Journal of Animal Ecology 2002
71, 547–557

ESSAY REVIEW

Review of methodology of quantitative reviews using meta-analysis in ecology

SIMON GATES
National Perinatal Epidemiology Unit, Institute of Health Sciences, Old Road, Oxford OX3 7LF, UK

Summary

1. Statistical methods for combination of independent results have been well publicized in the ecological literature, and have begun to be used for reviewing research. They provide a considerable advance in scientific rigour over traditional narrative or 'vote-counting' reviews.
2. However, other methodological developments for research synthesis have not yet been widely adopted.
3. This review briefly summarizes some of the techniques used for carrying out rigorous reviewing and synthesis of results in medical science, and surveys techniques used by ecological meta-analyses.
4. Many of the methods used to reduce bias and enhance the accuracy, reliability and usefulness of reviews in medical science have not yet been widely used by ecologists.
5. The quality of ecological reviews could be improved by adoption of some of these methods, such as specifying the methods used for literature searching, stating the types of study combined in the review and the strength of evidence they provide, presenting results as a point estimate with a confidence interval, investigating bias in selection of studies using funnel plots, making a clear distinction between the main analysis and subsidiary analyses and interpreting the latter with caution, and performing sensitivity analyses.

Key-words: meta-analysis, methodology, systematic review.

Journal of Animal Ecology (2002) **71**, 547–557

Introduction

Reviewing of research has an important place in scientific progress, as it is the means by which the generality of the results of studies can be assessed. Traditionally, reviewing has been done either by narrative reviews, where results are simply described, and a consensus emerges from the description, or by 'vote counting', where the number of statistically significant results for and against a hypothesis are counted. Both of these methods have potential for serious bias, and may yield misleading conclusions. To avoid these biases, the statistical techniques of meta-analysis have been developed. These allow formal combination of the results of independent studies. The effect size found in each individual study can be considered an independent estimate of the underlying true effect size, subject to random variation. Thus, a better estimate of the true effect size can be obtained by combining the estimates from independent experiments. Whether the result of each study is 'statistically significant' does not matter; all studies contribute to the overall estimate of the treatment effect. In the combination of results, results from more precise studies are usually given more weight, so that they have a greater influence on the overall estimate.

Meta-analysis overcomes some of the problems with narrative reviews, but other potential sources of bias remain. Bias in the selection of studies for the review (for example, by including only studies supporting a hypothesis), or poor methodology in primary studies, leading to bias in their results, may still render the results of a meta-analysis misleading. Hence it is important to avoid bias as far as possible in the conduct of the review.

The science of reviewing and synthesizing research using meta-analysis has developed very rapidly in medicine and health services research in recent years. The term 'systematic review' is now generally used for a rigorously conducted quantitative review, to emphasize the distinction between the scientific investigation

Correspondence: Simon Gates, National Perinatal Epidemiology Unit, Institute of Health Sciences, Old Road, Oxford OX3 7LF. E-mail: simon.gates@perinat.ox.ac.uk

and the statistical methods for combination of results. A systematic review may or may not involve meta-analysis, and meta-analyses need not be confined to systematic reviews. Systematic reviews of healthcare research are co-ordinated by the Cochrane Collaboration, an international network of individuals and institutions that promotes systematic reviewing of health care research, using a set of standard methods. The reviews are published electronically and periodically updated (Chalmers & Haynes 1995; see also the Cochrane Collaboration website http://www.cochrane.org). Cochrane reviews have become a cornerstone of the evidence on which medical practice is based.

In recent years, meta-analysis has begun to be used in ecology. As in medical research, reviews are fundamental to ecology. Ecologists seek to find general patterns that will explain results across different taxa or different environments, and as each study can deal with only one or a few species or situations, some form of quantitative combination of results is necessary to draw general conclusions. Meta-analysis provides a means of formally combining results, and estimating a general result. Moreover, it increases the statistical power available to test hypotheses, and allows the possibility of testing for differences in response between groups of organisms or different ecological settings. For these reasons meta-analysis will probably increase in importance in ecology in the future.

The advantages of statistical combination of independent results and the methods for carrying out such analyses have been reviewed in the ecological literature (Gurevitch *et al*. 1992; Gurevitch & Hedges 1993; Fernandez-Duque & Valeggia 1994; Arnqvist & Wooster 1995) and there is a growing body of literature on techniques for dealing with the peculiar problems of ecological data (e.g. Hedges, Gurevitch & Curtis 1999; Adams, Gurevitch & Rosenberg 1997). However, the wider developments in the science of reviewing research have not yet been so widely disseminated. This review aims: (1) to provide a summary of the procedures used for conducting and reporting systematic reviews in medicine; (2) to describe the methods currently being used by ecological reviews which use meta-analytical techniques, and to compare their methodology with that of medical reviews; and (3) to highlight areas where adoption of the methods of medical systematic reviews could improve the quality of ecological research synthesis. Reviews of methodology have played a very important role in improving the scientific quality of reviews in healthcare research and providing better evidence for clinical practice (Sacks *et al*. 1987; Mulrow 1987; McAlister *et al*. 1999).

Methods used in medical systematic reviews

PROTOCOL

The protocol for a systematic review should specify the criteria that will be used for deciding whether studies should be included, how their quality will be assessed, what data will be extracted, what comparisons will be made and the statistical methods that will be used. In particular, any 'subgroup analyses' (i.e. separate analyses of subgroups of studies or experimental subjects) should be specified in the protocol, to help to reduce problems of Type I errors caused by multiple unplanned comparisons. For medical reviews published by the Cochrane Collaboration, the protocol is peer reviewed and published, to allow more feedback to the authors before the review is conducted.

