

**Applying Systematic Evidence Reviews  
in Oregon Forest Policy:  
Opportunities and Challenges**

**December, 2005**

Institute for Natural Resources  
210 Strand Ag Hall  
Oregon State University  
Corvallis OR 97330

Lead Preparer: Jeff Behan  
Graduate Research Assistants: Seth Crawford and Erica Kleiner

## **Acknowledgements**

The staff at Institute for Natural Resources thanks the Oregon Department of Forestry for supporting the preparation of this report. I personally thank Gail Achterman, Executive Director of the Institute for Natural Resources, for reviewing an earlier draft. Thanks are also due to Seth Crawford and Erica Kleiner for their efforts in locating reference material and drafting sections of the report.

Several reviewers on staff at Oregon Department of Forestry provided helpful comments on an earlier draft.

Any factual errors, omissions or other flaws that remain in this document are my responsibility.

Jeff Behan  
December, 2005

## **Executive Summary**

Use of “best available science” to inform natural resource policies is codified in federal and state statutes. Forest management stakeholders consistently agree that the best available science should be used in policymaking. But conflicts over what is, and is not “good” science and selective use of studies with different conclusions by competing interest groups continue to challenge public land managers. These conflicts point to a need to develop a method of synthesizing technical information that relates to particular forest management questions in a way that will be more readily accepted as objective and definitive.

In June 2004, former Oregon Governor John Kitzhaber presented testimony to the Oregon Board of Forestry (BOF) on a number of forest policy issues, including the problem of "dueling science". Dr. Kitzhaber introduced the *Systematic Evidence Review* (SER) process and explained how it is used to rigorously evaluate evidence on treatment efficacy in clinical medicine. He suggested that this process could be adapted and brought to bear on developing a more credible evidence base for forest policy making.

The BOF subsequently incorporated exploration of the SER process into the Oregon Department of Forestry (ODF) State Forests Program work plan. ODF contracted with the Institute for Natural Resources (INR) at Oregon State University (OSU) to prepare a report on SERs. The Institute works to provide Oregon leaders with ready access to current, science-based information and methods for better understanding our resource management challenges and developing solutions. The BOF and ODF requested that INR develop:

- Background information on SERs;
- A comparison of medical research studies and natural resource research studies to identify any differences that affect the ability to develop methods for evaluating the quality of research evidence; and
- Proposed principles for a simplified SER-like research evidence evaluation process for ODF to use to organize, present, and synthesize scientific information for use in BOF decision making.

### **How does ODF currently gather and assess scientific information to use in forest management policies?**

- ODF utilizes scientific knowledge in its duties to (1) manage Oregon state forests for the “greatest permanent value” to the people of Oregon, and (2) regulate commercial forest operations on non-federal forests through the Oregon Forest Practices Act.

- ODF policies are informed by science through (1) internal science reviews, (2) external reviews commissioned by the agency to assess the scientific validity of its planning documents and regulatory proposals.
- Despite well-intentioned and in many cases quite involved efforts to use the “best science available,” ODF is regularly challenged by groups suggesting that they really are not doing so.
- These challenges may stem from disputes over which pieces of technical evidence were, or were not considered, or over how particular pieces of evidence were interpreted, weighed and applied in policymaking.
- The core of disputes over use of technical information may involve broader disagreements over forest policy goals, and the appropriate course of action when outcomes are uncertain, rather than disputes over scientific evidence *per se*.

### **What is a “Systematic Evidence Review” and how do SERs work?**

- An SER is rigorous, transparent, reproducible process for assessing scientific and technical information, used primarily in clinical medicine.
- An SER focuses tightly on a specific question, or small set of questions, which frame decisions about what evidence is relevant to the review, and what is not.
- SERs differ from traditional literature reviews in their use of pre-established, explicit protocols for finding, screening, grading and integrating primary research studies.
- SERs are designed to be as comprehensive, exhaustive and objective as possible, which means they are typically time consuming and expensive.
- Systematic Evidence Reviews and evidence based medicine have been described as a “paradigm shift” in healthcare, but there is still considerable debate about how SERs are conducted and used.
- Costs of medical SERs range from \$50,000-\$250,000. A natural resource SER may cost considerably less because the evidence base is likely to be smaller, but this would depend on the nature of the question.

### **History of Systematic Evidence Reviews**

- Since emerging in the 1980’s the SER approach has been widely adopted in the fields of clinical medicine and public health, and continues to expand rapidly.

- The largest international entity that conducts and disseminates SERs is the Cochrane Collaboration, which maintains a database of over 2000 SERs, develops and refines review methods, and offers training on conducting SERs.
- In the United States, SERs are conducted and disseminated by the Agency for Healthcare Research and Quality (AHRQ) through its 15 designated Evidence-based Practice Centers. Oregon Health Sciences University in Portland is one such center.
- Interest is growing in adapting the SER approach to other areas of public policy, including wildlife conservation, but SERs are not well suited for all policy areas, or to all questions within a particular field.

### **The Systematic Evidence Review process in clinical medicine**

- SERs require specific, tightly focused review questions to (1) clarify the purpose and delimit the scope of the review, and (2) strengthen linkages between the questions and subsequent steps in the review process.
- SERs use explicit protocols that spell out in advance exactly how evidence will be gathered, assessed, collated and summarized. Documenting all steps in the SER ensures that it is transparent, replicable and can be updated later if necessary.
- SERs are characterized by vigorous and thorough efforts to compile all available research and technical information that pertains to the review question(s), including unpublished and “gray” literature.
- Quality assessment of individual studies is a key characteristic of SERs, providing more rigor than traditional reviews. Quality assessments are used to (1) decide whether a relevant study should be included in the review, and (2) to rank included studies in an *evidence quality hierarchy*, usually based on study design and methods.
- Quality assessment is labor intensive and remains controversial. Random controlled trials and other tightly controlled study designs are favored in medical SERs, but there is no general consensus on standardized quality assessment criteria.
- Synthesis consists of tabulation of study characteristics, quality and outcomes for the primary purpose of investigating whether results are consistent across included studies, and if not, investigating reasons for apparent differences.
- A *narrative synthesis* is used to qualitatively compare and synthesize included studies. Qualitative synthesis may be all that is possible if differences in

population, intervention, outcome measures, designs and quality preclude meta-analysis.

- Quantitative synthesis, such as *meta-analysis* may be used to statistically combine study findings, as long as the studies are similar with regard to population and intervention under study, outcome measures, study design and quality.
- The strength of a body of evidence (all included studies) is assessed by examining aggregate study *quality*, the *quantity* of evidence (number of studies, sample size), *consistency* of findings across studies, and *coherence* of the evidence as a whole.
- SERs obtain their rigor partly through efforts to identify and explicitly acknowledge ways in which bias could enter into and affect results in primary studies, and to reduce bias during selection and review of these studies.
- SERs can be, and often are updated when new information emerges that will significantly strengthen or change the outcome of an existing review. Explicit documentation of how the review was conducted makes updating possible.
- A key use of SER information is developing *clinical practice guidelines*: “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.”
- More progress has been made in rigorously assessing available evidence through SERs than in incorporating this evidence into medical policy due to problems accessing clinical guidelines and because many patients prefer tailored care.

### **Critiques of SER methods and cautions about using SER information**

- Research quality assessment criteria are a cornerstone of SERs, but there is still no consensus on what these criteria should be and how different types of quantitative evidence should be weighed.
- There is growing criticism within the medical community of rigid study exclusion criteria and the practice of ranking evidence quality on the basis of research methodology alone.
- There are growing calls for including a broader range of evidence in SERs, including qualitative evidence and expert opinion, but finding ways to include, and weigh, such disparate types of evidence are major challenges.
- Framing questions in the tightly focused, specific way that current systematic review methods require may skew the review away from important issues that are more difficult to focus.

- Absence of evidence regarding the effectiveness or safety of a health care treatment or medication does not mean that the intervention is not safe or effective. In other words, “absence of evidence is not evidence of absence”.
- Strength of evidence *in and of itself* is not related to the magnitude of effectiveness of an intervention. There may be very strong evidence that the intervention has little impact or, conversely, the apparent impact of the intervention may be large but the evidence regarding the impact may be weak.
- How strong the evidence needs to be when making a particular type of decision should depend, at least in part, on the potential consequences if assumptions about the outcomes of an intervention turn out to be wrong.

### **Applying SERs to Natural Resource Issues: Challenges and Opportunities**

#### *Disentangling questions about evidence from those about values and preferences*

- Forest and ecosystem managers usually lack the widely agreed upon single objective (better human health) that clinical medicine practitioners enjoy.
- Clarifying and obtaining consensus on the underlying objective (e.g. the proposed management action about which the information is being collected) may be critical to conducting a successful natural resource SER.

#### *Differences between medical science and ecosystem science as fields of inquiry*

- Compared to clinical medicine, ecology is a younger science that involves a greater proportion of observational field-based studies to laboratory experiments.
- Ecological research typically involves greater methodological diversity, fewer laboratory controls, less replication and more “nuisance” variables than medical research.
- Ecology deals with larger spatial scales, longer timeframes and more complexity associated with multiple interacting species, habitats and ecological processes than medicine, which is focused on a single species.
- In general, there is less certainty about scientific conclusions in ecological studies than in clinical medicine.

### *The challenge of delimiting evidence that applies to a forest management question*

- SERs are best suited to synthesizing research focused on whether a single medical intervention “works” or “doesn’t work”. In the natural resources arena, SERs would be best suited to analogous, single variable questions.
- Complex, multifaceted forest management questions might be difficult to address using the SER approach, and for simpler questions, there may be little focused research evidence available.
- Much evidence concerning forest ecosystems consists of studies in which several variables are considered simultaneously in order to accurately describe real world ecological relationships.
- Synergies among species and processes are common in forest ecosystems. Thus, it may be impossible or misleading to isolate a single ecosystem component or single outcome of a management action as the focus of an SER.
- Delimiting the evidence that applies to a forest management question may also be challenging due to uncertainty about extrapolating results from studies to other areas with significant biological, physical, climatic or land use history differences
- For a forest management question structured as suggested by SER guidelines, there is likely to be a range of tangentially related research that falls somewhere between direct relevance and complete irrelevance.
- A feasible natural resource SER may require a compromise between a holistic approach that is closer to reality, but impractical for defining relevant studies, and a reductionist approach that may limit the review's relevance.
- A related challenge may lie in structuring a question with a degree of specificity that allows inclusion of enough evidence to make the review worthwhile, but also limits it to a manageable scope.

### *The challenge of assessing evidence quality in forest ecosystem science*

- There is likely to be a paucity of relevant, focused experimental research and a greater proportion of potentially diverse observational evidence available to address a natural resource question.
- This type of evidence is typically graded as “low” quality in medical SERs, e.g. often observational with few controls, frequently with confounding interactions.

- If the same criteria used to assess evidence quality in medical SERs are deemed appropriate for a natural resource SER, these criteria would probably need to be applied *less stringently* to assess forest ecosystem research.
- Depending on the nature of the SER question and evidence available to address it, it may also be necessary to develop *significantly different* criteria for assessing the quality of forest ecosystem research.
- There is a lack of consensus on quality assessment criteria used in medical SERs. Achieving consensus on if, and how, such criteria should be used in forest ecosystem SERs may prove difficult.
- There may be no single set of quality assessment criteria that will work for all natural resource SERs. Assessment criteria may need tailoring to fit the evidence that pertains to a particular SER question.
- In cases where the evidence consists of studies and monitoring with disparate methods, locations and outcome measures, there may be no clear rationale for saying that one piece of evidence is “higher” or “lower” quality than another.
- With all else being equal, studies that involve relatively larger spatial scales, more replication, more controls and longer timeframes are likely to produce the most reliable results.
- However, “all else equal” assumptions often don’t hold true. Greater complexity and diversity at larger scales can introduce more, rather than less uncertainty.

#### *Locating the evidence*

- Archiving of medical research abstracts and peer reviewed papers is more organized and standardized than in ecosystem science. Comprehensive literature searches in a natural resource SER may be harder to achieve than in clinical medicine.

#### *The role of qualitative research, expert judgments and experience*

- Expert knowledge and experience play a greater role in ecosystem research than in medicine because investigators must rely more heavily on expert judgment when interpreting results.
- Natural resource management also involves high levels of expert judgment because scientific information is often not available.

- Experiential knowledge may constitute an important part of the overall evidence base, but incorporating this evidence into quality assessment and ranking framework in the context of an SER remains problematic and controversial.
- One potential way around this debate is to understand scientific and expert knowledge as complementary, and equally important in ecosystem and natural resource management.

#### *Opportunities for applying Systematic Evidence Reviews to natural resource issues*

- Applying SERs to natural resource issues will be challenging, but early proponents of the SER approach in medicine faced significant challenges as well.
- Despite differences, there are a number of similarities between clinical medicine and aspects of conservation and natural resource management.
- These similarities include the common use of interventions (essentially experiments in progress), the need to make decisions on the basis of imperfect information, and the complementary role of evidence and experience.
- Some components of SERs could be incorporated into science reviews (e.g. better documentation of how studies were selected for inclusion, investigation of quality differences) in order to increase their objectivity and transparency.
- Conducting a synthesis of available science on a natural resource topic using SER techniques could highlight gaps in the evidence base and suggest relevant areas for future research.
- In combination with ecological monitoring and incremental updates, a synthesis of available science using SER techniques would mesh well with landscape-level adaptive management.

#### **Principles, Guidelines and Considerations for Applying Systematic Evidence Reviews to Forest Management in Oregon**

- “Evidence,” in an ecosystem management context, is more than just data and hard facts. It involves contextual information and interpretation. *Scientific* evidence consists of scientifically guided empirical observations combined with background information, logic, and scientific expertise.
- Sweeping generalizations about the appropriateness of particular statistical or research methods over others are unwarranted, and laboratory experiments do not necessarily carry greater weight than field experiments in forest ecology. All types of data can add to the evidence base.

- Evidence does not have to be quantitative or gathered by a scientist. The key is that the information was collected and interpreted as objectively as possible and can somehow be verified.
- The weight given to a particular piece of evidence should not depend on the type of observation but on the match between the observation and the question being asked.
- Medical SERs are, by design, rigorous, exhaustive and comprehensive, and thus time consuming and costly. A “small scale, practical approach” to science assessment is fundamentally different than SERs as they are defined in the medical field.
- Despite this inconsistency, some *aspects* of SERs could be readily adopted and incorporated into the internal and external science reviews that ODF already conducts.
- A systematic review of evidence pertaining to a forest management question *may* be feasible for ODF if (1) the question is tightly focused, (2) the evidence base pertaining to the question is not large, and (3) there is consensus on the boundaries of the evidence base.
- Natural resource SERs are more likely to be feasible for focused questions involving a single intervention and/or a single species. Multifaceted questions that involve more than one species or more than one outcome would be more difficult to address using the SER process.
- Independence from stakeholders is a fundamental aspect of SERs. An SER is more likely to be considered objective by all stakeholders if it is conducted by an independent entity, rather than internally by ODF.
- An SER process *may* reveal general consensus on the scientific evidence that is masked by fundamental differences of opinion on what outcomes are most important and what actions are appropriate in the face of imperfect evidence.

### **Three options for using the SER approach in Oregon forest management**

A full-scale SER on a complex forest ecosystem science question could be a major undertaking compounded by the need to recruit and train an external SER team before commencing the review itself. The feasibility of natural resource SERs and circumstances in which they would be most useful are not clear. Much could be learned by testing the SER approach, which offers some clear improvements over traditional literature reviews.

ODF could take an **incremental approach** to adapting the SER process to forest policy making in Oregon. Three tiered options for doing so are outlined below. These options roughly parallel the three existing approaches to science review at ODF: (1) routine internal reviews, (2), external reviews commissioned by ODF to review long-term planning documents, and (3) other external reviews completed as part of broader policy initiatives such as IMST reports for the Oregon Plan.

The form and details of each option are provided as a starting point and could benefit from further management and stakeholder review and discussion. The agency could develop a hybrid approach tailored to its needs in a particular circumstance.

**Option 1: Incorporate SER techniques into ODF’s “in-house” science assessments and any external review of this work.** The primary aim here would be to make existing ODF internal science review processes more transparent. This could be achieved with adjustments to what is already being done, primarily by adopting components of SERs to better document how science information is gathered and reviewed.

Under this scenario, ODF would not be rigidly bound to assuring that the review was an absolutely exhaustive and complete examination of all available evidence. As with all science reviews however, credibility would be predicated on perceptions of the degree to which the review was thorough and objective. This option would be best suited to cases where the available evidence is relatively clear, uncontroversial and limited in scope.

*Option 1a:* Conduct the science assessment “in house” as described above, with the additional step of soliciting external review of the draft final document. This process would approximate that used during the Independent Scientific Review of the Draft Western Oregon State Forests Habitat Conservation Plan, as described in Section II. The key differences would be use of SER procedures during the “in house” phase, and that external reviewers would be asked to assess the quality of evidence used and upon which they base their review and comments.

*Key components of Option 1:*

- To the degree possible, develop tightly focused, specific questions to delineate the purpose and scope of the review.
- Develop a simplified SER-like protocol to explain how the review will be conducted, using the example shown in Appendix 2 as a starting point. Development and use of a formal evidence quality hierarchy and ranking system is probably not be feasible at this level of review, but a narrative discussion and comparison of study quality could strengthen the review.

- Document in a systematic way which studies were included, what they said and how the information was interpreted. If studies were identified as relevant, but not included for quality or other reasons, document these reasons.
- If/when documents are sent out for external review, include in the review process an expectation that reviewers will also provide a quality assessment of the information upon which they based their review comments.

More specific guidelines for how this option might be implemented are offered in Appendix 4.

**Option 2: Commission an SER by an external, independent entity.** Under this scenario, a review of evidence would be contracted to a qualified independent entity. Such a review might be triggered by politically sensitive or difficult scientific questions about which ODF staff sought external scientific review. External review should assume impartiality and take advantage of academic expertise in specialized subdisciplines within ecological science.

The overall aim would be to prepare a defensible SER for a natural resource question, or a limited set of questions, with corresponding effort to obtain all relevant evidence and review it in formal, documented fashion. As with Option 1, this would not require an entirely new process. Existing external science review entities would consider using the SER approach.

*Key components of Option 2:*

- ODF would develop tightly focused questions to frame the purpose and scope of the SER. The additional step of vetting the SER questions with stakeholders could be considered. Questions would be refined in collaboration with stakeholders and SER review team.
- Develop a protocol that explicitly lays out how the review will be conducted. The external SER team should take the lead, or at least be included, in this process. If the review team believes it is feasible, develop and apply a formal set of evidence quality assessment and ranking criteria to the included studies.
- Publish results of review on ODF website and in academic journals.

**Option 3: Collaborate with other state and federal agencies to address regionally significant, highly policy relevant questions of using the SER process.** Many forest management issues transcend agency boundaries and should be addressed at the landscape scale. Some of these issues are controversial and challenging, and more than one agency could benefit from synthesis of all available evidence into a package of “best available science” that all participating agencies could then use. Post-wildfire “salvage”

logging and restoration is an example of a topic for which it may be worthwhile for ODF to initiate and/or participate in multi-agency efforts to identify key questions and support an SER process to address them.

Topics would need to be carefully considered because of the time and effort that would likely be required to coordinate a multi-agency SER. Various approaches are possible. For example:

*Option 3a:* Bring together an SER team comprised of technical specialists from within different agencies (e.g. ODF, Oregon Department of Fish and Wildlife, USDA Forest Service, USDI Bureau of Land Management, Fish and Wildlife Service, National Park Service, NOAA Fisheries) to develop questions, a protocol and conduct the review.

*Option 3b:* Conduct an external SER as described in Option 2, but solicit and coordinate support from other agencies.

*Key components of Option 3:*

- Similar to Option 2, but with interagency collaboration in (1) identifying and refining questions and vetting them with stakeholders, and (2) locating evidence, particularly unpublished monitoring data and other agency-specific information that may not be widely available, and (3) providing support to conduct the SER.

### **Conclusions and looking ahead**

A pilot test of a modified SER process could shed light on the accuracy of many of the untested perspectives and assumptions in this report regarding the potential for SERs in natural resources. There is no way to really know how accurate the analysis contained here is without testing it in practice. The best way to do this is by applying a modified SER process in a pilot test on a carefully selected but relevant natural resource question. It would be important to start with a question that is limited in scope and for which it is reasonably certain that enough evidence exists to conduct a useful review.

## Table of Contents

Executive Summary .....	3
Table of Contents .....	15
I. Introduction .....	17
II. How does Oregon Department of Forestry gather and assess scientific information? 20	
Oregon Statutes That Address the Use of Science in Forest Planning and Management .....	20
ODF Processes to Assess and Incorporate Science .....	21
Internal science reviews by ODF technical specialists .....	21
The Forest Practices Advisory Committee .....	23
Independent scientific reviews of state forest planning documents .....	23
Integrating Scientific Knowledge into ODF State Forest Management Plans .....	25
Uncertainty and adaptive management in Oregon forest policy .....	27
III. What is a “Systematic Evidence Review” and how do SERs work?.....	29
Overview of the SER process .....	29
History of Systematic Evidence Reviews .....	31
The Systematic Evidence Review Process in clinical medicine .....	34
A. Focus on answering a specific question(s) .....	35
B. Developing an SER protocol .....	36
C. Identifying relevant information .....	37
D. Assessing research quality .....	38
E. Summarizing and synthesizing the evidence .....	42
F. Acknowledging, reducing and removing bias.....	43
G. Updating SERs .....	45
H. Incorporating SER Information into Medical Policy and Practice.....	45
IV. Critiques of SER methods and cautions about using SER information.....	47
Appraising research quality .....	48
Including a broader range of evidence .....	48
Where does experiential knowledge fit in? .....	49
Cautions to bear in mind when considering the strength of a body of evidence .....	50
V. Applying SERs to Natural Resource Issues: Challenges and Opportunities .....	51
Disentangling questions about evidence from those about values and preferences .....	51
Differences between medical science and ecosystem science as fields of inquiry.....	52
The challenge of delimiting evidence that applies to a forest management question .....	54
The challenge of assessing evidence quality in forest ecosystem science .....	56
Locating the evidence .....	58
The role of qualitative research, expert judgments and experience .....	59
Disparities in healthcare and ecological research funding .....	60
Opportunities for applying SERs to natural resource issues .....	62
Applying SERs to conservation issues in the United Kingdom.....	63
Adapting components of SERs .....	65
SERs and adaptive management .....	65

VI. Principles, Guidelines and Considerations for Applying Systematic Evidence Reviews to Forest Management in Oregon.....	67
Key Points.....	67
Three options for using the SER approach in Oregon forest management .....	68
Responses to ODF Questions about SERs.....	71
VII. Conclusions and looking ahead.....	79
Appendix 1: Examples of Evidence Quality Hierarchies and Quality Assessment Criteria Used in Medical Systematic Evidence Reviews.....	80
Appendix 2: Sample Protocol for a Natural Resource SER .....	84
Appendix 3: Guidance for Adapting Components of Systematic Evidence Reviews to Small-Scale Technical Information Reviews.....	89
References Cited .....	98

## **I. Introduction**

### *Science as a foundation for forest management policies*

Use of science to inform natural resource policy has been implicit but standard practice in the United States for many decades, and is now codified in federal and state statutes. This practice reflects a general acceptance of science as the source of the most accurate knowledge that is available to humans about physical and biological characteristics of the world.

From the beginning of forestry in the United States Gifford Pinchot strongly advocated scientific management of forest resources as the foundation of “conservation”: the preservation and wise use of all natural resources. Today, federal, state and private forestlands in the U.S. are managed to meet an array of objectives, but science-based management has been widely adopted across all ownerships. Use of “best available science” is a relatively consistent point of consensus among forest management stakeholders. Diverse interests usually agree that scientific knowledge is a (more or less) “objective” basis for policies.

While there is general agreement about the scientific basis of forest management, complexity and multiple scale interactions and processes present significant hurdles to developing a robust understanding how forest ecosystems operate. This has led to difficulties in achieving consensus on what scientific research tells us about the potential outcomes of various forest management choices. Scientists rarely produce unequivocal, black and white answers to the questions that land managers pose. Uncertainty about outcomes of management actions is the norm, especially at larger spatial scales and longer time frames. Policymakers must consider legal mandates, social and political policies along with science-based recommendations, and stakeholders with different values and perspectives on policy goals and risk, even on topics where the scientific evidence is relatively clear.

