecology (2006) **94**, 7–16

doi: 10.1111/j.1365-2745.2005.01066.x

Introduction

‘Stress’ is used loosely by ecologists to describe the relative abiotic circumstances affecting species and

communities (Grime 1977). Although not without controversy, understanding the effects of stress has been useful for predicting net plant–plant interactions at points on abiotic gradients, and thus for estimating the relative importance of factors that organize plant communities. The relationship between stress and plant interactions has been conceptually formalized in the

'stress-gradient hypothesis', which posits that net competitive effects are more important, or at least more intense (Brooker *et al.* 2005), in relatively benign, low-stress environments, whereas facilitative effects are more important in relatively harsh, high-stress environments (Bertness & Callaway 1994). Within this context, Maestre *et al.* (2005) used meta-analysis to explore quantitatively the generality of the stress-gradient hypothesis in semi-arid environments and concluded that the magnitude of the net effect provided by neighbours, either positive or negative, is not higher under high abiotic stress conditions in semi-arid environments. This conclusion is intriguing in light of frequent spatial associations reported between species in these systems (Arriaga *et al.* 1993; Cavieres *et al.* 2000, 2002; Arroyo *et al.* 2002; Armas & Pugnaire 2005), experiments on single target species (Greenlee & Callaway 1996; Pugnaire & Luque 2001; Lortie & Turkington 2002), large-scale analyses of whole communities (Tewksbury & Lloyd 2001) and another extensive meta-analysis in a semi-arid system (Gomez-Aparicio *et al.* 2004), all of which clearly indicate that facilitation increases in importance in the more stressful places in arid and semi-arid systems.

Admittedly, a cursory review of the literature suffers from many limitations, including statistical rigour. However, the meta-analysis by Gomez-Aparicio *et al.* (2004) on an integrated set of experiments at a diverse collection of study sites supported the stress-gradient hypothesis. This suggests to us that different meta-analyses, although still quantitative, are of course also not entirely free from potential constraints. Computational or interpretative limitations can include but are not limited to incomplete reporting, non-independence of effect sizes, biases, data-selection criteria or lack of sensitivity. Finally and perhaps most importantly, the selection criteria for inclusion in the meta-analysis can exclude compelling studies or alternatively include comprehensive sets of studies that only tangentially address a hypothesis. Differences in data-selection criteria can sometimes even lead to opposite conclusions within the same delimited research area.

A synthesis of the stress-gradient hypothesis is important because its relevance is assumed for arid and semi-arid systems. Furthermore, the general value of meta-analysis as a method of ecological synthesis also deserves deeper consideration. To do this, we applied previously established conceptual and quantitative criteria for meta-analyses to the recently published analysis of Maestre *et al.* (2005). Our re-analyses explored the following questions: (i) is a hypothesis for a particular system generally testable *post hoc* via meta-analysis, (ii) was the original meta-analysis robust or sensitive to selective study exclusions, (iii) were the papers chosen true tests of the stress-gradient hypothesis and (iv) do additional meta-analyses using alternative selection criteria or more appropriate studies alter the outcome?

Defining the conceptual limits of meta-analysis

As Maestre *et al.* (2005) recognized, the use of conceptual models to predict the conditionality of plant–plant interactions is a current 'theme' in ecology. Such themes are rapidly being explored with meta-analysis, a relatively new technique allowing quantitative assessment of trends within a set of studies. However, neither conceptual models nor meta-analysis, even though quantitative through comparison of effect sizes in the latter case, determine the 'predictability' of an idea. Although hypotheses themselves are not predictable, they can be used to make testable predictions. Hypotheses can be general, specific or perhaps incorrect. Because individual researchers choose from a subset of models and hypotheses for a particular system and then use the selected hypothesis to generate testable predictions, one limitation is the scope of the environments and predictions tested by the primary studies. The appropriateness or 'fit' of the model (the stress-gradient hypothesis) was already decided *a priori* by the primary researcher. The *post hoc* meta-analysis is thus testable only to the extent that the collection of studies both sampled the full range of conditions within an environment and explored a full set of testable predictions. This issue is an extension of the 'research bias' limitation of meta-analysis, in which researchers are likely to select organisms for study that will generate statistically significant results. We propose that a 'hypothesis bias' also functions through researchers selecting or generating hypotheses that will most likely be able to explain the patterns at hand (Bauchau 1997) and testing only those predictions that are directly relevant. Hypotheses are not selected at random from a pool of available hypotheses, nor is the pool of ideas we explore free from bias. As such, a test of the predictability in changes in plant–plant interactions with abiotic stress cannot be achieved through *post hoc* delineation of two levels of stress and comparison of relative effect sizes across studies at these arbitrary levels of stress.

Meta-analysis does offer us new synthetic possibilities by combining published studies for further analyses and suggesting whether or not 'the magnitude of the treatment effect observed in a particular experiment is subject to chance variation' (see the conceptual illustration of this approach in Fig. 1a). Hence, comparing multiple tests of a hypothesis could include many different criteria in order to evaluate the relative importance of the content of the hypothesis – including its predictability (*sensu* Popper 1934, and the traditional hypothetico-deductive method). However, other factors such as testability, generality, consistency, accuracy, precision, an economy of propositions, or whether or not an explanation applies to broad questions and provides general information about similar events should also be considered (Fig. 1b). We propose that meta-analyses are best used to determine either generality or consistency of a conceptual model within a clearly delineated and restricted region (probably much more limited than