TYPES OF STUDY INCLUDED

Statistical combination of results can be applied to many types of study design. The most reliable systematic reviews in medicine are those based on 'randomized controlled trials' (RCTs: controlled manipulative experiments in which subjects are allocated to treatments at random), which should, if properly conducted, exclude any possibility of bias. However, where data from RCTs do not exist, some reviews have combined data from observational studies, in which individuals were not assigned to experimental treatments at random, but by the preference of the patient or clinician or some other (usually unknown) process. Because non-random allocation can often yield groups that differ in ways that affect the response variables being measured, their results are much more likely to be biased than those of randomized studies (Kunz & Oxman 1998). Hence a meta-analysis of observational studies is also much more likely to be biased. For example, meta-analysis of RCTs suggested that increased intake of beta carotene brought a small increase in the risk of death, but analysis of observational studies suggested a considerable reduction (Egger, Schneider & Davey Smith 1998). Where meta-analysis of non-randomized data is necessary, this should be made explicit in the review and the results should be interpreted cautiously.

LOCATION AND SELECTION OF RELEVANT STUDIES

Ideally, a review should include all of the relevant primary research that has been carried out, no matter when or where. The problem for the reviewer is to locate all this research, so that the analysis will be unbiased. If studies' results influence how likely they are to be included in the review, the results of the meta-analysis may be biased.

Locating studies has been aided by the development of computer databases, but these alone do not provide a comprehensive search (Dickersin, Schere & Lefebvre 1995). For example, Knipschild (1995) describes searching for studies of the effects of Vitamin C on the common cold: using the Medline database, 22 studies were found, but by checking the references of these studies and hand searching through conference abstracts

and non-computerized databases, another 39 studies were located.

Criteria for inclusion of studies in the review should be drawn up in advance and the finished review should contain lists of studies that were included or considered for inclusion but rejected.

ASSESSMENT OF BIAS

Publication bias, or the 'file drawer problem', has long been recognized (Rosenthal 1979); studies that show large and statistically significant effects may be more likely to be published than those that show no difference. Studies agreeing with currently fashionable hypotheses may also be more likely to be published (Simmons *et al*. 1999). Other biases may also operate; studies reporting statistically significant results may be more likely to be published in English, more likely to be published in a journal indexed in an electronic database, more likely to be cited and more likely to be published more than once. All of these may lead to studies with statistically significant results being more likely to be located and hence included in the meta-analysis, and all have been demonstrated in the medical literature (Egger *et al*. 1997a). For a comprehensive and unbiased search for relevant studies, foreign language journals and grey literature (such as conference proceedings, theses and reports) should all be included, although in practice it is virtually impossible to search all of these sources comprehensively.

Selection bias may be introduced at the stage where studies are selected for inclusion in the review. The decision about whether to include or exclude a study should ideally be made by looking only at its methods and not at its results. This would avoid any possibility that the researchers may unwittingly select studies by whether they support their hypothesis rather than by whether they satisfy the predetermined selection criteria.

Another bias may be introduced when extracting data from published studies. Observers may intentionally or unintentionally interpret data in different ways depending on the study's results. Ideally, to exclude any possibility of bias, data should be extracted by more than one person, and they should be blinded to the study and the treatment group.

There are several methods for investigating whether there is bias in the inclusion of studies. First, the number of unpublished studies averaging a zero effect that would need to exist to overturn the conclusions of the analysis can be calculated (the 'fail-safe' number). There are several criticisms of this method. First, there is no consensus on what constitutes 'overturning the conclusions of a review'. Some have used reducing the result to non-significance ($P > 0{\cdot}05$), whereas others have used reducing the effect size to a 'small' effect or zero. Second, the method assumes that the hidden studies average a zero effect, which may not be the case, especially if publication bias exists. If the studies with positive effects have been published but those with negative or zero effects have not been, then the hidden studies will average a negative effect.

The second method is the funnel plot (Light & Pillemer 1984; Egger *et al*. 1997b; Palmer 1999), a graphical method in which the effect size of each study is plotted on the horizontal axis with a measure of the precision of its estimate on the vertical axis. Generally, the sample size has been used for the vertical axis, but as the precision depends (for a binary response variable) on both its sample size and the proportion having the event of interest, it is preferable to use the standard error of the effect estimate instead (Sterne, Egger & Davey Smith 2001a). In the absence of bias, a characteristic symmetrical 'funnel' shape will be seen; studies with low precision (generally small studies) show a large scatter of effect sizes around the true value, whereas those with high precision (generally large studies) have an effect size close to the true value. Asymmetry in a funnel plot may indicate that there is bias in the studies included, although there may be other possible explanations (for example, there could be differences in the true treatment effect between small and large studies).

Several statistical methods have been proposed for detecting and adjusting for bias. These include 'trim and fill' (Duval & Tweedie 2000), a method based on the funnel plot, and two methods investigating the relationship between study size and estimated treatment effects; a rank correlation method (Begg & Mazumdar 1994) and a linear regression method (Egger *et al*. 1997b).

ASSESSMENT OF QUALITY OF STUDIES

Studies carried out using poor methodology are more likely to be biased, and hence inclusion of them in a systematic review may influence its results and conclusions. It is desirable to assess each primary study to determine whether its results are likely to be biased. Assessment may include the quality of randomization (were experimental subjects allocated to treatments at random?), blinding (were experimental treatments concealed from participants and investigators?) and appropriate analysis (Clarke & Oxman 2001).

STATISTICAL METHODS FOR COMBINATION OF DATA

There has been a great deal of development of statistical methods for combination of data from independent studies, and methods now exist for most types of data, including continuous, proportional, count and survival data (Deeks, Altman & Bradburn 2001). A wide range of software is now available for carrying out meta-analyses (reviewed by Sterne *et al*. 2001b).