Despite these challenges, scientific knowledge remains the best foundation for natural resource policy guidance that has yet been found. There is also broad agreement that it is wise to seriously consider predictions that science makes about the possible outcomes of management actions, even when these predictions involve significant uncertainty. For these reasons, there is a great deal of interest in finding effective methods for applying scientific knowledge when formulating natural resource policies, and for modifying such policies when new knowledge emerges. Federal agencies and some states are now required to consider the best science available when generating management plans.

Yet these legal mandates do not answer the question of what constitutes “best available science.” Conflicts over what is, and is not “good” science and selective use of studies with different conclusions by competing interest groups continue to challenge public land managers. These issues point to the need for synthesis of research results applicable to a particular question in a way that diverse stakeholders will more readily accept as

objective and definitive. There is also a pressing need to make best available science more accessible to land managers.

### *Systematic Evidence Reviews*

Similar issues in the fields of clinical medicine and public health, as well as the need to reduce regional variation in medical practices, prompted development in the early 1990's of a process known as Systematic Evidence Review (SER). SERs are independent, objective science assessments characterized by rigorous, well documented and replicable criteria for locating and synthesizing scientific information. SERs focus tightly on questions concerning the efficacy of medical interventions, e.g. a particular medication or treatment. SERs are used to search for, identify and "package" this information so that it is (1) more defensible and widely accepted as the "best" science available, and (2) more accessible to policymakers and the public. SERs help inform stakeholders and improve consensus on the state of knowledge and consequences of medical interventions.

In June 2004, former Oregon Governor John Kitzhaber presented written and verbal testimony to the Oregon Board of Forestry (BOF) on a number of forest policy issues, including the problem of "dueling science," in which stakeholders bring to the table only those scientific findings that support their policy views while ignoring or discounting findings that do not. Dr. Kitzhaber suggested that applying the SER process in such cases could aid in developing a more credible evidence base for forest policy making.

In a December 2004 workshop and March 2005 meeting, the BOF discussed these ideas, and decided to explore "systematic evidence review" through the State Forests Program work plan. The BOF and Oregon Department of Forestry (ODF) are interested in developing a transparent process that allows them and their stakeholders to have greater confidence in the nature and quality of scientific and technical information used in policy decision making. The BOF requested:

- Background information on SER;
- A basic comparison of medical research studies and natural resource research studies to identify any differences that affect the ability to develop methods for evaluating the quality of research evidence; and
- Proposed principles for a simplified SER-like research evidence evaluation process for ODF to use to organize, present, and synthesize scientific information for use in BOF decision making.

With this guidance, ODF contracted with the Institute for Natural Resources (INR) at Oregon State University (OSU) to investigate SERs. ODF asked INR to prepare a report on SERs, toward the goal of developing a small scale, practical SER-like approach that could be tested and adapted for use by ODF programs to objectively review and characterize technical information for decision making.

This report is the product of that effort. It is organized into sections, each of which begins with a series of statements that summarize the main points.

**Section II** reviews Oregon state mandates to use “best available science” in policymaking and summarizes ODF efforts to address these mandates.

**Section III** provides an overview of Systematic Evidence Reviews and how they are conducted in the fields of clinical medicine and public health.

**Section IV** summarizes some relevant critiques of SER methods and cautions concerning their use.

**Section V** addresses differences and commonalities between clinical medicine and ecosystem science, and the potential for applying SERs to conservation and natural resource policy questions.

**Section VI** offers some principles and guidelines about how ODF could apply components of SERs to their science assessments.

**Section VII** reflects on where efforts to conduct and utilize SERs are likely to go in the future.

The report also includes three appendices and a bibliography for further detail and reference.

## **II. How does Oregon Department of Forestry gather and assess scientific information?**

- ODF utilizes scientific knowledge in its duties to (1) manage Oregon state forests for the “greatest permanent value” to the people of Oregon, and (2) regulate commercial forest operations on non-federal forests through the Oregon Forest Practices Act.
- ODF policies are informed by science through (1) internal science reviews, (2) external reviews commissioned by the agency to assess the scientific validity of its planning documents and regulatory proposals.
- Despite well-intentioned and in many cases quite involved efforts to use the “best science available,” ODF is regularly challenged by groups suggesting that they really are not doing so.
- These challenges may stem from disputes over which pieces of technical evidence were, or were not considered, or over how particular pieces of evidence were interpreted, weighed and applied in policymaking.
- The core of disputes over use of technical information may involve broader disagreements over forest policy goals, and the appropriate course of action when outcomes are uncertain, rather than disputes over scientific evidence *per se*.

This section summarizes Oregon state legal mandates that direct forest managers to use the “best available science” when formulating policies, and reviews processes that ODF has used to carry out these directives as part of its duties to manage state forests and regulate forest practices.

### **Oregon Statutes That Address the Use of Science in Forest Planning and Management**

#### *Science and the Oregon Forest Practices Act*

Commercial forest operations on non-federal forests are regulated by the Oregon Forest Practices Act (FPA). Under the FPA, ODF is tasked with “encourage[ing] economically efficient forest practices that assure the continuous growing and harvesting of forest tree species and the maintenance of forestland for such purposes as the leading use on privately owned land, consistent with sound management of soil, air, water, fish and wildlife resources and scenic resources within visually sensitive corridors as provided by ORS 527.755 that assures the continuous benefits of those resources for future generations of Oregonians” (ORS 527.630 Policy, Oregon Forest Practices Act).

ORS 527.714 requires that any rule changes made to the FPA by the Board of Forestry must “reflect *available scientific information*, the results of relevant monitoring and, as

appropriate, adequate field evaluation at representative locations in Oregon” and that “there is monitoring or *research evidence* that documents that degradation of resources managed under ORS 527.710 (2) or (3) is likely” (emphasis added, ORS 527.714 [C]; ORS 527.714 [A]).

#### *Science and state forest planning*

OAR 629-035-0020 describes the values that Oregon state forests are to be managed for, and methods that must be employed to ensure that those values are achieved. ODF is tasked with managing state forests for “greatest permanent value”, which broadly includes “healthy, productive, and sustainable forest ecosystems that over time and across the landscape provide a full range of social, economic, and environmental benefits to the people of Oregon”.

OAR 629-035-0020 (3) requires that ODF state forest management practices “be based on the best science available” and “incorporate an adaptive management approach that applies new management practices and techniques as new scientific information and results of monitoring become available”. These legal directives are addressed through the ODF state forest management planning process.

### **ODF Processes to Assess and Incorporate Science**

ODF technical specialists regularly conduct “in-house” technical information review and synthesis work for legislative committee meetings, reports, forest plan development and presentations to the Board of Forestry. In other cases, the agency voluntarily commissions external science reviews, typically for long-term planning documents such as Forest Management Plans and Habitat Conservation Plans. Scientific evidence that pertains to ODF forest management practices has also been assessed by the Independent Multidisciplinary Science Team and the Forest Practices Advisory Committee as part of efforts to implement the Oregon Plan for Salmon and Watersheds. The ways that scientific information is reviewed in each of these contexts are summarized below.

#### **Internal science reviews by ODF technical specialists**

##### *Evidence gathering*

Most technical specialists hold advanced academic degrees in science, and gather scientific evidence in a manner that mirrors standard academic literature reviews. A typical ODF literature review includes searching databases through the State of Oregon Library (which has access to 73 separate databases) consulting topical subject libraries developed in-house by ODF staff (Procite bibliographies and hard-copy journal collections), and informal queries to regional university and agency staff members for other pertinent sources.

Technical specialists also attempt to stay abreast of new scientific information by attending scientific conferences and workshops, participating in topical listserv exchanges between experts and other agencies, subscribing to professional journal table of contents email lists (where the titles of articles in major journals are distributed via email every month), and by participating in projects with external agencies/organizations, such as the Cooperative Forest Ecosystem Research group. A substantial amount of new information is disseminated through informal social contacts with other researchers.

#### *Evidence evaluation*

ODF technical specialists rely on their academic training and work experience—“individual expertise”—to identify the most pertinent and methodologically sound research, which is then considered the “most current scientific information”. Specialists interviewed were reluctant to characterize any research or research summary as the “best available science,” because the term seems to imply a value judgment. This seemingly semantic issue characterizes the evidence evaluation problem experienced by ODF specialists and academics who focus on forest ecology- there is no widely accepted framework for them to systematically characterize the merits of a particular piece of research or body of evidence.

Internal and external peer review of ODF technical specialists’ evaluations of collected scientific evidence is common and expected, but not required. The organizational culture within ODF research nodes promotes transparency, especially between ODF and other organizations. Oregon State University College of Forestry professors regularly serve as external reviewers of ODF work products, but other agencies such as the Oregon Department of Fish and Wildlife, USDI Bureau of Land Management, and the USDA Forest Service Pacific Northwest Research Station also provide informal, *pro bono* peer reviews, often by highly qualified scientists who are generally well-respected in their academic fields. Although ODF documents rarely receive comprehensive assessment for science consistency, they are reviewed by qualified individuals based on knowledge those individuals have at the time of the review. This level of review is an accepted cultural norm within the academic community.

#### *Evidence presentation*

Interviews with ODF technical specialists and their supervisors revealed that the presentation of scientific information is often contentious. The staff takes much care to present only the conclusions of their scientific evidence evaluations to the BOF; no policy *recommendations* are included. Technical specialists do, however, often try to provide several different policy *options* that fit within the current regulatory framework, along with limited forecasts of potential risks associated with each option (to the extent that such risk evaluation is possible). The purpose of this approach is to give the BOF several feasible options when grappling with a policy decision and some characterization of potential risks associated with each. Public stakeholders regularly characterize this approach as an attempt by ODF staff to narrowly define the policy decision space available to the BOF.

## **The Forest Practices Advisory Committee**

The Forest Practices Advisory Committee (FPAC) was a citizens' advisory group created to help the BOF determine “to what extent changes to forest practices are needed to meet state water quality standards and to protect and restore salmonids” (Oregon Executive Order No. EO 99-01. 1999, Section 3 [c]). The FPAC was in existence from January, 1999 to June, 2000 and consisted of 13 members representing small and large forest landowners, environmental groups, sport fishing organizations, logging and commercial fishing interests, local governments and a labor union.

Four “issue papers” were completed jointly by FPAC and ODF. These papers addressed the major technical issues that the committee was tasked with—fish passage, forest roads, landslides, and riparian function—by outlining current scientific findings, watershed-scale effects, a description and evaluation of current applicable voluntary and regulatory measures, and suggestions for additional measures. Each paper was peer reviewed by a number of scientists—selected by FPAC with the ODF assistance—from across the Pacific Northwest with expertise specific to each issue. The IMST also reviewed each section and offered input and recommendations. The FPAC also considered the IMST<sup>1</sup> forestry report (IMST 1999) as part of their technical review process.

The FPAC deliberations and resulting technical papers were used to answer specific questions and build institutional knowledge. In 2000, FPAC proposed 24 final recommendations following public and scientific input. Based on these recommendations (and those from a similar group focused on eastside watersheds), as well as recent monitoring information, the BOF instituted regulatory and management changes including landslide mitigation regulations, new forest road rules, and updated harvesting guidelines for erosion-prone areas (ODF 2002). The BOF is now considering riparian buffer rule changes.

## **Independent scientific reviews of state forest planning documents**

ODF has sought and conducted independent scientific reviews of several major forest management proposals. A good example is the 1998 review of its Draft Western Oregon State Forests Habitat Conservation Plan (HCP). Dr. John Hayes, who chaired the HCP review panel, stated that “in assembling the team [of reviewers], an attempt was made to have a diversity of perspectives represented, preferably within each disciplinary area; this was an important criterion in the selection process” (Hayes 1998, p. 2). Four independent scientists selected the reviewers, without consideration of ODF or external interest group input. Twenty-six individuals served as reviewers for the Draft HCP (Hayes 1998, p. 3).

---

<sup>1</sup> In 1997, the bill creating the Oregon Plan for Salmon and Watersheds (Oregon Plan) established the Independent Multidisciplinary Science Team (IMST) to serve as a standing independent review panel on science related to salmon protection and watershed health. The IMST is comprised of seven scientists with expertise in soil science, stream ecology, fish biology, environmental physiology, resource management and forestry, and two support staff researchers (IMST 2005). Team members are jointly selected by the Governor, Senate President, and Speaker of the House. IMST reports to the Governor's Natural Resource Office and the House Agriculture and Natural Resource Committee's Subcommittee on Water.

In directing his team of reviewers, Hayes noted that “In some cases, there are few data available to answer some of the questions. In many cases, the best we have to go on is the professional judgment of leading scientists in the field. Your opinion is valued. In these cases, please identify to what extent your judgment is based on data.” (Hayes 1998, p. 248.) Reviewers accommodated this request to varying degrees.

Overall, “the reviewers expressed general favor with the overall direction presented in the HCP,” but “in every aspect reviewed some suggestions for improvement or areas o[f] concern were noted by the reviewers” (Hayes 1998, p. 6). For example, reviewers specifically identified aquatic and riparian conservation sections as “based on scientific principles, but...inadequately or superficially presented,” if presented at all (Hayes 1998, p. 29). Other critiques found the literature reviews, as presented, inadequate. Hayes also indicated that “in a minority of cases, reviewers argued that scientific justification for the approach used [was] lacking” (Hayes 1998, p. 29). Comments provided by the reviewers suggest that, in these cases, either a selective approach to scientific evidence was used or that review of the evidence was incomplete (Hayes 1998, p. 31).

Reviewers also commented that the HCP and the Forest Management Plan for Oregon are complementary and seem to integrate ecosystem management well, but the mechanisms for their overall synthesis were unclear. One reviewer in particular, Dr. Thomas Hinckley, noted that “without the [Forest] Management Plan document, it is not possible to have confidence in the Habitat Conservation Plan document” (Hayes 1998, p. 31).

Despite concerns raised by the reviewers, it is notable that ODF sought rigorous external review of its Draft HCP.

#### *Summary of ODF Process for Independent Scientific Reviews*

Based on descriptions of reviews conducted for the Elliott State Forest Management Plan (2004), Southwest Oregon State Forest Management Plan (2001), and the Northwest Oregon State Forests Management Plan (2001), the ODF process for conducting independent scientific reviews is, generally, as follows:

- ODF technical specialists create a draft planning document that is informed by scientific evidence they identify using standard academic literature search techniques.
- An ODF panel recommends qualified scientists to review the draft document and creates guiding questions for the reviewers.
- Reviewers are selected (e.g. by an ODF panel, steering or plan oversight committee) and drafts of the document are distributed.
- Reviewer comments are submitted to ODF.

- ODF staff make any changes to the document they deem necessary.
- Public drafts are distributed for further comments.

### **Integrating Scientific Knowledge into ODF State Forest Management Plans**

To direct management of Oregon’s state forests, ODF produces three nested levels of plans: (1) long-range *forest management plans* and *habitat conservation plans* that provide core values and management strategies that are used to develop (2) intermediate-range *district implementation plans*, and (3) annual *district operations plans* and *budgets* (Figure 1, ODF 2001, p. S-3):

**Figure 1: Structure of Oregon Department of Forestry Management Planning Process**



The ways in which scientific and technical information are assessed and incorporated into these plans are outlined below.

#### *Long Term Plans*

Long term Forest Management Plans (FMP) and Habitat Conservation Plans (HCP) are the foundation for ODF forest management decisions, and provide guidance on “best available science,” monitoring regimes, timber harvest levels and landscape-level ecosystem management.

Guiding scientific principles are established by ODF technical specialists in a document that forms the basic framework for an FMP or HCP, which is refined internally by ODF and then evaluated by an independent, external scientific review committee. ODF solicits these reviews voluntarily, doing so is not mandated. ODF then incorporates or addresses any changes suggested by reviewers that they feel are warranted, and releases an updated draft plan for public comment. External state and federal agencies may also provide input at this time. ODF then incorporates or addresses suggested changes as they feel

necessary and releases a compilation of public comments received and the agency's responses to them.

For example, the 2004 Elliot State Forest Management Plan began with an ODF wildlife biologist, silviculturist, and forest ecologist, the *Core Planning Team*. This team drafted sections of the *Resource Management Strategies*, which presented relevant scientific information identified through “extensive review of the scientific literature and many meetings and discussions with experts from other management agencies (e.g. ODFW) and universities” (ODF 2004, p. 2). Using the Resource Management Strategies as a guide, the Core Planning Team, other ODF technical specialists, and resource managers constructed a draft FMP. The Core Planning Team then compiled a list of potential reviewers to perform an independent scientific review of the draft FMP. Selection criteria were reviewers who:

- Were recognized within the scientific community as an authority in a field appropriate to the review.
- Had a background with recent research in an area appropriate to the review.
- Had a record that includes recent publications pertaining to an area appropriate to the review, published in peer-refereed publications.
- Were available to conduct a review in a timely fashion.

Eight reviewers with expertise in forest ecology, silviculture, aquatic ecology, wildlife and landscape management were selected by Elliot State Forest Steering Committee, which consisted of ODF state and district level staff, representatives from Oregon Departments of Fish and Wildlife, State Lands, and Justice, and a Coos County Commissioner. The Core Planning Team and Steering Committee then “prepared a set of questions to help structure the review,” where “reviewers were encouraged to critique any aspect of the [FMP] not covered by the questions if they so desired, and to refrain from answering any question that they felt unqualified, unable, or unwilling to answer” (ODF 2004, p. 3). After reviewers submitted their comments, a working group of ODFW and ODF technical specialists synthesized the responses, addressed criticisms and released the results as the *Scientific Review of the Elliot State Forest Management Strategies*.

### *Medium Term Plans*

District-level Implementation Plans (IPs) are developed “periodically” using established Forest Management Plans as their primary guide for “best science available” and resource management decisions. The first versions of these plans were approved in 2003. These plans guide how the overarching FMPs will be implemented in a given district. Management activities for the next ten years and estimates of a district's progress towards achieving FMP goals are described in IPs.

### *Short Term Plans*

Annual Operations Plans (AOPs) are developed every year by each management region (“district”). AOPs are based on the guidelines established in IPs, and are used to describe specific action plans for a given year of operations in a district. Public comment periods are a means for the public to see what activities are being planned, and to review whether plans are consistent with higher order planning documents (IP and FMP). Federal, state, and local agencies may post comments during this time as well.

### *Responding to public comments*

At the end of each comment period, comments from the public as well as other state and federal agencies are summarized, along with ODF responses to each, into a matrix that is then publicly distributed (e.g. ODF 2005). In the past, despite the transparency of this process, a number of concerns over the scientific validity of ODF proposals and operations have been raised by the public, sometimes without direct response from ODF on the scientific merits of the concerns.

## **Uncertainty and adaptive management in Oregon forest policy**

ODF management plans and forest practice regulations are scientifically based and have considerable public, external agency and scientific community input. Still, there are usually significant levels of uncertainty and risk associated with each policy decision. To some extent, uncertainty simply reflects the non-linear, stochastic way ecosystems function, but characterizing this uncertainty to stakeholders and policymakers who seek deterministic solutions remains a major challenge (Bradshaw and Borchers 2000). In most cases, accurate risk assessment is difficult or impossible because of the inherent complexity of landscape-level management and the paucity of relevant scientific information. Uncertainty is a “given”, so ODF management plans incorporate tenets of adaptive management (ODF 2001, p. 5-14):

- (1) Assess the problem
- (2) Design experiments and monitoring plans
- (3) Implement plans
- (4) Monitor
- (5) Evaluate outcomes
- (6) Adjust activities and policies

In addition to the basic scientific *research evidence* used to develop management plans, adaptive management also requires *monitoring evidence* to assess whether science-based assumptions about management outcomes turn out to be valid. When scientific evidence is equivocal, an adaptive management strategy can make it easier to move forward with management decisions, as long as robust monitoring regimes are in place to provide the necessary feedback. Availability of sufficient funding and personnel to conduct effective monitoring is a challenge for ODF. Over time this may result in a shortage of the kind of specific, relevant, timely scientific evidence upon which adaptive management depends.

### *Challenges to ODF use of science*

Detailed reviews of the scientific merits of ODF proposals occur regularly, but the manner in which technical specialists identify and interpret relevant evidence could be more transparent. The agency also does not have a process for assessing the scientific validity of their final management decisions. Despite honest and in many cases quite involved efforts to use the “best science available,” ODF is often challenged by groups suggesting that they really are not doing so. These challenges may stem from:

- Disputes over which pieces of technical evidence were, or were not considered,
- Differences of opinion about how to interpret and weigh particular pieces of evidence, or
- Disagreements over goals and the appropriate course of action when outcomes are uncertain, rather than disagreements about scientific evidence *per se*.

The Systematic Evidence Review (SER) process has been suggested as a potential method for assessing and synthesizing scientific and technical information in a way that will be more readily accepted by stakeholders. SERs and their components would also mesh well with effective adaptive management strategies.

The remainder of this report is dedicated to examining SERs and their potential applicability to natural resource management in Oregon.

### **III. What is a “Systematic Evidence Review” and how do SERs work?**

- An SER is rigorous, transparent, reproducible process for assessing scientific and technical information, used primarily in clinical medicine.
- An SER focuses tightly on a specific question, or small set of questions, which frame decisions about what evidence is relevant to the review, and what is not.
- SERs differ from traditional literature reviews in their use of pre-established, explicit protocols for finding, screening, grading and integrating primary research studies.
- SERs are designed to be as comprehensive, exhaustive and objective as possible, which means they are typically time consuming and expensive.
- Systematic Evidence Reviews and evidence based medicine have been described as a “paradigm shift” in healthcare, but there is still considerable debate about how SERs are conducted and used.
- Costs of medical SERs range from \$50,000-\$250,000. A natural resource SER may cost considerably less because the evidence base is likely to be smaller, but this would depend on the nature of the question.

#### **Overview of the SER process**

A systematic evidence review (SER) is a process for summarizing all available evidence that pertains to a specific question, usually about a healthcare treatment, diagnosis or preventive service. SERs utilize rigorous and explicit methods to identify, select and appraise relevant primary studies, and to collate and report their findings. A completed SER is used to transfer reliable science-based information to a wider audience of healthcare practitioners, professionals and consumers, and to develop standardized clinical practice guidelines for the medical treatment reviewed. SERs are the key component of evidence-based medicine (EBM), “the integration of best research evidence with clinical expertise and patient values” (Sackett et al. 2000).

SERs are scientific investigations in themselves, with pre-planned methods and a set of original studies as their "subjects." (Greenhalgh 1997.) Their purpose is not to just to be comprehensive, but to answer a specific question, to reduce bias in the selection and inclusion of studies, to appraise the quality of included studies, and to summarize them objectively (Cook et al. 1997). SERs are characterized by documented protocols with specific criteria for evaluating (1) whether a study is relevant to the question posed and should be included in the review, and (2) the robustness and quality of included studies and explanatory potential of each study design.

SERs attempt to mitigate the potential subjectivity of traditional literature reviews, which are typically unsystematic, not comprehensive, and susceptible to selection bias or publication bias (the tendency of statistically insignificant or inconclusive results to not be published in scientific journals). In addition to peer-reviewed publications, an SER strives to minimize bias by exhaustively searching for, recognizing and assembling any unpublished grey literature and negative evidence that can be identified. SERs also differ from literature reviews because several reviewers often work independently to screen papers for inclusion and assess their quality. Multiple reviewers can then “compare notes” to increase the consistency and reliability of the review.

In the medical field, even “small” SERs are likely to involve several reviewers screening hundreds of abstracts. As a result, SERs commonly require more time, staff, and money than traditional reviews. But SERs are not simply bigger than traditional literature reviews, they are qualitatively different in ways that make them more robust, transparent and reproducible (Table 1).

**Table 1. Systematic reviews and traditional reviews compared (Petticrew 2001).**

	<b>Good quality systematic reviews...</b>	<b>Traditional reviews...</b>
<i>Deciding on review question</i>	Start with clear question to be answered or hypothesis to be tested.	May also start with clear question to be answered, but more often involve general discussion of subject with no stated hypothesis.
<i>Searching for relevant studies</i>	Strive to locate all relevant published and unpublished studies to limit impact of publication and other biases.	Do not usually attempt to locate all relevant literature.
<i>Deciding which studies to include and exclude</i>	Involve explicit description of what types of studies are to be included to limit selection bias on behalf of reviewer(s).	Usually do not describe why certain studies are included and others excluded.
<i>Assessing study quality</i>	Examine in a systematic manner methods used in primary studies, and investigate potential biases in those studies and sources of heterogeneity between study results.	Often do not consider differences in study methods or study quality.
<i>Synthesizing study results</i>	Base their conclusions on those studies which are most methodologically sound.	Often do not differentiate between methodologically sound and unsound studies.