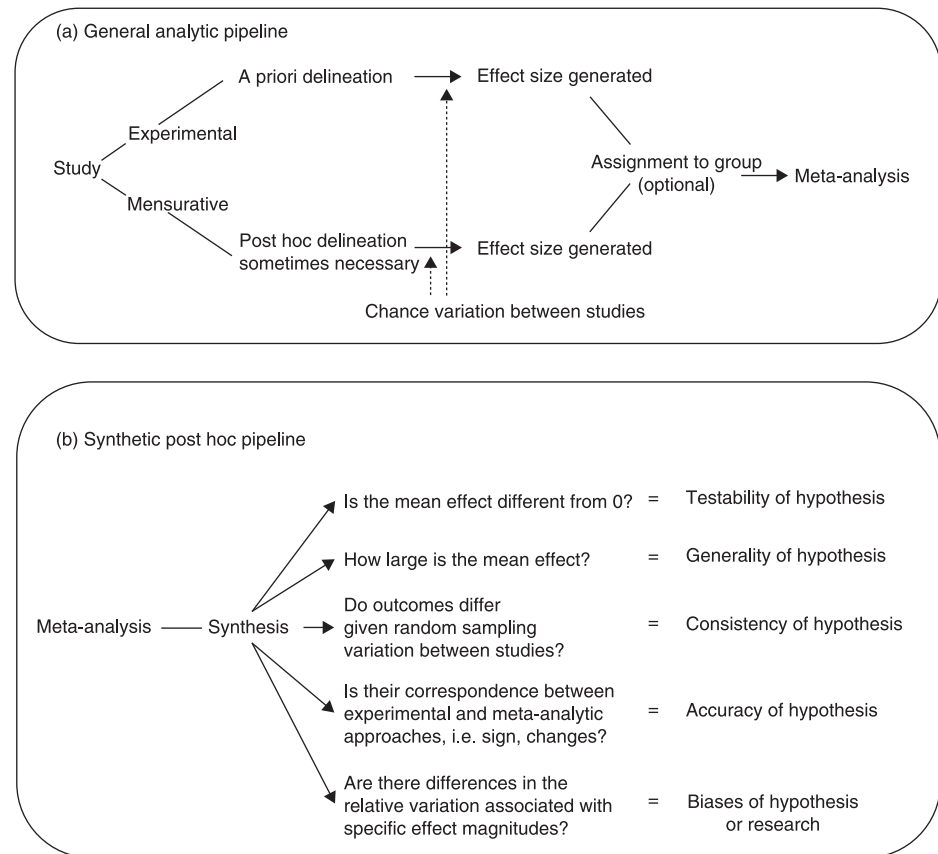


Fig. 1 Epistemological pipelines to consider when applying meta-analyses to synthesize the importance of a particular hypothesis: (a) the conventional approach to applying a meta-analytical comparison of effect sizes between studies with random variation influencing effect sizes between studies, and (b) different ways in which the general effect sizes can be compared and potentially used.

semi-arid systems) or domain (i.e. how important is the chance variation between related experiments?), but that predictability and perhaps testability be left to direct empirical within-study tests of an idea (Fig. 1b). A meta-analysis is not an experiment that can be used to reject null hypotheses but, instead, is a tool or more accurately a potential method of synthesis used *post hoc* to explore trends and generality without necessarily invoking falsification. It has been proposed as a method both to infer and to describe, and perhaps in some instances meta-analysis is best suited to description for the purpose of explaining variation directly ‘rather than seeking a global answer ...’. This is not to say that the quantitative comparisons generated through meta-analysis are not relevant to explore the value of an idea, simply that meta-analysis cannot be used to reject the applicability of an idea entirely as Maestre *et al.* (2005) have done in this instance for the stress-gradient hypothesis in arid environments. This is particularly true as meta-analyses are inherently limited to the design, scope and generalizability of the studies selected for inclusion.

A quantitative re-exploration of stress-gradient meta-analyses

In general, the analyses performed by Maestre *et al.* (2005) were technically sound, fully satisfying 10 of the

13 criteria proposed by Gates (2002) to assess the methodology of meta-analyses in ecology. Three criteria that could have been addressed more rigorously in the methodology of Maestre *et al.* (2005) included: (i) a more thorough exploration of the heterogeneity between groups (i.e. either the characteristics of the stress positions or the species through weighted regression techniques of continuous variables); (ii) an assessment of the ‘appropriateness’ of the gradients within the studies that were included in the final analyses (e.g. how different were the two ends of the continuum, was species assignment random, was the experimental design appropriate, and were different species tested?); and (iii) an assessment of the sensitivity of the meta-analyses (i.e. how important were common or large species and did multiple replicates of influential studies included alter the outcome of analyses?). In the first and second instances, further exploration of the importance of relevant ecological characteristics of participants and studies included in the meta-analyses would have been appropriate because tests of homogeneity often have low power and do not necessarily mean that important covariates within groups do not exist even when non-significant statistics are generated. It would have been extremely informative to test whether the ‘internal validity’ of the experiments themselves was sufficient; for instance, were the differences in stress in

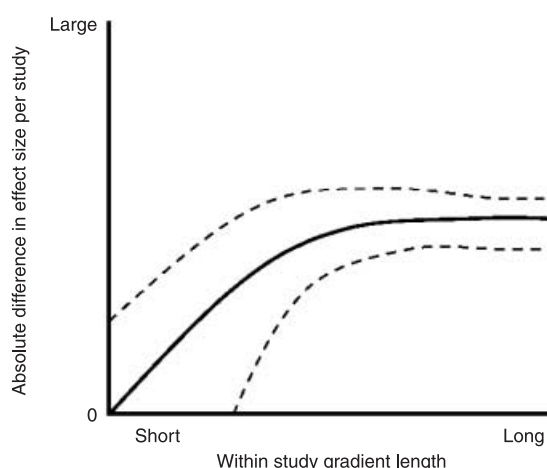


Fig. 2 An illustration of the predicted relationship between gradient length within a study and the absolute difference in mean effect size per study. The within-study gradient length could be calculated by taking the difference in the stress index for the two most extreme points along the gradient tested. The absolute difference in mean effect size is the difference in the effect sizes of a response variable at various points tested within a study. The dotted lines depict a representative 95% CI.

specific studies sufficient to test the stress-gradient hypothesis? In Fig. 2, studies that fall in the lower left of the plot have minimal ability to test this hypothesis and are more variable because only limited environmental variation is sampled. We explore the validity of specific studies below. Without having the continuous variables used by Maestre *et al.* (2005) to assign arbitrary 'low' vs. 'high' stress positions within each study or the mean plant size per species, we cannot quantitatively evaluate the efficacy of their meta-analyses for determining the consistency of the stress-gradient hypothesis as an organizing concept for all arid systems

comprising '3.9 billion ha worldwide' or whether it applies to local changes within a particular habitat.