PRESENTATION OF RESULTS

The most useful method of presentation of the results of a meta-analysis is an estimate of the combined effect

size, together with a confidence interval. This gives a range of plausible effect sizes that are consistent with the data. Frequently results are presented in a 'forest plot', which includes a point for each study with its confidence interval, plotted horizontally, with the combined estimate and confidence interval plotted below them.

SUBGROUP ANALYSES

Often it is of interest to carry out the analysis for subgroups of studies or experimental subjects. For example, in a review of a treatment for cancer it may be of interest to analyse separately its effects on cancers in different sites, and in different groups of patients. This increase in the number of analyses carried out raises the problem of multiple testing; if many tests are carried out, some 'significant' effects will be found simply by chance. To guard against this possibility, analyses of subgroups of studies or experimental subjects should be limited to a small number specified in advance in the protocol, and interpretation of subgroup analyses should be cautious. It is common practice in medical research to present 99% instead of 95% confidence intervals for subgroup analyses, to emphasize that there is greater uncertainty about the results.

SOURCES OF HETEROGENEITY

There are many factors that may influence the effect size for each study, and an important part of reviewing is to identify these factors. Some will be biological and therefore of direct interest, such as different taxa, different study locations, and whether the study population was increasing or declining. Where possible, hypotheses about these factors should be formulated in advance and prespecified subgroup analyses performed to test them. Other factors may be methodological, such as different study designs, differences in the populations studied or in the duration of the experiment. For a comprehensive summary of research findings it is desirable to explain differences in the results of studies and hence an investigation into the sources of heterogeneity is very useful. Tests of homogeneity among studies included in a meta-analysis are often performed; however, these often have low power and a non-significant result does not necessarily mean that hetero-geneity does not exist.

The effects of the characteristics of participants or studies on treatment effects is often investigated by 'meta-regression' (Thompson & Sharp 1999; Thompson 2001), a weighted regression technique that relates features of each study to the treatment effect.

SENSITIVITY ANALYSIS

Because subjective decisions must be taken in the course of a quantitative review, it is important to ask how sensitive the results are to the way the review has been carried out. It may be possible to combine the same set of data in different ways to reach different conclusions (Rosenthal 1978). Depending on the exact nature of the review being conducted, it may be relevant to examine the effects of changing the entry criteria, including or excluding studies whose qualification is questionable, excluding unpublished or low-quality studies, reanalysis using different methods or substituting a range of reasonable values for doubtful or missing data.

Review of methodology of ecological meta-analyses

METHODS

Identification of studies

Studies were identified by searching the databases Zoological Record, Biological Abstracts and Science Citation Index, by checking the references of studies retrieved and from personal reference collections. Studies were included if:

1. they reviewed a subject by quantitative synthesis of the results of several independent experimental or observational studies;

2. they were published in an ecological journal (I have used a broad definition of ecological journals, including one journal primarily concerned with plant physiology, but excluding fisheries);

3. they were published before the end of 1998.

Assessment of methodology

The assessment of the methodology and reporting was based on the criteria for assessment of reviews by Sacks *et al*. (1987), the NHS Centre for Reviews and Dissemination's DARE manual (NHS 1998), Oxman & Guyatt (1988) and Oxman (1995). The following data were recorded from each of the reviews:

1. the method used to search for studies relevant to the review;

2. whether the review included studies published in non-English language journals or grey literature, or unpublished material, and whether the authors of primary studies were contacted for missing data;

3. whether there were explicit criteria for inclusion of studies;

4. whether lists of studies included and excluded were published;

5. the type of studies included;

6. the type of data combined;

7. whether the quality of primary studies was assessed, and the criteria for quality assessment;

8. whether sources of heterogeneity were investigated;

9. whether recognized statistical techniques were used to combine results;

10. the methods used for presentation of results, particularly whether confidence intervals were used;

Table 1. Studies included in the review

Author	Subject	Organisms
Arnqvist *et al*. (1996)	Assortative mating	Water striders
Bender, Contreras & Fahrig (1998)	Patch size and population density	Animals
Brett & Goldman (1996)	Food-web structure regulation	Plankton
Brett & Goldman (1997)	Trophic cascade	Plankton
Côté & Poulin (1995)	Parasitism and group size	Social animals
Côté & Sutherland (1997)	Predator removal to protect bird populations	Birds
Curtis (1996)	Effects of elevated carbon dioxide	Trees
Curtis & Wang (1998)	Effects of elevated carbon dioxide	Trees
Dahl & Greenberg (1996)	Effects of predators on population density	Freshwater invertebrates
Dolman & Sutherland (1997)	Spatial patterns of resource depletion	Vertebrates
Fiske, Rintamäki & Karvonen (1998)	Mating success of lekking males	Animals
Gurevitch *et al*. (1992)	Competition	All organisms
Hamilton & Poulin (1997)	Parasite-mediated sexual selection	Animals
Hartley & Hunter (1998)	Nest predation rates	Birds
Jarvinen (1991)	Effects of female age on laying date and clutch size	Great tit and pied flycatcher
Koricheva, Larsson & Haukioja (1998a)	Secondary metabolism	Woody plants
Koricheva *et al*. (1998b)	Insect performance on stressed plants	Woody plants
Leung & Forbes (1996)	Fluctuating asymmetry	Animals
McCurdy *et al*. (1998)	Sex-biased parasitism	Birds
Møller & Thornhill (1997)	Heritability of developmental stability	Animals and plants
Møller & Ninni (1998)	Paternity	Birds
Poulin (1994)	Parasite-induced behavioural changes	Animals
Poulin (1996)	Sex-biased parasitism	Helminths
Proulx & Mazumder (1998)	Impact of grazing on plant species richness	Plants/herbivores
Schalk & Forbes (1997)	Sex-biased parasitism	Mammals
Tonhasca & Byrne (1994)	Crop diversification	Insects
Vander Werf (1992)	Clutch size	Birds
Van Zandt & Mopper (1998)	Adaptive deme formation	Insects
Wooster (1994)	Effects of predators on population density	Benthic invertebrates