SERs aim to provide the best available evidence on likely outcomes of various actions and, if evidence is unavailable, highlight areas where further original research is required. SERs *support* decision-making by providing the most independent, unbiased and objective assessment of evidence possible. They are not designed to make decisions on behalf of users. SERs have particular value for those lacking the knowledge to critically appraise primary research findings themselves, and in synthesizing results of separate studies addressing the same research question, which may have conflicting findings or commonalities that may otherwise not be widely recognized (Cook et al. 1997).

Several different methods are used to synthesize existing research in medical SERs, primarily *meta-analysis* of randomized controlled trials and *narrative synthesis*. Meta-analysis uses statistical techniques to combine findings of studies that meet the inclusion criteria. Meta-analysis is not possible or appropriate when studies are limited in number or very diverse, or when outcome measures used in the studies differ.

A narrative synthesis brings together study results and looks at similarities and differences between the studies and their outcomes. A narrative synthesis does not quantify an average effect size, although it can look at effectiveness and takes study methodological quality into account. Narratives typically use tables to provide a systematic and consistent record of information from the studies in the review (Boaz et al. 2002).

### **History of Systematic Evidence Reviews**

- Since emerging in the 1980's the SER approach has been widely adopted in the fields of clinical medicine and public health, and continues to expand rapidly.
- The largest international entity that conducts and disseminates SERs is the Cochrane Collaboration, which maintains a database of over 2000 SERs, develops and refines review methods, and offers training on conducting SERs.
- In the United States, SERs are conducted and disseminated by the Agency for Healthcare Research and Quality (AHRQ) through its 15 designated Evidence-based Practice Centers. Oregon Health Sciences University in Portland is one such center.
- Interest is growing in adapting the SER approach to other areas of public policy, including wildlife conservation, but SERs are not well suited for all policy areas, or to all questions within a particular field.

SERs and evidence-based medicine are rooted in the 1970's, when British epidemiologist Archie Cochrane spurred a sea change in medicine by pointing out that many health care treatment decisions were not based on reliable reviews of available evidence but on unsystematic and selective interpretation of the scientific literature, expert opinion and trial and error (Cochrane 1979). The crux of the problem was poor transfer of the huge volume of medical science to medical practice: "There is sufficient evidence to suggest that most clinicians' practices do not reflect the principles of evidence-based medicine but rather are based upon tradition, their most recent experience, what they learned years ago in medical school, or what they have heard from their friends. The average physician is said to read scientific journals approximately two hours per week, and most are likely overwhelmed by the volume of material confronting them." (Eisenberg 2001, p. 1.)

The first SERs were conducted on care during pregnancy and childbirth. By the 1980s, the benefits of SERs were being acknowledged by significant and growing numbers of practitioners and researchers. The Cochrane Collaboration (CC) was established in 1993 and now includes about 50 international collaborative Cochrane Review Groups (CRGs). Each group focuses on an important area of health care, and critically reviews all clinical trial evidence related to that topic. The CC is non-profit and strives for independence from commercial and other interests that may interfere with honest appraisal of evidence. The major product of the CC is the electronic Cochrane Database of Systematic Reviews,

in which new and updated SERs are published by the CRGs four times each year. SERs of randomly controlled trials (RCTs) are prepared and published in regularly updated databases with abstracts available free on the internet ([www.cochrane.org](http://www.cochrane.org)).

The Cochrane Database was first published in 1995 with 36 reviews and now contains the full text of more than 2000 reviews, each of which is kept up-to-date as new evidence accumulates. Currently, there are also more than 1400 published protocols for reviews in progress. The protocols set out how the reviews will be done and explicitly describe the methods to be used. Hundreds of newly completed reviews and protocols are added each year. In addition, a few hundred existing reviews are updated substantively enough to be equivalent to new reviews, and several hundred more are brought up to date by the addition of new information.

Another group of experts (the Cochrane Methods Group) researches methods, sets protocols and runs training on how to conduct SERs. Cochrane centers are usually funded through government grants, charitable donations and other noncommercial funding sources, and coordinate and support the work of the review groups and methods groups. Individual CC entities are funded by a large variety of governmental, institutional and private funding sources, and are bound by organization-wide policy limiting uses of funds from corporate sponsors. Today, there are more than 12,000 people working within the CC in over 90 countries. Half are authors of Cochrane Reviews and most are volunteers. About 1,000 Cochrane reviewers are Americans. The number of people involved with the CC has increased by 20% every year for the last five years. (Fazey and Salisbury 2002, Cochrane Collaboration 2005.)

Through the use of SERs, individual and policy-level health care decisions are increasingly being made on research-based evidence rather than on expert opinion or clinical experience alone. For example, SERs can help consumers decide whether to have a particular diagnostic procedure or select among several treatment alternatives. Policymakers can use SERs to help choose what types of health care to provide. (Bero and Jadad 1998.)

Prospects for wider use of SERs in fields such as natural resources are also being explored. A group at Australia National University investigated applying SERs to environmental management (Fazey et al. 2002) and biological conservation management (Fazey et al. 2004). In 2003, the Center for Evidence-Based Conservation (CEBC) was established at the University of Birmingham in the U.K., pioneering the use of SERs to support decision making in these fields. Section VI covers this topic in greater depth.

#### *Evidence-based practice centers in the United States*

The largest North American sources of high-quality SERs are the federal Agency for Healthcare Research and Quality (AHRQ) and its 15 designated Evidence-based Practice Centers (EPCs). Since the early 1990s the AHRQ has provided research support and policy guidance in healthcare research through the EPC program, which includes the Center for Evidence-based Policy at Oregon Health and Sciences University (OHSU). The AHRQ seeks to advance the medical field's understanding of how best to ensure that

reviews of the clinical or related literature are scientifically and clinically robust. (Fox 2005, West et al. 2002.)

Including the OHSU center, the 15 EPCs at various public and private universities and institutions in the U.S. and Canada have conducted over 100 SERs for professional societies, health plans, insurers, employers, and patient groups on topics such as cancer, pediatrics, endocrinology, cardiology, mental health, health care quality improvement and patient safety. Federal and state agencies, professional societies, health delivery systems, providers, payers and consumers can all benefit from SERs.

#### *SER Applications in Oregon*

An example of integration of evidence-based practice into Oregon state agency operations is a 2001 law which required the state to establish a Medicaid preferred drug list based first on clinical effectiveness as demonstrated in peer-reviewed research and secondly on cost. Another example is a 2003 legislative requirement that the Oregon Office of Mental Health and Addiction Services show during the budgeting process that an increasing percentage of its funding was spent on cost-effective, evidence-based practices.

The Oregon EPC is currently involved in three major projects. The Drug Effectiveness Review Project is a collaborative effort between the EPC and several states to obtain best available effectiveness comparisons between drugs in the same class and apply them to public policy, especially development of Medicaid preferred drug lists. The preventive service project assesses the merits of preventive measures, including screening tests, counseling, immunizations, and preventive medications. Finally, the food-related health claims project supports a federal Food and Drug Administration initiative to give consumers scientifically based information about the likely health benefits of food.

#### *Costs and funding of SERs*

Cochrane Collaboration leaders estimate that the cost of an average SER—mainly the time and direct expenses of reviewers and support staff—is roughly \$50,000. Reviews conducted by EPCs can cost \$250,000 or more because they address broader questions (Fox 2005). John Santa of the OHSU Center for Evidence-based Policy estimates that average costs for full medical SERs with three questions are \$150,000-\$200,000 (Santa 2005, personal communication). These estimates do not include the costs of sustaining the groups of researchers who produce and edit reviews. The Cochrane Collaboration Funders Forum recently estimated that, worldwide, core funding explicitly earmarked for SERs totals around \$20 million a year, with most of it supplied by the governments of the United Kingdom, Australia and Scandinavia.

## The Systematic Evidence Review Process in clinical medicine

- SERs require specific, tightly focused review questions to (1) clarify the purpose and delimit the scope of the review, and (2) strengthen linkages between the questions and subsequent steps in the review process.
- SERs use explicit protocols that spell out in advance exactly how evidence will be gathered, assessed, collated and summarized. Documenting all steps in the SER ensures that it is transparent, replicable and can be updated later if necessary.
- SERs are characterized by vigorous and thorough efforts to compile all available research and technical information that pertains to the review question(s), including unpublished and “gray” literature.
- Quality assessment of individual studies is a key characteristic of SERs, providing more rigor than traditional reviews. Quality assessments are used to (1) decide whether a relevant study should be included in the review, and (2) to rank included studies in an *evidence quality hierarchy*, usually based on study design and methods.
- Quality assessment is labor intensive and remains controversial. Random controlled trials and other tightly controlled study designs are favored in medical SERs, but there is no general consensus on standardized quality assessment criteria.
- Synthesis consists of tabulation of study characteristics, quality and outcomes for the primary purpose of investigating whether results are consistent across included studies, and if not, investigating reasons for apparent differences.
- A *narrative synthesis* is used to qualitatively compare and synthesize included studies. A narrative synthesis may be all that is possible if differences in population, intervention, outcome measures, designs and quality preclude meta-analysis.
- Quantitative synthesis, such as *meta-analysis* may be used to statistically combine study findings, as long as the studies are similar with regard to population and intervention under study, outcome measures, study design and quality.
- The strength of a body of evidence (all included studies) is assessed by examining aggregate study *quality*, the *quantity* of evidence (number of studies, sample size), *consistency* of findings across studies, and *coherence* of the evidence as a whole.
- SERs obtain their rigor partly through efforts to identify and explicitly acknowledge ways in which bias could enter into and affect results in primary studies, and to reduce bias during selection and review of these studies.

- SERs can be, and often are updated when new information emerges that will significantly strengthen or change the outcome of an existing review. Explicit documentation of how the review was conducted makes updating possible.
- A key use of SER information is developing *clinical practice guidelines*: “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.”
- More progress has been made in rigorously assessing available evidence through SERs than in incorporating this evidence into policy due to problems accessing clinical guidelines and because many patients prefer tailored care.

What follows is a description of steps involved with setting up and conducting an SER to address a question in the medical field, based primarily on descriptions by Boaz et al. (2002) and Khan et al. (2003). Understanding these details should be useful to groups who wish to apply an SER-like process to ecosystem management questions.

#### **A. Focus on answering a specific question(s)**

SERs target specific questions, which are formulated before the review commences. Developing the question(s) is a critical but often complex and time consuming part of the SER process. Review questions need to be clearly defined and answerable in scientific terms because the questions also generate the search terms and determine relevance criteria. Reviewers must make a dichotomous (yes/no) decision as to whether each potentially relevant paper will be included or, alternatively, rejected as "irrelevant." (Greenhalgh 1997.) Concentrating on specific questions or problems gives SERs clarity of purpose and content that enhances their usefulness. Review questions sometimes need to be modified in light of accumulated research as the review progresses.

Review questions can be formulated explicitly according to 1) a *specific population and setting* (e.g., elderly outpatients), 2) the *condition of interest* (e.g., hypertension), 3) a *specific test or treatment* (e.g., pharmacologic management), and 4) one or more *specific outcomes* (e.g., cardiovascular and cerebrovascular events and mortality). For example, a well-formulated SER question is, “Does pharmacologic treatment of hypertension in the elderly prevent strokes and myocardial infarctions or delay death?” (Cook et al. 1997.)

Khan et al. (2003) suggest similar components for an SER question: *population, intervention, outcome* and *study design*. Use of these components to develop an example targeted question is illustrated in Figure 2.

**Figure 2: Developing a Targeted Systematic Evidence Review Question**

<b>Population</b>	<b>Intervention</b>	<b>Outcome</b>	<b>Study Design</b>
Description of the group of participants or patients about whom the evidence is being sought in the review.	Actions or alternatives being considered for the population.	Measures of what the interventions are intended to achieve, e.g. avoiding illness or death, accurate detection or prediction of disease.	Consider how a study could be designed to examine the effect of the interventions.
<b>Restructure the “free form” question, “among postmenopausal women with abnormal bleeding, does a pelvic ultrasound scan exclude uterine cancer accurately?” to specifically address each component:</b>			
“In postmenopausal women, within a community setting, with bleeding...	...does a uterine ultrasound scan test accurately predict...	...histological diagnosis of uterine cancer?”	Study that recruits women from a relevant population, uses the test (scan) and a reference standard investigation to confirm/refute presence of cancer (histology), and determines the accuracy with which it identifies cancer.

It may also be useful to specify *comparators* that the review is focusing on, i.e. is the intervention (e.g. a particular drug) being compared to *no intervention at all* (no drug; or a placebo) or one or more *alternative interventions* (other drugs or treatments designed to mitigate the same condition)? Specificity about comparators will help delimit the scope of the review, and with decisions about which studies are relevant and which are not.

A review question that is framed in this explicit manner strengthens linkages between the question and subsequent steps of the SER process. It is important to consider how populations, interventions, and outcomes might vary among existing studies because these differences are key to defining study selection criteria and planning tabulation of findings. They are also relevant in understanding the reasons for variation in effects of interventions from study to study and in exploring the applicability of findings.

## **B. Developing an SER protocol**

The SER protocol explicitly spells out how the review will identify, appraise and collate evidence, helping reviewers think through the different review stages *a priori*, to anticipate problems and plan for them. Documenting what the reviewers intend to do and how also promotes transparency and replicability, and allows the SER to be updated by others at a later date.

Development and approval of the protocol comes first, before proceeding with the review itself. An SER will be less biased if the reviews questions are well developed, and the methods which will be used to answer them are decided before gathering evidence and drawing inferences. A protocol reduces the chance that study selection and analysis will be influenced by a presumption of the findings (Khan et al. 2001). Protocol components that should be documented include:

1. Background- context for the questions and why they are important
2. Objective of the Review
  - 2.1 Primary question
  - 2.2 Secondary questions (if applicable)
3. Methods
  - 3.1 Search strategy
  - 3.2 Study inclusion criteria and procedures
  - 3.3 Study quality assessment- e.g. checklists and procedures
  - 3.4 Data extraction strategy
  - 3.5 Data synthesis
  - 3.6 Project timetable
4. Potential Conflicts of Interest and Sources of Support
5. References

SER protocols are typically open for external review by stakeholders, subject experts and the user-community prior to being finalized, and are often published (e.g. in the Cochrane Library, on the web site of an electronic journal, or on the reviewers' web site). The protocol will form part of the SER final report and will help with management of the review (Khan et al. 2003).

A key purpose of the SER protocol is to determine a search strategy that will identify all research relevant to the question. The search must be sufficiently rigorous and broad to ensure that all studies eligible for inclusion are identified even though they may be excluded during the quality assessment phase of the review. Search protocols therefore balance *sensitivity* (identifying all relevant information) with *specificity* (excluding irrelevant information). A systematic, comprehensive search improves credibility and is key to the validity of review conclusions.

If unforeseen exigencies occur, the SER protocol may require amendment during the course of the review. If so, the reviewers should document where and why modifications were made to ensure that the process is transparent and repeatable.

### **C. Identifying relevant information**

SERs take a wide ranging, comprehensive approach to searching for relevant studies. Current technology allows for global searches of research databases to identify as many relevant studies as possible, not just those that are the most well known or have the most significant results. Initiatives such as the Cochrane Collaboration Methods Group have invested in improving search techniques and tools to facilitate this process. It may not be possible to locate all research on a given topic, but a good SER explains specifically how studies were identified and obtained, and highlights any known knowledge gaps.

*Literature search process:*

1. Generate a list of potentially relevant citations from sources such as
  - electronic bibliographic databases
  - search using a combination of terms
  - search for particular study designs of interest
  - avoid language restrictions (e.g. English only searches)
  - search multiple databases
  - reference lists of known primary and review articles
  - peer reviewed journals
  - gray literature
  - conference proceedings
  - ongoing research
  - knowledgeable professionals
  
2. Select relevant studies, i.e. those that address the questions being posed in the review. Obtain full manuscripts of all potentially relevant studies. Count studies with duplicate publications only once. Track down less accessible studies to reduce publication bias.
  
3. Sift through the manuscripts and make final inclusion/exclusion decisions based on the explicit study selection criteria defined in the SER protocol.

#### **D. Assessing research quality**

Peer review is a hallmark of scientific practice that is intended to (and does) improve the quality of research and reporting on it. However, as Steinberg and Luce (2005) remind us, publication in a peer-reviewed journal does not guarantee that the study was methodologically sound and well-conducted, that data analysis was performed correctly, or that study results were interpreted properly. In addition, a means is needed to compare different studies on the same topic. For these reasons, most SERs have protocols for examining study quality. Quality assessment gives SERs more rigor than traditional literature reviews, but requires considerable expertise and is labor intensive.

Using criteria that are clearly described in the SER protocol, reviewers appraise the quality of relevant studies to 1) decide which warrant inclusion in the review and, 2) rank those studies that are included for their relative quality. Randomly controlled trials (RCTs) are widely regarded as the highest quality evidence concerning the efficacy of a medical practice and most SERs focus on these. However comparative and observational studies (e.g. cohort, case-control and case-series studies) may also add important, usable evidence to the SER. In other cases, RCT data may simply be unavailable.

Recognition of the need to include different types of evidence in the overall evidence base spurred development of ranking systems that give more weight to the best evidence, and less weight to less certain evidence. These *evidence quality hierarchies* are based on research methodology and are commonly used as study screening protocols in SERs. "Methodologic quality" is defined as "the extent to which all aspects of a study's design and conduct can be shown to protect against systematic bias, nonsystematic bias and

inferential error" (Lohr and Carey 1999) or alternatively, "the extent to which a study's design, conduct, and analysis have minimized selection, measurement, and confounding biases" (West et al. 2002).

Study design is important, but it is only one quality criterion. How the study was conducted is also critical- results of poorly conducted randomized experiments may be less helpful than better conducted quasi-experimental studies. Moreover, in many medical sub-disciplines, there may be numerous individual studies using a single, widely agreed upon design. In these cases, more specific criteria are used to assess study quality.

No study, or set of quality criteria, is perfect. Quality hierarchies based on study design, and other, more specific criteria should be used to guide and focus what are ultimately judgment calls on the part of reviewers regarding whether to include or exclude a study, and then rank studies that are included. Table 2 compares several evidence quality hierarchies used in clinical medicine SERs, and one that was slightly adapted for use in conservation field SERs. Appendix 1 shows a more detailed summary of the evidence quality hierarchy provided by Khan et al. (2003) and quality criteria used by the British Medical Journal to compare different SERs and clinical trials on the same topic.

**Table 2: Hierarchies of Evidence Quality**

Kahn et al. (2003)	Stevens & Milne (1997)	UK Center for Evidence-based Conservation: “Hierarchy of quality of evidence modified for conservation use”	D.C. Hadorn, D. Baker, J.S. Hodges, & N. Hicks, 1996. Rating the Quality of Evidence for Clinical Practice Guidelines, <i>Journal of Clinical Epidemiology</i> , 49.	
<p>I: Experimental study: use of different interventions among participants is allocated by researcher</p> <p>a. Randomized controlled trial</p> <p>b. Experimental study without randomization</p> <p>II: Observational study with control group</p> <p>a. Cohort study</p> <p>b. Case-controlled study</p> <p>III: Observational study without control group</p> <p>a. Cross-sectional study</p> <p>b. Before and after study</p> <p>c. Case series</p> <p>IV: Case reports</p>	<p>I: Strong evidence obtained from at least one properly designed, randomized controlled trial of appropriate size</p> <p>II-1: Evidence obtained from well-designed controlled trials without randomization</p> <p>II-2: Evidence obtained from well-designed co-host or case-controlled analytic studies, preferably from more than one center or research group</p> <p>II-3: Opinions of respected authorities based on clinical evidence</p> <p>III: Opinions of respected authorities based on clinical evidence, descriptive studies, or reports of expert committees</p> <p>IV: Evidence inadequate owing to problems of methodology e.g., sample size, length or comprehensiveness of follow-up, or conflicts of evidence</p>	<p>I: Strong evidence obtained from at least one properly designed, randomized controlled trial at appropriate size</p> <p>II-1: Evidence from well designed controlled trials without randomization</p> <p>II-2: Evidence from a comparison of differences between sites with and without (controls) a desired species or community</p> <p>II-3: Evidence obtained from multiple time series or from dramatic results in uncontrolled experiments</p> <p>III: Opinions of respected authorities based on qualitative field evidence, descriptive studies or reports of expert committees</p> <p>IV: Evidence inadequate owing to problems of methodology e.g. sample size, length or comprehensiveness of monitoring, or conflicts of evidence</p>	<p>I-1. Meta-analysis or systematic reviews of multiple well-designed, randomized, controlled clinical trials</p> <p>I-2. Well-controlled, randomized clinical trials with adequate sample size</p> <p>I-3. Well-designed trial without randomization (single group pre/post, cohort, time series studies)</p> <p>II-1. Well-conducted, systematic review of non-experimental design studies</p> <p>II-2. Well-conducted case-control studies</p> <p>II-3. Poorly controlled (flawed randomized studies) or uncontrolled studies (correlational descriptive studies)</p> <p>II-4. Conflicting evidence or meta-analysis showing a trend that did not reach significance. NIH Consensus Report. Published practice guidelines, e.g. by professional organizations, healthcare organizations, or federal agencies</p> <p>III-1. Qualitative designs: case studies, opinions of expert authorities, agencies or committees</p>	<p>Research-Based Evidence</p> <ul style="list-style-type: none"> <li>• Meta-analysis of multiple controlled clinical trials</li> <li>• Experimental studies, such as well-controlled randomized clinical trials</li> <li>• Systematic reviews of all types of research</li> <li>• Multiple non-experimental studies, including descriptive, correlational, and qualitative research</li> <li>• Published Evidence Based Practice guidelines, e.g. by professional organizations</li> </ul> <p>Non-Research-Based Evidence</p> <ul style="list-style-type: none"> <li>• Case studies</li> <li>• Program evaluation, quality improvement data, or case reports</li> <li>• Opinions of experts (e.g., standards of practice, practice guidelines)</li> </ul>

*Quality assessment checklists* are a straightforward way to present information about study quality. Information is shown in table form, with quality criteria, or domains listed along one axis and included studies listed along another. Each study is then checked for whether or not each criterion is met. For example, study questions and population under study should be adequately specified, randomization and blinding properly conducted, interventions and outcomes properly specified, statistical analysis properly conducted, and results and discussion should be supported by the analysis.

Studies are then ranked within each study type according to the proportion of total quality items they comply with. If more detail is needed (e.g. caveats or comments) separate tabulation forms can be generated for each study. Studies identified in the search phase but not included in the review can be listed in the same or separate tables, with explanations of why they were excluded from consideration. Table 3 shows examples of quality assessment checklist criteria.

**Table 3: Important Domains and Elements for Systems to Rate Quality of Individual Articles**

Randomized Clinical Trials	Observational Studies
<ul style="list-style-type: none"> <li>• Study question</li> <li>• <i>Study population</i></li> <li>• <i>Randomization</i></li> <li>• <i>Blinding</i></li> <li>• <i>Interventions</i></li> <li>• <i>Outcomes</i></li> <li>• <i>Statistical analysis</i></li> <li>• Results</li> <li>• Discussion</li> <li>• <i>Funding or sponsorship</i></li> </ul>	<ul style="list-style-type: none"> <li>• Study question</li> <li>• Study population</li> <li>• <i>Comparability of subjects</i></li> <li>• <i>Exposure or intervention</i></li> <li>• <i>Outcome measurement</i></li> <li>• <i>Statistical analysis</i></li> <li>• Results</li> <li>• Discussion</li> <li>• <i>Funding or sponsorship</i></li> </ul>

(Key items are in italics. Agency for Healthcare Research and Quality/West et al. 2002)

Note that the AHRQ does not list “results” and “discussion” as key quality assessment domains. This is because it is assumed that reviewers will extract the original data and decide for themselves if the results were interpreted correctly as part of the quality assessment. In rigorous SERs, the original data are extracted and subjected to further quantitative analysis, including meta-analysis, whenever possible.

*Funding source and research quality*

In addition to parameters for study rigor, West et al. (2002) state that funding and/or sponsorship is a critical criterion for any system that assesses research quality. Linkages between a study and an entity with a vested interest in its outcome may not be reason to automatically exclude the study, but can be an important factor for interpreting results.

For example, Barnes and Bero (1998) found that out of 106 relevant studies, 37% concluded that passive smoking is not harmful to human health, but 74% of these were written by authors with tobacco industry affiliations. In multiple logistic regression

analyses controlling for article quality, peer review status, article topic, and year of publication, they found that the only factor associated with concluding that passive smoking is not harmful was whether an author was affiliated with the tobacco industry.