Despite the inherent limitation of using meta-analysis to 'test' a hypothesis, we explored the conclusions of Maestre *et al.* (2005) by conducting the following additional statistical tests: repeated analysis of the random effect survival model as initially reported without multiple repetitions from within a study (to discount the potential influence of studies from the same site or researcher – between-study heterogeneity, Q_B , was not significant but results may have been overextended or highly sensitive to non-independence); repeated fit of the random effect survival model for each functional grouping of species; analysis of the coefficient of variations (CV) for survival effects to determine if relative variance is an important attribute to consider directly in meta-analyses; and finally an exploration of whether the magnitude of effect sizes or variance in survival is associated with the responses associated with growth responses (16 studies reported both variables). Meta-analyses were performed with the same application, Meta Win version 2.1.4 (Rosenberg *et al.* 2000), and all other statistics were done with JMP 5.1 (SAS 2003). All comparative differences are summarized briefly in Table 1 and explained in full below.

The importance of species and relative variation

Before proceeding to the crucial question of data-selection criteria, it is insightful to explore further the original meta-analyses. In light of the non-significant total heterogeneity of the initial survival model, Q_H , repeated analysis of each functional grouping independently did not generate significant differences in heterogeneity; however, not all species exhibited the change in sign of effect from positive to negative as one moved from low to high stress categories. For instance, the response of

Table 1 A summary of changes in the outcome and interpretation of modified meta-analyses from Maestre *et al.* Modifications were carried out sequentially in the order shown

Additional analyses relative to Maestre <i>et al.</i>	Outcome
Multiple species from single studies excluded (retaining random effects model).	Maestre <i>et al.</i> supported. No effect of stress, and overall effect sizes actually increased by 11% with exclusions.
Meta-analyses repeated separately for each functional grouping (retaining random effects model).	Maestre <i>et al.</i> supported. However, the change in sign with stress was not the same for each functional grouping.
Compared the relative variation (CV) of effect sizes by functional grouping and relative position on stress gradient.	Maestre <i>et al.</i> not supported. Similar variances at all points on the gradient detected which suggests that gradients did not sample different stress levels.
Growth as indicator of stress instead of precipitation also tested.	Maestre <i>et al.</i> not supported. Magnitude and variance in growth positively predicted survival effects. Strongly suggests research or hypothesis bias.
More rigorous data-selection criteria applied and meta-analysis repeated (retaining random effects model).	Two outcomes generated. (i) Maestre <i>et al.</i> supported. No change in the effect of changing stress on survival detected. However, alternative interpretation proposed, namely changes in gradient length probably important. (ii) Maestre <i>et al.</i> not supported. Effect of neighbours was now significantly different from 0, which suggests that facilitation is important.

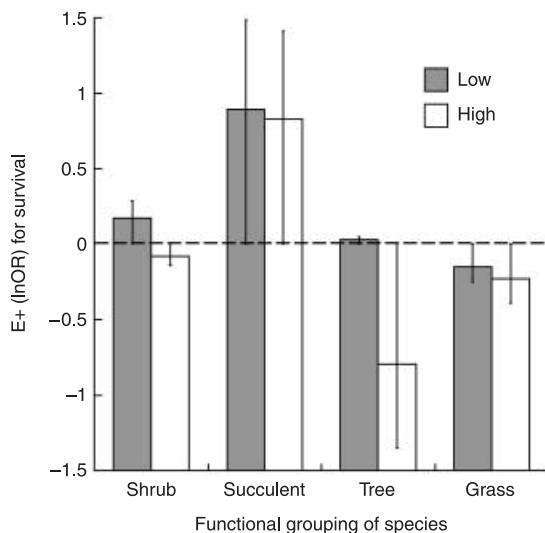


Fig. 3 The mean magnitude of survival effect sizes (E) analysed by functional grouping of species and relative stress in a random effects meta-analysis. The dotted line depicts zero net effects and error bars were generated via bootstrapping at the 95% CI. Data from Maestre *et al.* (2005).

succulents to neighbour removal was always strongly positive whereas grasses as targets always responded negatively (Fig. 3). Admittedly, the error associated with these responses often overlaps zero; however, functional group specificity is common in facilitative and competitive interactions (Callaway 1998), and the results in Fig. 3 indicate that the small number of studies within each group limits the ability (and probably the power) to test whether the change in interaction sign is a general ecological phenomenon within the studies (and thus species) selected by Maestre *et al.* (2005). More interestingly, the relative variation (CV) between survival effects did not differ by functional grouping nor by relative position on the stress gradient (ANOVA, all $P > 0.05$), indicating that the variance associated with either positive or negative plant–plant interactions was roughly equivalent at all points, both low and high, on the presumed stress gradient. Equivalent variance means that either plant–plant interactions are unimportant at all local points within arid systems (which would be at odds with the vast majority of experiments used in the meta-analyses) or that the delineation of two arbitrary points on gradients in each study was not sufficient (not ecologically different enough) to allow meaningful shifts between competition and facilitation. Without testing a factorial meta-analysis, which includes both the potential statistical interaction between species identity and variation in stress levels, it is reasonable to suggest that species-specific conclusions are premature at this time, i.e. until there are more than 26 studies with few representative species for each different functional grouping.