Table 2. Studies considered for the review but not included

Study	Reason for exclusion
Abouheif & Fairbairn (1997)	Data from literature, not experiments
Adams *et al*. (1997)	Methodological
Andow *et al*. (1995)	Not a review; combination of only a few studies
Burnham, Anderson & White (1996)	Combination of data to estimate survival parameters, not estimation of combined effect sizes
Fernandez-Duque & Valeggia (1994)	Analysis is illustrative only
Hechtel & Juliano (1997)	Combines data from only one study
Isbell & Young (1993)	Combines data from only one study
Moller & Raffaelli (1998)	Does not combine experimental results
Murray (1998)	Combines data from only one study
Osenberg, Sarnelle & Cooper (1997)	Methodological
Swanson & Johnson (1996)	Not a review of independent experiments
Venier & Fahrig (1998)	No combination of independent studies

11. whether prespecified subgroups were analysed and the results presented and interpreted appropriately;
12. whether publication bias, selection bias and data extraction bias were addressed;
13. whether sensitivity analyses were performed.

RESULTS

Forty-one studies were considered for inclusion. Twelve were rejected because they did not satisfy the criteria given above, and 29 were included in the review (Tables 1 and 2). Two studies did not use meta-analytic techniques, but were included in the review because they carried out syntheses of independent results.

Types of study included

Twelve reviews combined data from controlled manipulative experiments, 13 used data from observational studies, and in 4 reviews it was unclear what kinds of studies were included. One review used data from quasi-experimental studies as well as experiments. These made comparisons between 'experimental' and 'control' areas, not allocated at random by the experimenters, or between different time periods within one area. Most reviews did not contain an explicit statement of the type of studies that were being combined and did not discuss the possible biases involved in combination of non-randomized data.

Table 3. Methods used to locate studies for inclusion in reviews

Method	Number of studies (n = 29)
No information	8
Personal knowledge only	3
Computerized databases	13
Limited set of journals	7
Bibliographies of narrative reviews	6
Reference lists of primary studies	4
Conference abstracts	1
Advertisement for unpublished studies	1
One strategy used	10
Two strategies used	8
Three or more strategies used	3

Location of studies

Eight studies provided no information on the methods used to locate primary studies. Eight different methods were used by one or more of the remaining 21 reviews (Table 3).

Eleven reviews included at least one paper in a language other than English, or included one or more named non-English journals in their search, and 12 included primary studies published in grey literature. Eleven reviews included unpublished material, including three where the unpublished data were the authors' own. Four stated that authors of primary studies were contacted for missing data.

Criteria for inclusion

Nineteen reviews stated explicit criteria for inclusion of studies, and 27 gave lists of studies that were included. However, lists of studies included were often presented as lists of species included, which did not always make clear how many independent studies were involved. Only one review presented a list of studies that were considered for inclusion but rejected, and two others mentioned specific studies that were rejected, but did not give a full list.

Assessment of quality of primary studies

Only five reviews contained any evaluation of the quality of primary studies, although one of these simply noted that the review included some studies of poorer quality. One review classified studies as 'free from design problems' or not, and excluded one primary study because of design problems. Another review divided studies into those with sound design and those with design problems, and explored the effect of good or poor design on the results. The remaining two reviews assessed internal and external validity of the primary studies. Internal validity is concerned with aspects of the study that might render the results unreliable, such as differences in conditions between experimental and control organisms. External validity involves aspects that affect the generalizability of the results, such as whether the study was carried out on a declining or stable population.

None of the meta-analyses explicitly considered random allocation of subjects to treatments, concealment of experimental allocations or quality and appropriateness of the analytical methods used by the primary studies.

Methods for statistical combination of results

Twenty-six reviews used established methods to combine the results of independent studies, citing references to one or more standard texts (Glass, McGraw & Smith 1981; Hedges & Olkin 1985; Cooper 1989; Hunter & Schmidt 1990; Rosenthal 1991; Gurevitch & Hedges 1993; Cooper & Hedges 1994). Three of these reviews used the MetaWin statistical program (Rosenberg, Adams & Gurevitch 1996). Three reviews did not use established methods. One review used a simple ANOVA on a standardized measure of response to combine the results of different studies, one summarized results by vote counting, and the third used a linear regression, using different studies as independent data points (see Gurevitch & Hedges 1999).

Three measures of effect size were used: the Pearson correlation coefficient (9 reviews), a standardized measure of difference between the means of control and experimental groups (16 reviews, including one that also used the Pearson correlation coefficient), and the response ratio (the ratio of the mean of the experimental group to the mean of the control group; 2 reviews). The response ratio has the advantage that it requires only the means and sample sizes to be extracted from the primary studies, but methods for combining response ratios have been developed only recently (Curtis & Wang 1998). No reviews combined binary, proportional or survival data. Three reviews did not calculate combined effect sizes because of lack of data in the primary studies, but instead used statistically correct vote count methods (Hedges & Olkin 1985).

Presentation of results

Estimates of the combined effect sizes together with confidence intervals were presented by 14 reviews, with another two presenting confidence intervals for only some of the results. Other methods of presentation included using either the standard deviation, standard error or variance instead of a confidence interval, or box-and-whisker plots. One study confusingly displayed confidence intervals and standard errors for different results on the same graph. Graphs were used to display the main results in 11 studies, with 15 using tables and 3 using only text.