### **E. Summarizing and synthesizing the evidence**

All SERs contain a summary of study characteristics (populations, interventions, and outcomes), design, quality and results. This information is synthesized in an organized fashion using *narrative synthesis* and, if appropriate, *meta-analysis* (quantitative synthesis). Meta-analysis uses statistical techniques to combine quantitative findings of multiple studies. More weight is given to more informative studies (large studies with precise effect estimates) and less weight is given to less informative studies (small studies with imprecise effect estimates). Effects observed across all studies are then pooled to produce a weighted average effect.

Meta-analysis requires considerable subjective judgment about how to combine data into datasets, which datasets are relevant and methodologically sound enough to be included, and whether and how to investigate sources of heterogeneity. Decisions must also be made about the relative robustness of analyses with high quality data but small sample sizes or lower quality data with larger sample sizes.

Meta-analysis is attractive because it helps improve precision, and there is much discussion of it in the medical SER literature. But meta-analysis may not be possible or appropriate when there are important differences between studies in terms of populations, interventions, outcomes, designs and quality. As Petticrew (2001) points out, “one of the allures of meta-analysis is that it gives an answer, no matter whether studies were combined meaningfully or not. Systematic reviews should not therefore be seen as automatically involving statistical pooling...narrative synthesis of included studies is often more appropriate and sometimes all that is possible” (p. 100).

A narrative synthesis forms an important part of SERs, especially those in which meta-analysis was not conducted. A narrative synthesis tabulates study characteristics and results, and looks at similarities and differences between the studies and their outcomes. Although it does not quantify average effect size, a narrative allows informal evaluation of the effect of the intervention and how it may have been impacted by measured study characteristics and study quality. Narrative syntheses typically make use of tables to allow comparison and provide a systematic and consistent record of information from studies in the review, including:

- the population under study
- the intervention (e.g. drug or medical treatment)
- settings where the intervention was applied
- environmental, social and cultural factors that may influence findings
- nature of the outcomes measures used, their relative importance and robustness
- the validity of the evidence
- the sample sizes and the results of the studies included in the review.

(Center for Reviews and Dissemination 2001).

### *Assessing the strength of the body of evidence*

To rate the strength of evidence from a group of studies, reviewers must not only identify all relevant studies and evaluate the quality of each, but also assess the consistency of results and heterogeneity of study designs to determine their comparability. This is more difficult than assessing the quality or strength of an individual study. Key considerations include whether the approach used to identify potentially pertinent literature was comprehensive and unbiased, and whether bias was avoided in evaluating, synthesizing and interpreting available evidence (Steinberg and Luce 2005).

West et al. (2002) submit three criteria for assessing the strength of a body of scientific evidence:

- Quality: The aggregate of quality ratings for individual studies, predicated on the extent to which bias was minimized. Aggregate quality of a body of evidence is based on the internal validity of each included study.
- Quantity: Magnitude of effect, numbers of studies, and sample size or power.
- Consistency: For any given topic, the extent to which similar findings are reported using similar and different study designs.
- Steinberg and Luce (2005) add to this list coherence, i.e. do the findings of a body of evidence make sense as a whole?

## **F. Acknowledging, reducing and removing bias**

Systematic Evidence reviews strive to both detect bias in original studies, and reduce bias in the search for, selection of and analysis of these studies. Khan et al. (2003) state that the quality of an individual study depends on the degree to which it employs measures to minimize bias and error in its design, conduct and analysis. They provide a list of 4 types of potential bias in medical research:

*Selection Bias* occurs when subjects studied are not representative of the target population about which conclusions are drawn. Arises when participants are allocated to groups. It is important to check if appropriate measures were designed and implemented to prevent or minimize it.

*Performance Bias* occurs when there are systematic differences in care provided to participants in the comparison groups other than the intervention under investigation. May arise due to unintended interventions or co-interventions.

*Measurement Bias* is failure to control for the effects of data collection and measurement e.g. tendency of people to give socially desirable answers. Arises

particularly if the outcomes assessed were subjective, and if participants and researchers involved in ascertaining the outcomes were not blind to group allocation.

*Attrition Bias*- systematic differences between comparison groups in withdrawals or exclusions of participants from the results of a study. For example, patients may drop out of a study because of intervention side effects. Excluding these patients from the analysis could result in overestimation of the effectiveness of the intervention.

*Publication Bias* can occur because research with negative, statistically insignificant or inconclusive results is less likely to be published than research with positive results. Users of the published information may then be misled about the effectiveness of the intervention. Including gray literature and unpublished research in an SER is important because it helps reduce publication bias and increase the accuracy of the review.

*Cognitive bias* refers to observer effects and the limits of human perception. Humans must rely on previous experiences and exposures as their interpretive framework for examining an issue or situation. To the extent that no single human being is capable of omniscience, the evaluation of any “evidence” must inherently contain researcher bias.

*Systemic bias* arises from known or unknown constraints that are inherent to institutions, processes, or individuals. This type of bias is difficult to recognize and also to control for. An SER-relevant example is the socio-cultural bias that is prevalent in many medical and psychological studies. Large numbers of seemingly high-quality clinical trials have been conducted in the past fifty years; however, because of systemic biases within each field, many of those studies failed to adequately control for the potentially varying impacts of gender, ethnicity, socioeconomic class, or culture on treatment efficacy. Similarly, many studies of anti-depressant drugs are conducted by pharmaceutical manufacturers; most of these studies do not include alternative therapies as comparative treatments. Instead, the drug is simply compared to no treatment at all.

#### *Methodological and Clinical Heterogeneity*

Heterogeneity refers to variability or differences between studies in estimates of effects. An important distinction is made between *methodological heterogeneity* (differences in study design) and *clinical heterogeneity* (differences between studies in key characteristics of the participants, interventions or outcome measures). Large clinical or methodological differences between studies may preclude use of meta-analysis, and may help explain heterogeneity in results. (Bandolier 2005.)

#### *Analysis of heterogeneity*

Heterogeneity can be examined by constructing tables which group studies with similar characteristics and outcomes together. Studies can be stratified into subgroups based on populations, interventions, outcomes and methodology. Differences in the subgroups of studies can then be explored. If sufficient numbers of studies exist, meta-analysis can be undertaken on subgroups and significance of differences between subgroups can be assessed. Such analyses must be interpreted with caution as statistical power is limited

(Type I errors possible) and multiple analyses of numerous subgroups could result in spurious significance (Type II errors possible).

### **G. Updating SERs**

A key feature SERs is that they can be updated when new information emerges that will significantly strengthen or change the outcome of an existing review. A detailed protocol and careful documentation of how the initial SER was conducted are what make updates feasible. Updates may also be conducted within specified timeframes. For example, the Center for Evidence-based Policy at OHSU may update SERs at six, twelve, and twenty-four month intervals.

### **H. Incorporating SER Information into Medical Policy and Practice**

One of the primary ways that SER information is used to implement evidence-based medicine is through *clinical practice guidelines*: “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances” (Field and Lohr 1990). Guidelines are typically developed by an expert committee, often the same group that conducted the SER, but stakeholders and practitioners may also be involved. Guidelines consist of coherently sequenced recommendations based on best available evidence aimed at everyday decision-making situations and can be applied to any aspect of clinical care: how to and when to order diagnostic screening tests, when to provide certain medical services, how these should be performed, and how long patients should remain hospitalized following a procedure (Timmermans and Mauck 2005).

Evidence-based medicine is premised on the assumption that medical practitioners and patients will use SER information to improve the effectiveness and efficiency of healthcare. However, more progress has been made in rigorously assessing available evidence through SERs than in incorporating this evidence into policy. There are still significant delays (often many years) from the time strong evidence about an intervention becomes available and when that practice is widely adopted in the medical field. Most clinicians don’t use clinical guidelines because their patients expect tailored care and because they can’t easily reference guidelines in a busy practice setting. Use of guidelines may increase through development of clinical decision-support systems and electronic medical records and prescribing that could provide real-time, evidence-based treatment recommendations (Mendelson and Carino 2005).

#### *Perspectives on evidence based medicine*

EBM is very attractive to many health care players, especially state-level policy makers, large healthcare providers and the insurance industry while others, primarily practicing medical professionals, have mixed feelings about it. Critics of EBM fear that clinical guidelines will lead to stagnation of individual creativity and the “art” of medical practice. They fear that EBM will lower standards of safety by encouraging use of protocols that treat all patients as essentially interchangeable in place of clinical

judgment, resulting in “cookbook medicine” rather than the tailored care based on physician experience that many patients expect. There are also concerns that reliance on clinical practice guidelines will gradually reduce practitioner skill sets and that the impetus toward EBM is being driven by large health care providers seeking to cut costs.

EBM proponents argue that clinical experience and better evidence should be seen as complementary, and definitions of EBM have evolved to account for this perspective (e.g. Sackett et al. [2000] who state that EBM is “the integration of best research evidence with clinical expertise and patient values”). Proponents see EBM as the key to solving problems of rising costs, inequity and variability healthcare. Supplementing practitioner experience with scientific evidence should promote greater effectiveness of medical interventions, and greater efficiency by filtering scarce resources away from ineffective clinical practices. EBM should create better informed patients and practitioners through collectively agreed upon and publicly available information about treatment options. EBM should also provide a scientific basis for healthcare policies rather than reliance on opinions from policymakers and lobbyists for the pharmaceutical and insurance industries. (Timmermans and Mauck 2005.)

#### **IV. Critiques of SER methods and cautions about using SER information**

- Research quality assessment criteria are a cornerstone of SERs, but there is still no consensus on what these criteria should be and how different types of quantitative evidence should be weighed.
- There is growing criticism within the medical community of rigid study exclusion criteria and the practice of ranking evidence quality on the basis of research methodology alone.
- There are growing calls for including a broader range of evidence in SERs, including qualitative evidence and expert opinion, but finding ways to include, and weigh, such disparate types of evidence are major challenges.
- Framing questions in the tightly focused, specific way that current systematic review methods require may skew the review away from important issues that are more difficult to focus.
- Absence of evidence regarding the effectiveness or safety of a health care treatment or medication does not mean that the intervention is not safe or effective. In other words, “absence of evidence is not evidence of absence”.
- Strength of evidence *in and of itself* is not related to the magnitude of effectiveness of an intervention. There may be very strong evidence that the intervention has little impact or, conversely, the apparent impact of the intervention may be large but the evidence regarding the impact may be weak.
- How strong the evidence needs to be when making a particular type of decision should depend, at least in part, on the potential consequences if assumptions about the outcomes of an intervention turn out to be wrong.

Systematic evidence reviews and evidence-based medicine have indisputably improved clinical practice and made scientific information about medical treatments more accessible to end users. Some have even described the EBM movement as a “new paradigm” for medical practice (Evidence-Based Medicine Working Group 1992). As Steinberg and Luce (2005) put it, “if you are doing almost anything related to health care today, being ‘evidence-based’ is *de rigueur*. Even when it is not obligatory to do so, claiming to be ‘evidence-based’ conveys a measure of credibility nowadays that is valuable to have” (p. 80).

Despite this prominence, there is still lively debate within the medical community about some basic tenets and particular methodologies in SERs and evidence-based medicine. Indeed, some have commented that medical SER methods and applications are still in their infancy (Moja et al. 2005). Nevertheless, use of SERs is far advanced in medicine

compared to any other field, so lessons learned there are important to consider as a precursor to examining how SERs might be applied to natural resource management.

### **Appraising research quality**

Appraisal of methodological quality of primary studies is seen as essential by the principal entities that conduct SERs, but quality assessment is resource intensive and scientifically controversial, and there remains a lack of consensus within the medical research community on the ideal quality assessment hierarchy and checklist (Moja et al. 2005).

Wide variation in frameworks for weighing evidence quality illustrates this point. The RTI International-University of North Carolina Evidence-based Practice Center found that out of 121 different approaches for rating the quality of an individual study, only 19 met the EPC's a priori standards for such assessments. A similar study identified 40 approaches for rating the strength of an overall body of evidence, only 8 of which met the EPC's standards. (Steinberg and Luce 2005.)

### **Including a broader range of evidence**

Fox (2005) elucidates several impediments to incorporating SER information into U.S. health policy decision making. One such obstacle is excessive zeal and a narrow view on the part of some EBM proponents concerning the value of randomly controlled trials compared to other forms of evidence. Critics argue that a broader range of evidence from observational studies, analysis of administrative data and a variety of qualitative methods should be included in evidence reviews. Fox notes that this criticism is echoed by many Cochrane Collaboration members and Evidence Practice Center researchers, and that groups in several countries are exploring methods to expand the types of evidence used in SERs while maintaining their rigor and persuasiveness.

Kelly and Swann (2004) cite several concerns about what counts as evidence and about how questions are framed. They note that trial data reported in SERs, while very useful, covers only part of the picture and deals with only some of the issues in clinical practice. They found recurrent biases during quantitative synthesis work on SERs for the UK Health Development Agency. These problems ranged from a simple absence of information (only a small amount of medical research actually deals with interventions) to content bias (certain research questions are not investigated because they are not amenable to being answered by a randomized controlled trial). They also found artifact bias, in which the method used obscures the meaning of data because research questions have been asked in particular ways. Contextual issues that are of central importance to patients – their families, for example – are simply not addressed. They conclude that qualitative methods, which can shed light on such things, must be a part of the lexicon of the evidence base. But the question remains, how best to do this?

Similarly, Edwards et al. (1998) note that clinical practice frequently generates questions that are not easily answered by randomized controlled trials and that much research is not

ideally designed, yet it is still important to make the best use of all available evidence. They argue that using rigid study exclusion criteria in SERs risks underestimating the evidence base and distorting the true message that the SER is trying to identify. In place of research quality hierarchies that reject studies on the basis of design alone, they suggest that the message or “signal” within each piece of research should also be assessed. Methodological quality of the study would still be considered but would be balanced against the weight of its message. Fundamentally flawed research would still be rejected, but a study with some flaws would be included if its “signal” (potential value to the clinical question) was judged to be strong in relation to its “noise” (methodological drawbacks). The balance of these two elements is termed the “signal to noise ratio”.

Such discussions point to the need for a clearer definition of what might count as evidence in a more inclusive systematic review (Box 1).

**Box 1: What do we mean when we talk about evidence?**

A forensic attitude pervades science today, with scientists seen as detectives, hypotheses as suspects and data as clues. The concept of “evidence” is central to this formulation. But what, specifically, do we mean when we talk about evidence in this context? Evidence is more than just data and hard facts- it involves contextual information and interpretation. *Scientific* evidence consists of:

- Scientifically guided empirical observations combined with background information, logic, and scientific expertise.
- Observations from both manipulative and observational experiments designed to test a particular theory, or designed just to look at something potentially interesting.

Evidence does not have to be quantitative, or gathered by a scientist. The key is that the information was collected and interpreted as objectively as possible and can somehow be verified. An observation can be considered scientifically guided even after the fact by subjecting it to the same scrutiny and skepticism accorded to observations within a scientific study. Somewhat broader definitions of evidence:

- Knowledge based on credible investigation, calculation, or analysis.
- Information collected in ways that are as free as possible from personal or vested interest or belief.

The totality of evidence concerning a particular question or topic is the *evidence base*. The evidence base is:

- All observations plus background information plus logical relationships among all this information.
- Dynamic, not static, and evolves as understanding develops, as new information emerges and as relationships between issues previously thought to be unconnected come to light.

*Primary source: Taper, M. and S. Lele (eds.) 2004. The Nature of Scientific Evidence: Statistical, Philosophical, and Empirical Considerations. Chicago: University of Chicago Press.*

**Where does experiential knowledge fit in?**

Greenhalgh et al. (2003) investigated the potential for applying principles of evidence-based medicine to improve educational quality. They found that independent qualitative analysis of students' and staff experience was invaluable when testing the validity and transferability of published research evidence and quality standards.

Based on this they concluded that evidence in education should include not only formal, research derived knowledge but also tacit knowledge (informal knowledge, practical wisdom, and shared representations of practice). These types of experiential knowledge

are a potentially important source of evidence that pertains to natural resource management questions as well.

### **Cautions to bear in mind when considering the strength of a body of evidence**

Steinberg and Luce (2005) note that while methods for rating evidence quality and strength have become more sophisticated over the past decade, synthesis of results from multiple studies still involves substantial subjectivity in the form of judgments regarding relevance, admissibility (i.e. is the study “in” or “out” of the review?) and importance of individual pieces of evidence. These authors provide several cautions to bear in mind while considering strength of evidence during healthcare policy making that are also relevant to application of the SER approach in other disciplines:

- 1) Absence of evidence regarding the effectiveness or safety of a health care treatment or medication does not mean that the intervention is not safe or effective. In other words, “absence of evidence is not evidence of absence”. To illustrate this point, the authors cite work showing that at least 50% of medical treatments have never been validated by clinical trials, and that more than half of all medical services have weak or no evidence at all to support or refute their effectiveness.
- 2) Strength of evidence in and of itself is not related to the magnitude of effectiveness of an intervention. There may be very strong evidence that the intervention has little impact or, conversely, the apparent impact of the intervention may be large but the evidence regarding the impact may be weak.
- 3) It is very important to critically assess the relevance and/or generalizability of a study or body of evidence for populations or settings other than those that were studied (external validity).
- 4) How strong the evidence needs to be when making a particular type of decision should depend, at least in part, on the consequences of drawing the wrong conclusion. Willingness to act on the basis of weak or inconclusive evidence should include consideration of potential consequences if assumptions about the outcomes of an intervention turn out to be wrong.

In the event that SER-like processes are applied to natural resource questions, issues of what does, and what does not constitute relevant evidence, how different kinds of evidence are weighed, and the role of a synthesized body of evidence in policymaking are likely to be areas of significant interest to stakeholders. Goals of reducing controversy about what science says and how science is used to inform management strategies may not be realized without some degree of consensus on these issues. Sorting them out may pose significant challenges, some of which are explored in the next section.

## V. Applying SERs to Natural Resource Issues: Challenges and Opportunities

Clinical medicine and public health care have benefited greatly from the use of SERs to synthesize scientific information that may be available but difficult to access and understand as a body of evidence. This has prompted interest in adapting SERs for use in other areas of public policy, including forest management in Oregon.

However, there are significant differences between clinical medicine and ecosystem management that will need to be addressed if SERs are to be effectively applied to natural resource management. Table 4 summarizes a number of these differences.

**Table 4: Differences between Clinical Medicine and Natural Resource Management**

<i>Dimension</i>	<i>Clinical Medicine</i>	<i>Natural Resource Management</i>
Primary objective	Usually single objective- improve human health; rarely contested	Often multiple objectives- commodity production, wildlife, biodiversity, watershed health, recreation; frequently contested
Types of evidence	Often experimental; easier to control explanatory variables	Rarely experimental; usually difficult to control explanatory variables
Sample sizes	Easier to obtain large sample sizes	Harder to obtain large sample sizes
Outcomes	Can be easier to define and measure	Usually harder to define and measure
Number of species	Concentrates on well being of single species	Deals with multiple species and habitats
Application of research	Conclusions of studies can have global implications (all humans)	Conclusions of studies are often landscape or problem specific
Funding and resources	Significantly greater than ecological research, with strong interest from the private sector	Much less funding than in medicine, with relatively little interest from the private sector
Influence of politics	Generally supportive and not contentious	Often negative and contentious

(Adapted from a similar table in Fazey et al. [2004]).

What follows is a more detailed discussion of these differences and other challenges to applying the SER process to natural resource and ecosystem management issues. Opportunities for adapting components of SERs to natural resource questions are introduced at the end of this section.

### **Disentangling questions about evidence from those about values and preferences**

- Forest managers usually lack the kind of widely agreed upon single objective (better human health) that clinical medicine practitioners enjoy.

- Clarifying and obtaining consensus on the underlying objective (i.e. what the information that will result from the SER will be used for) may be critical to conducting a successful natural resource SER

The primary objective of clinical medicine- maintaining or improving human health- is rarely contested and there is little argument about what this means in the clinical sense. In contrast, there are usually multiple objectives for managing forests, and these objectives are unlikely to be ranked the same by all stakeholders. Depending on their vested interests and values, a particular stakeholder group might place primary emphasis on reducing wildfire risk, or watershed protection, enhancing biodiversity, habitat for a single species, timber production, recreation access or some other objective.

In short, forest managers usually lack the kind of overarching, widely agreed upon single objective that clinical medicine practitioners enjoy. For many who work at the science/natural resource policy interface, the “elephant in the living room” is the degree to which acrimonious disagreements over the science that relates to forest management objectives are really debates over values and perceptions of risk. As Atkins et al. (2005) note, what on the surface appear to be debates over evidence upon closer inspection often turn out to be fundamental differences of opinion on what outcomes are most important and what actions are appropriate in the face of imperfect evidence.

The importance of disentangling debates about scientific evidence from underlying values and preferences when applying SER techniques cannot be overstated and may be critical to planning and conducting a successful natural resource SER. Because an overall policy objective is what frames review questions, lack of consensus about this objective (or the primary objective among multiple objectives) may hamper efforts to delineate the review’s purpose and scope. Atkins et al. (2005) argue that clarifying this objective is a critical precursor to an SER, and that the SER process itself can help further differentiate technical and scientific evidence from values concerning how this evidence is applied in policymaking.

### **Differences between medical science and ecosystem science as fields of inquiry**

- Compared to clinical medicine, ecology is a younger science that involves a much greater proportion of observational field-based studies to laboratory experiments.
- Ecological research thus typically involves greater methodological diversity, fewer laboratory controls, less replication and more “nuisance” variables than medical research.
- Ecology deals with larger spatial scales, longer timeframes and more complexity associated with multiple interacting species, habitats and ecological processes than medicine, which is focused on a single species.
- In general, there is less certainty about scientific conclusions in ecological studies than in clinical medicine.

The nature of scientific research differs according to the field of inquiry and constraints imposed by the system being investigated. For example, tightly controlled and replicated laboratory experiments form the basis of most research on the physiology and genetics of organisms, including humans. However, this type of experimental approach is often impractical for investigating the population biology and ecological relationships of plants and animals in forest landscapes because of the temporal and spatial scales, and complexity of these systems. Forest ecology studies are usually field-based, and often designed to look at multiple species and processes interacting at local to landscape scales. Consequently, much ecological knowledge is derived from careful observation and analysis of empirical, rather than experimental data. (National Research Council 2004.)

Also, many ecological processes and interactions unfold over relatively long timeframes, which can make interpretation of outcomes of shorter term studies more equivocal. Researchers try to account for this with longer term studies, but funding cycles and practical exigencies can hinder such efforts. In contrast, it is usually possible to achieve a higher degree of certainty about the efficacy of a drug or other medical intervention (the focus of most SERs) over a shorter timeframe.

Replication is also problematic in many ecological studies, especially at the landscape-level. Where replication is possible, sample sizes are likely to be much smaller than is typical in clinical medical research, with correspondingly lower statistical power. As Lee (1993) points out, “it is difficult to observe the state of the ecological system and the human economy interacting with it...measurements of the natural world, such as the size of migrating populations, are inexact; and natural systems often yield only one data point per year, such as the number of fish in a run.” (p. 58.)

In addition to differences related to subject matter, forest ecology and medical science are also in different stages of development. Clinical medicine is in many ways a more “mature” science, in that it has better developed, more specific, overall stronger theories that have been subjected to a multitude of experimental tests and refined over many decades. Medical SERs usually address deductive, “top down” research- tightly controlled and focused experiments or clinical trials to test the efficacy of a single drug or treatment, couched in a detailed theoretical understanding of how the overall system (the human body) works and accompanied by statistical techniques that can be used as strong evidence of causality (Maurer 2004). This makes it easier to control or account for “nuisance” variables and to attribute causal effect to the variable under study.

In contrast, ecology is overall a younger science with theories that are less well developed, which makes prediction harder. Much of ecology involves “bottom up”, or inductive studies that are often observational and exploratory in nature. Rather than being experiments in the classic sense, this work typically involves describing what is going on, or seeing what happens in response to a natural disturbance or management action. Studies are often conducted to search for patterns in nature, rather than provide a strong test of a specific question. Statistical techniques used tend to be those that allow identification of repeated patterns that can be synthesized into generalities (Maurer 2004).

Field research makes it harder to control for nuisance variables, and study types tend to be more diverse, making it harder to compare findings. However, field studies by their nature have more real world validity than laboratory studies.

Adapting the SER process, which favors tightly focused, deductive research to forestry and natural resource sciences, where such evidence is less available and potentially less relevant to the questions at hand, may prove challenging. In this vein, Greenhalgh et al. (2003) argue that the linear and formulaic link between evidence and practice implicit in evidence based medicine may be inadequate for the complexities of education policy and that conceptual models designed for multifaceted problems may be more appropriate. This may also apply to the use of SERs to address multifaceted natural resource questions.