The difficulty in assigning stress levels

An index of aridity such as annual precipitation/potential evapotranspiration (P/PET) may be a good

surrogate for general available soil moisture in a particular place, but is not a valid surrogate for relative 'stress' across ecosystems and communities that vary in species composition (Lortie *et al.* 2004). In other words, what is stressful for one species may be ideal for another because of differences in adaptation to particular environments (Lortie *et al.* 2004). This is the reason Grime (1977) originally defined stress in terms of productivity. Using this view, stressful environments are defined as those in which producers are limited in their ability to convert energy to biomass (Grime 1977), and it is this approach that most ecologists adopt to examine stress across communities (Underwood 1989; Goldberg *et al.* 1999; Parker *et al.* 1999; Kammer & Mohl 2002). Furthermore, a potential measure of the stress gradient covered in a study would be a comparison of the difference in growth of the controls (i.e. plants growing without neighbours) between the low- and high-stress sites. As we could not evaluate relative differences in overall productivity as a surrogate for stress in the studies used by Maestre *et al.* (2005), we examined the relationship between the magnitude of effect sizes of growth responses (biomass in all cases, except one in which diameter was measured) and survival. In most instances, the magnitude and variance of effect sizes associated with growth positively predicted survival effect sizes or variance within each study (Fig. 4; regression analyses, all $P < 0.05$, $r^2 = 0.25$ in all fit instances). Although this might seem obvious due to potential collinearity between the two variables, it is important that the relationships were direct, i.e. within a study either both characters responded positively or both characters varied more strongly. We propose that variation due to 'research bias' or 'hypothesis bias' in selecting the study species, study sites, scales of manipulation, predictions directly tested and how 'stressful' the two extremes were classified may be generating the positive relationships presented by Maestre *et al.* (2005). For instance, species which responded most strongly to neighbour manipulation in terms of growth also responded strongly in terms of survival, suggesting that the research bias worked favourably in these studies, so that large effect sizes were generated consistently. Conversely, it appears that a group of other studies were less 'successful' in generating large effects and perhaps were not adequate experimental tests of this hypothesis. Within these studies, the variance in effect sizes generated per response variable was also positively correlated, thereby further supporting the 'hypothesis bias' in that the stress-gradient hypothesis clearly 'fit' some studies less well than others. This is a critical point because Maestre *et al.* (2005) implicitly assumed that all studies equally represented an adequate test of the stress-gradient hypothesis. Statistically this is correct as effect sizes are scale free and meta-analyses are weighted by the variance. Practically, however, it is clear that some studies were more variable, for whatever reason, thereby reducing the ability of meta-analysis to test the importance of this hypothesis. Direct examination of the frequency of the studies

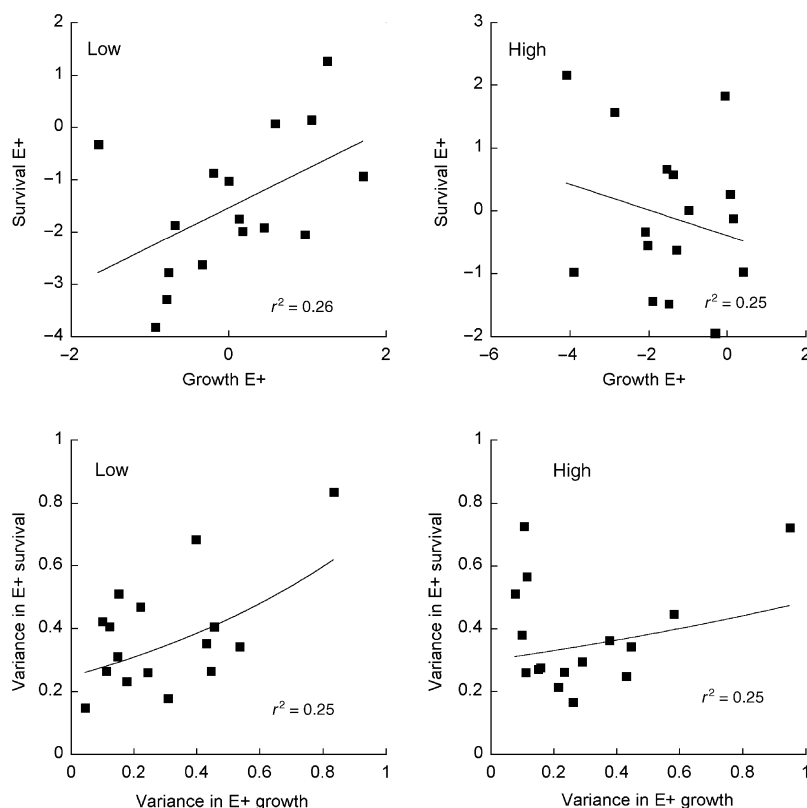


Fig. 4 Regression analyses of magnitude (or variance) in effect size of growth variables (primarily biomass) vs. the magnitude (or variance) of survival. Within gradient classifications of 'low' vs. 'high' generally refer to the relative available soil moisture at two different points tested in a study. The effect size measures $E+$ for survival and growth are $\ln(OR)$ and Hedge's d , respectively.

performed at different levels of aridity within this general semi-arid classification might help elucidate this variation.

Ecologically, it is also possible that larger effect sizes (and variances) in growth (i.e. biomass) are correlated with relative differences in primary productivity at the different endpoints of the sites. This potential interaction is not addressed in the meta-analyses but could have been easily incorporated and tested as an alternative classification for the low- vs. high-stress comparisons. It would also be interesting to determine whether the stress categorizations correlate with the productivity assignments within this set of studies. Even more proximately, it is likely that survival and growth are correlated owing to initial height effects with larger individual plants having increased survival, and these correlations should also be tested (Gomez-Aparicio *et al.* 2004). This relationship may be particularly important in stressful habitats because it is possible for large individual plants to occur in low-productivity habitats due to decreased potential competition and lower plant densities. Incorporation of mean plant size (initial or final) per species would be another informative covariate to screen in the meta-analyses and is probably important on the basis of the correlations we detected in subsequent analyses.