Subgroup analyses

Eighteen reviews reported results for various subgroups of the studies or individuals included. In most

cases these appeared to have been prespecified, as they were natural divisions into different classes of study (for example, different trophic groups, different habitats or different environmental stresses), although it was rarely stated at the outset what subgroups were to be investigated along with the main analysis. Most reviews made no corrections for multiple testing, but used the same $P < 0{\cdot}05$ criterion or 95% confidence interval as was used for the main analyses. One study used 99·9% confidence intervals for presentation of the results, and one review used a Bonferroni-type adjustment for multiple comparisons. Generally, no clear distinction was made between the main analysis and subgroup analyses, and it was difficult to disentangle the two.

Investigation of heterogeneity

Twenty-two reviews contained some exploration of the heterogeneity of results, using the Q statistic (Hedges & Olkin 1985) or another measure of homogeneity of effect sizes. Seven did not investigate heterogeneity further, and two found no heterogeneity. Nine reviews investigated the effects on heterogeneity of characteristics of the organisms studied, nine investigated methodological variables, and six investigated both. Explorations of the causes of heterogeneity generally took one of two forms; either splitting of the data set on some criterion and recalculating the heterogeneity statistic, and repeating this process until the heterogeneity was non-significant, or calculating the relationship (correlation or regression) between the effect size for each study and potential mediating variables.

Investigation of sources of bias

Publication bias was discussed by only 10 reviews. These all calculated a 'fail-safe' number of studies; the number of 'hidden' studies not included in the analysis that would have to exist to overturn the result. The fail-safe sample size was calculated in two different ways: some reviews calculated the number of 'hidden' studies necessary to reduce the overall effect size (measured by the d statistic) to 0·2, which is considered to be a small effect, whereas others calculated the number necessary to render the result non-significant at the $P = 0{\cdot}05$ level. The fail-safe sample sizes were not always presented in a way that made it easy to assess robustness of the results; for example, in two reviews the fail-safe numbers were presented in the text without a clear indication of which analysis they referred to. One review used the rule of thumb that the result should be considered robust if the fail-safe number exceeded $5k + 10$ (Rosenthal 1979), where k is the number of studies in the analysis, but conclusions about robustness were otherwise variable. For example, one review stated that their conclusions were 'very robust' with fail-safe sample sizes that were 2·36 and 3·77 times the sample size, while another stated that 'most results were quite robust' with fail-safe sample sizes varying from 3·45 to 57·8 times the number of studies in the analysis.

No studies used a funnel plot to identify biases in the studies included in the review, and none considered other sources of bias, such as selection bias and data extraction bias.

Sensitivity analyses

Only five reviews performed any sensitivity analysis. Two general types were performed. The first was investigation of the effects of non-independence between data points from the same study, location or species. These sensitivity analyses involved restricting the data to just one data point from each species, study or location. Second, there were investigations of the effect on the results of the analysis of exclusion of a study or group of studies that might have been particularly influential.

Discussion

Reviewing of research is of fundamental importance in a field such as ecology where most studies are neither large enough nor cover a sufficiently wide range of conditions to give definitive general answers. Traditional narrative reviews have been shown to be inadequate for drawing reliable conclusions, and there is growing awareness of the need for rigorous and objective reviews that evaluate the totality of the evidence in a quantitative way. Syntheses of research in ecology have not yet started to use the full range of techniques used to ensure robustness and lack of bias in medical systematic reviews. This is partly because the recent developments in medical systematic reviewing have not yet been publicized in the ecological literature, and partly because some of these developments are more difficult to carry out in ecology.

DIFFICULTY OF LOCATING RELEVANT STUDIES

Locating all relevant studies is a major challenge for any systematic review, as omission of any studies could lead to biased results. Computerized databases of published research are very useful for locating studies published in journals, but they may miss studies published in languages other than English, or in conference proceedings, reports or theses, which are much more difficult to search systematically. Studies with non-statistically significant results are less likely to be published in English and less likely to be published in a journal indexed by a bibliographic database, and hence may be less likely to be found and included in a review (Egger *et al.* 1997a). A significant proportion of medical research that appears in conference abstracts does not appear as a full publication, so the abstract may be the only clue to the existence of the study (Scherer & Langenberg 2001). It is likely that this problem also

exists in ecology. Because of the potential for a significant number of studies to exist that will not be located simply by searching journals, it is highly desirable to make the search for relevant studies as comprehensive as possible. This is difficult, but the best recommendation at present appears to be to use multiple search strategies.

DIFFICULTY OF EXTRACTING DATA

Several authors have noted the difficulty of extracting data from reports of primary studies (Gurevitch & Hedges 1993; Gurevitch & Hedges 1999; Osenberg, Sarnelle & Cooper 1999). As quantitative reviews become more widely conducted, the statistics necessary for combination of results are likely to become quoted routinely. Along with the increase in systematic reviewing in medical research, the standards of reporting of primary studies have improved (Moher *et al.* 2001), and many journals now require the statistics necessary for performing meta-analyses to be reported.

TYPES OF STUDY INCLUDED AND QUALITY ASSESSMENT

Existing ecological meta-analyses have combined data from experimental and observational studies, but few have discussed the quality of evidence provided by these different sorts of study. Reviews combining data from rigorously conducted, randomized experiments provide much stronger evidence than those using non-randomized observational data, or experiments with poor methodology. For this reason, reviews using meta-analysis should be explicit about the type of studies that are being synthesized and the implications for the conclusions of the review.

Randomization and blinding are two aspects of experimental quality that are given little weight in ecological reports but are considered of fundamental importance in medical research. In RCTs, a secure method of random allocation of subjects to groups is considered essential, and concealment of experimental allocations, for example by the use of a placebo control in a drug trial, is always included in the design if possible. In contrast, most ecological papers do not mention the methods of random allocation of subjects to groups or concealment of experimental allocations. Both of these aspects of experimental design deserve more attention, as they reduce the potential for bias and increase the strength of evidence that the experiment provides. Systematic reviews may have a useful role to play in encouraging these good practices, by highlighting the lack of randomization and blinding in existing research.

There has been some criticism of quality assessment of studies included in reviews, on the grounds that people vary in their assessments (Cooper 1984), and there is a potential for bias, as observers tend to rate the quality of papers agreeing with their own views as higher (Mahoney 1977). However, it would seem sensible to assess the methodological quality of papers and possibly exclude some, rather than to include potentially misleading results in the review. Where there is variation in methodological quality, it is useful to carry out sensitivity analyses to discover the effects on the overall conclusions of including or excluding certain studies.

ASSESSMENT OF PUBLICATION BIAS AND OTHER BIASES

No studies used funnel plots to assess publication bias and other biases in the sets of studies included in the review. Those that assessed publication bias calculated 'fail-safe' numbers of hidden studies. Different reviews used different methods to calculate them, so they are not always comparable between reviews. These numbers may be difficult to interpret, as they assume that the hidden studies have an average effect size of zero, which may not be the case if publication or other biases exist.

NON-INDEPENDENCE IN ECOLOGICAL META-ANALYSES

A problem that may affect ecological reviews to a greater extent than medical reviews is non-independence. For example, many reviews combine data from several species, with the aim of producing a general result. If several studies have been carried out on the same species, they may have results that are more similar than if the studies had been carried out on a random selection of species, and they may therefore bias the overall result. Similar arguments may apply to studies conducted at the same location or by the same research group. One of the reviews included 45 data sets, of which 33 were the authors' own, and 11 of the remaining 12 derived from one publication. It would be useful to explore the effects of potential non-independence in sensitivity analyses.

SENSITIVITY ANALYSES

A recent ecological paper gives a good example of the influence of data selection criteria on the results of a meta-analysis (Englund, Sarnelle & Cooper 1999). Other factors, such as quality of studies or non-independence, may also influence the results, so are useful to explore in further sensitivity analyses. With the variable quality of studies and standards of reporting in the ecological literature, sensitivity analyses may be very important in assessing the robustness of conclusions from meta-analyses.

CONCLUSION AND SUGGESTIONS FOR CONDUCT OF REVIEWS

Ecology shares with medicine the need to summarize research in order to reach general conclusions. Systematic

reviewing of research and meta-analysis provide the means to do this. The methodology of ecological systematic reviews falls short of the rigour of medical reviews, but medical reviews have had a much longer development. There is scope to improve the rigour of ecological research reviews by taking advantage of recent developments in medical systematic reviewing. Several features could be implemented relatively easily:

1. Multiple methods of searching for relevant studies should be used, and search methods should be reported in detail.

2. The types of study included in the review should be stated (e.g. randomized experiment, non-randomized experiment, observational study).

3. Estimates of the overall effect size and a confidence interval should be presented.

4. Bias in the inclusion of studies should be investigated by means of funnel plots.

5. A clear distinction should be made between main and subgroup analyses, and the latter should be interpreted with greater caution.

6. Sensitivity analyses should be carried out.

Acknowledgements

I thank Dr Jeremy Wilson, Prof. Joe Perry and two anonymous referees for comments on earlier drafts, which have greatly improved the manuscript.

References

Abouheif, E. & Fairbairn, D.J. (1997) A comparative analysis of allometry for sexual size dimorphism: assessing Rensch's rule. *American Naturalist*, **149**, 540–562.

Adams, D.C., Gurevitch, J. & Rosenberg, M.S. (1997) Resampling tests for meta-analysis of ecological data. *Ecology*, **78**, 1277–1283.

Andow, D.A., Klacan, G.C., Bach, D. & Leahy, T.C. (1995) Limitations of *Trichogramma nubiale* (Hymenoptera: Trichogrammatidae) as an inundative biological control of *Ostribia nubialis* (Lepidoptera: Crambidae). *Environmental Entomology*, **24**, 1352–1357.

Arnqvist, G., Rowe, L., Krupa, J.J. & Sih, A. (1996) Assortative mating by size: a meta-analysis of mating patterns in water striders. *Evolutionary Ecology*, **10**, 265–284.

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

Begg, C.B. & Mazumdar, M. (1994) Operating characteristics of a rank correlation test for publication bias. *Biometrics*, **50**, 1088–1101.

Bender, D.J., Contreras, T.A. & Fahrig, L. (1998) Habitat loss and population decline: a meta-analysis of the patch size effect. *Ecology*, **79**, 517–533.

Brett, M.T. & Goldman, C.R. (1996) A meta-analysis of the freshwater trophic cascade. *Proceedings of the National Academy of Science USA*, **93**, 7723–7726.

Brett, M.T. & Goldman, C.R. (1997) Consumer versus resource control in freshwater pelagic food webs. *Science*, **275**, 384–386.

Burnham, K.P., Anderson, D.R. & White, G.C. (1996) Meta-analysis of vital rates of the northern spotted owl. *Studies in Avian Biology*, **17**, 92–101.

Chalmers, I. & Haynes, B. (1995) Reporting, updating and correcting systematic reviews of health care. *Systematic Reviews* (eds I. Chalmers & D. G. Altman). BMJ, London.

Clarke, M. & Oxman, A.D. (2001) *Cochrane Reviewers Handbook 4.1.4. The Cochrane Library, Issue 4*. Update Software, Oxford.

Cooper, H.M. (1984) *The Integrative Research Review: a Systematic Approach*. Sage Publications, Newbury Park, CA.

Cooper, H.M. (1989) *Integrating Research: A Guide for Literature Reviews*, 2nd edn. Sage Publications, Newbury Park, CA.

Cooper, H.M. & Hedges, L.V. (1994) *Handbook of Research Synthesis*. Sage Publications, Newbury Park, CA.

Côté, I.M. & Poulin, R. (1995) Parasitism and group size in social animals: a meta-analysis. *Behavioral Ecology*, **6**, 159–165.