### **The challenge of delimiting evidence that applies to a forest management question**

- SERs are best suited to synthesizing research focused on whether a single medical intervention “works” or “doesn’t work”. In the natural resources arena, SERs would be best suited to analogous, single variable questions.
- Complex, multifaceted forest management questions might be difficult to address using the SER approach, and for simpler questions, there may be little focused research evidence available.
- Much evidence concerning forest ecosystems consists of studies in which several variables are considered simultaneously in order to accurately describe real world ecological relationships.
- Synergies among species and processes are common in forest ecosystems. Thus, it may be impossible or misleading to isolate a single ecosystem component or single outcome of a management action as the focus of an SER.
- Delimiting the evidence that applies to a forest management question may also be challenging due to uncertainty about extrapolating results from studies to other areas with significant biological, physical, climatic or land use history differences
- For a forest management question structured as suggested by SER guidelines, there is likely to be a range of tangentially related research that falls somewhere between direct relevance and complete irrelevance.
- A feasible natural resource SER may require a compromise between a holistic approach that is closer to reality, but impractical for defining relevant studies, and a reductionist approach that may limit the review's relevance.
- A related challenge may lie in structuring a question with a degree of specificity that allows inclusion of enough evidence to make the review worthwhile, but also limits it to a manageable scope.

The evidence base available to forest managers and policymakers is very different than that available to medical practitioners. There is a tremendous volume of clinical research available today, but the number of different study designs is relatively small. It is not uncommon to have several studies on the same treatment, conducted using the same methods. Study designs are primarily reductionistic, allowing researchers to focus tightly on a particular drug, treatment or human subpopulation and exclude all other variables. Drawing boundaries around a discrete body of this kind of evidence that relates to a medical SER question is thus relatively uncomplicated.

Ecological research with a similar narrow focus (e.g. studies designed to assess the effects of a single variable on a single species or environmental factor) could perhaps lend itself to such compartmentalization and be fairly straightforward to classify as pertinent to a natural resource SER. However, much of the evidence available to land managers consists of studies that examine several variables simultaneously using multivariate statistical analyses. This type of evidence may be more complicated to clearly specify as relevant (or irrelevant) to an SER question, especially for an SER question that addresses a single species or process, the form suggested by SER guidelines. Overall, the evidence available in forest ecology is less extensive, but at the same time more varied in research topic and design, and more holistic in nature (Box 2).

**Box 2: What constitutes scientific evidence in forest ecology?**

All sciences, but ecology in particular, rely on a range of types of data or empirical observations. Those observations may occur as part of a designed experiment, or as part of a survey of the natural world. This is not a simple dichotomy, but a continuum. At one end are experiments in which all extraneous factors are held constant, as far as possible, and the scientist varies one or more factors of interest in a controlled fashion. At the other end are observational surveys with no manipulation. Intermediate are field experiments in which some factors are manipulated, some are controlled, and others vary naturally. All observations coming from this range of situations are potentially important in understanding ecosystem processes and function, and all contribute to the evidence base.

In other words, scientific evidence in ecology comes from a *spectrum* of experimental and observational data. Use and interpretation of statistics depends on whether the type of science being done is primarily observational (inductive) or experimental (deductive). Some statistical methods can be useful in one kind of inquiry but not in another. Other methods can be used in either type of science, but require different assumptions and imply different things about the data depending on whether an investigation is primarily inductive or deductive.

Thus, sweeping generalizations about the appropriateness of any particular method of statistical inference over others are unwarranted, and laboratory experiments do not necessarily carry greater weight than field experiments. Laboratory experiments are more powerful for exploring the mechanistic bases of ecological processes, but field experiments have greater potential to be generalizable because, by incorporating more potentially interacting species, they can encompass a greater ecological scale.

All types of data play a role. The weight given to a particular piece of evidence should not depend on the type of observation but on the match between the observation and the question being asked. Evidence can include observations made prior to questions being asked and do not have to be made by a scientist. Scientific decisions in ecology are based on bringing together of different, even disparate kinds of evidence.

*Paraphrased from Scheiner, S. "Experiments, observations, and other kinds of evidence", and Maurer, B. "Models of scientific inquiry and statistical practice: Implications for the structure of scientific knowledge" in The Nature of Scientific Evidence, Taper, M. and Lele, S. (eds.) 2004. University of Chicago Press.*

Because synergies among species and processes are common in complex forest ecosystems, it may be impossible or seriously misleading to isolate a single ecosystem component, or a single potential outcome of a management action as the focus of an SER. (Fazey et al. 2004.) For many natural resource management issues, assumptions necessary to structure questions so they can be answered by an SER could result in a loss of focus on the way an ecosystem actually functions, and on issues that managers and stakeholders believe are most relevant to the policy decision.

Even a relatively simple silvicultural prescription that involves timber harvesting or thinning, and slash piling and burning could result in concurrent and interrelated changes in water quality, soil characteristics, insect populations, wildfire risk, understory plant communities and habitat for a number of different species. The science that bears on such multiple-outcome treatments may be challenging to assess using the SER process, which was developed primarily to address questions concerning whether an individual medical treatment is effective or not.

Extrapolation of research findings to larger populations and different locales is also much more problematic in ecology than in medicine (Lee 1993). Most research included in a medical SER concerns subpopulations of humans with similar characteristics or medical conditions. Unless an environmental variable is the focus of the study, where the participants live makes little difference when interpreting results and applying them to similar humans, even in other countries. On the other hand, location of an ecological study is critical when trying to ascertain whether findings apply in other areas, due to geographic variability in climate, geology, plant and animal communities, land use history and other factors. This variability can also present difficulties when comparing otherwise similar studies.

Drawing clear boundaries between studies that are relevant to the question at hand and those that are not could present challenges in a natural resource SER because of the interrelated nature of species and ecosystem processes, the diverse, multivariate nature of research conducted to investigate these relationships, and lack of certainty about when it is appropriate to extrapolate results to other geographic areas.

The SER process defines the evidence base through decisions (yes/no) as to whether each potentially relevant piece of evidence will be included or rejected as "irrelevant". Trying to make these decisions for natural resource questions might be controversial, especially those questions with significant implications for stakeholders. For a given question, there is likely to be a range of tangentially related research that falls somewhere between direct relevance and complete irrelevance.

### **The challenge of assessing evidence quality in forest ecosystem science**

- There is likely to be a paucity of relevant, focused experimental research and a greater proportion of potentially diverse observational evidence available to address a natural resource question.

- This type of evidence is typically graded as “low” quality in medical SERs, e.g. often observational with few controls, frequently with confounding interactions.
- If the same criteria used to assess evidence quality in medical SERs are deemed appropriate for a natural resource SER, these criteria would probably need to be applied *less stringently* to assess forest ecosystem research.
- Depending on the nature of the SER question and evidence available to address it, it may also be necessary to develop *significantly different* criteria for assessing the quality of forest ecosystem research.
- There is a lack of consensus on quality assessment criteria used in medical SERs. Achieving consensus on if, and how, such criteria should be used in natural resource SERs may prove difficult.
- There may be no single set of quality assessment criteria that will work for all natural resource SERs. Assessment criteria may need tailoring to fit the evidence that pertains to a particular SER question.
- In cases where the evidence consists of studies and monitoring with disparate methods, locations and outcome measures, there may be no clear rationale for saying that one piece of evidence is “higher” or “lower” quality than another.
- With all else being equal, studies that involve relatively larger spatial scales, more replication, more controls and longer timeframes are likely to produce the most reliable results.
- However, “all else equal” assumptions often don’t hold true. Greater complexity and diversity at larger scales can introduce more, rather than less uncertainty.

Conducting SERs in the natural resource arena in the same manner that they are conducted in medicine will involve use of quality criteria to make inclusion and ranking decisions about each piece of evidence identified as relevant during the search phase of the review. As explained previously, evidence quality hierarchies in medical SERs are based on experimental design and privilege randomized controlled trials over all other methodologies, a well-established and accepted practice from the perspective of reductionist, laboratory controlled science. This approach to quality assessment works relatively well in medical SERs because many medical practice questions fall into a simple and logical taxonomy (e.g., prevalence of the condition, prognosis, or therapy). Each type of question has a corresponding preferred research design (survey, cohort study, randomized controlled trial, etc.) with widely accepted criteria for assessing the statistical validity of results. (Greenhalgh 2003.)

Tightly controlled, randomized clinical trials are the “gold standard” in medical SERs, but this type of evidence does not hold such an elevated status in forest ecology, where

the evidence base is more varied and holistic. (See Box 2, p. 56.) Technical questions that relate to forest management are wider ranging and harder to categorize, with fewer direct links to particular preferred study designs and less agreement on criteria for assessing the validity of results. Study methodologies are more diverse, and randomized controlled trials, non-randomized controlled trials and case-controlled analytic studies are difficult if not impossible. Even where feasible, these study designs may not be relevant or useful for developing the kind of knowledge that forest managers need. Research with some degree of experimental control is conducted regularly in ecology, but uncontrolled interactions and nuisance variables are common, so study results in ecology tend to be more equivocal compared to clinical medicine.

In addition to hypothesis tests addressed by studies with some experimental controls, many other important research questions in forest ecology are open-ended, i.e. they begin with stems such as “how” or “what”. To address this type of question, studies or monitoring are often conducted simply to observe what happens in response to a disturbance or management action or simply to find out more about what is going on in a particular area, rather than to compare two or more situations.

There is likely to be a paucity of relevant, focused experimental research and a greater proportion of potentially diverse observational evidence available to address a natural resource question. Most observational evidence is rated as low-grade in medical SERs, and often not even included. Evidence that shows interactions also tends to be discounted in medical SERs but in ecology such evidence may be a relatively accurate representation of reality. These considerations suggest that criteria to assess evidence quality in natural resource SERs should be less restrictive and also different in some ways than criteria used in medical SERs.

Depending on the nature of the questions and evidence being assessed, developing “coarse-scale” quality standards specific to ecosystem science might be relatively straightforward. With all else being equal, studies that involve relatively larger spatial scales, more replication, more controls and longer timeframes are likely to produce higher quality evidence. However, “all else” is rarely equal and increasing complexity and variability at larger scales often complicates study results.

In cases where the evidence consists of studies and monitoring efforts with disparate methods, locations and outcome measures, there may be no clear rationale for saying that one piece of evidence is “higher” or “lower” quality than another. It may only be clear that they are different.

### **Locating the evidence**

- Archiving of medical research abstracts and peer reviewed papers is more organized and standardized than in ecosystem science. Comprehensive literature searches in a natural resource SER may thus be harder to achieve.

One outcome of widespread adoption of SERs and EBM in the medical community has been greater emphasis on clinical trial registries and in placing study reports into databases such as Medline. Standardization of terminology and search terms has also improved. Information technology and searchable electronic databases in all fields of study are evolving and expanding at a rapid rate but the field of ecological science lags behind the medical field in this regard. Electronic databases such as Agricola are useful, but there is less consensus on a single best database or central clearinghouse for all ecosystem research and data. There is also greater diversity in research designs and a less well developed theoretical framework in the field of forest ecology. These considerations suggest that certainty about the comprehensiveness of literature searches may be harder to achieve for natural resource SERs.

### **The role of qualitative research, expert judgments and experience**

- Expert knowledge and experience play a greater role in ecosystem research than in medicine because investigators must rely more heavily on expert judgment when interpreting results.
- Natural resource management also involves high levels of expert judgment because scientific information is often not available.
- Experiential knowledge may constitute an important part of the overall evidence base, but incorporating this evidence into quality assessment and ranking framework in the context of an SER remains problematic and controversial.
- One potential way around this debate is to understand scientific and expert knowledge as complementary, and equally important in ecosystem and natural resource management.

Qualitative information such as expert opinion and experience plays a larger role in ecosystem research than it does in medical research. Because ecological research is less reductionist and less tightly controlled than most medical research, and the theoretical knowledge base is less well developed, investigators are left with the alternative to rely more heavily on expert judgment when interpreting results. Deducing cause and effect is more problematic in the observational studies that are common in ecosystem science. The need for expert opinion to interpret what evidence is available will likely be a major component of SERs conducted for natural resources.

Expert opinion and experience on the part of ecosystem scientists and managers may form an important part of the overall evidence base. Given the paucity of research in many areas, this type of evidence is likely to be even more important in natural resource management than in medicine. Identifying ways to include a broader range of evidence in SERs, such as qualitative research and expert judgment, is an area of intense discussion both within and outside of the medical community.

Boaz et al. (2002) argue that tools to appraise qualitative information must be developed in order to capture the full complexity of an intervention, its impact and its transferability to other contexts. Dixon-Woods et al. (2004) conclude that there is a strong rationale for multi-strategy research and persuasive arguments for integrating data generated by a range of strategies within SERs. However, they note that how, or even whether, the quality of qualitative evidence should be appraised is still being debated.

One potential way around the debate over whether expert opinion and experience should or do constitute “evidence” is to view expert and scientific knowledge as complementary and equally necessary in natural resource management. Haas (2003) argues that while scientific evidence is important, managers today are often held to an unattainably high evidentiary standard when trying to implement management actions, and that the legitimacy of professional judgment in such cases needs to be restored. Fazey et al. (2005) provide a useful comparison of these two categories of knowledge (Table 5).

**Table 5. Some of the differences between expert and experimental knowledge**

	<b>Expert (experiential) knowledge</b>	<b>Experimental (scientific) knowledge</b>
Perspective	Greater capacity for a holistic perspective	Greater capacity for a reductionist perspective
Historical perspective (Temporal context)	Has some capacity to take into account the historical trajectory of something in order to make better predictions about the future by interpreting the present with respect to past experiences. For environmental systems, this requires extensive experience of the same phenomenon or system (e.g., some longtime ODF district managers and staff).	Has less capacity to take into account the influence of a historical trajectory because predictions are based only on what is occurring in the present.
Learning from long-term outcomes	Less capacity to learn from interventions whose outcomes take a long time to become apparent because an individual’s experience is finite and relies more on immediate feedback.	Has greater capacity to learn from interventions that have long-term outcomes because experiments can run over long periods of time.
Dealing with confounding factors	Has less capacity to deal with confounding factors when trying to distinguish between cause and effect.	Has greater capacity to deal with confounding factors when trying to distinguish between cause and effect.
Accessibility	Difficult for others to access and pick up because it is either inarticulate (tacit), or very difficult to articulate (implicit)	Easier for others to access and pick up because it is formalized and made explicit
Requirement	Requires experimental knowledge as a check and balance to ensure accurate connections between cause and effect.	Requires expert knowledge to identify appropriate questions, interpret results and maintain a more holistic perspective.

(Table slightly adapted from Fazey et al. 2005.)

### **Disparities in healthcare and ecological research funding**

- Funding for forest ecosystem research and monitoring is only a fraction of that available in the field of clinical medicine.
- This disparity affects that overall amount of high quality evidence available to land managers as well as potential funding to conduct natural resource SERs.

Health care is one of the most salient and important issues for U.S. citizens, which translates into major investments in health care research by government agencies, charities and the pharmaceutical industry. In 2003, an estimated \$94 billion was spent on health care research in the U.S., double the amount spent in 1993 (New York Times 2005). Another source put the total for 2004 at \$109 billion (Research America 2004). Pharmaceutical companies often fund SER evaluations of drug efficacy because of the enormous potential payoff they can receive if findings are positive. This is an inherent conflict of interest and major flaw in the medical SER process. However, it also assures that enough funding is available to conduct lengthy and expensive reviews.

Estimates of total annual spending for ecosystem research could not be located, but such expenditures probably amount to only a fraction of spending for health care research. This disparity is the major reason for the overall lower volume of scientific evidence available to natural resource managers, despite widespread acknowledgement of the need for better scientific information to support land management decision making.

It is not clear how much a natural resource SER would cost. This would depend on the nature and number of questions asked, the volume of research judged to be relevant to the review, the accessibility of this research, and the human resources necessary to systematically review it. An organization of SER practitioners in the natural resource field that is analogous to the Cochrane Collaboration in the medical field does not yet exist, although existing ecosystem and forest science organizations could provide a basis for forming one.

A network of conservation professionals familiar with SER techniques has been established in the United Kingdom (the Center for Evidence-based Conservation, discussed below). However, this effort is in its infancy and few professionals in the Pacific Northwest with the necessary qualifications to conduct a natural resource SER have even heard of the term. Thus, there would be costs associated with training a review team. If the question was tightly focused, and the evidence base limited and not controversial, a natural resource SER might be possible for a reasonable cost. These are major assumptions and would need to be tested in practice.

## Opportunities for applying SERs to natural resource issues

- Applying SERs to natural resource issues will be challenging, but early proponents of the SER approach in medicine faced significant challenges as well.
- Despite differences, there are a number of similarities between clinical medicine and aspects of conservation and natural resource management.
- These similarities include the common use of interventions (essentially experiments in progress), the need to make decisions on the basis of imperfect information, and the complementary role of evidence and experience.
- Some components of SERs could be incorporated into science reviews (e.g. better documentation of how studies were selected for inclusion, investigation of quality differences) in order to increase their objectivity and transparency.
- Conducting a synthesis of available science using SER techniques could highlight gaps in the evidence base and suggest relevant areas for future research.
- In combination with ecological monitoring and incremental updates, a synthesis of available science using SER techniques would mesh well with landscape-level adaptive management.

There will be significant challenges to adapting SERs to natural resource management questions, but it is important to bear in mind that proponents of SERs and evidence-based medicine also faced considerable skepticism during the early years of work in this arena. Pullin and Knight (2001) argue that differences between ecological and medical studies are not fundamental in *practice*, but are instead primarily differences in *scale* and *emphasis*. Despite some clear differences, there are a number of similarities between clinical medicine and conservation (Table 6).

**Table 6: Similarities between dimensions of clinical medicine and conservation**

<i>Dimension</i>	<i>Similarity</i>
Overall goal	Common goal of doing more good than harm
Applied science	Interaction and communication between researchers and practitioners is essential to achieve effective outcomes
Intervention	Procedures and interventions are common, and are essentially experiments in progress
Monitoring of outcomes	Essential for informing future practice
Decision space & timeframe	Decisions often need to be made quickly and with incomplete or equivocal information
Experience	Has an important role and is widely used by practitioners; can complement evidence in decision making

(Slightly adapted from a similar table in Fazey et al. [2004].)

In 2002, a group at Australia National University investigating the application of SERs to environmental management (Fazey et al. 2002) identified some advantages and disadvantages to what they termed *evidence-based environmental management* (EBEM):

*Advantages:*

- Provides a process for reviewing evidence.
- Facilitates increased data collection, monitoring, and access to data.
- Helps us understand what we don't know.
- Provides institutional and practical working framework.
- Breaks some of the entrenched 'experientially' based approaches.
- Facilitates international collaboration and understanding across systems.

*Disadvantages*

- Will not work in all scenarios.
- Systematic review process is only as good as the evidence itself.
- Difficult to compare lower levels of evidence and review them.
- Difficult to measure some environmental management outcomes.
- Sets expectation that there is a 'right' answer.
- May create a funding bias towards more robust studies.

The group concluded that although some environmental management scenarios are more suited to an evidence-based approach than others, all have some components that could benefit from a more rigorous approach to the collection and assessment of evidence. They identified two main areas where EBEM might be very helpful:

- To provide a framework for collecting better evidence (i.e. to increase the quality and quantity of evidence collected).
- To develop methods for systematically reviewing existing evidence.

However, they also noted that a number of big questions remain about whether it is really possible to formulate natural resource management and conservation questions that can be answered using an EBEM approach and, if so, whether it will ever be possible to collect and collate data into a form that can be systematically reviewed.

### **Applying SERs to conservation issues in the United Kingdom**

To date, application of SERs to natural resource questions has occurred primarily in the United Kingdom. The Center for Evidence-Based Conservation (CEBC) was established in 2003 at the University of Birmingham, U.K. to support decision making in conservation and environmental management through production and dissemination of SERs on the effectiveness of management and policy interventions. With support from a range of environmental and academic entities, the CEBC acts as a source of evidence and coordinator of a growing collaborative network undertaking SERs.

The CEBC was formed because despite the growing amount of scientific information produced by researchers, conservation practice and policy remain largely experience-based with limited evaluation of what works and what does not. This is mainly due to poor accessibility of evidence on the effectiveness of different interventions- the provision and delivery of scientific evidence is not tailored to the user-community. The CEBC believes that the fundamental problem of evidence accessibility can be addressed through use of SERs and active dissemination of subsequent findings from an independent, unbiased source. The CEBC is the first institution worldwide to apply SERs to the fields of conservation and environmental management.

The objective of the CEBC is to conduct and widely disseminate SERs pertaining to questions identified by decision-makers in the conservation and environmental sectors, following a methodology modified from that established in the field of health care research and practice. In addition, the CEBC offers workshops that provide training in conducting, interpreting and utilizing the results of SERs. The CEBC maintains that the SER framework is not intended to negate individual experience or expertise. It is a tool to be utilized by decision-makers, designed to support their needs by presenting an independent, unbiased and objective assessment of all available evidence. The CEBC research and dissemination unit at the University of Birmingham provides the Center with both an academic infrastructure and continued impartiality, supported by a consortium of conservation and environmental organizations.

A collaborative project between the UK Population Biology Network and the CEBC brings together existing networks of scientists and decision-makers to develop and implement a model for knowledge transfer in biodiversity management using the CEBC evidence-based protocol. A trained team will systematically search for and access relevant information, interpret the findings and present them on an open-access web site in a format that is usable by decision-makers. Organizations representing decision-makers (English Nature, Scottish Natural Heritage, Countryside Council for Wales, Royal Society for the Protection of Birds and The National Trust) will then assess the information produced and provide feedback on its utility.

A sample SER protocol developed by the CEBC to investigate the effectiveness of instream habitat improvements on salmonids is provided in Appendix 2. Additional protocols as well as completed conservation SERs can be viewed on the CEBC website: <http://www.cebc.bham.ac.uk/>

In fall 2005, the British Ecological Society (BES) initiated the Ecology into Policy Grant (EIPG) program to support researchers to conduct SERs of the ecological science in areas of policy importance. Grants of £5000 (~\$8900) are given towards the cost of accessing information (including travel), employment of short-term assistance, training to conduct systematic reviews and purchasing statistical software. The BES recommends using the CEBC protocol for conducting an SER.

### **Adapting components of SERs**

Fazey et al. (2004) found no reason why attempts could not be made to apply at least some *components* of SERs to conservation issues. Our analysis supports this conclusion. For example, a clear protocol describing the approach taken by the reviewer is often lacking in traditional literature reviews, leaving the reader wondering how studies included were identified, appraised and synthesized. Developing tightly focused review questions and clearly defined protocols for how the review will be conducted are aspects of SERs that could be applied to science reviews in natural resources. Documenting in a more systematic way which studies were included, what they said and how the information was interpreted also appear to be relatively straightforward goals.

Gates (2002) investigated the use of meta-analysis in ecology and suggested that several features of SERs could be implemented in reviews of ecological research relatively easily:

- Multiple methods of searching for relevant studies should be used, and search methods should be reported in detail.
- The types of study included in the review should be stated (e.g. randomized experiment, non-randomized experiment, observational study).
- Estimates of the overall effect size and a confidence interval should be presented.
- Bias in the inclusion of studies should be investigated by means of funnel plots.
- A clear distinction should be made between main and subgroup analyses, and the latter should be interpreted with greater caution.
- Sensitivity analyses should be carried out.

### **SERs and adaptive management**

Identifying clear questions and protocols for science reviews that are conducted as part of forest management planning in Oregon should help to:

- Increase the transparency of the research approach and the methods used
- Encourage an increased focus on the impact and outcomes of interventions ('what works') while continuing to explore processes involved and the perspectives of stakeholders ('why it works')
- Encourage development of improved information systems, including databases of research and associated search tools
- Generate future research questions (Boaz et al. 2002.)

One of the major strengths of the SER process in terms of integration with current ODF management strategies is that a completed SER would provide a snapshot of “best science available”—which is required by OAR 629-035-0020 (3)—at a given point in time. In combination with ecological monitoring and incremental updates, a synthesis of available science using SER techniques would mesh well with landscape-level adaptive management as mandated by OAR 629-035-0020 (3).

Adaptive management confronts scientific uncertainty directly by specifically trying to learn from management interventions, which are essentially experiments in progress (Fazey et al. 2004). Provided that adequate monitoring programs are in place, new information should emerge that could be integrated into an existing SER through an SER update. This stepwise, incremental improvement of a solid and reliable knowledge base could be used to incrementally refine management strategies, which is very consistent with tenets of the adaptive approach.

Principles, guidelines and considerations for making ODF science reviews more systematic and transparent are described in greater detail in the next section.