An assessment of the data-selection criteria

When we attempted to improve consistency by excluding multiple species from single large studies (i.e. potential

lack of independence), the outcome of the random effects survival model was not altered, and the grand means were very close to the absolute total effect sizes in all cases (i.e. an 11% increase in effect sizes with selective repeated exclusions). Hence, the survival meta-analyses of Maestre *et al.* (2005) were robust. However, this does not yet consider whether the studies chosen provide a fair and adequate test of the hypothesis.

We also explored the conclusions of Maestre *et al.* (2005) through a careful case-by-case examination of each of the studies we were able to access. We found many papers that clearly did not incorporate gradients of any sort, others that did not test different points on a stress gradient and others for which ascertaining gradients was very difficult (Table 2). These data did not constitute a 'clearly delineated and restricted region' for analysis. As an example, Maestre *et al.* (2005) used a north-facing slope as a low-stress endpoint and a south-facing slope as a high-stress endpoint from a study by Callaway *et al.* (1991). However, the paper clearly reports no difference in productivity among sites, and if a classification need be made, productivity tended to be higher on the south-facing slope. Detailed searches for explicit evidence for true gradients significantly reduced the number of appropriate studies (Table 2). Conversely, we searched the same literature as Maestre *et al.* (2005) (i.e. all major ecological journals) by visually inspecting the table of contents for each journal published within the last 10 years and

Table 2 A list of papers subject to additional inspection and meta-analysis. 'No gradient' refers to papers that did not adequately identify a gradient within the study system tested or that violated one of the initial Maestre *et al.* (2005) meta-analysis inclusion criteria. The list of 'gradient present and shift detected' papers refers to papers that adequately identified an abiotic gradient and showed a significant shift from negative to positive plant–plant interactions with increasing stress. The list of 'gradient present and no shift detected' refers to papers that adequately identified a gradient but did not detect a significant shift in plant–plant interactions. The latter list is of course still appropriate for meta-analysis. The full bibliography for the first and third columns is provided in Maestre *et al.* (2005). Citations listed in the second column are listed herein

Used by Maestre <i>et al.</i> but no gradient	Gradient present and shift from competition to facilitation detected	Gradient present and no shift detected
Arizaga & Ezcurra 2002	Castro <i>et al.</i> 2004	Anderson <i>et al.</i> 2001
Berlow <i>et al.</i> 2003	Chambers 2001	Casper 1996
Brooks 2003	Eckstein 2005	Donovan & Richards 2000
Brown & Archer 1999	García-Fayos & Gasque 2002	Goldberg <i>et al.</i> 2001
Burger & Louda 1995	Gomez-Aparicio <i>et al.</i> 2004	Gurevitch 1986
Callaway <i>et al.</i> 1991	Greenlee & Callaway 1996	Lortie & Turkington 2002
Facelli & Temby 2002	Gutierrez <i>et al.</i> 1993	McClaran & Bartolome 1989
Friedman & Orshan 1975	Halvorson & Patten 1975	Tielborger & Kadmon 1997
Friedman & Orshan 1977	Hastwell & Facelli 2003	
Gebauer <i>et al.</i> 2002	Hoffman 1996	
Lenz & Faceli 2003	Ibáñez & Schupp 2001	
Peltzer 2001	Kitzberger <i>et al.</i> 2000	
Tirado 2003	Maestre <i>et al.</i> 2001	
Van Auken & Lohstroh 1999	Maestre <i>et al.</i> 2002	
	Peek & Forseth 2003	
	Pugnaire & Luque 2001	
	Ratliffe <i>et al.</i> 1991	
	Schade <i>et al.</i> 2003	
	Tewksbury & Lloyd 2001	
	Valiente-Banuet & Ezcurra 1991	

generated a very different list of studies appropriate for meta-analyses on the stress-gradient hypothesis, with at least 10 additional studies (including the crucial paper by Gomez-Aparicio *et al.* 2004). This data-selection assessment of meta-analysis further emphasizes our conclusion that inherent difficulties in defining stress levels *post hoc* and accurate within-study identification of stress are critical for a more effective synthesis of the value of this hypothesis. Interestingly, application of more rigorous data-selection criteria on the original collection of studies detected a facilitative effect of neighbours (Table 1, final summary point). This demonstrates that even with low sample sizes, data-selection criteria are a critical consideration.

The importance of mixed gradients

In light of the potential biases in searching for and selecting which data to include in the original meta-analyses, we examined the potential of meta-analyses for testing the stress-gradient hypothesis under the most favourable circumstances possible. We repeated the survival meta-analysis using the subset of Maestre *et al.*'s (2005) studies that showed a statistically significant switch in net interactions from competition to facilitation with stress within each respective study system. We chose to test survival because it is probably the most important response variable that can be influenced by facilitation, and because it was the largest set

of suitable cases available in the original meta-analyses. The studies we selected were those papers listed in Table 2 that showed a significant switch from positive to negative interactions and were also included in the original meta-analysis (i.e. all the papers that overlapped between our list of appropriate papers and those of Maestre *et al.*). A total of 52 suitable cases were generated (hence approximately a 50% reduction from the original meta-analysis), and we fit the same random-effects meta-analysis model with confidence intervals estimated using bootstrapping procedures (and with 9999 iterations) to these cases. Despite the fact that each study independently detected a switch in net interactions with stress level and supported the stress-gradient hypothesis, the meta-analysis detected no significant difference between low and high stress on the effect of neighbours on the survival of target species (Fig. 5). This unequivocally demonstrates that variation in gradient lengths between studies introduces variation into the meta-analysis that renders it incapable of detecting significant differences between low- and high-stress sites across studies, and that the arbitrary classification *post hoc* of low vs. high stress does not work without accounting for potential statistical interaction effects. Inclusion of the studies listed in Table 2 that had an adequate gradient but did not detect a significant switch from negative to positive also did not change the outcome of our meta-analysis (or the assumptions and results reported below). Similar to

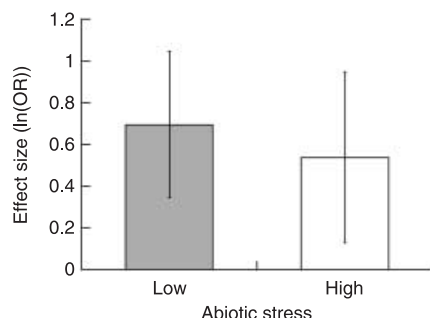


Fig. 5 A random effects model for survival in the presence of neighbours under low and high stress for studies that detected a significant switch in net interactions. A total of 52 cases were tested. Mean and bootstrapped confidence intervals were reported.

the initial Maestre *et al.* (2005) survival analysis, we checked all assumptions. There was no evidence of bias in reporting results within this reduced set of studies (funnel plots and weighted histograms were checked), there was no significant difference between the size of effects by abiotic stress ($Q_B = 0.31$, d.f. = 1, $P = 0.58$) and overall heterogeneity of the model was not significant ($Q_H = 43.2$, d.f. = 51, $P = 0.77$). Furthermore, fail-safe tests indicated that a very large number of additional studies would be necessary to change this meta-analytical result (Rosenthal's method: 203 studies). Hence, reduction in the number of cases was not the issue, but rather the fact that meta-analysis, at least in this instance, could not effectively test for relative differences between two undefined and arbitrary classifications of stress. We propose that changes in gradient length between studies and failure of *post hoc* classification to identify stress preclude any meaningful test of the stress-gradient hypothesis. As an aside on the general prevalence of facilitation in 'stressful' environments, this survival meta-analysis did indicate that target species had significantly increased survival (i.e. means greater than 0 at 95% CI) in the presence of neighbours at both stress levels (Fig. 5), which suggests

that facilitation is more important than competition at many points within semi-arid ecosystems.

Ecological lessons of meta-analyses

Meta-analyses are not experiments. Conclusions must be drawn cautiously at best and alternative interpretations should include an assessment of the adequacy of the experimental data. In this particular instance, we explored the generality of the meta-analyses testing the stress-gradient hypothesis for arid systems and revisited the primary conclusions. We found that the meta-analyses were technically sound but poorly applied and that the interpretations overextended the importance of non-significant mean effect sizes at higher levels of stress. Furthermore, we interpreted these trends as evidence for equivalency of the importance of facilitative interactions to those of competition, at least in the arid systems within this data set, and consistency of the finding that large effect sizes of one variable may be related to other covariates, including productivity. The general value and utility of meta-analysis as a powerful and appropriate method of synthesis is clear, but a more broad approach to interpreting differences in effect sizes as a test of a conceptual model is encouraged [see Table 3 and Gates (2002) for criteria]. A more rigorous understanding of the specific studies included in the analysis would have better contributed to accurate interpretations.

This of course leads to the intriguing consideration of when to reject a conceptual model (if at all). Maestre *et al.* (2005) view a 'contrasting body of evidence' as grounds for suggesting that 'the models proposed so far may not be generally applicable.' Although this seems to be a reasonable assumption for the evaluation of a model, it is also possible that the experiments are not adequate tests of the model (i.e. have poor 'fit' due to within-study variability), that although each study might, for instance, include some measure of stress, the experiment itself was not designed to test the hypothesis in question, that changing species and gradients (and

Table 3 Recommendations for meta-analyses. A brief list of recommendations based upon our exploration and consideration of the meta-analysis data presented by Maestre *et al.* (2005)

1. The statistical package MetaWin is capable of fitting mixed models and continuous variables. As such, instead of generating rules for assigning positions on a gradient to low or high, use the actual index of aridity, the salinity for the site or, as appropriate, some measure of primary productivity. This would provide for a much richer analysis and broaden the scope of the conclusions, e.g. does the mean index of stress for the sites predict the magnitude of effect expressed in the presence/absence of neighbours?
2. Just as primary experimental studies should report sample sizes, the effect size data tables should also include the 'n' used within each study to calculate effect size so as to facilitate additional meta-analyses or explorations of the data set in potentially novel ways.
3. Test alternative classifications or assignments of sites to explore how sensitive the results might be to these *post hoc* classifications.
4. Repeat the analyses without multiple entries per study by the same authors to ensure that the trends detected are truly general in the sense that they apply to an entire environment and not just to certain types of site (i.e. test for independence).
5. Report the fail-safe numbers as an indication of robustness, i.e. how many studies would it take to overturn the current result.
6. Explicitly identify which aspect of synthesis is being tested (*sensu* Ford 2000). For instance, is consistency, generality or some other aspect of importance being explored through these meta-analytical comparisons? Predictability of a hypothesis is not the most appropriate nor sole attribute to consider in meta-analyses.

gradient lengths) concurrently might be important, or that the model cannot be used to explain patterns explored at the scale of the meta-analytical comparisons. For instance, the stress-gradient hypothesis predicts that increasing stress shifts net interactions toward a more frequent positive net outcome (Bertness & Callaway 1994), but is the most appropriate scale within-study local comparisons or across-study broad comparisons? Without controlling for differences between local gradients in overall moisture, it has been shown that topographically driven differences in water stress can completely mask significant facilitative effects occurring at the landscape level, with very large differences in apparent facilitation at xeric sites and competition at mesic sites (Tewksbury & Lloyd 2001). As such, a meta-analytical comparison might explore whether the most extreme end of the abiotic gradient tested within each study generated effect sizes consistently different from zero instead of comparing the low- vs. high-sites across studies or alternatively control for large-scale differences in moisture and test each point independently. The latter approach removes the second level of hypothesis biasing in a meta-analysis by eliminating the need to assign each site to a low or high category because in reality the sites within studies are likely to fall along a larger-scale moisture gradient.

Furthermore, a statistical gradient effect using the current meta-analytical approach may also be due to differences between plasticities or responsiveness of target species and not to stress (similar to the concern of Gomez-Aparicio *et al.* 2004). To correct for these potentially confounding effects, a reasonable approach is to check the response of more than one species to manipulation on a single gradient (thereby controlling gradient length) or to compare the response of a single species to different gradients repeatedly (thereby controlling species). Under this alternative scenario, we might instead interpret a lack of increased relative effect sizes at higher stress levels as evidence for species effects, poor fit of the experimental tests to the hypothesis, or significant 'chance' variation between experiments in arid environments due to different local stressors or more importantly variable year-to-year effects, i.e. wet vs. dry years shifting the relative balance between competition and facilitation (Greenlee & Callaway 1996; Tielborger & Kadmon 2000; Gomez-Aparicio *et al.* 2004). The importance of positive interactions in arid systems has been repeatedly demonstrated experimentally and are thus 'real' in some sense; however, without a reasonable collection of studies that experimentally test for the shift from positive to negative interactions along sufficiently different gradient endpoints within a study (and preferably control for species effects, local conditions, gradient length and year-to-year variability), it is difficult to accept that a meta-analysis could discount the stress-gradient hypothesis entirely. More appropriately of course, a factorial meta-analysis that includes interaction effects would be preferable because interaction effects, with year or gradient length

for instance, may not be apparent by comparing the effects at each level independently (Gurevitch *et al.* 2000). To conclude, Maestre *et al.* (2005) have not provided acceptable synthetic evidence for the rejection of the stress-gradient hypothesis, and our reconsideration of the arid and semi-arid literature strongly suggests that the number of studies which adequately sample different points on gradients is relatively limited at this time.

References

- Armas, C. & Pugnaire, F. (2005) Plant interactions govern population dynamics in a semi-arid plant community. *Journal of Ecology*, **93**, 978–989.
- Arriaga, L., Maya, Y., Daiz, S. & Cancino, J. (1993) Association between cacti and nurse perennials in a heterogeneous dry forest in northwestern Mexico. *Journal of Vegetation Science*, **4**, 349–356.
- Arroyo, M.T.K., Cavieres, L., Penaloza, A. & Arroyo-Kalin, M.A. (2002) Positive associations between the cushion plant *Azorella monantha* (Apiaceae) and alpine plant species in the Chilean Patagonian Andes. *Plant Ecology*, **169**, 121–129.
- Bauchau, V. (1997) Is there a file drawer problem in ecological research? *Oikos*, **79**, 407–409.
- Bertness, M.D. & Callaway, R. (1994) Positive interactions in communities. *Trends in Ecology and Evolution*, **9**, 191–193.
- Brooker, R.W., Kikvidze, Z., Pugnaire, F., Callaway, R.M., Choler, P., Lortie, C.J. & Michalet, R. (2005) The importance of importance. *Oikos*, **109**, 63–70.
- Callaway, R.M. (1998) Are positive interactions species-specific? *Oikos*, **82**, 202–207.
- Callaway, R.M., Nadkarni, N.M. & Mahall, B.E. (1991) Facilitation and interference of *Quercus douglasii* on understorey productivity in central California. *Ecology*, **72**, 1484–1499.
- Castro, J., Zamora, R., Hódar, J.A. & Gómez, J.M. (2004) Seedling establishment of a boreal tree species (*Pinus sylvestris*) at its southernmost distribution limit: consequences of being in a marginal Mediterranean habitat. *Journal of Ecology*, **92**, 266–277.
- Cavieres, L., Arroyo, M.T.K., Penaloza, A., Molina-Montenegro, M. & Torres, C. (2002) Nurse effect of *Bolax gummiifera* cushion plants in the alpine vegetation of the Chilean Patagonia Andes. *Journal of Vegetation Science*, **13**, 547–554.
- Cavieres, L., Penaloza, A. & Arroyo, M.T.K. (2000) Altitudinal vegetation belts in the high-Andes of central Chile (33°S). *Revista Chilena de Historia Natural*, **73**, 331–334.
- Chambers, J.C. (2001) *Pinus monophylla* establishment in an expanding pinon-juniper woodland: environmental conditions, facilitation and interacting factors. *Journal of Vegetation Science*, **12**, 27–40.
- Eckstein, L.R. (2005) Differential effects of interspecific interactions and water availability on survival, growth and fecundity of three congeneric grassland herbs. *New Phytologist*, **167**, 1–11.
- Ford, E.D. (2000) *Scientific Method for Ecological Research*. Cambridge University Press, Cambridge.
- García-Fayos, P. & Gasque, M. (2002) Consequences of a severe drought on spatial patterns of woody plants in a two-phase mosaic steppe of *Stipa tenacissima* L. *Journal of Arid Environments*, **52**, 199–208.
- Gates, S. (2002) Review of methodology of quantitative reviews using meta-analysis in ecology. *Journal of Animal Ecology*, **71**, 547–557.
- Goldberg, D.E., Rajaniemi, T., Gurevitch, J. & Stewart-Oaten, A. (1999) Empirical approaches to quantifying interaction

- intensity: competition and facilitation along productivity gradients. *Ecology*, **80**, 1118–1131.
- Gomez-Aparicio, L., Zamora, R., Gomez, J.M., Hodar, J.A., Castro, J. & Baraza, E. (2004) Applying plant facilitation to forest restoration: a meta-analysis of the use of shrubs as nurse plants. *Ecological Applications*, **14**, 1128–1138.
- Greenlee, J.T. & Callaway, R.M. (1996) Abiotic stress and the relative importance of interference and facilitation in montane bunchgrass communities in western Montana. *American Naturalist*, **148**, 386–396.
- Grime, J.P. (1977) Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *American Naturalist*, **111**, 1169–1194.
- Gurevitch, J., Morrison, J.A. & Hedges, L.V. (2000) The interaction between competition and predation: a meta-analysis of field experiments. *American Naturalist*, **155**, 435–455.
- Gutierrez, J.R., Meserve, P.L., Contreras, L.C., Vasquez, H. & Jaksic, F.M. (1993) Spatial distribution of soil nutrients and ephemeral plants underneath and outside the canopy of *Portleria chilensis* shrubs (Zygophyllaceae) in arid coastal Chile. *Oecologia*, **95**, 347–352.
- Halvorson, W.L. & Patten, D.T. (1975) Productivity and flowering of winter ephemerals in relation to Sonoran Desert shrubs. *American Midland Naturalist*, **93**, 311–319.
- Hastwell, G.T. & Facelli, J.M. (2003) Differing effects of shade-induced facilitation on growth and survival during the establishment of a chenopod shrub. *Journal of Ecology*, **91**, 941–950.
- Hoffman, W.A. (1996) The effects of fire and cover on seedling establishment in a neotropical savanna. *Journal of Ecology*, **84**, 383–393.
- Ibáñez, I. & Schupp, E.W. (2001) Positive and negative interactions between environmental conditions affecting *Cercocarpus ledifolius* seedling survival. *Oecologia*, **129**, 624–628.
- Kammer, P.M. & Mohl, A. (2002) Factors controlling species richness in alpine plant communities: an assessment of the importance of stress and disturbance. *Arctic, Antarctic, and Alpine Research*, **34**, 398–407.
- Kitzberger, T., Steinaker, D.F. & Veblen, T.T. (2000) Effects of climatic variability on facilitation of tree establishment in northern patagonia. *Ecology*, **81**, 1914–1924.
- Lortie, C.J., Brooker, R.W., Kikvidze, Z. & Callaway, R.M. (2004) The value of stress and limitation in an imperfect world: a reply to Korner. *Journal of Vegetation Science*, **15**, 577–580.
- Lortie, C.J. & Turkington, R. (2002) The facilitative effects by seeds and seedlings on emergence from the seed bank of a desert annual plant community. *Ecoscience*, **9**, 106–111.
- Maestre, F.T., Bautista, S., Cortina, J. & Bellot, J. (2001) Potential for using facilitation by grasses to establish shrubs on a semiarid degraded steppe. *Ecological Applications*, **11**, 1641–1655.
- Maestre, F.T. (2002) *La restauración de la cubierta vegetal en función del patrón espacial de factores bióticos y abióticos*. PhD Thesis, Universidad de Alicante.
- Maestre, F.T., Valladares, F. & Reynolds, J.F. (2005) Is the change of plant–plant interactions with abiotic stress predictable? A meta-analysis of field results in arid environments. *Journal of Ecology*, **93**, 748–757.
- Parker, E.D., Forbes, V.E., Ritter, C., Barata, C., Admiraal, W., Levin, L., Loeschke, V., Lyytikäinen-Saarenmaa, P., Høgh-Jensen, H., Calow, P. & Ripley, B.J. (1999) Stress in ecological systems. *Oikos*, **86**, 179–184.
- Peek, M.S. & Forseth, I.N. (2003) Microhabitat dependent responses to resource pulses in the aridland perennial, *Cryptantha flava*. *Journal of Ecology*, **91**, 457–466.
- Popper, K.R. (1934) *The Logic of Scientific Discovery*, New edition (March 29, 2002) edn. Routledge, New York.
- Pugnaire, F.I. & Luque, M.T. (2001) Changes in plant interactions along a gradient of environmental stress. *Oikos*, **93**, 42–49.
- Ratcliff, R.D., Duncan, D.A. & Westfall, S.E. (1991) California oak–woodland overstory species affect herbage understory: management implications. *Journal of Range Management*, **44**, 306–310.
- Rosenberg, M.S., Adams, D.C. & Gurevitch, J. (2000) *Meta Win. Statistical Software for Meta-Analysis*. Version 2.1.4. Sinauer Associates, Inc, Sunderland, MA.
- SAS (2003) *Jmp 5.1.1*. SAS Institute Inc, Cary, NC.
- Schade, J.D., Sponseller, R., Collins, S.L. & Stiles, A. (2003) The influence of Mesquite canopies on understory vegetation: effects of landscape position. *Journal of Vegetation Science*, **14**, 743–750.
- Tewksbury, J.J. & Lloyd, J.D. (2001) Positive interactions under nurse-plants: spatial scale, stress gradients and benefactor size. *Oecologia*, **127**, 425–434.
- Tielborger, K. & Kadmon, R. (2000) Temporal environmental variation tips the balance between facilitation and interference in desert plants. *Ecology*, **81**, 1544–1553.
- Underwood, A.J. (1989) The analysis of stress in natural populations. *Biological Journal of the Linnean Society*, **37**, 51–78.
- Valiente-Banuet, A. & Ezcurra, E. (1991) Shade as a cause of the association between the cactus *Neobuxbaumia tetetzo* and the nurse plant *Mimosa luisana* in the Tehuacan Valley, Mexico. *Journal of Ecology*, **79**, 961–971.

Received 14 May 2005

revision accepted 28 July 2005

Handling Editor: David Gibson