Côté, I.M. & Sutherland, W.J. (1997) The effectiveness of removing predators to protect bird populations. *Conservation Biology*, **11**, 395–405.

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.

Curtis, P.S. & Wang, X. (1998) A meta-analysis of elevated CO_2 effects on woody plant mass, form, and physiology. *Oecologia*, **113**, 299–313.

Dahl, J. & Greenberg, L. (1996) Impact on stream benthic prey by benthic vs drift feeding predators: a meta-analysis. *Oikos*, **77**, 177–181.

Deeks, J.J., Altman, D.G. & Bradburn, M.J. (2001) Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. *Systematic Reviews in Health Care: Meta-analysis in Context* (eds M. Egger, G. Davey Smith & D. G. Altman). BMJ Books, London.

Dickersin, K., Scherer, R. & Lefebvre, C. (1995) Identifying relevant studies for systematic reviews. *Systematic Reviews* (eds I. Chalmers & D. G. Altman). BMJ, London.

Dolman, P.M. & Sutherland, W.J. (1997) Spatial patterns of depletion imposed by foraging vertebrates: theory, review and meta-analysis. *Journal of Animal Ecology*, **66**, 481–494.

Duval, S.J. & Tweedie, R.L. (2000) Trim and fill: a simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, **56**, 455–463.

Egger, M., Davey Smith, G., Schneider, M. & Minder, C. (1997b) Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, **315**, 629–634.

Egger, M., Schneider, M. & Davey Smith, G. (1998) Spurious precision? Meta-analysis of observational studies. *British Medical Journal*, **316**, 140–144.

Egger, M., Zellweger-Zahner, T., Schneider, M., Junker, C., Lengeler, C. & Antes, G. (1997a) Language bias in randomised controlled trials published in English and German. *Lancet*, **350**, 326–329.

Englund, G., Sarnelle, O. & Cooper, S.D. (1999) The importance of data selection criteria: meta-analyses of stream predation experiments. *Ecology*, **80**, 1132–1141.

Fernandez-Duque, E. & Valeggia, C. (1994) Meta-analysis: a valuable tool in conservation research. *Conservation Biology*, **8**, 555–561.

Fiske, P., Rintamäki, T. & Karvonen, E. (1998) Mating success in lekking males: a meta-analysis. *Behavioral Ecology*, **9**, 328–338.

Glass, G.V., McGraw, B. & Smith, M.L. (1981) *Meta-analysis in Social Research*. Sage Publications, Beverly Hills, CA.

Gurevitch, J. & Hedges, L.V. (1993) Meta-analysis: combining the results of independent experiments. *Design and Analysis of Ecological Experiments* (eds S. M. Scheiner & J. Gurevitch.). Chapman & Hall, New York.

Gurevitch, J. & Hedges, L.V. (1999) Statistical issues in ecological meta-analyses. *Ecology*, **80**, 1142–1149.

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

Hamilton, W.J. & Poulin, R. (1997) The Hamilton and Zuk hypothesis revisited: a meta-analytical approach. *Behaviour*, **134**, 299–320.

Hartley, M.J. & Hunter, M.L. (1998) A meta-analysis of forest cover, edge effects and artificial nest predation rates. *Conservation Biology*, **12**, 465–469.

Hechtel, L.J. & Juliano, S.A. (1997) Effects of a predator on prey metamorphosis: plastic responses by prey or selective mortality? *Ecology*, **78**, 838–851.

Hedges, L.V., Gurevitch, J. & Curtis, P.S. (1999) The meta-analysis of response ratios in experimental ecology. *Ecology*, **80**, 1150–1156.

Hedges, L.V. & Olkin, I. (1985) *Statistical Methods for Meta-analysis*. Academic Press, New York.

Hunter, J.E. & Schmidt, F.L. (1990) *Methods of Meta-analysis: Correcting Error and Bias in Research Findings*. Sage Publications, Newbury Park, CA.

Isbell, L.A. & Young, T.P. (1993) Social and ecological influences on activity budgets of vervet monkeys and their implications for group living. *Behavioural Ecology and Sociobiology*, **32**, 377–385.

Jarvinen, A. (1991) A meta-analytic study of the effects of female age on laying-date and clutch-size in the great tit *Parus major* and the pied flycatcher *Ficedula hypoleuca*. *Ibis*, **133**, 62–66.

Knipschild, P. (1995) Some examples of systematic reviews. *Systematic Reviews* (eds I. Chalmers & D. G. Altman). BMJ, London.

Koricheva, J., Larsson, S. & Haukioja, E. (1998a) Insect performance on experimentally stressed woody plants: a meta-analysis. *Annual Review of Entomology*, **43**, 195–216.

Koricheva, J., Larsson, S., Haukioja, E. & Keinänen, M. (1998b) Regulation of woody plant secondary metabolism by resource availability: hypothesis testing by means of meta-analysis. *Oikos*, **83**, 212–226.

Kunz, R. & Oxman, A.D. (1998) The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *British Medical Journal*, **317**, 1185–1190.

Leung, B. & Forbes, M.R. (1996) Fluctuating asymmetry in relation to stress and fitness: effects of trait type as revealed by meta-analysis. *Ecoscience*, **3**, 400–413.

Light, R.J. & Pillemer, D.B. (1984) *Summing Up – The Science of Reviewing Research*. Harvard University Press, Cambridge, MA.

Mahoney, M.J. (1977) Publication prejudices: an experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, **1**, 161–175.

McAlister, F.A., Clark, H.D., van Walraven, C., Straus, S.E., Lawson, F.M., Moher, D. & Mulrow, C.D. (1999) The medical review article revisited: has the science improved? *Annals of Internal Medicine*, **131**, 947–951.

McCurdy, D., Shutler, D., Mullie, A. & Forbes, M.R. (1998) Sex-biased parasitism of avian hosts: relations to blood parasite taxon and mating system. *Oikos*, **82**, 303–312.

Moher, D., Schulz, K.F., Altman, D.G. & Lepage, L. (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*, **357**, 1191–1194.

Møller, A.P. & Ninni, P. (1998) Sperm competition and sexual selection: a meta-analysis of paternity studies of birds. *Behavioural Ecology and Sociobiology*, **43**, 345–358.

Moller, H. & Raffaelli, D. (1998) Predicting risks from new organisms: the potential of community press experiments. *Statistics in Ecology and Environmental Monitoring: Risk Assessment and Decision Making in Biology* (eds D. J. Fletcher, L. Kavalieris & B. J. F. Manly). Otago University Press, Dunedin.

Møller, A.P. & Thornhill, R. (1997) A meta-analysis of the heritability of developmental stability. *Journal of Evolutionary Biology*, **10**, 1–16.

Mulrow, C.D. (1987) The medical review article: state of the science. *Annals of Internal Medicine*, **106**, 485–488.

Murray, B.R. (1998) Density dependent germination and the role of seed leachate. *Australian Journal of Ecology*, **23**, 411–418.

NHS (1998) *Writing Abstracts for the Database of Abstracts of Reviews of Effectiveness*. NHS Centre for Reviews and Dissemination, University of York, York.

Osenberg, C.W., Sarnelle, O. & Cooper, S.D. (1997) Effect size in ecological experiments: the application of biological models in meta-analysis. *American Naturalist*, **150**, 798–812.

Osenberg, C.W., Sarnelle, O. & Goldberg, D.E. (1999) Meta-analysis in ecology: concepts, statistics and applications. *Ecology*, **80**, 1103–1104.

Oxman, A.D. (1995) Checklists for review articles. *Systematic Reviews* (eds I. Chalmers & D. G. Altman). BMJ, London.

Oxman, A.D. & Guyatt, G.H. (1988) Guidelines for reading literature reviews. *Canadian Medical Association Journal*, **138**, 697–703.

Palmer, A.R. (1999) Detecting bias in meta-analyses: a case study of fluctuating asymmetry and sexual selection. *American Naturalist*, **154**, 220–233.

Poulin, R. (1994) Meta-analysis of parasite-induced behavioural changes. *Animal Behaviour*, **48**, 137–146.

Poulin, R. (1996) Sexual inequalities in helminth infections: a cost of being a male? *American Naturalist*, **147**, 287–295.

Proulx, M. & Mazumder, A. (1998) Reversal of grazing impact on plant species richness in nutrient-poor vs. nutrient-rich ecosystems. *Ecology*, **79**, 2581–2592.

Rosenberg, M.S., Adams, D.C. & Gurevitch, J. (1996) *Metawin. Statistical Software for Conducting Meta-analysis: Fixed Effect Models, Mixed Effect Models and Resampling Tests*, Version 1.0. Sinauer, Sunderland, MA.

Rosenthal, R. (1978) Combining results of independent studies. *Psychological Bulletin*, **85**, 185–193.

Rosenthal, R. (1979) The 'file drawer' problem and tolerance for null results. *Psychological Bulletin*, **86**, 638–641.

Rosenthal, R. (1991) *Meta-analytic Procedures for Social Research*. Sage Publications, Newbury Park, CA.

Sacks, H.S., Berrier, J., Reitman, D., Ancona-Berk, V.A. & Chalmers, T.C. (1987) Meta-analyses of randomized controlled trials. *New England Journal of Medicine*, **316**, 450–455.

Schalk, G. & Forbes, M.R. (1997) Male biases in parasitism of mammals: effects of study type, host age and parasite taxon. *Oikos*, **78**, 67–74.

Scherer, R. & Langenberg, P. (2001) *Full Publication of Results Initially Presented in Abstracts (Cochrane Methodology Review). The Cochrane Library, Issue 4*. Update Software, Oxford.

Simmons, L.W., Tomkins, J.L., Kotiaho, J.S. & Hunt, J. (1999) Fluctuating paradigm. *Proceedings of the Royal Society Biological Sciences Series B*, **266**, 593–595.

Sterne, J.A.C., Egger, M. & Davey Smith, G. (2001a) Investigating and dealing with publication and other biases. *Systematic Reviews in Health Care: Meta-analysis in Context* (eds M. Egger, G. Davey Smith & D. G. Altman). BMJ Books, London.

Sterne, J.A.C., Egger, M. & Sutton, A.J. (2001b) Meta-analysis software. *Systematic Reviews in Health Care: Meta-analysis in Context* (eds M. Egger, G. Davey Smith & D. G. Altman). BMJ Books, London.

Swanson, B.J. & Johnson, D.R. (1996) Spatial and temporal trends and effects of population size on the frequency of color phenotypes in the wild red fox (*Vulpes vulpes*). *Canadian Journal of Zoology*, **74**, 1622–1631.

Thompson, S.G. (2001) Why and how sources of heterogeneity should be investigated. *Systematic Reviews in Health Care: Meta-analysis in Context* (eds M. Egger, G. Davey Smith & D. G. Altman). BMJ Books, London.

Thompson, S.G. & Sharp, S.J. (1999) Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, **18**, 693–708.

Tonhasca, A. & Byrne, D.N. (1994) The effect of crop diversification on herbivorous insects: a meta-analysis approach. *Ecological Entomology*, **19**, 239–244.

Van Zandt, P.A. & Mopper, S. (1998) A meta-analysis of adaptive deme formation in phytophagous insect populations. *American Naturalist*, **152**, 595–604.

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

Venier, L. & Fahrig, L. (1998) Intraspecific abundance–distribution relationships. *Oikos*, **82**, 483–490.

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

Received 1 November 2001; accepted 20 February 2001