## **VI. Principles, Guidelines and Considerations for Applying Systematic Evidence Reviews to Forest Management in Oregon**

ODF commissioned this project to develop a “small scale, practical approach that can be tested and adapted for use by ODF programs to objectively review and characterize technical information for decision making” (ODF 2005). Developing reliable summaries of research results and other evidence will address the statutory requirement that ODF state forest management practices “be based on the best science available” (OAR 629-035-0020 [3]). There is a critical need for an information synthesis process that is credible enough to reduce conflicts among forest management stakeholders who come to the table with different ideas about what constitutes “best science,” and a tendency to offer only studies that support their point of view.

Components of the SER process may help reduce this conflict by helping ODF to better clarify and document the technical questions they address, and how they define, locate and evaluate the evidence that is relevant to these questions. With adequate time and resources, the SER process may aid in providing a more comprehensive, credible summary of “best science available” at a given point in time. Despite the effort and resources necessary to develop accurate, trustworthy information, such efforts provide significant payoffs in additional credibility among stakeholders (Shindler and Neburka 1997).

### **Key Points**

After reviewing how SERs are used in clinical medicine and public health, and challenges and opportunities regarding the application of SERs to ecosystem and natural resource management, several key points emerge that frame options for incorporating tenets of SERs into ODF science assessment efforts:

- “Evidence”, in an ecosystem management context, is more than just data and hard facts- it involves contextual information and interpretation. *Scientific* evidence consists of scientifically guided empirical observations combined with background information, logic, and scientific expertise.
- Sweeping generalizations about the appropriateness of particular statistical or research methods over others are unwarranted, and laboratory experiments do not necessarily carry greater weight than field experiments in forest ecology. All types of data can add to the evidence base.
- This evidence does not have to be quantitative, or gathered by a scientist. The key is that the information was collected and interpreted as objectively as possible and can somehow be verified.
- The weight given to a particular piece of evidence should not depend on the type of observation but on the match between the observation and the question being asked.

- Medical SERs are, by design, rigorous, exhaustive and comprehensive, and thus time consuming and costly. Thus, a “small scale, practical approach” to science assessment is fundamentally different than SERs as they are defined in the medical field.
- Despite this inconsistency, some *aspects* of SERs could be readily adopted and incorporated into the internal and external science reviews that ODF already conducts.
- A systematic review of evidence pertaining to a forest management question *may* be feasible (for ODF) if 1) the question is tightly focused, 2) the evidence base pertaining to the question is not large, and 3) there is consensus on the boundaries of this evidence base.
- Natural resource SERs are more likely to be feasible for focused questions involving a single intervention and/or a single species. Multifaceted questions that involve more than one species or more than one outcome would be more difficult to address using the SER process.
- Independence from stakeholders is a fundamental aspect of SERs. An SER is more likely to be considered objective by all stakeholders if it is conducted by an independent entity, rather than internally by ODF.
- An SER process *may* reveal general consensus on the scientific evidence that is masked by fundamental differences of opinion on what outcomes are most important and what actions are appropriate in the face of imperfect evidence.

### **Three options for using the SER approach in Oregon forest management**

A full-scale SER on a complex forest ecosystem science question could be a major undertaking compounded by the need to recruit and train an external SER team before commencing the review itself. The feasibility of natural resource SERs and circumstances in which they would be most useful are not clear. Much could be learned by testing the SER approach, which offers some clear improvements over traditional literature reviews.

ODF could take an **incremental approach** to adapting the SER process to forest policy making in Oregon. Three tiered options for doing so are outlined below. These options roughly parallel the three existing approaches to science review at ODF: (1) routine internal reviews, (2), external reviews commissioned by ODF to review long-term planning documents, and (3) other external reviews completed as part of broader policy initiatives such as IMST reports for the Oregon Plan.

The form and details of each option are provided as a starting point and could benefit from further management and stakeholder review and discussion. The agency could develop a hybrid approach tailored to its needs in a particular circumstance.

**Option 1: Incorporate SER techniques into ODF’s “in-house” science assessments and any external review of this work.** The primary aim here would be to make existing ODF internal science review processes more transparent. This could be achieved with adjustments to what is already being done, primarily by adopting components of SERs to better document how science information is gathered and reviewed.

Under this scenario, ODF would not be rigidly bound to assuring that the review was an absolutely exhaustive and complete examination of all available evidence. As with all science reviews however, credibility would be predicated on perceptions of the degree to which the review was thorough and objective. This option would be best suited to cases where the available evidence is relatively clear, uncontroversial and limited in scope.

*Option 1a:* Conduct the science assessment “in house” as described above, with the additional step of soliciting external review of the draft final document. This process would approximate that used during the Independent Scientific Review of the Draft Western Oregon State Forests Habitat Conservation Plan, as described in Section II. The key differences would be use of SER procedures during the “in house” phase, and that external reviewers would be asked to assess the quality of evidence used and upon which they base their review and comments.

*Key components of Option 1:*

- To the degree possible, develop tightly focused, specific questions to delineate the purpose and scope of the review.
- Develop a simplified SER-like protocol to explain how the review will be conducted, using the example shown in Appendix 2 as a starting point. Development and use of a formal evidence quality hierarchy and ranking system is probably not be feasible at this level of review, but a narrative discussion and comparison of study quality could strengthen the review.
- Document in a systematic way which studies were included, what they said and how the information was interpreted. If studies were identified as relevant, but not included for quality or other reasons, document these reasons.
- If/when documents are sent out for external review, include in the review process an expectation that reviewers will also provide a quality assessment of the information upon which they based their review comments.

More specific guidelines for how this option might be implemented are offered in Appendix 3.

**Option 2: Commission an SER by an external, independent entity.** Under this scenario, a review of evidence would be contracted to a qualified independent entity. Such a review might be triggered by politically sensitive or difficult scientific questions about which ODF staff sought external scientific review. External review should assume impartiality and take advantage of academic expertise in specialized subdisciplines within ecological science.

The overall aim would be to prepare a defensible SER for a natural resource question, or a limited set of questions, with corresponding effort to obtain all relevant evidence and review it in formal, documented fashion. As with Option 1, this would not require an entirely new process. Existing external science review entities would consider using the SER approach.

*Key components of Option 2:*

- ODF would develop tightly focused questions to frame the purpose and scope of the SER. The additional step of vetting the SER questions with stakeholders could be considered. Questions would be refined in collaboration with stakeholders and SER review team.
- Develop a protocol that explicitly lays out how the review will be conducted. The external SER team should take the lead, or at least be included, in this process. If the review team believes it is feasible, develop and apply a formal set of evidence quality assessment and ranking criteria to the included studies.
- Publish results of review on ODF website and in academic journals.

**Option 3: Collaborate with other state and federal agencies to address regionally significant, highly policy relevant questions of using the SER process.** Many forest management issues transcend agency boundaries and should be addressed at the landscape scale. Some of these issues are controversial and challenging, and more than one agency could benefit from synthesis of all available evidence into a package of “best available science” that all participating agencies could then use. Post-wildfire “salvage” logging and restoration is an example of a topic for which it may be worthwhile for ODF to initiate and/or participate in multi-agency efforts to identify key questions and support an SER process to address them.

Topics would need to be carefully considered because of the time and effort that would likely be required to coordinate a multi-agency SER. Various approaches are possible. For example:

*Option 3a:* Bring together an SER team comprised of technical specialists from within different agencies (e.g. ODF, Oregon Department of Fish and Wildlife, USDA Forest Service, USDI Bureau of Land Management, Fish and Wildlife Service, National Park Service, NOAA Fisheries) to develop questions, a protocol and conduct the review.

*Option 3b:* Conduct an external SER as described in Option 2, but solicit and coordinate support from other agencies.

*Key components of Option 3:*

- Similar to Option 2, but with interagency collaboration in (1) identifying and refining questions and vetting them with stakeholders, and (2) locating evidence, particularly unpublished monitoring data and other agency-specific information that may not be widely available, and (3) providing support to conduct the SER.

### **Responses to ODF Questions about SERs**

ODF submitted several questions on the topic of applying an SER-like process to forest policy issues for INR to consider. Several aspects of these questions have been discussed elsewhere in this report but some have not, so each question is specifically revisited below. Most responses hinge on the complexity and depth of the science review.

#### **1. What circumstances might trigger an SER-type process for ODF?**

- Ascertain, if possible, whether or not the issue is primarily about science. If it is not, then an SER is probably not appropriate.
- Also ascertain whether adequate evidence exists that could be evaluated using an SER approach. SERs can also illuminate information gaps, but if it is already clear that such a gap exists, an SER might not be worth the time and expense.
- A politically high-profile forest management policy decision or topic for which scientific evidence will be a primary consideration.
- A situation in which controversy over interpretation of scientific knowledge about a particular topic is hindering progress on policy making, and in which initial assessment indicates that clarifying science can break the logjam.
- A topic or management action for which there are significant risks of negative consequences if conclusions about the science on which the action is based turn out to be wrong.

- An emerging policy issue for which little scientific information has been previously collected and assessed by ODF, or a situation in which stakeholders and policymakers are looking to the agency for technical expertise.
- A situation in which other science review processes have left significant uncertainty concerning management direction, and in which initial scoping indicates that additional evidence is available.

It may also be useful to think about different *levels* of science review (Guldin et al. 2003). The agency could conduct “in house” reviews on less complex, less controversial topics with a better evidence base. For more complex questions, or where the science is less clear, stakeholder involvement and consultation with experts from other agencies and academia may be warranted or necessary to improve consensus, transparency and rigor. Circumstances that prompted the review will help guide such decisions (Table 7).

**Table 7:** Factors to consider when determining the appropriate level of science review

<b>Factor To Consider</b>	<b>Potential Effects (Risks &amp; Benefits)</b>	
	<b>Lower</b> level of science review <i>(Review probably not needed or met by smaller, perhaps internal effort)</i>	<b>Higher</b> level of science review <i>(Review helpful, larger, perhaps external effort required)</i>
Level of public or science controversy	<ul style="list-style-type: none"> <li>• Limited in scope and action</li> <li>• Lesser controversy</li> </ul>	<ul style="list-style-type: none"> <li>• Highly disputed and/or arguably insufficient data</li> <li>• Greater controversy</li> </ul>
Legal mandate	<ul style="list-style-type: none"> <li>• No clear mandate requiring science evaluation</li> </ul>	<ul style="list-style-type: none"> <li>• Clear mandate requiring science evaluation</li> </ul>
Spatial and temporal scales	<ul style="list-style-type: none"> <li>• Localized site conditions</li> <li>• Small watershed</li> </ul>	<ul style="list-style-type: none"> <li>• Broad geographic ranges; multiple agencies</li> <li>• Transcend organizational boundaries</li> </ul>
Duration of effects	<ul style="list-style-type: none"> <li>• Short-term effect on communities economy, and/or environment</li> </ul>	<ul style="list-style-type: none"> <li>• Long-term effect on communities, economy, and/or environment</li> </ul>
Scope of decision	<ul style="list-style-type: none"> <li>• Routine management actions (site-specific)</li> </ul>	<ul style="list-style-type: none"> <li>• Large-scale regional and forest plans or plan amendments</li> </ul>
Clarity of state of knowledge	<ul style="list-style-type: none"> <li>• Well-developed routine analysis</li> <li>• Professionally recognized science findings</li> </ul>	<ul style="list-style-type: none"> <li>• Emerging science and technology</li> <li>• Disputed findings and interpretations</li> <li>• Ambiguous, diversity of opinion</li> </ul>
Data availability & quality	<ul style="list-style-type: none"> <li>• Quality data exists</li> <li>• Generally accepted</li> <li>• Associated risk small</li> </ul>	<ul style="list-style-type: none"> <li>• Substantial data gaps</li> <li>• Arguably insufficient data</li> <li>• Highly disputed</li> </ul>
Scope of effects	<ul style="list-style-type: none"> <li>• Limited effect on or change to communities and the environment</li> </ul>	<ul style="list-style-type: none"> <li>• Long-ranging associated risks to communities and the environment</li> </ul>

Based on similar tables in Guldin et al. (2003) and Gravenmier and Connelly (2005).

## 2. How should review questions be formed, and who should be involved?

*How should review questions be formed?*

- The review question should be management or policy relevant.
- The question should be clearly defined and answerable in scientific terms because it both generates search terms and determines relevance criteria. The challenge may lie in structuring the question with a degree of specificity that allows inclusion of enough evidence to make the review worthwhile, but also limits it to a manageable scope. This may require compromise between a holistic approach, involving many variables and therefore closer to reality, and a reductionist approach that may limit the review's relevance, utility and value.
- The question should specify the *subject*, or *population*, the *intervention*, or *management action*, the *outcome*, and any relevant *comparator*.
- Achieving clarity about essential subject, intervention and outcome elements of the question can be challenging but it is important to be explicit at an early stage.

*Who should be involved?*

- In medical SERs, experts as well as practitioners and end users of the information are involved. Some stakeholder involvement may be beneficial in a natural resource SER to help reach consensus on the question, with due caution in assessing input from stakeholders with vested interests in the review outcome.

Example Question: In *watersheds with populations of Coho salmon* (the subject) do *buffer strips on non-Coho-bearing streams* (the intervention) result in *significantly lower water temperatures in Coho-bearing streams* (the outcome) compared to *no buffer strips* (the comparator)?

### **3. What are the criteria for evaluating information, and who develops the criteria?**

Evidence quality assessment criteria and hierarchies used in medical SERs will probably need to be adapted as part of generating the protocol for a natural resource SER. Deciding how to weigh different types of evidence is likely to be challenging. Appendix 4 offers some guidance. (Depending on the nature of the questions and evidence base, this step would probably not need to be repeated from scratch for every review.)

Ecologists, wildlife biologists, silviculturists and other specialists from academia and natural resource agencies with relevant expertise are the logical pool of experts to consult for help developing evidence quality and ranking criteria. As appropriate or deemed necessary, SER evidence quality criteria could be made available for review by stakeholders and the user community prior to initiating work. For any monitoring data to be included, technicians responsible should be consulted to provide details about how sampling was conducted. This could shed light on the quality and reliability of the monitoring data, and add valuable contextual information to the evidence base.

Rigid quality assessment criteria based solely on experimental design may not be useful if the evidence base is limited to studies with few experimental controls. At “lower” levels of evidence (e.g. observational studies, particularly without controls), other dimensions of quality become even more critical. These dimensions include:

- How well the study was conducted (was bias taken into account, were there confounding factors?)
- Size/strength of the effect (if the effect is very large then the association is likely to be correct even if the studies are low level, e.g. smoking and cancer).
- Volume of evidence (are there multiple studies that show the same effect?)
- Consistency- is the same effect seen in different studies?
- Specificity- is the effect specific for a particular set of parameters and not for others?
- Temporality- is the causative agent always present before the effect occurs (i.e. does the ‘cart comes before the horse’?)
- Biological gradient- is there is a relationship between intervention and response?
- Plausibility- is there a plausible mechanism for the effect? (this factor may not be reliable on its own, but combined with the other criteria, a known mechanism for the effect seen adds considerable weight to the assessment).
- Coherence- does the effect comply with the known facts about the natural history and biology of the condition? (related to plausibility)
- Analogy- in some cases it may be reasonable to draw an analogy with another similar situation, particularly if the analogous situation had very disastrous consequences.

In this vein, it may also be useful to consider the “signal to noise” ratio in the study, as described by Edwards et al. (1998). In other words, if the effect size (the signal) is very strong in relation to the study design drawbacks or nuisance variables (the noise) the study might be included even if it has significant methodological weaknesses.

The review protocol may require amendment as the review progresses and additional considerations come to light. If so, the reviewers should be explicit about where and why modifications occurred to ensure that the process is transparent and to allow stakeholders to judge the validity of the review.

#### **4. How is synthesis of the information best handled so that the assessment remains objective?**

Synthesis of multiple scientific studies (and other relevant evidence), even using quantitative methods, requires judgment calls by reviewers at several junctures, e.g. which studies to include and which to exclude, how to assess the quality of and assign weights to different studies in the evidence base, and how to integrate the findings of all pieces of evidence (Steinberg and Luce 2005). Interpreting results and drawing conclusions also involves judgment calls by individual scientists in reports on their work.

The fundamental strength of the scientific method and also of SERs is that they involve checks and balances throughout the process (e.g. explicitly defined methods for conducting the study or review, peer review) to ensure that they are carried out *in as objective a manner as possible*. Compared to traditional literature reviews, SERs are considerably more objective in that study inclusion, assessment and synthesis criteria are much more explicit and transparent. Methods are continually being refined and improved in order to make SERs as objective as they can possibly be, with opportunities throughout the process for reviewers to be made aware of and address potential sources of bias.

Science reviews conducted independently by a party external to ODF and stakeholders with vested interests may be perceived as more objective and credible. (This is not a comment on the quality of internal ODF reviews.) If the review topic is high profile or contentious, involving stakeholders during development of review questions and protocols, and diligent documentation of steps in the review should enhance its overall credibility. Stakeholders should also be involved in the search for relevant literature. If studies identified by stakeholders are ultimately excluded for quality or relevance reasons, the reasoning behind these decisions should be documented.

Citizen involvement in science assessment also may complicate the policy development process, a potential tradeoff for increased credibility. Problems with separating values from what the evidence says may surface. SERs offer the potential to incrementally enhance social learning about scientifically guided knowledge of forest ecosystems, but in the end, complete consensus that a science review is “objective” may be difficult to achieve. Judgment calls are necessary when integrating and interpreting science, and there may be fundamental differences of opinion about these judgments.

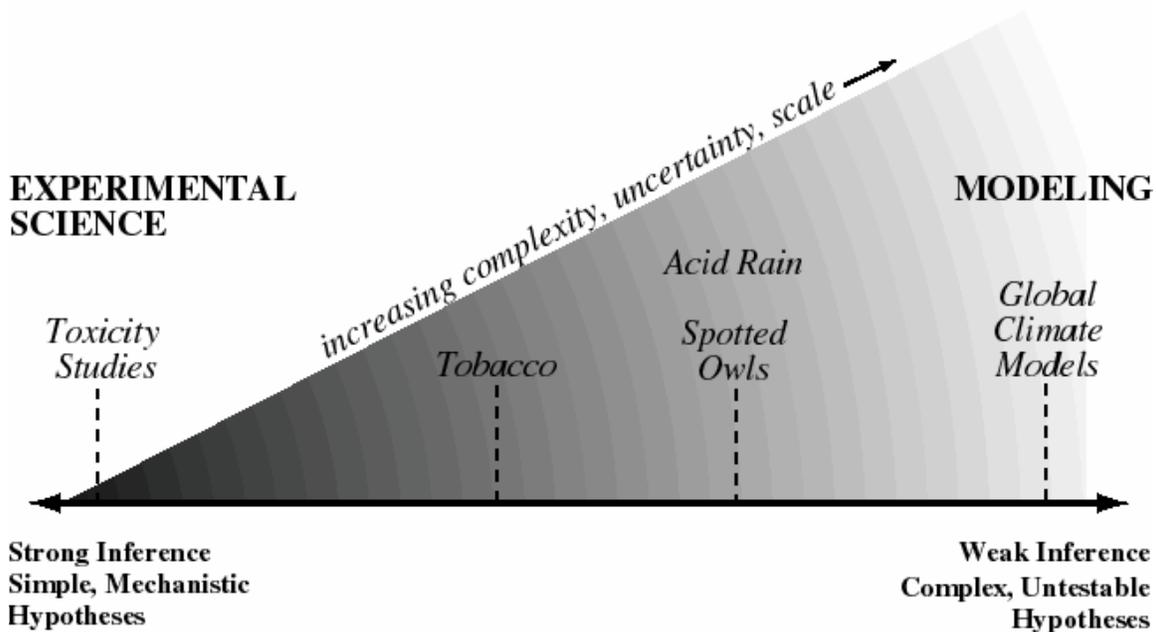
#### **5. How are uncertainties associated with the information best characterized?**

Bradshaw and Borchers (2000) point out that uncertainty among citizens who are not familiar with how science works poses one of the most difficult challenges of translating science into policy. Ecosystem scientists are familiar with uncertainty and complexity, but the public and policy makers often seek certainty and deterministic solutions for environmental problems. They often have little tolerance for uncertainty associated with complex ecosystems expressed in terms such as *p*-values of 0.05 or model prediction errors. Uncertainty often leads to oversimplification of complex issues and rejection of information that conflicts with existing beliefs and behaviors, fueling acrimonious public

debates. Science frequently ends up competing with the demands of economics and politics, except when the scientific evidence provides a high level of certainty.

Lack of confidence in scientific findings among the public and policy makers, and the strength of scientific evidence, vary according to scale. At the same time, the importance of uncertainty depends on the context of a decision and increases with larger landscapes and longer time horizons (Burgman et al. 2005). In general, complex ecosystem-level evidence is intrinsically more uncertain while traditional experimental science usually retains more credibility (higher certainty) because it is conducted at scales familiar to people, or at levels of complexity where scientific inference is rarely disputed. Uncertainty concerning evidence in ecosystem and natural resource management can thus be characterized as a continuum (Figure 3).

**Figure 3: Continuum of uncertainty, and acceptance of scientific findings (Bradshaw and Borchers 2000.)**



Bradshaw and Borchers (2000) suggest that the best way to deal with these challenges is through social learning to better align public understanding of scientific uncertainty with that of the science community. For example, ecological monitoring, designed and performed in partnership with citizens, scientists and managers, can enhance public and institutional learning. Such a strategy could increase understanding that:

- Scientific evidence about ecosystems is intrinsically uncertain, with new information continually altering our perceptions and beliefs.

- Natural resource decisions based on scientific information must be made in a context of uncertainty.
- Faster and better science as an adequate basis for policy formulation is inconsistent with the nature of scientific inquiry and resilient policy formulation.

The *precautionary principle* has been widely suggested and in some cases adopted as a means for preserving future options and reducing risks of irreversible ecosystem changes in the absence of certainty about outcomes of management actions. Because uncertainty is inherent in managing natural resources, the principle holds that where there are threats of serious or irreversible damage, lack of full scientific certainty should not be used as a reason to postpone cost-effective measures to prevent environmental degradation. The principle recognizes that it is usually easier and less expensive to prevent environmental damage than to repair it later, and accordingly shifts the burden of proof away from those advocating protection toward those proposing an action that may be harmful.

Some argue that careful consideration is warranted before invoking the precautionary principle and that it does not always work as intended. For example, Mealey et al. (2005) make the case that the principle, as embodied in the way the Endangered Species Act has been interpreted and implemented under the Northwest Forest Plan has traded lower short term risks for greater long term risks to the northwest spotted owl by preventing active management that would ultimately improve owl habitat but may involve some short term risk to individual birds. Similarly, Chase (1997) argues that the evidentiary standards for “no harm” that are often required when the principle is invoked are so high that they can never be met, and prevent action as long as they are adhered to.

Deciding how much certainty is sufficient is inherently a judgment about risks of acting too soon (promoting a policy that turns out to be ineffective or harmful) and acting too late (delaying a beneficial action to get better evidence, Atkins et al. [2005]). To confront uncertainty, complementary approaches such as quantitative risk assessment, safe minimum standards and the precautionary principle can be used. (Fazey et al. 2004.)

**6. If the objectives of an SER process are to improve decision maker and stakeholder confidence in the information used for decision making, and perhaps to reduce conflict, how can the success of this new model be evaluated over time?**

In the medical field systematic evidence reviews are evaluated based on whether or not they influence medical or health care decisions made by the general public and policy makers. There is a lack of research on the impact that SERs actually have on medical decisions, which may be the result of a lack of interest by the research community or the complexity of studying decision-making processes (Bero and Jadad 1997).

SERs also might be evaluated with a survey given to both decision makers and stakeholders to determine whether they have experienced increased confidence in the information on which natural resource decisions are based. The survey could also be

used to determine if the new model proved to be successful in reducing disagreement and conflict associated with natural resource decision-making.

The utility of an SER like process over time could also be evaluated by a committee to evaluate use of the SER process to determine:

- Whether it represents an improvement over other ODF science review processes
- If it increases objectivity and transparency
- Reduces instances of “dueling science”
- Fair and equitable in terms of access
- Situations where it works well and situations where it does not

The success of an SER-like process could also be evaluated by comparing its strengths, weaknesses and effectiveness to those of science assessment processes used by forestry departments in other states, and those that ODF has used in the past.

## VII. Conclusions and looking ahead

- A pilot test of a modified SER process could shed light on the accuracy of many of the untested perspectives and assumptions in this report regarding the potential for SERs in natural resources

This report investigated the potential for applying a process used in clinical medicine to systematically review evidence concerning the efficacy of medical interventions to questions in the fields of natural resource and ecosystem management.

