Ecological Restoration Institute

Fact Sheet: Systematic Reviews and the Quality of Evidence *August 2011*

Systematic Reviews as a Means of Assessing the Quality of Evidence in Scientific Studies

Federal land management policies direct agencies to use “the best available science” to inform agency decisions. However, reactions to the term “best available science” sometimes lead to assertions that the quality of such science is a matter of opinion. How then do managers evaluate and select scientific information that is defensible and supportive of management policies and decisions as well as environmental assessments? This is especially important for restoration and forest management projects at larger spatial scales where field experiments are too costly and/or slow to produce needed results. Recently, conservation science has begun employing a rigorous analytical approach, known as a systematic review, for evaluating the existing scientific information and then informing managers about treatment options.

Unlike conventional literature reviews, which often summarize studies by providing qualitative descriptions of research results without paying much attention to the quality of the sources or the rigor of the experimental design, a systematic review is replicable and strives to answer management questions based on the available evidence (Pullin and Stewart 2006). The goal in a systematic review is to exhaustively search and obtain all relevant, peer-reviewed journal publications as well as unpublished, often not peer-reviewed, grey literature (e.g., theses and dissertations, conference proceedings, agency reports, newspaper and magazine articles, web content) and research findings. The final review then quantitatively summarizes the findings, highlights areas where additional research is needed, and provides management recommendations that incorporate the quality (i.e., rigor and strength) of individual science findings. When quantitative analysis is not possible, a transparent and rigorous qualitative review summarizes the evidence.

Assessing the quality of evidence is a critical component of a systematic review. In developing such an approach, it is helpful for land managers and writers of environmental assessments and impact state-ments to know what legally constitutes “good science.” Currently, natural resource agencies enjoy a considerable amount of discretion in cases of scientific uncertainty under the National Forest Manage-ment Act, National Environmental Policy Act, and Endangered Species Act (Schultz 2008). As long as a decision has some documented scientific support, and is not “arbitrary or capricious,” the courts tend to defer to the agency’s judgment. A more rigorous definition of sound science is the Daubert standard, which is a rule pertaining to the admissibility of expert witnesses' testimony in U.S. federal courts. Ac-cording to the Daubert standard, a conclusion will qualify as “scientific knowledge” if the proponent can demonstrate that it is the product of sound scientific methodology (i.e., the scientific method). There are five components of sound scientific methodology. The study must be: 1) empirically tested, 2) use ex-perimental controls (i.e., treatment vs. no treatment), 3) peer-reviewed and published, 4) accepted and used by the relevant scientific community, and 5) include an assessment of how confident the authors are in the results (Table 1). These very components are considered when assessing scientific evidence in systematic reviews (Table 1). Quantitative approaches found in Table 1 incorporate formal weighting schemes, whereas qualitative reviews are more flexible in that there is no commonly accepted method for assessing the quality of studies.

The Ecological Restoration Institute is dedicated to the restoration of fire-adapted forests and woodlands. ERI provides services that support the social and economic vitality of communities that depend on forests and the natural resources and ecosystem services they provide. Our efforts focus on science-based research of ecological and socio-economic issues related to restoration as well as support for on-the-ground treatments, outreach and education.

Ecological Restoration Institute, P.O. Box 15017, Flagstaff, AZ 86011, 928/523-7182, FAX 928/523-0296, www.eri.nau.edu

Table 1. Approaches to assessing quality of evidence in quantitative and qualitative analyses in systematic reviews.

Daubert Principle Component	**Quantitative: Meta-analysis**	**Quantitative: Bayesian**	**Qualitative: ERI approach**
1. Empirical testing: the theory or technique must be falsifiable, refutable, and testable	Assess quality of experimental design, reject inadequate studies[1]	Can choose to reject inadequate studies and/or use a weighting scheme	Assess quality of experimental design (1-6) [1]
2. The existence and maintenance of standards and controls	Must have control	Not required	Assess quality of experimental design (1-6) [1]
3. Subjected to peer review and publication	Can be a covariate	Can use a weighting scheme	Journal Impact Factor (www.scimagofr.com)
4. Degree to which the theory and technique is generally accepted by a relevant scientific community	Not explicit	Not explicit	Number of citations per year (from Google Scholar)
5. Known or potential error rate	Weighting scheme is used based on reported variance or alternative measures of uncertainty	Reported variance is used	Not explicit
ANALYSIS	Confidence intervals generated around effect size means (Gurevitch and Hedges 1993)	Confidence intervals generated around predicted value of outcome; different probabilities can be generated based on weighting scheme (Newton et al. 2007)	Final scores are calculated as (Quality of design) + 2 (Impact Factor) + (Citations/year) and binned into three categories

[1] Based on quality of evidence ranking from Pullin and Knight (2003, Table 2)

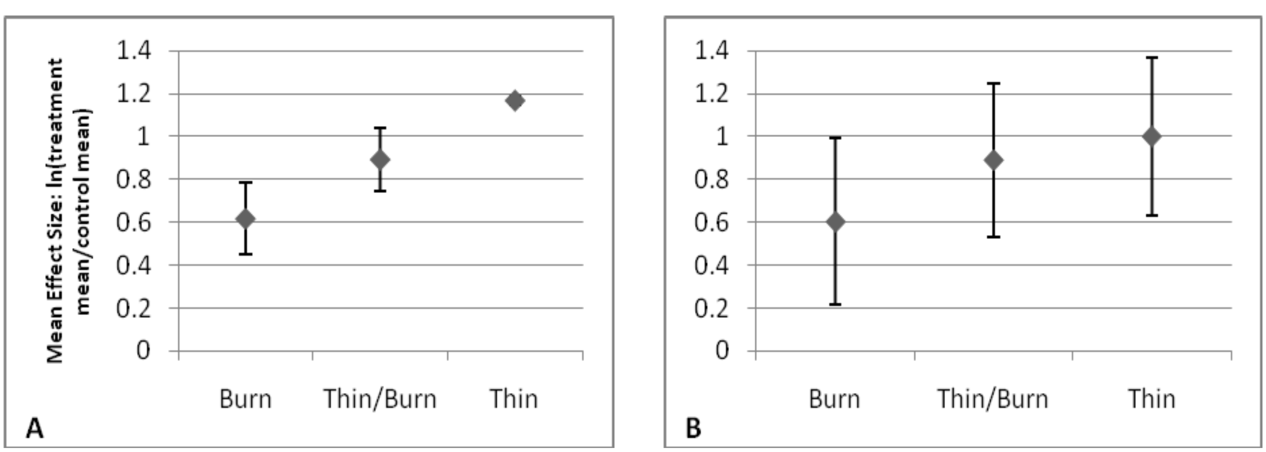
As a simple example, the Pullin and Knight (2003) ranking system for assessing the quality of a study's experimental design could be used to rate studies and eliminate inadequate studies (e.g., Table 2, category IV) from consideration (Table 2). This approach only addresses criteria 1 and 2 of the Daubert standard; an alternative approach that incorporates four of the Daubert criteria is provided as an additional example (Table 1, last column). Whatever the approach, the goal in preparing a systematic review is to clearly assess and present each line of evidence, and then draw conclusions that rely most heavily on the highest quality studies.

Table 2. Hierarchy of quality of evidence, based on adequacy of experimental design (from Pullin and Knight 2003).

Category	Quality of evidence
I	Strong evidence obtained from at least one properly designed randomized controlled trial of appropriate size.
II-1	Evidence obtained from well-designed controlled trials without randomization.
II-2	Evidence obtained from a comparison of differences between sites with and without (controls) a desired species or community.
II-3	Evidence obtained from multiple time series or from dramatic results in uncontrolled experiments.
III	Opinions of respected authorities based on qualitative field evidence, descriptive studies, or reports of expert committees.
IV	Evidence inadequate owing to problems of methodology (e.g., sample size, length or comprehensiveness of monitoring, or conflicts of evidence).

Weighting schemes are incorporated to assess the quality of evidence using either meta-analysis or Bayesian methods. These methods often use the reported standard deviation (although alternative measures of uncertainty can be used) of a particular study as a measure of its reliability. As a result, studies with a smaller standard deviation "count" more in the analysis than studies with larger variability. For example, a meta-analysis of studies that assessed crowning index responses to various forest treatments was conducted with and without a weighting scheme (Figure 1). Crowning index (CI) is the wind speed needed to start and sustain an active crown fire. A higher CI number means a forest is more resilient to crown fire. In this meta-analysis, a mean effect size greater than zero indicates that the CI was higher in the treatments compared to the controls, while the 95% confidence intervals show the range of variability in results across the studies (Figure 1). The weighted analysis (A) used a weighting scheme based on the amount of sampling effort involved (an alternative measure of uncertainty). In other words, larger studies were weighted more heavily than smaller studies in calculating mean effect size. The unweighted analysis (B) counts each study's results equally in calculating mean effect size (i.e., a study of two sites of 5 ha each would "count" the same as a study of 50 sites of 100 ha each).

Figure 1. A comparison of a weighted (A) and unweighted (B) meta-analysis of crowning index responses to various forest restoration treatments (Fulé et al. unpublished data)



Using a weighting scheme, the smaller (and in this case, more variable) studies did not affect the overall results as much as when no weighting scheme was employed. Thus, it is more difficult to draw conclusions from the unweighted study, as shown by the larger confidence intervals indicating no difference between treatments (Figure 1). While in this example, the weighting scheme was critical to the interpretation of the results, in other cases the weighting scheme may have very little impact (e.g., Kalies et al. 2010), indicating that there is more consistency in results across studies.

Conclusions

Quality of evidence assessments recognize that all studies are not equal, and provide an objective, repeatable means of identifying studies that do not meet the criteria of "scientific knowledge." Explicitly incorporating an assessment of evidence, such as a weighting scheme, can impact the conclusions of a systematic review. Whatever methods of quality assessment are used, it is important to present them clearly and transparently, so that managers and environmental assessment writer understand the differences in the quality of studies when interpreting results and making their decisions.

References

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

Kalies, E.L., C.L. Chambers, and W.W. Covington. 2010. Wildlife responses to thinning and burning treatments in southwestern conifer forests: a meta-analysis. *Forest Ecology and Management* 259:333-342.

Newton, A.C., G.B. Stewart, A. Diaz, D. Golicher, and A.S. Pullin. 2007. Bayesian Belief Networks as a tool for evidence-based conservation management. *Journal for Nature Conservation* 15:144-160.

Pullin, A.S. and T.M. Knight. 2003. Support for decision making in conservation practice: an evidence-based approach. *Journal for Nature Conservation* 11:83-90.

Pullin, A.S. and G.B. Stewart. 2006. Guidelines for systematic review in conservation and environmental management. *Conservation Biology* 20:1647-1656.

Schultz, C. 2008. Responding to scientific uncertainty in U.S. forest policy. *Environmental Science & Policy* 11:253-271.

For more information about systematic reviews, visit the Centre for Evidence-based Conservation at www.cebc.bangor.ac.uk

For more systematic reviews done by the Ecological Restoration Institute, link to http://www.eri.nau.edu/en/evidence-based-restoration-projects