Despite the rather cautious tone we have taken regarding this potential, the need to better incorporate science into policy is great, and a modified SER process holds promise as an additional tool to address this need. John Kitzhaber established Oregon as a leader in applying the SER process to state policies for medical coverage (Fox 2005). State and federal agencies in Oregon have an opportunity to take advantage of this experience and reputation by taking the lead on testing the application of the SER approach to of technical information related to natural resource questions.

There are a number of differences between medical and forest ecosystem science that pose significant challenges in adapting the SER process to forest management issues. However, Oregon is well positioned to take a leadership role in attempts to be more systematic in the way that science and scientifically guided information appraised and used in land management policies. A pilot-test that applied a modified SER process to a carefully chosen question could shed light on the accuracy of the mostly untested perspectives and assumptions outlined in this report about the potential for SERs in natural resource management.

Many of the forest management questions to which an SER-like process might be applied transcend ODF agency boundaries. They apply to all forests that are managed for multiple values, including those on federal and private lands. This suggests that there is great potential for an interagency effort to support and conduct a trial science review using the SER approach.

Adopting SER components incrementally, exploring options for collaborating with other state agencies and universities to conduct pilot SERs, and staying abreast of changes in the way SERs are applied in the medical field are all ways make further progress toward greater understanding and consensus on the science used and the way it is used in forest policy.

## Appendix 1: Examples of Evidence Quality Hierarchies and Quality Assessment Criteria Used in Medical Systematic Evidence Reviews

### Quality criteria for assessment of medical evidence at higher and lower levels of evidence

(Based on NHMRC 2000, How to Use the Evidence: Assessment and Application of Scientific Evidence.)

STUDY TYPE	QUALITY CRITERIA
<b>Higher levels</b>	
Systematic review	Was an adequate search strategy used? Were the inclusion criteria appropriate and applied in an unbiased way? Was a quality assessment of included studies undertaken? Were the characteristics and results of the individual studies appropriately summarized? Were the methods for pooling the data appropriate? Were sources of heterogeneity explored?
Randomized controlled trial	Was the study double blinded? Was allocation to treatment groups concealed from those responsible for recruiting the subjects? Were all randomized participants included in the analysis?
<b>Lower levels</b>	
Cohort study	How were subjects selected for the intervention? How were subjects selected for the comparison or control group? Does the study adequately control for demographic characteristics, clinical features and other potential confounding variables in the design or analysis? Was the measurement of outcomes unbiased (i.e. blinded to treatment group and comparable across groups)? Was follow-up long enough for outcomes to occur? Was follow-up complete and were there exclusions from the analysis?
Case-control study	How were cases defined and selected? How were controls defined and selected? Does the study adequately control for demographic characteristics and important potential confounders in the design or analysis? Was measurement of exposure to the factor of interest (e.g. the new intervention) adequate and kept blinded to case/control status? Were all selected subjects included in the analysis?

### Evidence quality assessment criteria based on study design (Khan et al. 2003).

*Experimental study*- A comparative study in which the use of different interventions among participants is allocated by the researcher. Level assigned to evidence based on soundness of design: I

- Randomized controlled trial (with concealed allocation): Random allocation of participants to an intervention and a control (e.g. the placebo or usual care) group, with follow-up to examine differences in outcomes between the two groups. Randomization (with concealment of allocation sequence from caregivers) avoids bias because both known and unknown determinants of outcome, apart from the intervention, are usually equally distributed between the two groups of participants.

- Experimental study without randomization (sometimes erroneously called quasi-experimental or quasi-randomized or pseudorandomized studies): Studies in which allocation of participants to different interventions is managed by the researcher but the method of allocation falls short of genuine randomization, e.g. alternate or even-odd allocation. Such methods fail to conceal the allocation sequence from caregivers.

*Observational study with control group*- comparative study in which the use of different interventions among participants is not allocated by the researcher (it is merely observed). Level assigned to evidence based on soundness of design: II

- Cohort study: Follow-up participants who receive an intervention (that is not allocated by the researcher) to examine the difference in outcomes compared to a control group, e.g. participants receiving no care.
- Case-control studies: Comparison of intervention rates between participants with the outcome (cases) and those without the outcome (controls).

*Observational study without control group*. Level assigned to evidence based on soundness of design: III

- Cross-sectional study: Examination of the relationship between outcomes and other variables of interest (including interventions) as they exist in a relevant population at one particular time.
- Before-and-after study: Comparison of outcomes in study participants before and after an intervention.
- Case series: Description of a number of cases of an intervention and their outcomes.

*Case reports*- Pathophysiological studies or bench research; expert opinion or consensus  
Level assigned to evidence based on soundness of design: IV

### British Medical Journal (BMJ) Guidelines

The British Medical Journal (BMJ) group publishes the online journal *Clinical Evidence* which summarizes evidence on medical interventions from high quality SERs and large well-designed randomly controlled trials (RCTs). (British Medical Journal 2005.) Evidence on prognosis or baseline risk may also come from such studies or from well-designed cohort studies. Studies are thought of in these categories:

- Methodology sound; INCLUDE
- Methodology suboptimal; INCLUDE if necessary but cite reservations
- Methodology unsound; DO NOT INCLUDE

The BMJ group uses the criteria listed below to assess study quality.

*BMJ quality criteria for systematic evidence reviews:*

- Questions and methods clearly stated
- Comprehensive search methods described.
- Explicit methods used to determine which studies were included in the review.
- Methodological quality of primary studies was assessed.
- Selection and assessment of primary studies reproducible and free from bias.
- Differences in individual study results adequately explained.
- Results of primary studies combined appropriately.
- Reviewers' conclusions supported by data cited.

*BMJ quality criteria for randomized controlled trials:*

- Were the setting and study patients clearly described?
- Was assignment randomized and similarity between groups documented?
- Was allocation to study groups adequately concealed from patients and investigators, including blind assessment of outcome?
- Were all clinically relevant outcomes reported?
- Were > 80% of patients who entered the study accounted for at its conclusion?
- Were they analyzed in the groups to which they were randomized (intention to treat)?
- Were both statistical and clinical significance considered?

*BMJ quality for cohort studies on prognosis of baseline risk*

- Was an inception cohort assembled?
- Recruitment setting, diagnostic criteria, disease severity, co-morbidity and demographic details should be documented
- Was the referral pattern described?
- Referral or diagnostics access bias avoided?
- Was an adequate follow up rate achieved?
- > 80% patients entered accounted for in results and clinical status known?
- Were objective outcome criteria developed and used?
- Was outcome assessment blind?
- Was adjustment for extraneous prognostic factors carried out?

*BMJ quality criteria for evidence of harm*

Rules of evidence on harm are the same as rules of evidence on beneficial effects of treatment in that the best evidence on harmfulness or otherwise of treatments comes from large RCTs (or reviews thereof). However group sizes have to be large and follow up prolonged for rare side effects to be detected and evidence on harm from RCTs may not be available. Bias due to non-comparability of groups is more likely in cohort studies and more likely still in case control studies. Case series or case reports are the weakest forms of evidence, though associations in case reports have often been subsequently confirmed.

- Was the study the strongest that could have been performed under the circumstances?
- Were study groups sufficiently comparable in respects other than exposure?
- Was determination of exposure free from bias?
- Was the determination of outcomes (in cohort studies) or the distinction between cases and controls (in case control studies) free from bias?

Were both clinical and statistical significance considered in reporting the strength of the association?

Is the association consistent in different studies?

Is the temporal sequence of exposure and outcome in the right direction?

Is there a dose response gradient?

Does the association make sense?

## **Appendix 2: Sample Protocol for a Natural Resource SER**

Retrieved from:

<http://www.cebc.bham.ac.uk/Documents/instreamdevicesProtocolfinal.pdf>

### **CENTRE FOR EVIDENCE-BASED CONSERVATION SYSTEMATIC REVIEW NO. 12**

#### **WORKING TITLE: DOES IN-STREAM HABITAT IMPROVEMENT INCREASE THE ABUNDANCE OF TROUT AND SALMON?**

**Lead Reviewer: Currently there is no lead reviewer.**

**Please contact Dave Showler regarding this review.**

Postal Address: Dave Showler.

Centre for Ecology, Evolution and Conservation,

School of Biological Sciences,

University of East Anglia,

Norwich,

NR4 7TJ.

**Email Address: [d.showler@uea.ac.uk](mailto:d.showler@uea.ac.uk)**

**Facsimile: +44 (0)1603 592250**

### **REVIEW PROTOCOL**

#### **1. BACKGROUND**

In-stream habitat improvement devices may be used in an attempt to redress habitat degradation and enhance trout and salmon stocks, in streams and rivers which have been detrimentally affected by anthropogenic influences. Interventions such as installment of flow deflectors, artificial riffles, and also the use of livestock fencing to reduce bankside erosion, aim to restore habitat to something approximating natural conditions.

An effect of installation of in-stream habitat devices is often the narrowing of over-wide streams with subsequent increased water velocities and turbulence resulting in beneficial impacts on water quality (Environment Agency (EA) 1996, Hendry *et al.* 2003, Milner pers.com 2005). Fencing and isolation of the river from livestock indirectly promotes beneficial salmon habitat reducing erosion and sediment inputs (Duff 1977, Platts *et al.* 1983, Platts & Nelson 1985), However the effectiveness of these devices is often not fully known as performance evaluations are rarely conducted (Harper & Quigly 2005). The EA is interested in ascertaining the impact of flow deflectors, artificial riffles and livestock fencing on the abundance of trout and salmon stocks and also bullhead (a UK BAP species of conservation concern).

The effectiveness of in-stream devices may be affected by local gradient and valley confinement (drivers of geomorphology), proportion of cobbles in substrate (driver of salmonid distribution, Armstrong et al. 2003), degree of existing modification (negatively related to salmonid abundance), and distance from source and water quality (effects carrying capacity, Armstrong et al. 2003). The impact of these potential effect modifiers also requires investigation.

A systematic review methodology will be used to retrieve data concerning the impact of in-stream habitat improvement devices. The review will limit bias through the use of comprehensive searching, specific inclusion criteria and formal assessment of the quality and reliability of the studies retrieved. Subsequent data synthesis will summarise empirical evidence, thereby assisting in the formulation of appropriate evidence-based management guidelines and highlighting gaps in research. The review should be of use to Environment Agency and Rivers Trusts practitioners and also have wider international relevance.

## 2. OBJECTIVE OF THE REVIEW

### 2.1 Primary question

*Does in-stream habitat improvement increase the abundance of trout and salmon?*

Table 1: Definition of components of the primary systematic review question.

Subject (Population)	Intervention	Outcome		
		Primary	Secondary	Tertiary
Trout and salmon populations	Devices that aim to restore stream habitats i.e. anti-livestock fences, artificial riffles and flow deflectors Vs No instream devices	Change in abundance of trout and salmon	Changes in the abundance of other species including bullheads.	Any other outcomes

### 2.2 Secondary question

*What influence does the local gradient and valley confinement, proportion of cobbles in substrate, degree of existing modification, distance from source, water quality, water flow and size of stream have on the impact of instream habitat devices?*

## 3. METHODS

### 3.1 Search strategy

The following electronic databases will be searched:

1. ISI Web of Knowledge
2. Science Direct
3. Directory of Open Access Journals (DOAJ)
4. Copac

5. Scirus
6. Scopus
7. Index to Theses Online (1970-present)
8. Digital Dissertations Online
9. Agricola
10. Europa
11. English Nature's "Wildlink"
12. JSTOR
13. BIOSIS via EDINA
14. SIGLE via ARC2WebSPIRS

The following English language and latin search terms will be used:

1. Trout\*
2. Salmo\*
3. Bullhead\*
4. Cottus gobio
5. River\* and flow
6. Stream\* and flow
7. flow deflector\*
8. graz\* and fish

Further terms may be added as the search progresses involving combination of the existing terms and the use of taxa-specific terms if necessary. Publication searches will be undertaken on conservation and statutory organization websites (Countryside Council of Wales, Department of Agriculture and Rural Development, Department of Environment, Food and Rural Affairs, English Nature, EA, Freshwater Fisheries Lab, Joint Nature Conservancy Council, National Parks and Wildlife Service, Scottish Natural Heritage) and using the meta-search engines Dogpile, Alltheweb and Google Scholar. Fishbase.org. will also be searched. The first 100 word document or PDF hits from each data source will be examined for appropriate data. Additional foreign language searches will be undertaken on Google to capture information from the following non-English speaking countries with significant managed salmonid fisheries: Denmark, Finland, France, Norway, Spain and Sweden. In addition bibliographies of articles viewed at full text will be searched. Authors, recognized experts and practitioners will also be contacted for further recommendations and for provision of any unpublished material or missing data that may be relevant. The EA will be asked for access to information retrieved from the Research and Development Project 603 which produced R&D Technical Report W18, Restoration of Riverine Trout Habitats: A Guidance Manual. Participants in the pan-European FAME project (Development, Evaluation and Implementation of a Standardized Fish-based Assessment Method for the Ecological Status of European Rivers) will be asked to provide access to data. Questionnaires will be circulated to practitioners in order to collate experience.

### 3.2 Study inclusion criteria

- Relevant subjects: Rivers and streams containing trout and salmon populations.
- Type of Intervention: Devices that aim to restore stream habitat e.g. anti-livestock fences, artificial riffles and flow deflectors vs. no treatment.
- Types of Outcome: The primary outcome is change in abundance of trout and salmon. However studies will not be rejected on the basis of outcome and outcomes other than change in fish abundance will be catalogued. Adverse outcomes have been defined by the EA as a reduction in abundance of bullheads by 50% or more.
- Types of Study: Type of study will not be used to define inclusion or exclusion criteria. It is envisaged that all information regarding the primary outcome will be collated within a Bayesian

framework. Appropriate spatial or temporal controls are a prerequisite for studies to be included in inferential meta-analysis.

Where there is insufficient information to make a decision regarding study inclusion when viewing titles and abstracts, then relevance to the next stage of the review process will be assumed. Reviewers will consider articles viewed at full text for relevance excluding or admitting them to different categories of relevance and quality. At least two reviewers will independently assess a random subset of 25% of articles viewed at full text. Disagreement will be resolved by consensus, or following assessment by a third reviewer.

### 3.3 Study quality assessment

Reviewers will consider articles viewed at full text excluding or admitting them to different categories of information quality. At least two reviewers will independently assess a random subset of 25% of articles viewed at full text. Disagreement will be resolved by consensus, or following assessment by a third reviewer.

### 3.4 Data extraction strategy

Data regarding study characteristics, quality and results will be recorded on a specially designed data extraction form. These forms may be amended after consultation with statisticians and piloting of the data extraction process.

### 3.5 Data synthesis

It is envisaged that all information will be collated within a Bayesian framework. This will incorporate meta-analysis where appropriate data exists. Reasons for heterogeneity in results including local gradient and valley confinement, proportion of cobbles in substrate, degree of existing modification, distance from source and water quality will be investigated by meta-regression where appropriate data exists.

### 3.6 Reasons for heterogeneity

The following potential reasons for heterogeneity have been formally identified a priori in order of importance by the EA.

1. Local gradient and valley confinement
2. Proportion of cobbles in substrate
3. Degree of existing modification
4. Distance from source
5. Water quality
6. Water flow
7. Size of stream

## 4. POTENTIAL CONFLICTS OF INTEREST AND SOURCES OF SUPPORT

No conflicts of interest to be declared. This systematic review is funded by NERC

## 5. REFERENCES

Armstrong, J.D., Kemp, P.S., Kennedy, G.J.A., Ladle, M., and Milner, N.J. (2003) Habitat requirements of Atlantic salmon and brown trout in rivers and streams. *Fisheries Research*. 62, 143-170.

Duff, D.A. (1977) Livestock grazing Impacts on Aquatic Habitat in Big Creek, Utah. Symposium on livestock interacting with wildlife fish and their environments. California USA, Sparks, Nevada, University of California, Davis.

Environment Agency (1996) Restoration of Riverine Salmon Habitats. A Guidance Manual. Fisheries Technical Manual 4. R&D Technical Report W44.

Harper D.J. & Quigley J.T. (2005) A comparison of the area extent of fish habitat gains and losses associated with selected compensation projects in Canada. *Fisheries*, 30 (2), 18-25.

Hendry, K., Cragg-Hine, D., O'Grady, M., Sambrook, H., and Stephen, A. (2003) Management of habitat for rehabilitation and enhancement of salmonid stocks. *Fisheries Research*. 62, 171-192.

Milner, N.J., Wyatt, R.J., Scott, M.D. (1993) Variability in the distribution and abundance of stream salmonids and the associated use of habitat models. *Journal of Fish Biology*. 43 (Suppl. A), 103-119.

Platts, W.S., Magahan, W.F. and Mishall (1983) Methods for Evaluating Stream, Riparian and Biotic Conditions. U.S. Forest Service General Technical Report. INT-138. *North American Journal of Fish Management*. 9, 332-340.

Platts, W.S. and Nelson, R.L. (1985) Stream Habitat and Fisheries Response to Livestock Grazing and Instream Improvement Structures, Big Creek, Utah. *Journal of Soil and Water Conservation*. 40, 374-378.

## 6. APPENDIX

EA provided the following contacts, stakeholders and experts. We intend to invite these people and organisations to comment on the protocol and attend a stakeholder meeting to verify and refine the protocol prior to undertaking the search for information.

Contacts: Miran Aprahamian and Mark Diamond

Stakeholders: Environment Agency fisheries staff, Fisheries staff in other public bodies in Scotland and Ireland, Fisheries Trusts, The Association of Fisheries Trusts and other angling organisations.

Experts (researchers): Nick Giles, Terry Langford, David Cragg-Hine, David Solomon, Ted Potter, Martin O'Grady.

### Appendix 3: Guidance for Adapting Components of Systematic Evidence Reviews to Small-Scale Technical Information Reviews

This appendix provides general guidance for structuring current ODF science review procedures in a manner consistent with tenets of systematic evidence reviews.

Science reviews can span a continuum from small-scale internal reviews to large-scale reviews conducted by independent entities with multiple agency participation. Decisions about the level of review that is needed or most appropriate can be aided by considering the factors shown in Table 1:

**Table 1:** Factors to consider when determining the appropriate level of science review

Factor To Consider	Potential Effects (Risks & Benefits)	
	Lower level of science review (Review probably not needed or met by smaller effort)	Higher level of science review (Review helpful, larger effort required)
Level of public or science controversy	<ul style="list-style-type: none"> <li>Limited in scope and action</li> <li>Lesser controversy</li> </ul>	<ul style="list-style-type: none"> <li>Highly disputed and/or arguably insufficient data</li> <li>Greater controversy</li> </ul>
Legal mandate	<ul style="list-style-type: none"> <li>No clear mandate requiring science evaluation</li> </ul>	<ul style="list-style-type: none"> <li>Clear mandate requiring science evaluation</li> </ul>
Spatial and temporal scales	<ul style="list-style-type: none"> <li>Localized site conditions</li> <li>Small watershed</li> </ul>	<ul style="list-style-type: none"> <li>Broad geographic ranges; multiple agencies</li> <li>Transcend organizational boundaries</li> </ul>
Duration of effects	<ul style="list-style-type: none"> <li>Short-term effect on communities, economy, and/or environment</li> </ul>	<ul style="list-style-type: none"> <li>Long-term effect on communities, economy, and/or environment</li> </ul>
Scope of decision	<ul style="list-style-type: none"> <li>Routine management actions (site-specific)</li> </ul>	<ul style="list-style-type: none"> <li>Large-scale regional and forest plans or plan amendments</li> </ul>
Clarity of state of knowledge	<ul style="list-style-type: none"> <li>Well-developed routine analysis</li> <li>Professionally recognized science findings</li> </ul>	<ul style="list-style-type: none"> <li>Emerging science and technology</li> <li>Disputed findings and interpretations</li> <li>Ambiguous, diversity of opinion</li> </ul>
Data availability & quality	<ul style="list-style-type: none"> <li>Quality data exists</li> <li>Generally accepted</li> <li>Associated risk small</li> </ul>	<ul style="list-style-type: none"> <li>Substantial data gaps</li> <li>Arguably insufficient data</li> <li>Highly disputed</li> </ul>
Scope of effects	<ul style="list-style-type: none"> <li>Limited effect on or change to communities and the environment</li> </ul>	<ul style="list-style-type: none"> <li>Long-ranging associated risks to communities and the environment</li> </ul>

Based on similar tables in Guldin et al. (2003) and Gravenmier and Connelly (2005).

Regardless of the level of review, a *protocol* should be developed describing how the review will be conducted. The protocol demonstrates ways that the review will be “systematic”. The level of detail in the protocol should reflect the complexity and depth of the review. A relatively simple protocol may suffice, the key is that it is explicit enough to reproduce the review steps later if necessary.

A sample protocol is shown in Appendix 2. Additional conservation protocols can be viewed at the Center for Evidence-based Conservation (CEBC) website:

<http://www.cebc.bham.ac.uk/>

Even if the review is not comprehensive, address some or all of the following SER elements:

1. **Try to frame *clear, concise questions***
2. **Document how and where evidence was searched for**
3. **Document information about included and excluded studies**
4. **Document how the evidence was interpreted and synthesized**

Each of these topics is covered in more detail below.

## 1. Framing Clear Questions

Asking the right question(s) is key to conducting a successful SER. Tightly focused questions are critical in order to delineate the scope of the review and reduce ambiguity when making judgments about which studies are relevant and which are not. The challenge may lie in structuring the question with a degree of specificity that allows inclusion of enough evidence to make the review worthwhile, but also limits it to a manageable scope. This may require compromise between a holistic approach, involving many variables and therefore closer to reality, and a reductionist approach that may limit the review's relevance, utility and value.

If possible and relevant to the issue, specify the

- *Subject*- the *species, process, or location* being assessed; the unit of study, defined in terms of the subject(s) to which the intervention will be applied, and appropriate geographical scale (e.g., ecosystem, habitat or species).
- The *intervention or management action* being proposed.
- The *outcome* and how it is measured. All relevant objectives of the proposed management intervention that can be reliably measured should be specified. Particular consideration should be given to the most important management outcome, and to any outcome that is critical in the judgment of whether the proposed intervention has greater benefits or disadvantages than any alternatives.
- A relevant *comparator* should also be defined. Is the intervention being compared to no intervention or are alternative interventions being compared to each other?

Example question: In *watersheds with populations of Coho salmon* (the subject) do *buffer strips on non-Coho-bearing streams* (the intervention) result in *significantly lower water temperatures in Coho-bearing streams* (the outcome) compared to *no buffer strips* (the comparator)?

## 2. Documenting Searches

Proper documentation of the evidence search phase helps ensure that the review is both replicable and transparent. It allows stakeholders to evaluate the thoroughness of the search for potentially relevant studies and how well they think bias has been minimized, which increases credibility. It also helps avoid repeating searches in future reviews.

For each electronic database searched, keep an accurate record of:

- the name of the database searched
- the name of the host/system used
- the date when the search was run
- the exact keywords or search terms used (standard keywords as well as Boolean operators and other search terms used in particular databases.)

Example: AGRICOLA on Ovid, searched January 2006, period 1980 to date. Terms:

Coho, buffers  
Coho, buffer  
Coho, buffer strip  
salmon, buffers  
salmon, buffer  
salmon, buffer strip  
salmonids, buffers  
salmonids, buffer  
buffer width  
Coho AND buffers  
salmon AND buffers

ODF technical specialists already use Procite bibliographic software tools as part of current science reviews. Such tools are indispensable in an SER for keeping track of which references are judged to be relevant, which have been ordered from sources, which have been received, and for documenting studies that were located but rejected. A hard copy file with a section for each study can be used to collect information about the study and all papers that report on it (Cochrane 2002).

Reference management software options:

ProCite  
Reference Manager  
EndNote  
IdeaList.

Reference management software also provides a good means for documenting how particular pieces of evidence were interpreted. For example, the “notes” field in Endnote is a logical place to keep track of such information for each reference. Procite and Endnote are popular reference management applications that are now owned by the same company and compatible with each other. Reference databases are increasingly

sophisticated, as is the reference management software. Reference information can often be imported directly without the need to cut and paste or type the information in.

Tips for saving time and effort:

- Identify terms used to index and describe a sample of studies already known to be relevant to the review, and use these terms in the search strategy.
- Pilot new search terms on part of the database to see whether they help identify relevant material before running them on the entire database.
- Use date limits for your search if appropriate. For example, if a technique has only been around since a certain date, there's no point searching before then

Important things to remember during documentation:

- Any modifications to the search (e.g. new search terms) should be documented.
- Reasons for changes and amendments should be noted at the time they are made.
- Unfiltered search results should be saved in their entirety and retained for future potential reanalysis.
- Details of the search may not appear in the finished report, but it is easy to note that full details are available on request

For journal hand searches, provide a list of journal titles in alphabetical order using full titles. State the earliest and latest year searched, and any missing journal issues that were not searched. Document conference proceedings titles in addition to the conference name. For efforts to identify unpublished studies, provide a brief summary including databases searched giving database details as above. Document efforts to contact investigators for information about unpublished studies, etc. Provide a brief summary of other sources searched (e.g. bibliographies, reference lists and internet web sites), giving date searched, search terms used, the URL if relevant, any specific features which might impact on the search process (e.g. 'only the titles were searchable').

### **3. Documenting information about included and excluded studies**

A narrative synthesis typically includes tables to provide a systematic and consistent record of information about all studies in the review and to allow comparison. A summary of included (and excluded) studies can be presented in a table with fields for some or all of the following:

- Study title and principal investigator(s)
- Study dates and *duration*\*
- Publications resulting from the study (if more than one)

- Study location, settings where the intervention was applied
- Ecosystem type; plant association group
- Watershed type (if applicable, e.g. 6<sup>th</sup> field)
- Research question(s), hypotheses
- Species studied (if applicable)
- *Study design, experimental controls\**
- *Pretreatment data (yes/no)\**
- Intervention, management action, or disturbance under study (e.g. fire)
- *Replications (if applicable)\**
- *Sample sizes\** and results
- Nature of the outcomes measures used, their relative importance and robustness
- Effect size

\*potential quality assessment criterion

Fields used may vary with the type of review question, depth and comprehensiveness of the review, and nature of the evidence base. Studies identified in the search phase but not included in the review can be listed in the same or separate tables, with explanations of why they were excluded from consideration. If more detail is needed (e.g. caveats or comments) separate tabulation forms can be generated for each study. As many details about documentation as possible are specified during the protocol development phase of the review.

#### *Assessing evidence quality*

Quality assessment can be difficult, time consuming and controversial. But when done in a transparent fashion using logical criteria, even a “coarse filter” quality assessment adds strength and credibility to the review. Fields in *italics* in the list above comprise a limited set of quality criteria that could be expanded as needed.

With all else being equal, studies that involve relatively larger spatial scales, more replication, more controls and longer timeframes generally produce the most reliable results. However, “all else equal” assumptions often don’t hold true. Greater complexity and diversity at larger scales can introduce more, rather than less uncertainty.

If a single table becomes unwieldy, a separate table can be generated for each study with space to provide more detail about how it rates for each quality criterion. By keeping a consistent table format, studies can be cross referenced and compared. If multiple similar studies are available they can be ranked according to the proportion of total quality items they comply with.

If appropriate, ranking can also be accomplished using an evidence quality hierarchy. Two examples of such hierarchies adapted for assessing evidence quality in ecosystem and conservation science are provided below:

**Hierarchy of possible types of studies to investigate effects of a management action on wildlife habitat** (Fazey et al. (2002).

Relative quality level	Study design	Analogue in clinical medicine	Relative level of confidence
1.	Study experimentally induced changes, with different management regimes assigned (at random) to different sites.	Randomized controlled trials	Very high
2.	Study experimentally induced changes, with different management regimes applied (by managerial choice) to different sites.	Cohort studies	High
2.	Compare sites subject to natural changes (e.g. flooding) or accidents (e.g. oil spills), with comparable (control) sites where they has been no management action.	Case-control study	High
3.	Before/after studies of effects of management or natural changes (e.g. flooding) or accidents (e.g. oil spills).	Case-series study, pre- and post-test	Moderate
4.	Gather observational data from a number of sites, spanning a range of management regimes. Use the data to determine conditions that lead to favorable outcomes.	Case-series study, post-test only	Low

**Hierarchy of possible types of studies for monitoring and evaluation of a rehabilitation project**

Relative quality level	Study design	Example Observation	Relative Level of confidence
1.	Replicated sampling, replicated controls, sampling before and after rehabilitation	'The increase in the number of salmon in the treated reach was greater than any increase at either control reach'	Very high
2.	Unreplicated, controlled, sampling before and after rehabilitation	'The number of salmon increased after rehabilitation in the treated reach, but not in the control reach'	High
3.	Unreplicated, uncontrolled, sampling before and after rehabilitation; OR Unreplicated, controlled, sampling after rehabilitation	'There were more salmon after the work than before'; OR 'After rehabilitation there were more salmon in the control reach than in the treated reach'	Moderate
4.	Unreplicated, uncontrolled, sampling after rehabilitation	'There was a gradual increase in the number of salmon in the two years after the work'	Low
5.	Unreplicated, uncontrolled, anecdotal observation after rehabilitation	'I saw lots of salmon after we had done the work'	Very low

Modified from a table in from Fazey et al. (2002), which was modified from Rutherford et al. (2000), *A Rehabilitation Manual for Australian Streams*, Vol 1, pp 164-73.

Rigid quality assessment criteria based on experimental design may not be useful if the evidence base is limited to studies with few experimental controls. At “lower” levels of evidence (e.g. observational studies, particularly without controls), other dimensions of quality become even more critical. These dimensions include:

- How well the study was conducted (was bias taken into account, were there confounding factors?)
- Size/strength of the effect (if the effect is very large then the association is likely to be correct even if the studies are low level, e.g. smoking and cancer).
- Volume of evidence (are there multiple studies that show the same effect?)

- Consistency- is the same effect seen in different studies?
- Specificity- is the effect specific for a particular set of parameters and not for others?

#### **4. Documenting how the evidence was interpreted and synthesized**

To rate the strength of a body of evidence (a group of studies), reviewers must not only identify all relevant studies and evaluate the quality of each, but also assess consistency of results and heterogeneity of study designs to assess their comparability. This is more difficult than assessing the quality or strength of an individual study.

Meta-analysis is preferred for evidence synthesis in the medical profession, but may be inappropriate or not possible for many natural resource SERs. A narrative synthesis is especially critical as the principal means of synthesizing evidence when meta-analysis is not conducted. Tabulated study characteristics and results in the narrative synthesis allow similarities and differences between the studies to be compared to arrive at an overall assessment. Quantitative synthesis may not be possible, but a narrative can document an organized qualitative evaluation of the strength of a body of evidence as a whole, and how it may have been impacted by study characteristics and quality.

Criteria for assessing the strength of a body of scientific evidence:

Quality: Aggregate quality of a body of evidence is based on the internal validity of each included study.

Quantity: Magnitude of effect, numbers of studies, and sample size or statistical power.

Consistency: For any given topic, the extent to which similar findings are reported using similar and different study designs.

Coherence: Do the findings of a body of evidence make sense as a whole?

#### **Guidance for disentangling questions about evidence from those about values and preferences in conduct and application of SERs**

For many of those interested in how scientific information is used to inform natural resource policy-making, the “elephant in the living room” is the degree to which debates about science are really disagreements over values and perceptions of risk. Atkins et al. (2005) note that what on the surface appear to be debates about evidence upon closer inspection often turn out to be fundamental differences of opinion on what outcomes are most important and what actions are appropriate in the face of imperfect evidence.

Atkins et al. (2005) argue that the SER process itself can help differentiate questions about evidence from those about values and preferences, and submit a set of questions to facilitate this:

*1. What is the ultimate goal, and how does the intervention achieve those ends?*

Explicitly examining the steps that link the policy or management action to the most important outcome serves three important purposes:

- It identifies distinct questions to which the evidence can be applied.
- It makes explicit the fact that policy decisions have a range of intended and unintended outcomes.
- It helps distinguish when the evidence directly addresses the important effects of policy from when the effects must be inferred from less direct evidence.

*2. How good is the evidence that the intervention can improve important outcomes?*

Understanding the scientific basis of debates about a particular policy or management action can be aided by addressing the hierarchy of questions suggested by Haynes (1999):

- *Can it work?* (Efficacy- or performance under ideal conditions.)
- *Will it work?* (Effectiveness- performance under real world conditions in this instance.)
- *Is it worth it?* (What are the costs and benefits?)

Conflict may arise because the evidence pertinent to the first question may not answer the second and third questions, which may be the most relevant for decision-makers. The difference between performance under ideal conditions (efficacy) can differ dramatically from actual practice (effectiveness).

*3. How good is the evidence that the intervention will work in my setting?*

An intervention or management action is more likely to work in practice when evidence can link it to desired final outcomes rather than intermediate outcomes. For example, evidence that a particular set of riparian buffer standards directly improves salmonid fry survival is more compelling than evidence that such standards moderate water temperature in such a way that is thought to improve fry survival. In addition, results that are consistent across studies provide more reassurance that findings are not the result of chance or applicable only to one locality or situation.

*4. How do the potential benefits compare with the possible harms of costs of the intervention?*

Deciding whether a policy is “worth it” requires considering outcomes affected by the policy, the size of individual benefits and harms, the degree of uncertainty surrounding these estimates, and importance of the different outcomes. These judgments are inherently expressions of values. To confront uncertainty, complementary approaches such as quantitative risk assessment, safe minimum standards, and the precautionary principle are typically applied (Fazey et al. 2004). Cost issues are a frequent source of conflict, but are often submerged in evidence debates.

*5. What constitutes “good enough” evidence for a policy decision?*

The judgment about what degree of certainty is sufficient is inherently a judgment about the risks of acting too soon (promoting a policy that turns out to be ineffective or harmful) and acting too late (delaying a beneficial intervention to get better evidence). Controversy is most common in areas where definitive studies are lacking or inherently difficult.

## References Cited

- Atkins, D., Siegel, J., & Slutsky, J. (2005). Making policy when the evidence is in dispute. *Health Affairs*, 24, 102-113.
- Bandolier. (2005). Online journal, "Evidence-based thinking about healthcare". Retrieved September 14, 2005.  
<http://www.jr2.ox.ac.uk/bandolier/booth/glossary/hetero.html>
- Barnes, D., & Bero, L. A. (1998). Why review articles on the health effects of passive smoking reach different conclusions. *Journal of the American Medical Association*, 280, 1566-70.
- Bero, L. A., & Jadad, A. R. (1997). How Consumers and Policymakers Can Use Systematic Reviews for Decision Making. *Annals of Internal Medicine*, 127, 37-42.
- Bero, L. A., & Jadad, A. R. (1998). From Systematic Reviews: Synthesis of Best Evidence for Health Care Decisions. *American College of Physicians*, 8, 21-30.
- Boaz, A., Ashby, D., & Young, K. (2002). *Systematic Reviews: What have they got to offer evidence based policy and practice?* Retrieved July 15, 2005 from <http://www.evidencenetwork.org/Documents/wp2.pdf>
- Bradshaw, G. A. & Borchers, J. G. (2000). Uncertainty as information: narrowing the science-policy gap. *Conservation Ecology* 4(1): 7. [online] URL: <http://www.consecol.org/vol4/iss1/art7/>. Retrieved 9-30-05
- British Medical Journal. 2005. Clinical Evidence. Online journal, accessed 9-14-05. <http://www.clinicalevidence.com/ceweb/about/appraisal.jsp> BMJ Publishing Group Limited 2005.
- Burgman, M. A., Lindenmayer, D. B. & Elith, J. (2005). Managing landscapes for conservation under uncertainty. *Ecology* 86:2007-2017.
- Center for Evidence-Based Policy (2003). Oregon Health and Science University. Retrieved July 26, 2005 from <http://www.ohsu.edu/policycenter>
- Centre for Evidence-Based Conservation. (2005). Retrieved July 15, 2005 from <http://www.cebc.bham.ac.uk/introSR.htm>
- Centre for Evidence-Based Conservation. (2005). *Systematic review template*. Retrieved August 19, 2005 from <http://www.cebc.bham.ac.uk/Documents/Protocol%20Template.doc>

- Centre for Evidence-Based Conservation. (2005). *Systematic review terminology*. Retrieved August 29, 2005 from <http://www.cebc.bham.ac.uk/terminology.htm>
- Centre for Evidence-Based Medicine-Oxford. (2005). *Glossary of terms in evidence-based medicine*. Retrieved August 29, 2005 from <http://www.cebm.net/glossary.asp>
- Centre for Reviews and Dissemination. (2001). *Undertaking Systematic Reviews of Research on Effectiveness*. CRD Report 4, 2<sup>nd</sup> Edition. Retrieved October 26, 2005 from: <http://www.york.ac.uk/inst/crd/report4.htm>
- Chase, A. (1997). Some cautionary remarks about the Precautionary Principle. Presentation and paper at 'Countdown to Kyoto': The Consequences of the Mandatory Global Carbon Dioxide Emissions Reductions, Australian APEC Study Centre, Canberra, 19–21 August 1997. Retrieved November 1, 2005 from: <http://www.apec.org.au/docs/chase.pdf>
- Cochrane A. L. (1979). 1931-1971: A critical review, with particular reference to the medical profession. In: *Medicine for the year 2000*. London: Office of Health Economics, 1-11.
- Cochrane Collaboration. (2005). <http://www.cochrane.org/docs/newcomersguide.htm>  
Retrieved 8-30-05.
- Cook, D. J., Mulrow, C. D., & Haynes, R. B. (1997). Systematic reviews: synthesis of best evidence for clinical decisions. *Annals of Internal Medicine* 126, 5, 376-380.
- Dixon-Woods, M., Agarwal, S. Young, B., Jones, D & Sutton A. (2004). *Integrative approaches to qualitative and quantitative evidence*. National Health Services Health Development Agency. London, UK.  
[http://www.med.umich.edu/csp/Course%20materials/Summer%202005/Goold\\_Integrative\\_Approaches.pdf](http://www.med.umich.edu/csp/Course%20materials/Summer%202005/Goold_Integrative_Approaches.pdf)
- Edwards, A. G., Russell, I. T. & Stott, N. C. (1998). Signal versus noise in the evidence base for medicine: an alternative to hierarchies of evidence? *Family Practice* 15, 4: 319-322.
- Eisenberg, J. M. (2001). What Does Evidence Mean? Can the Law and Medicine Be Reconciled? *Journal of Health Politics, Policy and Law*, 26:2, April 2001. Duke University Press. <http://www.ahrq.gov/clinic/jh ppl/eisenbrg.htm>
- Evidence-based Medicine Working Group. (1992). Evidence-based Medicine: A new approach to teaching the practice of medicine. *Journal of the American Medical Association* 268, no. 17: 2420-242

- Fazey, I., & Salisbury, J. G. (2002). Evidence-based environmental management: What can medicine and public health tell us? Proceedings of a workshop, June 24, 2002. National Institute for the Environment, Australian National University, Canberra, Australia.  
[http://eprints.anu.edu.au/archive/00002150/01/ebem\\_workshop\\_summary.pdf](http://eprints.anu.edu.au/archive/00002150/01/ebem_workshop_summary.pdf)
- Fazey, I., Salisbury, J. G., Lindenmayer, D. B., Maindonald, D., & Douglas, R. (2004). Summarizing and disseminating conservation research: can we use methods applied in medicine? *Environmental Conservation* 31 (3), 190–198.
- Fazey, I., J. A. Fazey, and D. M. A. Fazey. (2005.) Learning more effectively from experience. *Ecology and Society* 10(2): 4. [online] URL:  
<http://www.ecologyandsociety.org/vol10/iss2/art4/>
- Field, M. J. & Lohr, K. N. (1990). Clinical practice guidelines: directions for a new program. Washington D.C.: National Academies Press.
- Fox, D. M. (2005). Evidence of Evidence-Based Health Policy: The Politics of Systematic Reviews in Coverage Decisions. *Health Affairs*, 24, 114-122.
- Gates, S. (2002). Review of methodology of quantitative reviews using meta-analysis in ecology. *Journal of Animal Ecology* 71, 4, 547-557.
- Gravenmier, R. and Connelly, W. (2005). Levels of science consistency review for land use planning in region 6. United States Department of Agriculture Forest Service Region 6.
- Greenhalgh, T. (1997). How to read a paper: the basics of evidence based medicine. *British Medical Journal* 315, 672-675.
- Greenhalgh, T., Toon, P., Russell, J., Wong, G., Plumb, L., & Macfarlane, F. (2003). Transferability of principles of evidence based medicine to improve educational quality: systematic review and case study. *British Medical Journal*, 326 (7381), 142-146. [www.bmj.com/cgi/content/full/326/7381/142](http://www.bmj.com/cgi/content/full/326/7381/142)
- Guildin, J. M., Cawrse, D., Graham, R., Hemstrom, M., Joyce, L., Kessler., McNair, R., Peterson, G., Shaw, C. G., Stine, P., Twery, M., & Walter, J. (2003). *Science consistency reviews: a primer for application*. United States Department of Agriculture Forest Service. FS-771
- Haas, G. E. (2003). Restoring Dignity to Sound Professional Judgment. *Journal of Forestry* 101, 6, p. 38-43.
- Hayes, J. P. (1998). *An Independent Scientific Review of ODF Western Oregon Habitat Conservation Plan*. Department of Forest Science, Oregon State University.

- Independent Multidisciplinary Science Team. (1999). *Recovery of Wild Salmonids in Western Oregon Forests: Oregon Forest Practices Act Rules and the Measures in the Oregon Plan for Salmon and Watersheds*. Technical Report 1999-1 to the Oregon Plan for Salmon and Watersheds.
- Independent Multidisciplinary Science Team. (2002). *Lorensen-Hinkle Creek Research Proposal*. Retrieved August 1, 2005 from <http://www.fsl.orst.edu/imst/reports/Loresen-HinkleCr.pdf>
- Independent Multidisciplinary Science Team. (2005). IMST website. Retrieved from: <http://www.fsl.orst.edu/imst/index.html>
- Kelly, M. & Swann, C. (2004). Foreword in *Integrative approaches to qualitative and quantitative evidence*, Dixon-Woods, M., S. Agarwal, B. Young, D. Jones and A. Sutton. National Health Services Health Development Agency. London, UK. [http://www.med.umich.edu/csp/Course%20materials/Summer%202005/Good\\_Integrative\\_Approaches.pdf](http://www.med.umich.edu/csp/Course%20materials/Summer%202005/Good_Integrative_Approaches.pdf)
- Khan, K. S., Popay, J., & Kleijnen, J. (2001). *Undertaking systematic reviews of research effectiveness*. Retrieved August 26, 2005 from [http://www.york.ac.uk/inst/crd/pdf/crd4\\_ph3.pdf](http://www.york.ac.uk/inst/crd/pdf/crd4_ph3.pdf)
- Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). *From Systematic Reviews to Support Evidence-based Medicine: How to Review and Apply Findings of Healthcare Research*. London: Royal Society of Medicine Press Ltd.
- Kitzhaber, J. (2004). *Testimony to the Oregon Board of Forestry*. Retrieved July 5, 2005 from <http://law.lclark.edu/org/kcenter/testimonyobf.html>
- Lee, K. N. (1993). Compass and Gyroscope: Integrating Science and Politics for the Environment. Washington, D.C.: Island Press.
- Lohr, K. N. and Carey, T.S. (1999). Assessing 'best evidence': issues in grading the quality of studies for systematic reviews. *Joint Commission Journal on Quality Improvement* 25, 9, 470-479.
- Maurer, B. A. (2004). Models of scientific inquiry and statistical practice: Implications for the structure of scientific knowledge in The Nature of Scientific Evidence, Taper, M. & Lele, S. (eds.) Chicago: University of Chicago Press.
- Mealey, S. P., Thomas, J. W., Salwasser, H. J., Stewart, R. E., Balint, P. J. and Adams, P. W. (2005). Precaution in the American Endangered Species Act as a Precursor to Environmental Decline: The Case of the Northwest Forest Plan in Biodiversity and the Precautionary Principle: Risk and Uncertainty in Conservation and Sustainable Use, Dickson, B. and Cooney, R. (eds). London UK: Earthscan Press.

- Mendelson, D. & Carino, T. (2005). Evidence-based medicine in the United States —De rigneur or dream deferred? *Health Affairs*, 24, 133-136.
- Moja, L.P., Telaro, E., D'Amico, R., Moschetti, I., Coe, L. & Liberati, A. (2005). Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. *British Medical Journal* 330, 1053. Originally published online April 7, 2005; doi: 10.1136/bmj.38414.515938.8F
- National Research Council. (2004). Improving the Use of the "Best Scientific Information Available" Standard in Fisheries Management. Committee on Defining Best Scientific Information Available for Fisheries Management, Ocean Studies Board, Division on Earth and Life Studies. National Academies Press: Washington, D.C.
- New York Times. (2005.) *Medical Research Got More Money Over Last Decade*. September 21, 2005. Online version accessed 9-21-05. <http://www.nytimes.com/2005/09/21/national/21medical.html>
- NHS Centre for Reviews and Dissemination. (2001). Retrieved July 24, 2005 from <http://www.york.ac.uk/inst/crd/>
- Oregon Department of Forestry. (1999). *Forest Practices Advisory Committee on Salmon and Watersheds*. Retrieved August 3, 2005 from <http://www.odf.state.or.us/pcf/pub/fp/FPAC/FPACreport.pdf>
- Oregon Department of Forestry. (2001). *Northwest Oregon State Forests Management Plan*. Retrieved July 27, 2005 from [http://oregon.gov/ODF/STATE\\_FORESTS/docs/management/nwfmp/17-5-Implement\\_prn.pdf](http://oregon.gov/ODF/STATE_FORESTS/docs/management/nwfmp/17-5-Implement_prn.pdf)
- Oregon Department of Forestry. (2002). *Current Events in the Forest Practices Program*. Retrieved July 27, 2005 from <http://159.121.125.11/FP/CurrentEvents/CurrentEventsPg.htm>
- Oregon Department of Forestry. (2004). *Scientific Review of the Elliot State Forest Management Strategy*. Retrieved August 2, 2005 from [http://egov.oregon.gov/ODF/STATE\\_FORESTS/elliott.shtml#Scientific\\_Peer\\_Review](http://egov.oregon.gov/ODF/STATE_FORESTS/elliott.shtml#Scientific_Peer_Review)
- Oregon Department of Forestry. (2005). *Public Comment on AOPs*. Retrieved on July 12, 2005 from [http://oregon.gov/ODF/STATE\\_FORESTS/docs/management/aop/AOP\\_Public\\_Comment\\_Evaluation\\_2005.pdf](http://oregon.gov/ODF/STATE_FORESTS/docs/management/aop/AOP_Public_Comment_Evaluation_2005.pdf)

- Petticrew, M. (2001). Systematic reviews from astronomy to zoology: myths and misconceptions. *British Medical Journal* 13 Jan 322(7278) pp98-101. Available at: <http://www.bmj.com/cgi/content/full/322/7278/98>
- Pullin, A. & Knight, T. (2001). Effectiveness in Conservation Practice: Pointers from Medicine and Public Health. *Conservation Biology*, 15, 50-54.
- Research America. (2004). 2004 *Investment in U.S. Health Research*. Accessed online 9-20-05.  
<http://www.researchamerica.org/publications/appropriations/healthdollar2004.pdf>
- Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). Evidence based medicine: How to practice and teach EBM (2nd ed.). London: Churchill Livingstone.
- Santa, J. (2005). Phone conversation regarding the potential cost differences between medical and natural resource SERs. July, 2005.
- Scheiner, S. (2004). Experiments, observations, and other kinds of evidence in The Nature of Scientific Evidence, Taper, M. & Lele, S. (eds.) Chicago: University of Chicago Press.
- Shindler, B. & Neburka, J. (1997). Public participation in forest planning: 8 attributes of success. *Journal of Forestry* 95, 1, 17-19.
- Steinberg, E. P., & Luce, B. R. (2005). Evidence-based? Caveat Emptor! *Health Affairs*, 24, 80-92.
- Stevens, A. & Milne, R. (1997). *The Effectiveness Revolution and Public Health*. In G. Scally (ed.), Progress in public health (pp.197-225). London: Royal Society of Medicine Press.
- Stewart, G.B., Pullin, A.S., Davies, Z.G. (2005). Retrieved on August 12, 2005 from <http://cebc.bham.ac.uk/Documents/CEBC%20Systematic%20Review%20GuidelinesV1.pdf>
- Timmermans, S. & Mauck, A. (2005). The promises and pitfalls of evidence-based medicine. *Health Affairs*, 24, 18-28.
- Taper, M. & Lele, S. (2004). The Nature of Scientific Evidence. Chicago: University of Chicago Press.
- West, S., King, V., Carey, T., Lohr, K., McKoy, N., Sutton, S. & Lux, L. (2002). Systems to rate the strength of scientific evidence. AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality.