FREP

REPORT #37

February 2014

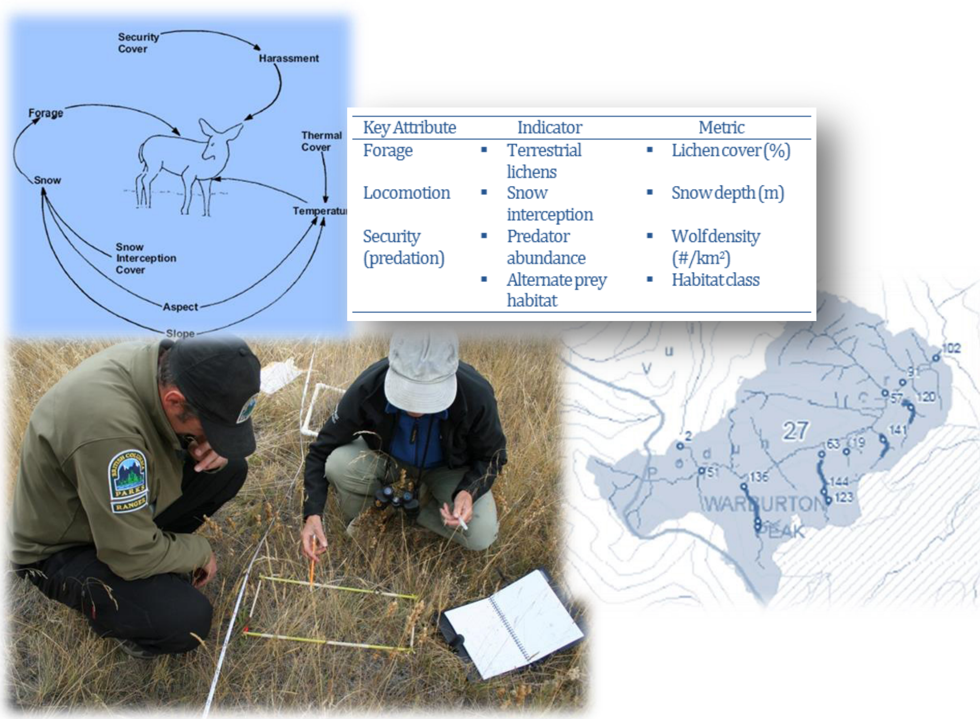
FOREST AND RANGE EVALUATION PROGRAM

A Guide to Wildlife Resource Value Effectiveness Evaluations

Prepared by:
Kathy Paige and Laura Darling, B.C. Ministry of Environment
and Darcy Pickard, ESSA Technologies

February 2014

VERSION 1.0



Key Attribute	Indicator	Metric
Forage	▪ Terrestrial lichens	▪ Lichen cover (%)
Locomotion	▪ Snow interception	▪ Snow depth (m)
Security (predation)	▪ Predator abundance ▪ Alternate prey habitat	▪ Wolf density (#/km²) ▪ Habitat class

FREP Forest & Range Evaluation Program

BRITISH COLUMBIA

Sustainability of Forest and Range Resources Through Science and Stewardship

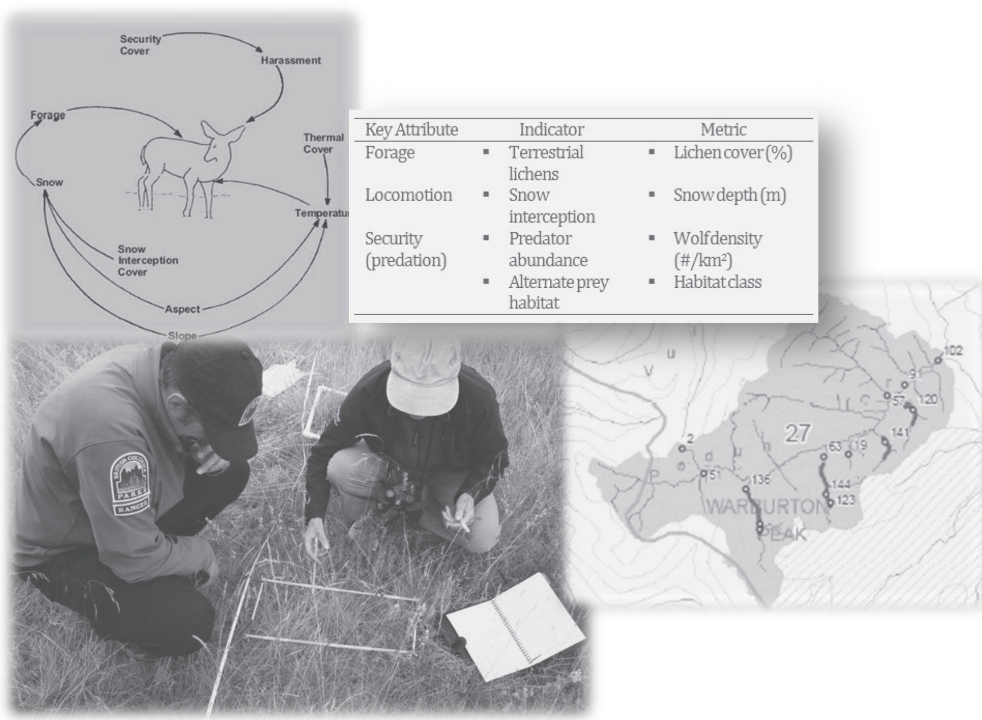
FOREST AND RANGE EVALUATION PROGRAM
MINISTRY OF ENVIRONMENT
MINISTRY OF FORESTS, LANDS AND NATURAL RESOURCE OPERATIONS

A Guide to Wildlife Resource Value Effectiveness Evaluations

Prepared by:
Kathy Paige and Laura Darling, B.C. Ministry of Environment
and Darcy Pickard, ESSA Technologies

February 2014

VERSION 1.0



Key Attribute	Indicator	Metric
Forage	▪ Terrestrial lichens	▪ Lichen cover (%)
Locomotion	▪ Snow interception	▪ Snow depth (m)
Security (predation)	▪ Predator abundance	▪ Wolf density (#/km²)
	▪ Alternate prey habitat	▪ Habitat class

Photo: Victoria Stevens, Deer Image: Armleder et al. 1986

The FREP Mission:
To be a world leader in resource stewardship monitoring and effectiveness evaluations; communicating science-based information to enhance the knowledge of resource professionals and inform balanced decision-making and continuous improvement of British Columbia's forest and range practices, policies and legislation. http://www.for.gov.bc.ca/hfp/frep/index.htm

FREP Forest & Range Evaluation Program

British Columbia

Citation:
Paige, K., Darling, L.M. and D. Pickard. A Guide to Wildlife Resource Value Effectiveness Evaluations. Ecosystems Branch, Ministry of Environment, Resource Practices Br., Ministry of Forests, Lands and Natural Resource Operations. Victoria, BC. FREP Report 37.
http://www.for.gov.bc.ca/hfp/frep/publications/index.htm

Cover photo: Armleder, H. M., R. J. Dawson, and R. N. Thomson. 1986. Handbook for Timber and Mule Deer Management Co-ordination on Winter Ranges in the Cariboo Forest Region. B.C. Min. For., Victoria, B.C. Land Management Handbook 13.

Copies of this report may be obtained, depending on supply, from:
Government Publications
PO Box 9452, Stn Prov Govt
Victoria BC V8W 9V7

In Victoria (250) 387-6409
Outside Victoria 1-800-663-6105

http://www.publications.gov.bc.ca

For more information on Resource Practices Branch publications, visit our web site at:
http://www.for.gov.bc.ca/hfp/meta/publications.htm

ACKNOWLEDGEMENTS

This document is based on the knowledge of many individuals working on monitoring and effectiveness evaluation projects. Many thanks to Scott Barrett, James Dalby, Todd Manning, Melissa Todd, and Richard Thompson for providing review comments.

PREFACE

The *Forest and Range Practices Act* and regulations embody a results-based forest framework. Under this approach to forest management, the forest industry is responsible for developing results and strategies, or using specified defaults, for the sustainable management of resources. The role of government is to ensure compliance with established results and strategies and other practice requirements, and to evaluate the effectiveness of forest and range practices in achieving management objectives.

The Forest and Range Evaluation Program (FREP) is a multi-agency program established to evaluate whether practices under the Act are meeting not only the intent of its current objectives, but also to determine whether the practices and the legislation itself are meeting government's broader intent for the sustainable use of resources. For more about FREP, see: http://www.for.gov.bc.ca/hfp/frep/index.htm.

The Minister of Forests, Lands and Natural Resource Operations is responsible for designation of important mechanisms described under the Act to address conservation of wildlife habitat, including Wildlife Habitat Areas, Ungulate Winter Ranges, and others. A program to monitor and evaluate the wildlife objectives established through the Act for these mechanisms is under development. A strategic framework for monitoring effectiveness of wildlife resource values, along with various guidance documents and reports that describe work completed under FREP's Wildlife Resource Value are available. For more about the Wildlife Resource Value program, see: http://www.for.gov.bc.ca/hfp/frep/values/wildlife.htm.

This Guide to Wildlife Resource Value Effectiveness Evaluations is one of the key guidance documents developed under the Wildlife Resource Value Framework. It is designed to provide guidance for establishment of projects that evaluate the effectiveness of tools or mechanisms available under the Act to conserve wildlife resource values, particularly Wildlife Habitat Areas and Ungulate Winter Ranges. This guide was developed with FREP wildlife resource value objectives in mind; the intended audience is FREP Wildlife Resource Value project leads. This version (Version 1.0) is considered by the authors as a work in progress and the intent is to update it as more is learned about conducting wildlife effectiveness evaluations.

TABLE OF CONTENTS

List of Tables

List of Figures

1.0 INTRODUCTION

Effectiveness evaluation is a process of checking, assessment, and reflection—a process that any program or project should adopt to determine whether program or project goals and objectives are achieved. Some of the main reasons to conduct effectiveness evaluations in resource management situations are to improve management actions, maximize the impact of limited resources, build support by documenting success, and show that efforts and policies are justified, or dollars well spent.

Monitoring is an important component of an effectiveness evaluation that involves collecting and analyzing information. The key components of an effectiveness monitoring and evaluation program include: clearly stating the goals; developing a conceptual model that links relevant ecosystem function components (i.e., species and habitats) and stressors/disturbances (risks); selecting suitable indicators to monitor; estimating the status and trends of indicators; determining values that trigger a management response; and linking monitoring results to management (Lint et al. 1999; Noon 2003).

Monitoring and evaluating wildlife objectives and management mechanisms established through the *Forest and Range Practices Act* falls under the Wildlife Resource Value of the Forest and Range Evaluation Program (FREP), a joint British Columbia initiative between the Ministries of Environment, and Forests, Lands and Natural Resource Operations. Effectiveness monitoring and evaluation of wildlife resource values are conducted to answer broad questions, such as:

- Are our actions (e.g., habitat designations) having the intended impact?
- Are we achieving our conservation goals and objectives?
- Are management actions implemented as intended?
- What improvements, if any, are needed to achieve goals and objectives?
- What can we do to enhance our success?

To answer these questions and assess the status and condition of a species, its habitat, or threats to species and their habitats, a suite of biological and (or) threat indicators is tracked over time. When a conservation intervention or management practice is initiated or under way, comparisons are made of the current versus desired condition, the relative impacts of different treatments, or the effects of pre- and post-treatment condition.

Several program assessment procedures have been developed for various types of projects and proponents, all with the goals of: promoting efficient and effective use of resources; collecting the right type of data in sufficient quality and quantity to support the goals of the study; and linking results of the monitoring with decisions for improved resource management.

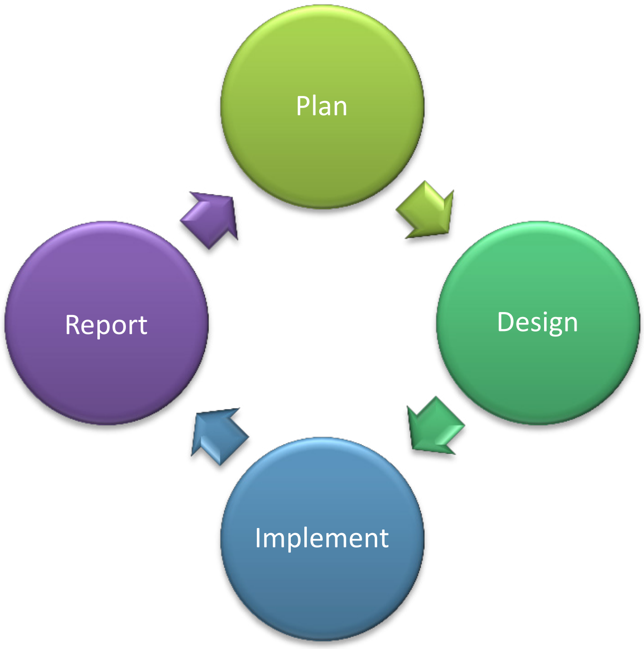
Examples of some of these procedures include the following.

1. The Environmental Protection Agency's "Data Quality Objectives Process" (2006) provides a method for planning the acquisition of environmental data to measure performance (compliance) and (or) acceptance criteria, and includes strong quality assurance concepts.
2. The National Park Service's "Vital Signs Monitoring" (2012) determines the status and trends in the ecological condition of selected park resources by integrating ecological principles.
3. The University of Michigan's Environmental Monitoring Initiative (Schueller et al. 2006) provides a user-friendly, step-by-step process based on the Logic Model for developing and implementing evaluation plans for ecosystem and community-based projects.
4. The Nature Conservancy of Canada's approach (Gordon et al. 2005) offers a structured process to monitor the effectiveness of their conservation strategies in meeting program planning and stewardship objectives on the lands they administer or manage.

Although none of the four assessment processes described above is fully compatible with the kind of evaluations FREP needs, the strengths of each were used to develop the wildlife resource value effectiveness evaluation process discussed in this guide. The reader is encouraged to review these other assessment processes to gain a deeper insight into the rationale for, and development of, sound effectiveness evaluation projects.

1.1 *Purpose of this Guidance Document*

This document provides guidance for the establishment of projects that will evaluate effectiveness of tools or mechanisms, available under the *Forest and Range Practices Act*, intended to conserve wildlife resource values, particularly Wildlife Habitat Areas (WHAs) and Ungulate



Winter Ranges (UWRs). This guidance was developed with FREP wildlife resource value objectives in mind and the intended audience is FREP wildlife resource value project leads.

The four parts of this document mirror the four major steps involved in effectiveness evaluation (Sections 2–5; see Table 1). Each step involves several tasks and output products, which together will lead to achieving the evaluation objectives. Not all projects will proceed to the end step: although all tasks are thought as necessary components of the compete process, progress will vary for different situations; for some species, achieving only some initial tasks (e.g., determine need and current state) will provide value.

Table 1. ***The four steps involved in developing and implementing effectiveness evaluations for FREP wildlife resource value projects, including tasks and outputs.***

Step	Task	Output Product
1. Plan the Project	Frame monitoring questions and objectives	Objectives and questions
	Consider priority of monitoring objectives and questions	Conceptual model
	Compile knowledge base	Indicators
	Build conceptual model of relationships	Draft project charter
	Review and select indicators	Evaluation framework
	Establish indicator thresholds or triggers	
	Establish how condition or effectiveness will be determined	
2. Design the Project	Establish study and sampling design	Monitoring plan
	Determine data collection methods	Pilot reports
	Test and refine methods (pilot testing)	
	Determine data analysis methods	
	Undergo statistical review	
3. Implement the Project	Project planning	Project charter
	Secure resources	Work plan
	Training	Field forms
	Outline quality assurance	Data
	Collect data	
	Establish data management procedures	
4. Report	Analyze data	Reports, extension notes, scientific papers, management recommendations
	Establish baselines for indicators	
	Report study results	
	Summaries and advice to decision makers	

1.2 *How to Use this Guidance Document*

Although the target audience for this guidance document has a common responsibility as project leads, other users will have different needs and therefore different ways they might use this document. Some users studying an animal for which very little is known, must begin by obtaining a basic understanding of the species and ways to monitor it. Other users may focus on species for which biological knowledge and field methods are well developed, or for which a preliminary effectiveness evaluation protocol already exists: in these situations, undertaking Steps 1 or 2 (Table 1) may not be required. Such users may ask: what pieces do I have; where in this process do I jump in; if steps 1 and 2 are done, what do I do next?

Sections 2–5 describe the four steps for effectiveness evaluation outlined in Table 1. The reader (i.e., project leader) can enter at any appropriate step in the process. To determine where a project falls in the process of developing and implementing an effectiveness evaluation protocol, users of this document should review the green-shaded "action checklist" at the end of key sections, and then insert their project into the appropriate step to match their need.

2.0 PLAN THE PROJECT

Monitoring and evaluation projects can be quite involved especially if a standard protocol does not exist (see sidebar, "What is a Protocol?"). The use of an established or standard protocol that outlines a procedure, practice, or set of rules, is essential to the success of monitoring projects; if one does not exist, it is important to develop one. Before deciding to embark on a project or to develop a protocol, do some background work to assess current knowledge and the priority of the project, and to define project objectives.

What is a Protocol?

In the realm of environmental surveys, a detailed step-by-step process for collecting field data might be labelled as a "protocol." In this guide, a broader definition of protocol is used that encompasses the steps necessary to develop the study design, collect and analyze the data, and report the results. Standardized step-by-step, data-collection procedures and techniques are referred to as "methods" in this guide. The goal of a monitoring protocol is to establish consistency, transparency, and minimize or balance measurement error and sampling error. Measurement error results from differences in collection methods, rather than actual changes in the environment (Oakley et al. 2003). Sampling error results from taking measurements from a subset of a population and is minimized through sampling design.

2.1 *Framing the Monitoring Objectives and Questions*

IN BRIEF . . .

This task involves the clarification of key management issues and questions, such as:

- *What is the expected or desired outcome of the management action? What is the objective of the Forest and Range Practices Act mechanism?*
- *What key management issues may be hindering the desired outcome of the management action?*
- *Who is the audience for the results of the study—the land manager, a statutory decision maker, government policy, the public, or all of these?*
- *What questions need to be answered to address the management issues and improve the management outcome?*
- *What is the focus of the question (and monitoring)—threat reduction, habitat quality, or population status?*
- *At what scale does the issue (and monitoring) apply—local, watershed, or regional?*

Development of a wildlife resource value effectiveness evaluation project begins with clarification of the key monitoring questions (the uncertainties) regarding the *Forest and Range Practices Act* habitat mechanisms (e.g., Ungulate Winter Range [UWR] or Wildlife Habitat Area [WHA]). Usually, some uncertainty surrounds whether the expected outcomes of the current management approach (i.e., the objectives of the mechanism) are being achieved. Pertinent management issues (or unknowns) are the source of the very specific questions and hypotheses examined in an effectiveness evaluation project. By focussing on recognized knowledge gaps and relevant management issues, an effectiveness monitoring project will address the area of greatest need, producing a sound evaluation of effectiveness.

Broad-scale priority evaluation questions for UWRs and WHAs (see sidebar, "Priority Evaluation Questions for Wildlife Values") were initially developed and then updated by the Forest and Range Evaluation Working Group.[1] Although these questions are both too complex and too general to serve as a focus for a single evaluation project, they point to several more specific topics that could be targeted for evaluation, including: amount, quality and distribution of UWRs and WHAs; survival and fitness of the species for which these areas are designed; or trends in threats to the habitat or species.

Priority Evaluation Questions for Wildlife Values

The key questions asked about Ungulate Winter Ranges and Wildlife Habitat Areas are:

- *Do Ungulate Winter Ranges maintain the habitats, structures, and functions necessary to meet the species winter habitat requirements, and is the amount, quality, and distribution of UWRs contributing effectively with the surrounding land base (including protected areas and managed land base) to ensure the winter survival of the species now and over time?*
- *Do Wildlife Habitat Areas maintain the habitats, structures, and functions necessary to meet the goal(s) of the WHAs, and is the amount, quality, and distribution of WHAs contributing effectively with the surrounding land base (including protected areas and managed land base) to ensure the survival of the species now and over time?*

1 The Forest and Range Evaluation Working Group established the FREP structure and framework and initiated development of resource value indicators and monitoring/evaluation protocols. A list of priority evaluation questions was developed in 2004. For the latest update, see: http://www.for.gov.bc.ca/hfp/frep/about/questions.htm - questions.

Specific monitoring questions investigated for WHAs and UWRs might be at small or large spatial scales and (or) address single or multiple environmental parameters. Examples of specific questions include:

- Does this WHA provide sufficient habitat to provide for survival of the species?
- Is habitat attribute X above a standard habitat quality threshold within the UWR?
- Do WHAs in this region provide better habitat for the species at risk than similar areas that are not designated as WHAs?
- Is the species population persistent in this UWR?

The desired outcomes of implementing a management tool should also be considered. Some WHA and UWR goals and objectives, particularly from the earliest designations, may be unclear, unspecific, or difficult to evaluate. As a result, they may need to be restructured into clear, measurable, and outcome-based objectives. Objectives, unlike goals, should be tangible, precise, and measurable. Original objectives of each established WHA or UWR should be reviewed to define the key issues and questions that require investigation and to determine whether effectiveness evaluation is appropriate.

Selection and refinement of monitoring questions requires consideration of other factors beyond the management issue itself (Gordon et al. 2005; Schueller et al. 2006), such as who regulates the management action, our ability to influence the management action, and whether appropriate legal changes are possible. For example, opportunities for partnerships, co-operation, and cost efficiencies need to be considered, as well as whether the target species or habitat is suited to cost-effective monitoring. Limitations and potential barriers such as budget constraints, logistics, and access issues also need to be considered. Although some of these factors are taken into account during the prioritizing process (see Section 2.2), considering them in a broad sense right from the start will serve as a coarse filter.

In some cases, identifying the unknowns and framing management questions is a straightforward process. Sometimes issues are less clear and thus determining specific management questions will be more complex. Therefore, one should identify concerns about effectiveness of WHAs or other mechanisms in detail, treating each issue separately and completely in an uncomplicated format (see Section 2.4).

To improve future management, effectiveness evaluation projects must be based on answerable questions and hypotheses (or objectives) that are specific and precise. Framing questions using the "SMART" concept—that is, ***S***pecific; ***M***easurable; ***A***ttainable, achievable, and appropriate; ***R***ealistic, relevant, and results oriented; and ***T***ime-bound (Doran 1981; Meyer 2003)—ensures that the project leader and co-operators agree on the most important information needed. Questions and hypotheses should be framed to provide the practitioner with a clear-cut answer and direction for management; that is, rather than asking for a "yes" or "no" answer, instead ask "how much?" or "to what extent?". Before monitoring starts, key questions and null hypotheses must be formulated and confirmed in order to apply the appropriate sampling design and data analysis and reduce the chance of reaching erroneous conclusions (Shaffer and Johnson 2008).

Thus, many types of monitoring questions are possible, such as those that:

- address issues of population abundance, habitat suitability, use of the area by wildlife, or mitigation of threats;
- examine the generic or specific goals of WHAs and UWRs (see sidebar, "Generic versus Specific Goals");
- assess the status of a species or habitat, or effectiveness of an action at a local, landscape, or regional level; or
- investigate progress toward meeting management goals, threat reduction, improving condition, or successful completion of activities.

However, focussing on one particular question may not provide sufficient evidence to fully assess effectiveness; instead, multiple lines of evidence from a suite of questions may be needed.

Generic versus Specific Goals

Goals are broad statements of intent. Examples of generic goals that might apply to all WHAs include:

- *Maintain the habitat suitability and key habitat attributes of the area that contribute to survival of the species in the area (e.g., microclimate, prey species, nest trees)*
- *Maintain the use of the area by the species*
- *Maintain the ecological processes (ecosystem functioning) that support the species*
- *Abate stresses from forest and range management practices*

Examples of specific goals that might apply to individual WHAs include:

- *Reduce road mortality*
- *Minimize soil disturbance*
- *Maintain stand structure*
- *Maintain security cover*

ACTION CHECKLIST . . .

✓ *Describe in detail the concerns or issues about the effectiveness of the WHA or UWR. Define each problem separately and completely, in an uncomplicated format.*

✓ *Frame the concerns in specific, focussed questions.*

✓ *Ensure that answers to these questions will provide direction for improved management.*

2.2 *Prioritizing the Monitoring Objectives and Questions*

IN BRIEF . . .

This task involves the determination of priorities, such as:

- *Which mechanisms, and which of the species they serve, are the highest priorities for effectiveness monitoring?*
- *Which objectives or questions (for those species or mechanisms) are the highest priorities for monitoring?*
- *Does the proposed project meet Ministry and regional mandates, goals, and priorities?*
- *Is the proposed project a high priority based on FREP's wildlife value priority criteria? Or is the project one of the highest priorities according to the Conservation Framework priority-setting process?*

When determining whether to pursue an effectiveness evaluation of a WHA, UWR, or other management tool, first consider and confirm the priority of undertaking the project. Consider, for example, provincial and regional goals and priorities for the species and the habitat designation, as well as the potential benefits of the evaluation project at the regional and provincial level.

At a provincial level, ensure that evaluation projects align with Ministry of Environment and Ministry of Forests, Lands and Natural Resource Operations (FLNRO) mandates and priorities. High-level goals such as Environment's ". . . to provide for healthy and diverse native species and ecosystems, . . ." and the FLNRO goal of ". . . ensuring environmental standards are upheld and environmental sustainability is achieved with resource use activities in British Columbia, . . ." provide the context for the habitat designations and management practices discussed here.

Provincial direction is also provided by explicit legal objectives for wildlife outlined under the Act's regulations; the objective is to conserve sufficient wildlife habitat (area, distribution of areas and attributes) for the survival of species at risk and the winter survival of specified ungulate species without unduly reducing the supply of timber from British Columbia's forests.[2]

Provincial priorities for FREP's Wildlife Resource Value are determined through a priority-setting process. Highest priority projects should be initiated before lower priority ones. Before undertaking a project, consider the following criteria.

- Is there sufficient information to develop an evaluation project?
- Does the project address a priority question (i.e., Wildlife Resource Value priorities)?
  - What is the species conservation priority based on the British Columbia's Conservation Framework?
  - What is the province's investment in management actions (e.g., the number and total area of UWRs or WHAs)?
  - What is the relative importance of the WHA or UWR to conservation of the species?
  - What uncertainty is associated with the effectiveness of the management action(s) and the risk to the value (e.g., species, habitat)?
- What is the likelihood of project success (i.e., project feasibility, benefits, support)?

2 *Forest and Range Practices Act*, Forest Planning and Practices Regulation. B.C. Reg. 14/2004, Section 7. http://www.bclaws.ca/Recon/document/ID/freeside/14_2004 - section7. Similar direction is supplied in the Woodlot Licence Planning and Practices Regulation. B.C. Reg. 21/2004, Section 9(2). http://www.bclaws.ca/Recon/document/ID/freeside/21_2004 - section9.

- Another example of how to set monitoring priorities is the procedures associated with British Columbia's Policy for Mitigating Impacts on Environmental Values, which recommends considering the uncertainty of the mitigation activity, combined with the risk to a value, to determine whether implementation or effectiveness monitoring is appropriate.[3]

ACTION CHECKLIST . . .

Priority evaluation questions provide the context for developing specific indicators for effectiveness monitoring.

✓ *Determine where the species ranks in higher-level priorities (e.g., ministry mandates, regional priorities) and demonstrate whether monitoring is a priority.*

2.3 *Compiling a Knowledge Base*

IN BRIEF . . .

This task involves the collation of current knowledge and data about:

- *Species biology*
- *Species habitat use and habitat requirements*
- *Known or potential stressors, threats, and disturbances*
- *Past, current, and expected habitat suitability and capability*
- *Past, current, and expected management practices and adjacent activities*

After key issues and questions are identified and an effectiveness evaluation is confirmed as a priority project, the next task is to gather existing biological data and information about management activities and threats (risks) in the area. This collated "current state" information will provide a solid baseline or reference point for the evaluation, and possibly confirm the validity of the proposed evaluation objectives or lead to refinement of them. As well, this information provides crucial input to the conceptual models used to refine key questions into monitoring questions and hypotheses.

3 See *Procedures for Mitigating Impacts on Environmental Values: Final Working Draft* (2012), http://www.env.gov.bc.ca/emop/docs/EMProceduresFinalWorkingDraft.pdf (Accessed January 2014).

2.3.1 Ecological knowledge

Collecting pertinent information related to the species, its natural history and ecological role, the particular location, and the management tool (WHA or UWR) is imperative for successful evaluations. Gathering and incorporating current knowledge ensures that the project design will address advances in research and improvements to indicators. To refine understanding of a species and to enlarge the knowledge base over time, include data and conclusions from baseline, pilot, and ongoing studies as these are collected.

Investigate potential sources of information (either text-based or spatial), such as: published journals, published and unpublished consultant reports to government or industry, internet publications, results from predictive ecosystem mapping and (or) habitat modelling, recognized species specialists, Ministry of Environment species and habitat specialists, and local experts. A review of this information should confirm the need for the proposed evaluation project, as well as existing information gaps and needs, and help to develop the study objective and design.

Some general information sources include:

- Conservation Data Centre: http://www.env.gov.bc.ca/cdc/
- BC Species and Ecosystems Explorer: http://www.env.gov.bc.ca/atrisk/toolintro.html
- Identified Wildlife Management Strategy Species Accounts http://www.env.gov.bc.ca/wld/frpa/iwms/accounts.html
- Wildlife Species Inventory – Species Inventory Database: http://www.env.gov.bc.ca/wildlife/wsi/index.htm
- eFlora: http://www.geog.ubc.ca/biodiversity/eflora/
- eFauna: http://www.geog.ubc.ca/biodiversity/efauna/
- Wildlife Resource Value effectiveness evaluation reports: http://www.for.gov.bc.ca/hfp/frep/values/wildlife.htm
- Resources Information Standards Committee: http://www.ilmb.gov.bc.ca/risc/

2.3.2 Management activity

The primary management activity of concern in wildlife resource value effectiveness evaluations is the management mechanism or tool used to protect the species of interest (i.e., a WHA or an UWR) and the management actions associated with this tool (i.e., general wildlife measures). Using text-based or spatial information sources, document the parameters of the tool and practices, such as: the size and shape of a WHA; the number and condition of important habitat features within the WHA; and the specifics of required management activities, such as no-harvest zones or minimum grazing stubble heights. Documentation of these parameters will help to determine whether the management tool was implemented as intended, and to evaluate effectiveness and validate relationships proposed in the conceptual model (see Section 2.4). Documenting past and present land or wildlife management activities in and around the study area is also important because these activities may have affected (or will affect) the habitat or the target species in a way that could confound the proposed evaluation, requiring a refinement of study objectives and consideration in the development of a study design or during data analysis.

2.3.3 Threats and risks

Activities or conditions that exist in or near the study area may upset the natural balance of the ecosystem. Stressors may exist at a low level with negligible or no effect, or may develop into major threats to ecological integrity, either quickly or over a long period of time. One-time or ongoing disturbances may be or become substantial threats. A review of threats in both a spatial context (GIS, mapping) and as a written description may provide differing perspectives. The magnitude of any threat manifests in the level of risk it poses to the integrity and well being of a population or habitat. In addition to the magnitude of individual threats, one must consider the cumulative effects of ongoing or concurrent disturbances and threats. A review of existing and potential stressors and threats to the ecosystem under investigation, as well as the cumulative effects of these stressors and threats, may indicate the need to refine study objectives and (or) to consider these factors in study design and data analysis.

ACTION CHECKLIST . . .

✓ *Collate data and reports*

✓ *Include spatial (GIS, mapping) data*

✓ *Summarize knowledge, activities, and stressors that affect the species or the habitat*

2.4 *Demonstrating Relationships: Conceptual Models*

IN BRIEF . . .

With this task, we ask:

- *What do we know or believe about the relationships between populations, habitats, and stressors (threats)?*
- *Are the relationships quantifiable, or are we limited to qualitative descriptions?*
- *What are the interactions and influences in the ecological system?*
- *What factors and interactions are most important?*

After compiling a knowledge base, the next task is to employ conceptual models in the examination of functional interactions and relationships between species and ecological system components, linking species and habitats (functions) to threats (risks) and their effects. Models are an important component of monitoring and evaluation projects. Nyberg (2010) discussed the strengths of various conceptual models and provided recommendations for their use in the wildlife resource effectiveness evaluation process. Here, we summarize some of the key aspects of Nyberg's report.

Usually, a conceptual model describes the effects of various stressors (risks) on the focal population. This general model is then broken down to show pathways through which particular stressors (risks) could affect habitat attributes and population processes (functions). In the management context of effectiveness evaluation, a conceptual model illustrates the proposed relationship between a management activity and a species response. Developing sound models is crucial in the effectiveness evaluation process because these models may be used to draft a list of candidate indicators for each potential interaction (see Section 2.5); indicator variables are selected for monitoring from this list.

Models can:

- describe functional relationships between system components, clarify the linkages between stressors (risks) and outputs or outcomes, and document the understanding of the evaluated system;
- highlight important knowledge and data gaps and identify areas of uncertainty to be addressed through monitoring or research;
- identify the most important monitoring questions;
- identify measurable indicators that will answer the questions; and
- form the basis for more informative and powerful quantitative models.

The interaction of biological factors within an ecosystem is complex, and modelling can improve our ability to understand these interactions. Because ecosystem parameters are usually interactive, models based on separate relationships between a species and its environment are generally not meaningful. Developing an explicit, all-encompassing model of important relationships allows us to compare the importance of changes in different parameters, determine the net effect of these changes, characterize and quantify cause-and-effect relationships, and set priorities for reducing uncertainty. Subsequently, key relationships are highlighted to illustrate the focus of an effectiveness evaluation.

Models of relationships can be developed at various levels of detail and scale, from simple graphs and diagrams to complex mathematical simulations. To develop sound effectiveness evaluations, the conceptual models built should document relationships between resource values, threats (risks), opportunities and stated goals, objectives, and strategies. Input to conceptual models may come from experts and (or) output of more complex mathematical modelling procedures.

Conceptual models are built using tabular, graphic, or diagrammatic formats. General concept maps (e.g., "situation maps" in Schueller et al. 2006), often presented in tabular format, provide a logical structure to guide thinking about important questions and linkages between goals, ecosystems, and population traits. Influence diagrams (e.g., see Nyberg 2010), using arrows to show direction (and sometimes relative magnitude) of influence, illustrate relationships among different ecosystem parameters and provide a clear understanding of how each contributes to the end result. Bayesian belief models (Marcot et al. 2001; Nyberg 2010) integrate biological data presented in an influence diagram with expert opinion and an estimate of certainty (belief weighting) of life history parameter values and relationships among parameters. Unlike concept maps and influence diagrams, Bayesian models are used to predict a range of possible outcomes and their probabilities for populations, habitats, threats, and prey. Therefore, Nyberg

(2010) recommended Bayesian models as the preferred format for conceptual models in wildlife effectiveness evaluations, although simpler models are also appropriate when data are lacking.

Use current knowledge to fully and correctly develop a model of relationships by including important interactions and elements along with definition of the importance and probability of different parameter values. Build models collaboratively with species and habitat experts as well as land managers. Building simpler conceptual models may require a day's workshop, whereas complex mathematical models may require longer preparation time and review. The complexity of a model will reflect the complexity and scale of the study's objectives and the questions asked.

An examination of the final conceptual model will reveal important knowledge gaps, uncertainties around important parameters, and interactions and threats that may influence achievement of the management objective (i.e., its effectiveness). The variables that describe the known conditions and relationships and the unknowns, uncertainties, and threats are candidate monitoring indicators in effectiveness evaluation. A list of measurable variables to consider as monitoring indicators is one output product of the conceptual model.

ACTION CHECKLIST . . .

- ✓ *Use gathered information of the current state*
- ✓ *Convene workshop of species and habitat experts*
- ✓ *Develop a conceptual model that describes functional relationships, influences, and interactions among species, habitats, and threats (risks) in the broad ecological system*
- ✓ *List all the measurable variables that are associated with the model relationships*

2.5 *Determining Indicators (Measures)*

After developing a sound conceptual model that describes species, habitat, and threat (functions and risks) interactions and relationships, and drafting a list of potential measurable monitoring variables, the next task is to evaluate which of these variables will provide the best indication that a management objective is being achieved and then to select a logistically feasible combination of these indicators. Of key importance in this indicator identification process, is that indicators can only measure effectiveness if they relate back to how the management tool or practice was undertaken.

IN BRIEF . . .

This task examines the conceptual model and asks:

- *What variables can (or should) we measure to determine whether the particular management mechanism is achieving its objective?*
- *What variables have the strongest link to directly answering the key monitoring questions?*
- *What attributes of the population or the habitat (if measured over time) will indicate the effectiveness of the mechanism or the associated practice?*
- *What factors contribute to the success of the practice?*
- *What balance of population, habitat, and threat indicator variables provides the best opportunity for successfully answering the key monitoring questions?*
- *What factors (parameters) are unsuitable for monitoring (perhaps because appropriate methods of evaluation do not exist, or difficult logistic constraints or high cost are issues) and must be excluded from consideration as an indicator variable?*
- *What protocols are available? What high-priority, practical, measurable variables should be selected to implement a pilot study?*
- *What are the predicted values of the selected indicator variables as a result of the management practice (e.g., implementation of the management mechanism)?*
- *What values of the indicator variables (e.g., an upper or lower threshold) will trigger a management response?*

2.5.1 Indicators to monitor

Ecosystem elements, risks, and the relationships demonstrated among them in conceptual models can be monitored through the use of indicators. Indicators are measurable parameters (attributes) that describe specific characteristics of the species of interest or the target ecosystem; indicators represent the key features of the desired outcome (i.e., the management objective), providing information on the state or condition of a system element or the relationship between elements in a system.

Some indicators are direct measures of ecosystem or biological parameters of concern, whereas others measure factors that are known or believed to represent parameters of concern (i.e., surrogate or proxy indicators). Surrogate indicators are useful when measuring the parameter of interest is difficult or not cost-effective. For example, habitat attributes can act as surrogates for species (or one species for another), and indicators at one scale used to monitor biological diversity at another scale. Although relationships are sometimes uncertain, a surrogate should reflect the condition of the unmeasured parameter. While surrogacy is commonly used in biological monitoring, most situations require validation of surrogacy assumptions.

Indicators are qualitative or quantitative. Quantitative indicators reflect a numeric relationship between certain habitat variables and species response, and are often of more value than qualitative indicators. Qualitative indicators, although sometimes providing a more complete understanding of a relationship, are prone to observer subjectivity and hindered by artificial categorization of the values.

Indicators are selected to make comparisons over time (i.e., trends), among places (e.g., different ecological sites or different treatments), or with an ideal (e.g., comparison with the desired condition or a control) or a target (e.g., ecological threshold or management trigger).

Although some indicators are assessed via field data collection, some spatial indicators are assessed using routine evaluations, such as geographic information systems (GIS) analyses or mapping exercises (see Section 3.1.5). In most cases, GIS-based assessment is less expensive to complete, but field data, often used to validate GIS assessments, is of relatively high value and crucial to a sound evaluation.

Indicators are characterized by:

- the realm that is measured (habitat, species/population, threat);
- the part of the ecosystem that is measured (ecosystem function, structure, or species);
- the intent of the measurement (status assessment or effectiveness of action); and
- the project "element" for which progress will be measured (objective – progress toward goal; threat – progress to reduce threats; assets – progress toward improving assets; and strategies – progress toward completing successful activities).

In addition, the indicators measured and inferences from them need to reflect the various spatial scales at which wildlife utilize landscapes. For example, the indicators selected can address:

- Large-scale management concerns, including maintenance of regional or subregional populations through a network of management tools in the region, distribution and abundance of wide-ranging or widely dispersed species, and quality of habitat in a large area.
- Medium-scale concerns, involving various spatial management tools (e.g., WHAs and UWRs) by watershed or management unit, such as assessing distribution and abundance of habitat, or barriers preventing movement or use of an area.
- Small-scale concerns, such as spatial management tools at a stand level or smaller that address forest structure and condition, or habitat elements (features) required for long-term survival.

The draft list of key questions and potential indicators for a study is usually longer than is feasible to evaluate. Selecting the most valuable, cost-effective indicators is one of the most important aspects of a monitoring project. Often some flexibility is required, as the indicators initially selected may prove less valuable as knowledge gaps are identified and new knowledge is gained, or when selected indictors fail to provide sufficient accuracy. After choosing (or developing) an indicator to answer a question, it may be important to test the indicator in a pilot study to determine whether it meets the criteria of a good indicator (see sidebar, "Characteristics of Good Indicators") and provides a sound answer to the management question. Pilot studies will also indicate whether the proposed study is feasible and the data from the selected indicators will make a valuable contribution to knowledge. Subsequently, adjust or replace indicators or add other indicators to ensure effective monitoring.

Most effectiveness evaluations involve monitoring more than one indicator. In the realm of wildlife resource values, most projects evaluate a combination of forest or range structure indicators and species indicators. Although measuring forest or grassland structure may appear more cost effective than measuring population parameters (e.g., resource constraints or logistics issues make it easier to sample elements that do not move or hide, that are not uncommon or vulnerable), measuring only habitat may fail to indicate whether providing habitat structure is sufficient to maintain productive animal populations over time. Measuring species population indicators can provide an early warning system for critical declines and identify

Characteristics of Good Indicators

Good indicators exhibit the following characteristics.

- *Focussed on answering a specific evaluation question; directly reflects and covers key aspects of the management goal.*
- *Clearly related (correlated) to the true parameter they represent.*
- *Easy to interpret; useful in describing the elements or relationships they represent, and easily communicated to the target audience and stakeholders.*
- *Practical, feasible, easy to measure, reliable, robust and cost effective, maximizing information gain while minimizing time, effort and expenditure; the data is obtained at an acceptable cost and in a reasonable time frame.*
- *Fine-scaled, but only as the question requires; addresses the scope of the question.*
- *Operationally relevant, addressing an issue of concern and the forest or range practice under study with the appropriate level of accuracy, precision, and scale to provide results that are useful in the decision-making process; relate back to implementation of the tool or practice; measure changes that can be averted by management actions.*
- *Responsive to change in a predictable way; sensitive to changes in management or stressors, reflecting well-understood causal relationships; consistently responsive to actions we are concerned about; shows differences between units; shows trends over time.*
- *Precise and unambiguous, not clouded by background processes such as weather, climate, stochastic events, or natural variation (low naturally occurring variability).*
- *Supported by science and widely used (rationale, methodology, and analysis well documented; measurable in a standard scientifically credible manner); results have acceptable variability; peer reviewed.*
- *Integrated, providing information about multiple levels or aspects of the system; relevant at site, feature and landscape scales (as appropriate).*

where finer-scale population monitoring is required. Measuring species or population indicators allows comparisons to habitat benchmarks and helps refine model relationships over longer time periods and larger areas. Nevertheless, selecting the correct species indicator is critical—monitoring indicators for all species is usually not possible, and therefore choosing indicators for the appropriate informative species (perhaps some focal species or a guild of species) is necessary to answer the important monitoring questions.

Sample evaluation questions and indicators for numerous objectives used in ecosystem monitoring projects have already been developed (see: Bunnell 2005; Gordon et al. 2005; Schueller et al. 2006; National Park Service 2012). The reader is urged to review the indicators presented in these documents and the reports available on FREP's Wildlife Resource Value webpage (http://www.for.gov.bc.ca/hfp/frep/values/wildlife.htm).

2.5.2 Thresholds and expected values

Several types of data, in various combinations, are used for comparisons in effectiveness evaluations (see sidebar "Baselines, Benchmarks, Controls, Targets, Thresholds, and Triggers"). For example, in response to a management activity, or over time, an indicator value may rise above, fall below, or fluctuate around some level that is significant to management. Thresholds determined from previous studies or values derived for the study will often trigger management decisions and actions, ensuring the survival of a species or conservation of an ecosystem. At other times, the management goal involves an indicator reaching (or stabilizing at) a target value, or an indicator is predicted to reach a certain level within a specific time frame, based on knowledge from previous studies or from benchmark or control sites.

Evaluators will want to compare observed conditions or trends in indicators with expected values, thresholds, triggers, and targets, and then predict expected indicator values resulting from the management practice. These predictions, ideally quantitative estimates with confidence limits, should reflect explicit hypotheses that address expected future conditions and trends tested in the effectiveness evaluation. Quantitative models provide valuable input to forecasts of future conditions of indicator values; simpler models may require adjustments to include alternative management scenarios. New information gathered during monitoring is used to update models and predictions for indicator values.

Baselines, Benchmarks, Controls, Targets, Thresholds, and Triggers

Baseline data reflect the condition of the habitat or species at a specified time, either at some time before establishment of the UWR or WHA, at time of establishment, or at some fixed time when the evaluation study begins. This information is incomplete for UWRs, WHAs, Wildlife Habitat Features, ungulates, and species at risk. Therefore, pilot studies and studies to collect baseline data are recommended as a first step when implementing effectiveness evaluation projects. Baseline data from a pilot study can help determine necessary sample sizes and improve field and analytical methods. If collected at a later stage using the same or a comparable methodology, baseline data can aid comparisons of subsequent measures of habitat or species biology and estimation of trends over time.

Benchmark data include measures of habitat or species in a reference condition other than the baseline value. Benchmarks are standards or targets aimed for, or a meaningful point of reference in time, space, or condition. Comparison of effectiveness evaluation data to benchmark data can provide insights into level of risk and the need for changes in management practices.

Sometimes a benchmark is a biological threshold value, a target, or a management trigger point that, if reached, would signal a need for a change in management practice. Targets are set through regulation (or policy), certification standards, or (sometimes) scientific evidence (e.g., the density of wildlife trees in a management zone, or number of bats using a roost tree in a WHA). Thresholds are usually upper or lower limits that serve as triggers, or early warning signals, for responsive action. It is possible to initially estimate thresholds from "a syntheses of available data, model projections of known relationships, or reasoned guesses" (Houde et al. 2005), and then refine these through adaptive management.

Control data describe a group of subjects (e.g., wildlife species) or conditions (e.g., habitat parameters) that is not exposed to any treatment (either experimental or operational) and that is matched as closely as possible (i.e., all else being equal) with an experimental group of subjects or conditions. The untreated control group acts as a standard or yardstick to detect and measure changes that occur in the treated group.

ACTION CHECKLIST . . .

- ✓ *Examine the conceptual model of relationships to determine potential measurable parameters and variables that will indicate achievement of wildlife resource value objectives*
- ✓ *Select highest priority, feasible indicators for the effectiveness evaluation study*
- ✓ *Determine critical threshold, trigger, or target values of the selected indicators*
- ✓ *Predict indicator values as a result of management practice*

2.5.3 Developing an evaluation tool for effectiveness indicators

IN BRIEF . . .

This task examines potential outcomes of all indicators combined to create an evaluation tool, by asking:

- *What levels of the population or habitat condition indicators show the population or habitat condition is secure, at risk, or not effective?*
- *What levels of the threat indicators show extreme, unacceptable, moderate, or low levels of risk to the population, its habitat, or the integrity of the WHA or UWR?*
- *What combination of condition and risk indicator values will lead to the conclusion that the WHA or UWR is effective, at risk, or ineffective at meeting management objectives?*

After indicators are selected for monitoring and evaluation, the project team must consider how it will evaluate them to determine effectiveness of the WHA or UWR; that is, how will they combine monitoring results of population and habitat condition indicators with threat (risk) indicators for an overall conclusion about effectiveness?

This task requires:

1. Pre-defined categories of condition for each indicator of population and habitat condition (e.g., poor, satisfactory, or good condition; declining, stable, or increasing trend)
2. Pre-defined categories of threat for each indicator of risk (e.g., low, moderate, extreme; long term or immediate)

3. Decision-making processes or protocols (Figure 1) that assign an overall condition rating to the combined suite of indicators (e.g., highly, moderately, or not functional) and overall risk rating to the combined suite of risk indicators (e.g., low, long term, or immediate)

4. A tool (e.g., see Table 2) that will assign effectiveness ratings to the study area based on overall ratings of condition and risk (i.e., a pre-defined effectiveness rating scale that reflects all potential combinations of condition and risk ratings)

Assigning condition (functionality) and threat (risk) rating categories to a WHA or UWR (Figure 1 and Table 2, above) requires *a priori* recognition of indicator critical values that have biological or management relevance, such as thresholds, targets, or trigger values (see Section 2.5.2). Critical values (or values relative to them, such as 50% of the critical value) are often the cut-off points for condition or threat rating categories. Categories may also reflect direction of trends (favourable vs. unfavourable) in condition or risk. The rating categories are defined from expert opinion or previous scientific evidence before the effectiveness monitoring begins. They are often key factors represented in preceding conceptual models.

Table 2. ***The standard wildlife resource value effectiveness ratings matrix[a] with typical functionality and risk categories and conventional management interpretation.***

Condition	Risk[b]		
	Low	*Moderate*	*High*
High	Effective	Effective	At risk
Moderate	Effective	At risk	At risk
Low	At risk	Not effective	Not effective

a See the FREP wildlife resource value monitoring and evaluating framework (2009) at: http://www.for.gov.bc.ca/ftp/hfp/external/!publish/frep/values/Wildlife_Framework_Paper.pdf

b Management interpretation: "Not Effective" and "At Risk" effectiveness ratings indicate need for management action. An "Effective" rating does not require a change in management regime or new actions aside from continued monitoring.

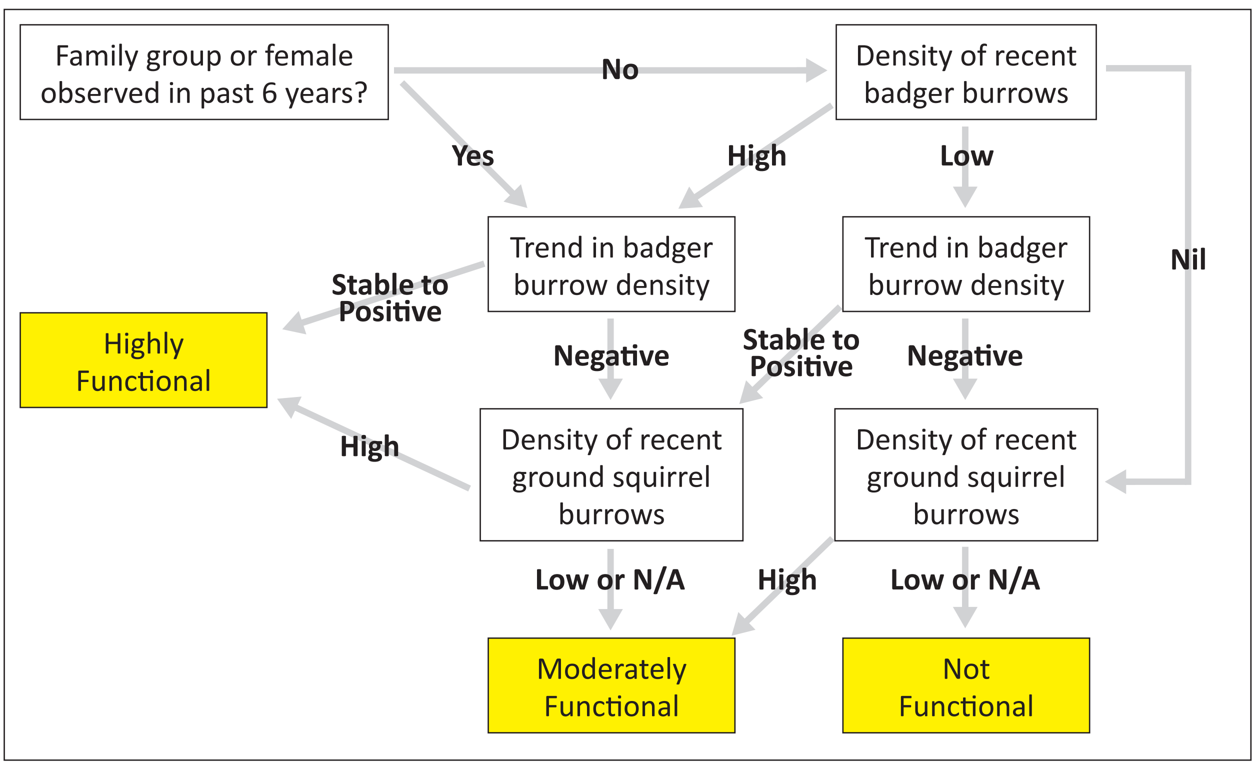


Figure 1. ***Example decision-making process: Assessment process for assigning badger WHA functionality ratings (from Kinley 2009, quoting Newhouse et al. [2007:6]): "Densities of recent (occupied during year of survey) badger or ground squirrel burrows considered to be 'high' are to be determined regionally based on the results of baseline WHA assessments. Ground squirrel densities are to be considered only in regions where ground squirrels are the predominant prey. The density of recent badger burrows must decline by at least 30% annually (50% biennially) to be considered 'negative'."***

Assigning an overall rating of condition or risk to the combined suite of condition or risk indicators can be complex and requires consideration by species experts. For example, an indicator may have greater importance for species survival or feasibility of related management action and therefore receive a heavier weighting than other indicators. In another case, a high risk rating for one indicator may override any lesser value of other risk indicators. It is important to consider all possible combinations of risk and condition ratings and any situation-specific interrelationships among the ratings and assign these to overall effectiveness ratings before monitoring starts.

The matrix shown in Table 2, which considers condition ratings in the context of risk ratings, is a useful tool for all effectiveness evaluations of wildlife resource value management practices. Although its categories of functionality and risk may be adapted (the wording changed) to reflect different projects and hypotheses (with sufficient knowledge for more categories), the 3 x 3 format is standard for the wildlife resource value effectiveness monitoring and evaluation program.

It is important to consider this final step of effectiveness evaluation before beginning data collection (*a priori*) in order that the study approach and sampling design provide the appropriate data to make the assessment; however, the project team should re-examine the evaluation tool and revise it (or its interpretation of it) as necessary following the outcome of a pilot study or several years of monitoring (*post hoc*).

ACTION CHECKLIST . . .

To prepare for the next step in an effectiveness evaluation:

- ✓ *Develop a flowchart or decision-tree protocol that determines whether the combined condition (functionality) indicators show the WHA or UWR is highly functional, moderately functional, or not functional (other categories could be good condition, moderate condition, and unacceptable condition).*
- ✓ *Develop a flowchart or decision-tree protocol that determines whether the combined threat (risk) indicators show the level of risk to the WHA or UWR is nil-to-low, long term, or immediate (other categories could be low, moderate, or extreme risk).*
- ✓ *Determine the overall effectiveness rating from a matrix (the standard wildlife resource value program "effectiveness rating matrix") that cross-references condition (functionality) and threat (risk) ratings to conclude the WHA or UWR is effective, at risk, or not effective.*

2.6 *Pilot-testing a Protocol*

During the development of an effectiveness monitoring protocol, pilot studies allow the project team to test assumptions, field and analytical methods, and data quality, and ultimately improve the protocol. The project leader needs to determine whether a proposed effectiveness evaluation protocol is in its final form or requires pilot testing before implementation.

Pilot studies involve the development of a preliminary study and sampling design (Section 3) before implementation (Section 4); the outcome of the pilot may result in changes to the study design, the sampling design, or the data collection or analytical methods. The process of feedback for improvement is the crucial product of pilot studies.

Most projects will require one (or possibly two) season to test study design, field methods, data forms, and analyses. A pilot season provides important feedback to improve project methods and outcomes and may provide new or updated baseline data (Table 3).

Table 3. ***Possible objectives and improvements from pilot-testing.***

Pilot-test objective	Recommended improvements
Feasibility of methods and delivery	Adjust methods
Validity of indicators	Drop or add indicators
Focus of questions	Refine questions
Utility of field forms	Revise field forms
Soundness of sampling design	Adjust design
Rigour of data, power of statistical tests	Amend rigour
Rigour of effectiveness evaluation matrix	Amend matrix
Baseline data	Adjust sampling design or effort

2.7 *Tracking Project and Program Efficiency*

IN BRIEF . . .

This task asks:

- *Is the project cost effective?*
- *Are the project (or program) resources (funds and people) well spent?*
- *Are we getting value for the dollars and effort spent?*

As part of developing and implementing effectiveness evaluation projects, it is important to assess and track project or program efficiency (e.g., is the money and effort well spent). Evaluating program efficiency determines the success of wildlife resource value program administration, co-ordination, and leadership in terms of costs and benefits. Information collected to track the efficiency of each project helps assess overall program efficiency and ensures continuing support for the monitoring project.

Performance measures to gauge the efficiency of effectiveness evaluation projects and programs fall into three categories.

1. Cost effectiveness measures, such as:
   - expenditures (dollars, staff time), showing totals and trends
   - feasibility of field logistics
2. Project (or program) efficacy measures, such as:
   - numbers and types of partnerships and co-operators
   - level of funding from other sources (partners)
   - amount of co-operative data collection (sharing data or data collection with other monitoring projects or programs)
   - volume, distribution, and timeliness of extension deliverables
3. Contribution to improved management measures, such as:
   - (trends in) public awareness of management issue and of management efforts
   - changes in management expenditures with gained knowledge

Measures of project efficiency will differ depending on the stage of development of the effectiveness evaluation—that is, whether the protocol is under development, a pilot study is under way, or project is in the implementation phase. Efforts to track project efficiency should be included in project work plans and reports.

3.0 DESIGN THE PROJECT

Designing a statistically rigorous monitoring project involves two key tasks:

1. designing the overall approach, referred to as the "study design"; and
2. establishing the rules that govern what is measured, referred to as the "sampling design."

Careful consideration of both is crucial to successfully addressing the management questions and monitoring objectives. It is also important to document decisions and details to ensure that the monitoring project will continue for many years, even if the project leader has moved on. Providing clear documentation of the "why, what, when, who, and how" will sustain the program well into the future. Thus, the details of the study design or overall approach, sampling design, field methods, and data analysis plan should be described in a monitoring plan. Even if a project is not a field project that requires a sample design, the details of the monitoring project should be documented in a monitoring plan.

The monitoring plan should be reviewed by a statistician, preferably one with biostatistical experience. Statistical review, undertaken early in project development, ensures that: the right parameters are measured, enough measurements are taken, effort is not wasted on extraneous sampling, the data are sound and credible, estimates of reliability can be calculated, the sampling scheme is appropriately rigorous for robust inferences, and the analyses provide an answer to the management issue or question.

3.1 *Designing the Approach*

IN BRIEF . . .

This task answers the following questions:

- *What key question(s) and hypotheses are under evaluation?*
- *What is the general delivery model; who will conduct the various aspects (planning, field work, analysis, reporting) of the project—headquarters, region, district, contractor, ministry, industry?*
- *What is the project's geographic location and scale—stand, landscape, region?*
- *What is the project's temporal scale—at what time of year and at what interval?*
- *At what intensity will the project be conducted—routine, extensive, intensive?*
- *What field methods will be used—existing standards or innovative procedures?*
- *What overlap or commonalities exist between this and other monitoring projects?*

The first task in designing the study is to decide on an overall approach. What are the study's objectives? Where will the study be conducted? In general, when will data collection take place and how will it be collected? These components of the study design are addressed in the following subsections.

3.1.1 Study design[4]

It is possible to approach an environmental field study in many ways (see Eberhardt and Thomas 1991 for an overview). Choosing the right approach requires careful consideration of the study objectives, the degree of control required, the desired level of inference, the effect size of interest, and the trade-offs surrounding issues of cost and feasibility of the various approaches. Cochran (1977) described two broad types of survey: (1) descriptive and (2) analytical. The objective of descriptive surveys is to obtain information about general categories of objects (e.g., the frequency of large woody debris pieces in a watershed), whereas, analytical surveys are used to make comparisons among groups within the population to test hypotheses (e.g., are there fewer large woody debris pieces in Fisheries Sensitive Watersheds than in undesignated watersheds?). Hurlbert (1984) categorized studies as either manipulative or mensurative experiments: manipulative studies are those in which the investigator has control over the factors in the study and mensurative studies are those in which only passive observations are used. Eberhardt and Thomas (1991) included replication as a key requirement for improving the strength of inference and described eight categories of environmental studies, ranging from the preferred approach of a controlled experiment with replication to a simple descriptive sampling approach. Schwarz (2012) provided an excellent summary of the trade-offs between different study approaches, from descriptive surveys to designed experiments. Figure 2 illustrates the relationship between the degree of control and the strength of inference possible for an array of study designs. Appendix 1 provides examples using WHA monitoring.

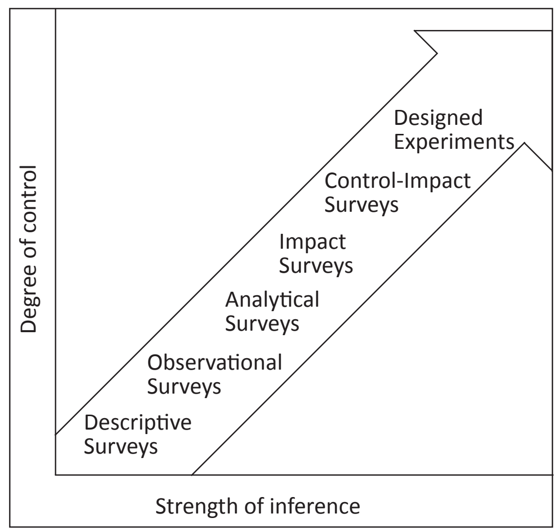


Figure 2. ***Relationship between degree of control, strength of inference (and ability to determine causation), and type of study design (from Schwarz 2012).***

As a first task, describe the overall approach and objective. Houde et al. (2005) discussed four general approaches for effectiveness monitoring: (1) operational, (2) experimental, (3) design-based, and (4) model-based. Operational approaches use real-life and readily available treatments or practices applied at an operational scale under operational conditions. An experimental approach creates and compares different treatments. Design-based approaches are grounded in a statistical sampling design, whereas model-based approaches intensively collect information at a few representative sites (sometimes

4 This section on study design was adapted from Wieckowski et al. (2008).

called "sentinel sites") to create an ecological model that is then applied to similar sites. The most appropriate approach, or combination of approaches, depends on the objective of the study (or monitoring question).

Also consider the kinds of comparisons, if any, that the evaluation will address: before- and after-treatment comparisons (such as before and after harvesting, or before and after implementing a mitigative management action), comparisons to control sites (e.g., no treatment sites), benchmark (desired condition) sites, or to important threshold or management trigger values.

3.1.2 Delivery model

The delivery model is another component of the overall approach. Project implementation is expressed in a general way, such as "led from Headquarters with assistance from District staff," or "conducted by consultant under contract administered from Coast Region." The delivery model also describes interagency, academic, and (or) industry partnerships and proposed or confirmed funding arrangements.

3.1.3 Spatial scale

Setting the spatial scale for an effectiveness evaluation project determines how widely the results of the monitoring project will apply. As a general rule, the spatial scale of effectiveness monitoring should match the scale of the project objectives and management strategies and actions. Extrapolation of study results beyond the area of inference is a common problem in biological studies—inferences drawn from the sampling population should apply to the broader target population (see Section 3.2).

Some projects are designed at a site-specific scale, focussing on a specific UWR or WHA (e.g., a study examines whether desired structural elements and habitat conditions exist or are maintained within a WHA). If a species targeted for conservation occurs entirely within a WHA and does not migrate outside its boundary, or conservation actions and objectives are restricted to the WHA, then a site-specific scale is appropriate.

Projects designed at a landscape-unit or watershed scale answer questions about multiple sites, such as:

- Do current UWR boundaries fit patterns of observed animal use?
- Is habitat distribution and abundance within a group of WHAs sufficient for species survival?
- Is additional protection needed for other life requisites?

If the distribution of a target species is not limited to a WHA, or if multiple reserves are in an area, then a landscape-scale assessment would be best.

At a regional or subregional scale, or even at a provincial scale, projects are designed to determine whether UWRs, for example, are in sufficient numbers and adequately distributed to maintain healthy regional populations. An evaluation may occur at different scales (e.g., provincial, regional, local) depending on what is of interest.

3.1.4 Temporal scale

The temporal scale provides a general descriptive component of the study's overall approach, addressing aspects of project timing and duration, such as:

- time of year for monitoring activities
- re-sampling interval (e.g., weekly, monthly, every 3 years)
- duration of monitoring
- expected completion date

Details concerning the project's temporal scale are more fully developed during the sampling design phase (see Section 3.2).

3.1.5 Intensity

The overall approach should include a description of evaluation intensity, of which FREP recognizes two levels: routine and extensive.[5] These intensity levels reflect the degree of effort, cost, and complexity, all of which depend on the monitoring questions, indicators, and methods, the resources available, and the circumstances. Routine evaluations generally involve relatively inexpensive and rapid (but statistically valid) measurements and visual assessments, including office-based computing exercises. Extensive evaluations are usually more focussed and rigorous, and involve detailed on-the-ground monitoring, including the collection of categorical data using visual estimates or relatively simple measurement. The resulting data are more quantitative and data analyses involve detailed comparisons.

Routine and extensive evaluations provide valuable information on current status, trends, and implementation issues and can be applied to different types of wildlife

5 See more details of the monitoring categories adopted by FREP, at: http://www.for.gov.bc.ca/hfp/frep/rsm/index.htm

resource value monitoring. Evaluations may be extremely detailed, long-term commitments, or quick assessments designed to identify key areas of concern. Levels of intensity are not sequential—projects must be initiated at the level that reflects program priorities and information need.

3.1.6 Location

The location component describes the project's geographic location, either broadly or specifically (e.g., northeastern British Columbia, Coast Forest Region, North Island District, ICHdw subzone, or WHA-999 and its specific location), and the rationale and factors considered in selecting this location, describing what makes it special or the best site for an evaluation.

3.1.7 Methods

A brief description of field and office procedures, methods, and techniques is also included as part of the overall project approach. It does not need to include details of the sampling design (see Section 3.2) but might briefly describe standard techniques that will be used or adapted, or innovative methods that will be developed. It might also include intentions to ensure data consistency, quality assurance, or data storage and analysis standards. Differences in methods selected for this project from other monitoring projects (especially FREP monitoring projects) should be noted and briefly rationalized.

3.1.8 Overlap with other monitoring

Finally, when describing the overall project approach, it is also important to note any overlap and commonalities of methods, data collection, and sites, or co-operative efforts with other projects taking place under FREP or by any other program.

ACTION CHECKLIST . . .

Determine the overall project approach:

- ✓ *Study design and delivery model*
- ✓ *Spatial and temporal scale*
- ✓ *Level of monitoring intensity*
- ✓ *Geographic location*
- ✓ *General description of methods*
- ✓ *Program overlap*

3.2 Sampling Design

IN BRIEF . . .

To develop a sound sampling design before project implementation, we examine the following questions:

- *What is the target population?*
- *What variable(s) will be measured and how will they be measured?*
- *What is the sampling unit for the measured variable(s)?*
- *Will samples be stratified or random?*
- *What sample size is required? Consider effect size.*
- *How many times a year will you sample a point? In how many years would you like to detect a change or trend?*
- *What analytical methods will be applied, and at what levels of significance (alpha level)?*
- *What data, if any, is available for measured variables to provide guidance on the natural variability in measured variables?*
- *How much money and personnel is available to project?*

This section[6] guides the user through the necessary tasks to create a sampling design. Thompson (2004) described two necessary components to an effective sampling design:

1. the spatial and temporal selection of sampling units, and
2. the measurement protocol within a given sampling unit.

The first component recognizes that, for most ecological studies, we cannot sample all possible areas in all possible time periods. This sampling variability is the subject of most traditional statistical sampling books. The second component recognizes that the probability of detecting species (especially rare or elusive species) is less than perfect. In other words, even if an animal or plant is present in a sampling unit, it may elude detection.

Consideration of subsequent data analysis should also take place when developing the sampling design; ask yourself: "What will I do with the data, if I have it?" Both spatial and temporal selection, as well as detectability, should be

6 Adapted from recommendations by Elzinga et al. (2001) and Vesely et al. (2006).

incorporated into a sampling design (Thompson and Seber 1996; Yoccoz et al. 2001; Thompson 2004) along with a discussion of the intended data analysis (Thompson and Seber 1996). The ability to create an effective sampling design depends on clearly articulated study objectives (see Section 2.1 and 2.2) and knowing the key questions and hypotheses for evaluation. If clear objectives are formulated, then a pilot study, past research, or expert knowledge should provide the information needed to work through the following tasks and design an effective study.

3.2.1 Life history characteristics

The life history characteristics of the subject species drive the decisions made at each phase of sampling design development. A description of life history should include: what is known or unknown about the spatial and temporal distribution of the species; some indication of abundance, reproductive strategy, detectability, and habitat use; and anything else related to the subject species' life history that might affect the sampling design. This task should follow easily from the conceptual model and indicator development in Section 2.4 and 2.5.

3.2.2 Target population

The *target population* has been described variously as:

- the complete collection of individuals we wish to study (Lohr 1999);
- the population about which information is wanted (Cochrane 1977); or
- the complete set of units about which we want to make inferences (Elzinga et al. 2001).

To make inferences about the entire target population, all individuals within the target population must have some chance of selection in the sample (Figure 3).

The *sampling population* is the collection of all possible sampling units that might have been chosen in a sample; that is, the population from which the sample was taken (Lohr 1999). The sampling population should be the same as the target population, although often time or budget constraints dictate that the sampling population is smaller than the target population. If this is the case, then any inferences drawn from the sample only apply to the sampling population (Cochrane 1977). Extending inference beyond the sample population requires additional information and expert judgement. If the sampling population does not cover the entire target population, then collecting supplemental information to describe how the two populations differ may help to understand the limitations of the inference. For example, if the target population is all fish-bearing streams in British Columbia but the sampling population only included streams up to a certain elevation, then we should not extend inferences based on lower-elevation streams to those at higher elevation. Similarly, if logistical constraints prevented the sampling of deep pools, then we should not extend inference to deep pools, as these are effectively removed from the target population.

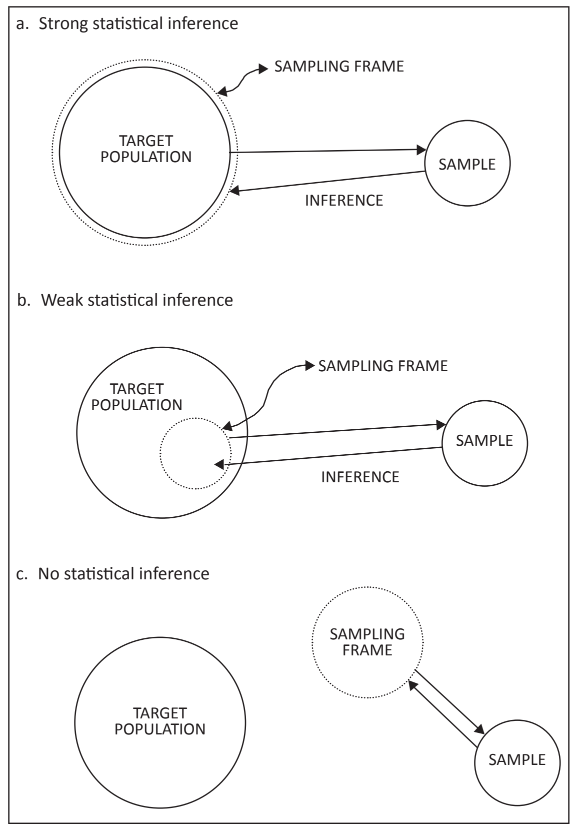


Figure 3. ***Three different conceptual relationships among the target population of inference, the sampling frame, and the sample of units drawn for characterization (excerpted from Skalski et al. 2005:364).***

3.2.3 Sampling unit

The *sampling unit* is the actual unit of measurement. The sampling population is divided into many sampling units. The list of all possible sampling units is called the *sampling frame* (Lohr 1999). Individual animals or plants generally represent the sampling units for measured organismal characteristics, such as height, weight, and health. Plots or similar area delineations are usually the sampling units for population measures, such as abundance, density, or biomass; however, multiple levels of sampling units may be necessary. For example, if interested in characteristics such as height or weight, then individual animals seem the obvious sampling unit, but without knowing how many animals exist or their locations, you can't actually create a sampling frame. Consequently, you may first need to define sampling units as plots or transects in space, and then measure characteristics of individual animals observed or captured within the plot or transect. This is an example of a multi-stage sampling design, in which the plot or transect is the primary sampling unit and the individual organism is the secondary sampling unit. Multi-stage sampling designs are very common and are discussed in more detail later in this section.

If the sampling unit is some measure of area, such as a plot or a quadrat, then the next thing to consider is the size and shape of the sampling unit. Generally, the aim is to choose the sampling unit that provides the most precision for the least effort. A good strategy to accomplish this is to choose the size and shape of sampling unit that will minimize the between-unit variability without adding too much sampling effort. A good idea, if feasible, is to choose a plot size large enough to ensure some observations or captures within the sample unit (i.e., avoid too many zeros). An excessive number of plots with zero individuals of interest will likely complicate associated analysis. Elzinga et al. 2001 provided a useful summary of key considerations when choosing the size and shape of the sampling unit for ecological studies, a condensation of which follows.

- *Travel and setup time versus searching and measuring time*: Evaluate the trade-off in effort and precision between measuring fewer large plots versus more small plots. How hard is it to find and measure the individuals of interest? Are individuals large and conspicuous (e.g., trees), or are they small and difficult to find?
- *Spatial distribution of the individuals in the population*: Many species of plants and animals have clustered distributions. If the distribution is clustered, then small plots will tend to result in frequent zeros, whereas larger plots are more likely to intersect a cluster. The orientation of the sampling units is also important. For example, if a possible gradient in the measurement of interest exists that is related to some environmental covariate such as sun exposure, slope, or moisture content, then the sampling units should cross this gradient. If moisture content was an issue, then laying out plots so that each plot had a similar range of moisture content should reduce between-plot variability; alternatively, stratify plots on the basis of the environmental differences (see discussion on stratified sampling).
- *Edge effects*: Determining whether an individual lies within or outside a plot boundary is sometimes difficult. Some observers may tend to include all boundary sightings, whereas others may ignore them all. Either way can lead to biased estimates and the sampling protocol should explicitly define a consistent rule for handling edge observations. Rectangular plots have the greatest edge per unit area, whereas circles have the least.
- *Abundance of target population*: An initial sense of abundance is useful in choosing an appropriately sized sampling unit. The goal is to avoid either creating a sample unit so large that you will have to count thousands of individuals or so small that you end up with many empty plots.
- *Ease in sampling*: Long, narrow quadrats are generally easiest to search (Krebs 1989; Elzinga et al. 2001). An observer can move in one direction through the entire plot with little likelihood of losing track of individuals already counted. For example, in a recent pilot study, a long rectangular plot was found preferable to a circular plot of a similar area for measuring grass height (L. Tedesco, Ministry of Environment, 2008, pers. comm.). The rectangular plot allowed clear sighting of all plants and measurement with little extraneous movement. Conversely, keeping track of which plant had been measured in the circular plot was difficult and more time consuming as the observer had to shift position frequently.
- *Disturbance effects*: Consider how the sampling may directly affect the species of interest and the consequent interpretation of monitoring results. This is important for two reasons: (1) it would be counterproductive to affect the species of interest (which in the case of the FREP program will often be a species at risk); and (2) sampling disturbance to sites (especially around permanent sampling sites) could affect the behaviour of the monitored species over time.

Results from pilot studies, information from other published research, or expert opinion based on biological characteristics should help in choosing the appropriate size and shape of the sampling unit, according to the above considerations.

3.2.4 Positioning of sampling units

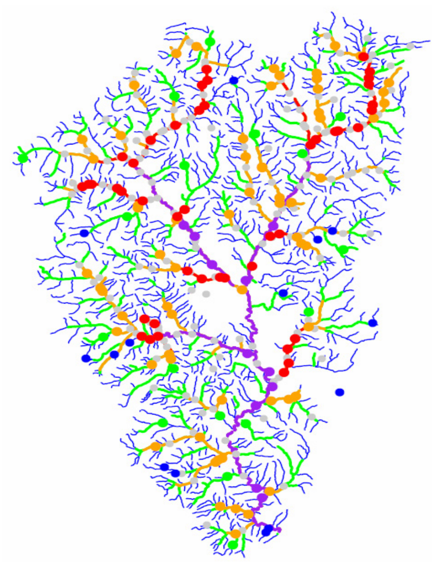
Now that the target population, sampling frame, and sampling unit are defined, how should one choose the actual sample? Probability sampling theory requires that each sampling unit has a known probability of being chosen and that the units are randomly selected (Cochrane 1977).

Standard probabilistic designs – Three basic probabilistic sampling designs are most commonly used:

1. simple random sampling,
2. systematic random sampling, and
3. stratified random sampling.

With simple random sampling, a random sample of all sampling units is selected within the sampling frame (e.g., drawing numbers from a hat). In systematic random sampling, sampling units are selected at regular intervals, using a randomly selected starting point (e.g., reading every tenth name from the phone book, or taking a sample every 10 m). With stratified random sampling, the sampling frame is divided into non-overlapping groups (strata) based on some characteristic such as sex or habitat type, and then a random sample is chosen from each of the strata. Although it is always possible to use simple random sampling, this sampling design is not always the most efficient choice because it may require a larger sample size to obtain comparable information about the population compared to another, more efficient sampling design. If the population of interest is randomly distributed, then systematic random sampling will approximate the simple random design (Lohr 1999). If the target population changes proportional to position (e.g., samples taken upstream vs. downstream), then a systematic random sample may ensure spatial coverage; however, if the target population displays regular or cyclical characteristics, then a systematic sampling design is a poor choice. A stratified random sample is more efficient when less variability is evident within strata than between strata (Cochrane 1977; Lohr 1999). A stratified random sample is also useful if estimates for individual strata are desired, as well as for the entire population.

Generalized Random–Tessellation Stratified (GRTS) designs – This design is a spatially balanced probabilistic survey design developed by the U.S. Environmental Protection Agency (EPA) under their Environmental Monitoring and Assessment Program. The GRTS overcomes some of the shortcomings of both simple random sampling (which tends to "clump" sampling sites) and systematic sampling by generating an ordered, spatially balanced, and unbiased set of sites that represent the population from which the sample sites will be drawn. Software required to create a GRTS design include: psurvey.design, R statistical package (both available from the EPA's Software Downloads for Spatial Survey Design and Analysis webpage[7]), and ArcGis. On request, the EPA's Monitoring and Design team[8] will also supply customized GRTS-generated sample points for selected areas, as long as the client indicates or provides the necessary GIS layers for the intended spatial sampling frame (e.g., stream networks, polygon boundaries). Figure 4 illustrates an example GRTS design, showing a random, spatially balanced draw of a pre-selected number of sampling points for each of five user-defined, stream-order strata within a watershed hydrology network.



Stream Order	% of sample sites
1st	10
2nd	10
3rd	40
4th	30
5th	10

Figure 4. ***Hypothetical GRTS draw of random, spatially balanced sampling points within stream-order classifications for the Methow Basin in Oregon (source: Phil Larsen, EPA).***

7 See: http://www.epa.gov/nheerl/arm/analysispages/software.htm

8 Email contact: ORD_EMAP_DESIGN@epamail.epa.gov

Judgement or non-probabilistic designs – Judgement samples are selected subjectively based on some prior belief about where individuals are located, on arbitrarily chosen representatives of the target population, or chosen just for convenience. So-called "representative reaches" in stream surveys are an example of a judgement sample; however, without a census of the target population (e.g., stream), it is impossible to insure that a representative sample is chosen (e.g., reach). McDonald (2004) described many examples in which allocating some effort outside supposed core areas considerably improved understanding of distribution and estimates of abundance for rare and elusive populations. For example, if core habitat had relatively consistent densities but marginal habitat density fluctuated according to population size, then ignoring the marginal habitat would mask the true health of the population. Sometimes locations thought of as particularly sensitive "sentinel sites" are selected for monitoring, but this too is a controversial issue (Edwards 1998). For example, inferences made strictly from judgement samples have resulted in some famous miscalculations in election result predictions (Edwards 1998).

Non-standard probabilistic designs – Many variations on these basic designs have been created to address particular situations (e.g., cluster sampling, adaptive sampling, and distance sampling; see Appendix 2 for other sampling designs). It is possible to combine any of these designs into a multi-stage sampling strategy. For example, a simple random sample of transects is chosen from each stratum within the target population, and then a systematic random sample of plots is selected within each transect, followed by a census of individuals within each plot. Calculating an estimate of the mean from a multi-stage sample is fairly intuitive, but the variance calculations are more complicated. A typical mistake is to treat all observations as if drawn randomly from the target population. In reality, the secondary sampling units (plots in this case) are a sample from the transect, not the population. Increasing the number of plots within transects only helps improve the precision of the estimate of the single transect. This is one form of pseudoreplication, as discussed by Hurlbert (1984).

Strategy – Stratification can be a very powerful tool for improving the efficiency of a design, and so assessing whether stratification is appropriate is a good place to start. The following series of questions, combined with an understanding of the strengths and weaknesses of each sampling design, should enable an appropriate allocation of sampling units within the population or strata.

1. Is a stratified design appropriate?
   - Are obvious groupings apparent in the target population, such as ownership boundaries, habitat types, or differences in density? If so, is it reasonable to assume that the variability between groupings is greater than the variability within these groupings?
   - Would it be useful to know the precision of estimates for particular groups within the populations?
2. How should sampling units be positioned within the population (or strata)?
   - How big is the total target population?
   - How much time does it take to move between sampling locations?
   - Are the objects distributed at random, uniformly (typical of territorial animals), or in clusters?
   - Are any regular landscape features, such as ridges, fence lines, roads, or riffle/pool sequences, related to the response variable of interest?
   - Are any known gradients, such as upstream vs. downstream, low moisture to high moisture, or elevational differences, apparent in the target population?
   - How time consuming is the actual measurement process within each sampling unit?

3.2.5 Sample size and statistical power

The next task is to determine the appropriate number of samples at each level of the sampling design to ensure sufficient precision for the study objective and to optimize the allocation of monitoring effort and cost (too many samples waste dollars and effort, too few samples will not answer the questions of interest).

Sample-size calculators use preliminary estimates of variability to assess the trade-off between effort and precision. Many free programs (e.g., Program TRENDS or MONITOR) can help determine the best sample size for your project and all function in essentially the same way.

In general, sample-size calculations require:

- estimates of variability within and between sampling units at each stage of the design (preliminary estimates of variability are obtained through pilot studies, similar studies conducted elsewhere, or theoretical behaviour of the system);

- the desired level of precision (biological effect size of interest) (see Appendix 3);
- acceptable Type I (α) and Type II (β) error rates (see Appendix 3);
- cost of sampling at each stage of the design;
- the question of interest and related statistical test (i.e., testing for a difference between two groups, or testing for a trend over time etc.); and
- knowledge about the distribution of the data of interest.

It is inadvisable to use a generic sample-size calculator without understanding the underlying assumptions of the calculator. Sample-size calculations depend on the study design (e.g., simple, stratified, multi-stage), the distribution of the data, and the study objective (e.g., comparing two groups vs. assessing trends). The most useful calculators include some measure of cost per effort to optimize the precision against cost trade-offs. Since the calculations are generally not difficult, ask a statistician to develop a sample-size calculator for your project, especially if the study or sampling design is unusual.

In statistics, *power* refers to the ability to detect an effect (e.g., a difference between two groups, a difference in distribution, a shift in location, a trend over time). Power analyses are based on the same concepts as sample-size calculations but are used to determine the size of an effect that can be detected for a given sample size. Typically, statistical power is calculated for a range of sample sizes, where increasing sample size results in increasing precision and, therefore, power. It is useful to display power as it changes with sample size and cost per effort.

In practice, determining the appropriate effort is an iterative problem. You start with the desired precision and Type I error rates and see what sample size would be necessary. If the cost is prohibitive you then complete power analyses to see how precise you can be within your budget for different error rates. Often there is a clear threshold where more effort no longer improves the precision of the estimate. This information helps to ensure you get the most cost-effective sampling program. If the precision and error rates you can afford are insufficient, then you may need to rethink the overall approach of your study. Appendix 3 summarizes important statistical sampling concepts.

3.2.6 Sampling frequency

Should sample collection take place daily, monthly, annually, every 10 years? The appropriate sampling frequency will differ by study objective and species. Is the study interested in a one-time estimate of status, or is the objective to monitor the species or community over time? In ecological studies, the temporal scale of interest is often annual (i.e., annual estimates of abundance, health, survival). These estimates are also combined into multi-year studies that allow for estimates of trends over time and comparisons between years. Several interesting and challenging statistical design questions may arise, depending on the temporal scale of interest.

Deciding when to sample within a year – When and how often should one conduct sampling within a year to obtain annual estimates? This depends on the life history of the species of interest and the target population. Generally speaking, the best time to sample the population is when the probability of detection is the greatest, because of behaviour, colour (i.e., mating coloration or vegetation coloration), time of day, or season.

Questions to consider:

- Does the species migrate or hibernate?
- Does the species' behaviour change with season? For example, does the species become more or less conspicuous during mating season?
- Is the species nocturnal or diurnal?
- Does the species only live for part of a year (plant or animal)?
- Will any major events within the year change the response variable of interest?
    - Hunting season
    - Significant overwinter mortality
    - Grazing, if interested in grass height
- If measuring an important habitat feature (e.g., cover availability) that may change seasonally, how will this affect your question of interest: during the best conditions, worst conditions, both, or only when the species of interest is present?
- What time period within the year do you wish to make inferences about?

Long-term monitoring programs – Sampling designs spanning multiple years are important when trying to consider changes to individuals or populations over time. Many conflicting opinions exist regarding how to implement sampling designs over many years. A major discussion point is whether to select permanent (long-term) or temporary sites. A typical response may be "use permanent sites for trend detection and temporary sites for status assessment," although real situations are rarely this simple. Each approach has pros and cons and researchers differ in their opinions of which approach is preferable. To provide some insight into this debate, a brief summary of the advantages and disadvantages to each of these approaches follows.

Temporary sites – Advantages

- Good procedure for locating rare items (McDonald 2003)
- Can use to estimate "net change" (McDonald 2003)
- Better strategy for estimating the average for all occasions (Cochrane 1977)
- Removes the cost of permanently marking sites (Elzinga et al. 2001)
- Can automatically account for changes in population composition (McDonald 2003)
- No conditioning impacts

Temporary sites – Disadvantages

- Lower efficiency method for net change (McDonald 2003)
- Does not allow estimation of individual change (McDonald 2003)

Permanent sites – Advantages

- Useful when a high degree of correlation is evident between sampling units from one period to the next, such as might occur with long-lived plants or long-lived and sedentary animals (Elzinga et al. 2001)
- Planning and survey design work are minimized (McDonald 2003)
- Well suited to estimate gross change and other components of individual change (McDonald 2003)
- Most powerful for detecting linear trend (Urquhart and Kincaid 1999)

Permanent sites – Disadvantages

- Possibility that, by chance, selected sites are the ones that change (a new sample each year would have a better chance of picking up true change, but the sample size needed in each time period is greater [Vesely et al. 2006])
- Only effective if the sampling effort is sufficient to represent the range of conditions across the area of inference (Vesely et al. 2006); this approach does not account for changes in population composition over time (McDonald 2003)
- Can be costly to permanently mark sites (Elzinga et al. 2001)
- Serious potential that site impacts will affect the measured variable through repeated years of sampling (Elzinga et al. 2001; Buckland et al. 1993; McDonald 2003)
- Permanent markers may provide a perch for raptors or songbirds
- Possible to trample vegetation
- Possible to disturb habitat or individuals

McDonald (2003) found that the term "trend" is used in different ways by different researchers, and perhaps this is the source of some controversy around "permanent versus temporary" sites. For example, at what scale is the change of interest: reach, stream, watershed, province? McDonald (2003) provided several useful definitions that can help to clarify the study objective and hence determine the best sampling frequency.

Net change: Measurement of total change in a parameter arising from all sources; can be a change in mean or total response. An individual change can happen without causing net change (as fish move from one stream segment to another), so individual stream segments could experience a trend, whereas the overall population of the watershed does not.

Individual change: Change experienced by an individual or particular member of the population. This is further divided into three categories:

1. *Gross change*: Change in response of a particular population unit (e.g., change in pH of a particular lake)
2. *Average gross change*: If all rivers in a collection of rivers have higher levels of suspended sediment (so the same change occurs to many individual units)
3. *Instability:* Variance of responses from individual population units

Another, recently popularized option (although this idea has been discussed by many; see Jessen 1942; Cochrane 1977) is to use a design with some combination of repeated visits and new sites. Such designs are referred to variously as "revisit," "repeated," "rotating," "split-panel," or "replacement." McDonald (2003) provided an excellent review of these sampling designs and developed consistent terminology. These designs provide a compromise between temporary and permanent sites and allow the estimation of both status and trends. McDonald (2003) found that "consensus opinion among the reviewed articles appeared to be that some sort of split-panel design had the best chance of satisfying the sometimes competing objectives inherent in many environmental monitoring projects." The following definitions, from McDonald (2003), are helpful.

Panel: A group of population units that are all sampled during the same sampling occasion or time period. Note that a population unit may be a member of more than one panel under this definition.

Revisit design: The rotation of sampling effort among survey panels; the pattern of visits to the panel; the plan by which population units are visited and sampled through time.

Membership design: The method used to populate the survey panels; the way in which units of the population become members of a panel.

Implementing and analyzing these designs is difficult. The analysis is complicated and consistent funding is required to avoid missing data. Although these designs are usually recommended, the specific frequency of return visits and the sample sizes required will depend on the specific study and the life history of the subject. Generally, if revealing trend is the primary study objective, then allocate more than 50% of sampling effort to the revisited sites; if both status and trends are important, then allocate the sampling effort equally to the revisited and new panels (McDonald 2003). More research is needed to address the optimal allocation of sampling effort among panels (McDonald 2003).

3.2.7 Field methods

Field methods are the actual methods used to collect data within a given sampling unit. Where possible, use existing established methods or protocols first (Vesely et al. 2006); several peer-reviewed and tested field protocols are available at the Resources Inventory Standards Committee website.[9]

Pollack et al. (2004) recommended that the design of field methods should minimize detection error. Questions related to detection include: Are field methods appropriate for the species? Are individuals hard to detect because of rarity or because the method is not efficient? Is the animal of interest "available" to the sampling method? Thompson (2004) described diverse examples where much of a population is not available for counting, such as:

- Aerial surveys of marine mammals, where some individuals are underwater;
- Surface counts of ants or other insects, where only the foraging component of the population is available;
- Salamanders or frogs, where a portion of the population is underground, underwater, or otherwise not visible;
- Bird point counts, where not all birds sing; and
- Terrestrial mammals in areas with dense vegetation (e.g., elk, moose).

Issues of imperfect detectability must be accounted for. Many available techniques tackle this issue, including: capture-recapture, which evaluates the probability of capture; distance sampling, which considers detection as a function of distance; and sightability models, which use observable covariates for factors that influence detectability.

Documentation of field methods is important for anyone who wishes to replicate the sampling. For example, document (or rationalize) variations from a standard method or protocol; during the study, record any changes made to a sampling method and ensure the changes are fully explained in the data analysis and any reports.

3.2.8 Planning the data analysis

Finally, one should consider *a priori* how to analyze and report the collected data. If study objectives are not clearly defined, then this task will be impossible. Use past research, pilot data, or expert opinion to describe the proposed approach. Once sampling is complete, describe the observed data for the approach taken (it may not exactly follow what was proposed). Although by no means an exhaustive list of possible analytical approaches, the following questions and examples provide an idea of planning the data analysis phase of the project.

9 Resources Inventory Standards Committee (RISC) website: http://www.ilmb.gov.bc.ca/risc/about.htm

What does the data look like?

- Are the data a direct measure or an index?
- Describe the response or index of interest (e.g., abundance, density, survival, diversity index, health index)

How are the data distributed?

- continuous (e.g., height, weight)
- categorical (e.g., colour, sex)
- strictly positive (e.g., count data)
- bounded between 0 and 100 (e.g., percent cover)
- binomial (e.g., each observation can only take one of two values)

What exactly is the statistic of interest?

- mean, median, maximum, minimum, rolling mean, range, variance

What questions are of interest?

- Does the response differ by year, group, or location?
- Does a relationship exist between environmental covariates and the response of interest?
- Does the response exceed a threshold or target?

What statistical tests are appropriate?

- Parametric tests, such as a *t*-test, assume a normal distribution of the data
- Non-parametric tests, such as the Mann-Whitney test, are appropriate when the data distribution is not obvious
- Regression analyses describe the relationship between environmental covariates and the response of interest
- Multivariate tests, such as the multi-response permutation procedure, are appropriate for community-level questions in which multiple responses are possible (i.e., the abundance of several species)

Determine effect size of interest and acceptable error rates (i.e., significance levels)

- Determine biologically meaningful effect size of interest
- Determine the significance level (α), or acceptable Type I error rate. Describe the consequences of a Type I error.
- Determine desired power ($1–\beta$), or the acceptable Type II error rate. Describe the potential consequences of a Type II error.

Check the assumptions

- Any analysis should include a discussion of the assumptions made about the data or the outcome of the protocol: what the assumptions were, whether they were met, and any weaknesses in the assumptions.

ACTION CHECKLIST . . .

Determine and describe the sampling design by:

- ✓ *Describing life history of the target species*
- ✓ *Defining the target population*
- ✓ *Choosing appropriate sampling unit(s)*
- ✓ *Determining how sampling unit(s) are chosen*
- ✓ *Determining sampling effort (sample size), frequency, and permanence*
- ✓ *Describing the details of the field methods*
- ✓ *Describing intended or proposed data analysis*

4.0 IMPLEMENT THE PROJECT

At this stage, you will have . . .

1. Determined where your project is positioned in the planning and design process—what has already been done; whether a protocol(s) already exists and whether pilot testing is needed;
2. Determined the need, and current knowledge, and developed the conceptual model and effectiveness indicators for monitoring (Section 2); and
3. Established a study design with a statistically rigorous sampling design appropriate to your project's position in the development process (Section 3).

The following tasks are important for implementing the project.

4.1 *Confirm Agency Support*

First, develop a project charter for approval by the appropriate supervisor or manager. The charter should indicate how the project is aligned with program priorities and include a title, objectives, project lead and team, an estimated budget or resources required, and an assessment of project's priority.

4.2 *Prepare a Project Work Plan*

Prepare a project work plan for each WHA or UWR effectiveness evaluation project before the project gets under way. Because effectiveness monitoring will continue over some period of time, this plan must include proposed activities, timelines, and deliverables for the duration of this period along with full details of the work planned for the first (or current) year. Also include equipment and training requirements, field logistics, and describe any extension products or communications.

Update work plans annually and report on project performance measures as the project progresses from pilot to full implementation, documenting obstacles and delays, results of pilot tests, and decisions such as amending data collection or analytical methods. Include annual updates from a stable project (i.e., one in which no amendments are made to methods) in an annual Progress Report.

4.3 *Data Management and Quality Assurance*

FREP has adopted a quality assurance framework[10] that establishes a process of checks and balances to ensure all efforts are made to meet data security and quality standards, from point of collection to final storage. Although data collected in wildlife resource value projects are not currently integrated into the FREP information management system, all wildlife resource value projects should adhere as much as possible to FREP's quality assurance and quality control commitments.

To meet the FREP commitment to quality management, wildlife resource value projects should adopt a process than ensures data collected is copied, transferred, and stored, such that:

- integrity of raw data is maintained;
- missing or erroneous data is addressed;
- data are entered, checked, archived, and analyzed on a regular annual cycle relevant to management, reporting, and further study; and
- details of all results from summaries and statistical analyses are available for review and archived.

5.0 REPORT

Knowledge gained from monitoring may range from procedural improvements to data collection and designs through to furthering our knowledge of the complexity of natural systems. The purpose of reporting is to share this knowledge with those responsible for making resource management decisions, those interested or affected by this knowledge, and the general public.

Reporting out evaluation results provides material for discussion and recommendations for changes to management practices and monitoring protocols. The wildlife resource value effectiveness evaluation program is one step in an adaptive management cycle, whereby decision makers are informed of new science, recommendations are made for any necessary changes to practices, and management is continually improved.

5.1 *Report Templates*

The FREP Communications Strategy [11] and the FREP Reporting Guidelines [12] provide overall direction for communications and standard layout and procedures for various types of FREP reports and publications.

Several types of reports could stem from effectiveness evaluation projects. Background and planning reports

10 See http://www.for.gov.bc.ca/HFP/frep/qmgmt/index.htm

11 http://www.for.gov.bc.ca/ftp/hfp/external/!publish/frep/archived/PM-FREP_Communications_Plan.pdf

12 http://www.for.gov.bc.ca/ftp/hfp/external/!publish/frep/qmgmt/QCProtocol4-ElectronicForms-Dec9-2008-MASTER-BLANKFORM.pdf

will come first. Results of a monitoring project in its early stages of development will include an assessment of the method(s) used to assess effectiveness, in addition to the actual evaluation the effectiveness of the WHA or UWR to protect habitat for a species. Final reports will summarize the effectiveness evaluation and make recommendations for improvements.

5.2 *Input to Decision Making and Resource Management*

It is very important that an effectiveness evaluation program and project results connect to resource management planning or practices. Providing clear documentation of the project's data and the evaluation information brought forward is essential because these results will ultimately guide decisions about resource management planning and practice.

Improvements to resource management may occur informally through decisions to change management practices (e.g., the size of a specific WHA) or formally through government decisions to amend policy or regulation as necessary (e.g., provincial policy that defines WHA size for a species). The presentation of project results to land managers, industry practitioners, and government staff and (or) executive will serve as the crucial "closing-the-loop" step in a cycle of adaptive management and continuous improvement.

APPENDIX 1. EXAMPLE STUDY DESIGNS

Based on definitions from Schwarz (2012), specific study designs are provided using Wildlife Habitat Area monitoring examples.

Descriptive Study

A WHA is selected and an indicator(s) (e.g., road density) is measured. The information collected is only relevant to the WHA sampled.

Observational Study

A non-designated area and a WHA are selected; selected indicators are measured in both. Comparisons between the two areas are possible, but the results are only applicable to the sampled areas. Using this approach, it is not possible to conclude whether any observed differences are representative of the differences between the areas. Descriptive and observational studies involve non-randomly selected sampling units; as a result, the information obtained is limited to the sites actually observed.

Analytical Survey

A random sample of WHAs in two or more categories is selected and road density is measured in each. An estimate of the mean road density with known precision is possible for each category. The estimates from the categories are comparable; however, it is possible that another unknown factor (besides WHA designation) is actually responsible for the difference.

Impact and Control-Impact Surveys

The goal of this approach is to assess the impact of some change, in this case the designation of a WHA. Various impact designs exist with increasing levels of effort and increasing degrees of inference. Mellina and Hinch (1995) provided a summary of different impact designs and described how each is used to assess watershed restoration. Schwarz (2012) and Underwood (1994) provided a good description along with examples for a range of impact studies, as well as an evaluation of respective strengths and weaknesses. The simplest impact studies look at a single location before and after some event. Obtaining multiple observations before and after an event improves the ability to determine whether an observed change is "real" by taking into account the natural year-to-year variability. Because obtaining "before" samples is often difficult, obtaining variance estimates by randomly sampling from similar but undisturbed habitats is a possible approach (Underwood 1994). This study design, termed a "Before-After-Control-Impact" (or BACI), considerably improves assessments by adding a control site with similar general characteristics to the treatment site (e.g., region, annual precipitation, size). Such designs are intended to examine whether a particular action results in a change at the treatment/impact site relative to the control site, while simultaneously adjusting for extraneous co-variables that may similarly affect both impact and control areas. In most cases, the use of controls greatly increases the power of detecting treatment–impact effects; however, poorly chosen control sites can decrease the power of detecting an effect (Korman and Higgins 1997; Roni et al. 2002). For example, a lack of randomization in assigning impact sites prevents us from inferring whether the impact will occur elsewhere. Alternatively, if only a single impact/

control pair exists, how do we know that the results are not just a consequence of the choice of sites?

Figure A1.1 illustrates the value of including a control site to assess the effects of an impact/treatment on populations that are naturally highly variable over time.

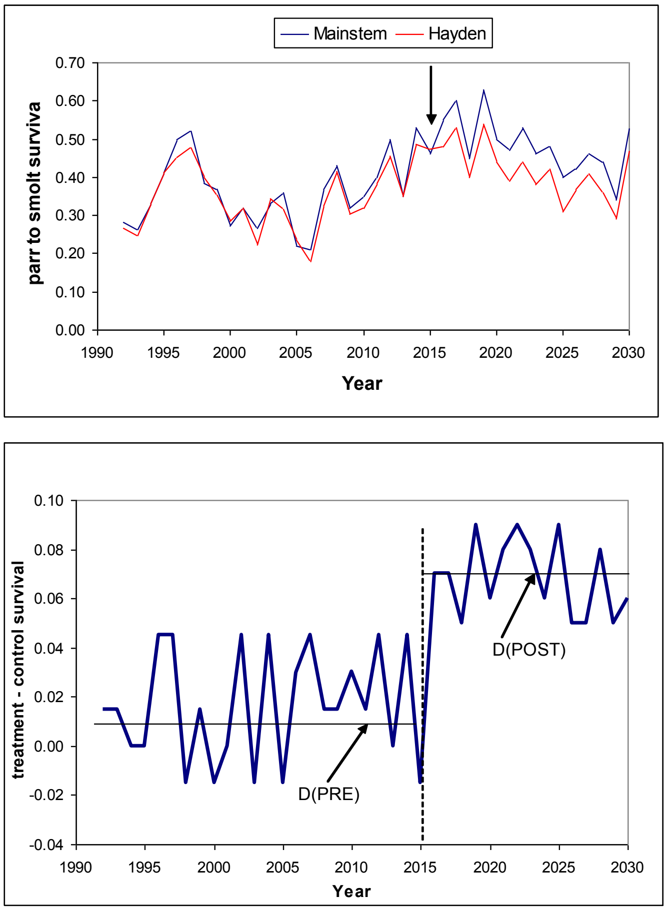


Figure A1.1. ***Value of a BACI design for inferences: (Top) Hypothetical time series of mainstem river and Hayden Creek (control area) parr-to-smolt survival rates, with habitat restoration actions assumed to happen simultaneously in 2015 (dark arrow) in the mainstem river. Time series for both impact and control streams track erratically over time. (Bottom) Hypothesized difference between mainstem river and Hayden Creek parr-to-smolt survival rates over the time series. D(PRE) and D(POST) represent average survival difference pre- and post-impact in 2015 in the mainstem river. Difference between control and impact streams shows much greater parr–smolt survival in the treatment stream versus the control, indicating a benefit of the restoration action over time that would not have been apparent given annual variation without a control system for comparison (source: Marmorek et al. [editors] 2006:77).***

In this example, the ability to detect improved salmon parr-to-smolt survival after a hypothetical habitat restoration treatment is not possible without a control (Hayden Creek) because of high variability in survival. Evaluation of the average annual *difference* between fish survival in the treatment versus the control stream, however, indicates that survival in the treatment stream is much greater relative to the control in the years subsequent to the restoration action.

Before-After-Control-Impact designs are of three types: (1) simple BACI, (2) BACIP (P for paired measurement through time), and (3) MBACI (multiple controls and, if possible, multiple impact sites; Downes et al. 2002). In simple BACI designs, measurements are made before an impact at control and treatment locations and then after the impact. However, BACIP designs, in which paired measurements are taken in both the impact and control sites at multiple random times before and after impact, are better able to avoid spurious results (Green 1979; Stewart-Oaten et al. 1986). By further extending the BACI design to include multiple controls and multiple impact sites, Keough and Mapstone (1995) created the MBACI design to address questions about the impacts of an action across broader scales. Multiple treatment and control locations are chosen randomly from a group of potential locations, thereby providing the means to extrapolate to a larger area. If it is not possible to randomly assign treatments and controls, but the same pattern is observed in multiple pairs, it is reasonable to assign a causal relationship (Schwarz 2012). The MBACI design compares a fixed period of time before the manipulation to (in ideal situations) a similar period of time after the manipulation.

Designed Experiments

In a designed experiment, the investigator has control over the treatment and can randomly assign experimental units to treatments. The degree of control the investigator has on a study affects the ability to show causation. The ability to make inference to other sampling units depends on random selection of samples or assignment of treatments.

Many good references on the subject of designing experiments are available, including: Schwarz's (2012) online course notes, which are geared to the environmental scientist and are probably the best place to start; Montgomery's (1997) solid introductory textbook, which is probably more than enough for most environmental studies; Box et al.'s (1978), traditional reference; and Wu and Hamada's (2000) more recent and extensive reference.

APPENDIX 2. EXAMPLES OF ADDITIONAL SAMPLING APPROACHES

Cluster Sampling

What: A random sample of clusters is selected and a census of all sampling units within the cluster is taken (Figure A2.1). For example, if we are interested in the average number of bicycles per household, we could select a sample of city blocks and then visit every household within the block. It is often relatively cheap to collect more information once at a site. This improves the estimate of the cluster (city block in this case) but may only have limited improvement to the overall precision. This is a special case of the two-stage sample in which the second stage is a census rather than a sample.

When: This type of strategy is useful when the cost of moving between sites is expensive relative to the cost of obtaining many more samples within a given site. If information about the cost of moving between clusters and preliminary estimates of variability within and between clusters is available, then we can determine whether this is a suitable approach.

Multi-stage Sampling

What: Similar to cluster sampling except that, rather than taking a census of all sampling units within the primary sampling unit, a secondary sample (e.g., simple random sample, systematic random sample, stratified random sample) is taken. For example, if a random sample of transects is taken from the population of possible transects, then in a multi-stage sample we might take a systematic sample every 10 m along the transect, thus creating secondary sampling units. In Figure A2.1, this would be illustrated by taking a sample of the squares within the three selected clusters (primary sampling units), rather than every small square (secondary sampling unit) within each cluster.

When: A multi-stage sampling strategy is often employed when the cost of measuring the units within the secondary sample is small compared to the cost of moving between the primary sampling units. One must be cautious to use the correct estimate of variance and therefore confidence intervals when using multi-stage estimates. The observations are not treated as a random sample from the target population but rather a random sample from the primary sampling unit (e.g., transect). Increasing the number of observations within each primary sampling unit will improve the precision of the estimate of the unit but will have limited affect on the precision of the overall estimate. Increasing the number of primary units is generally required to substantially improve the precision of the overall estimate. The optimum sample size for both primary and secondary sampling units is determined (as explained in Section 3.2.5) using preliminary estimates of the variability within and between primary units combined with the cost per effort of moving between these units versus sampling more units within each primary unit.

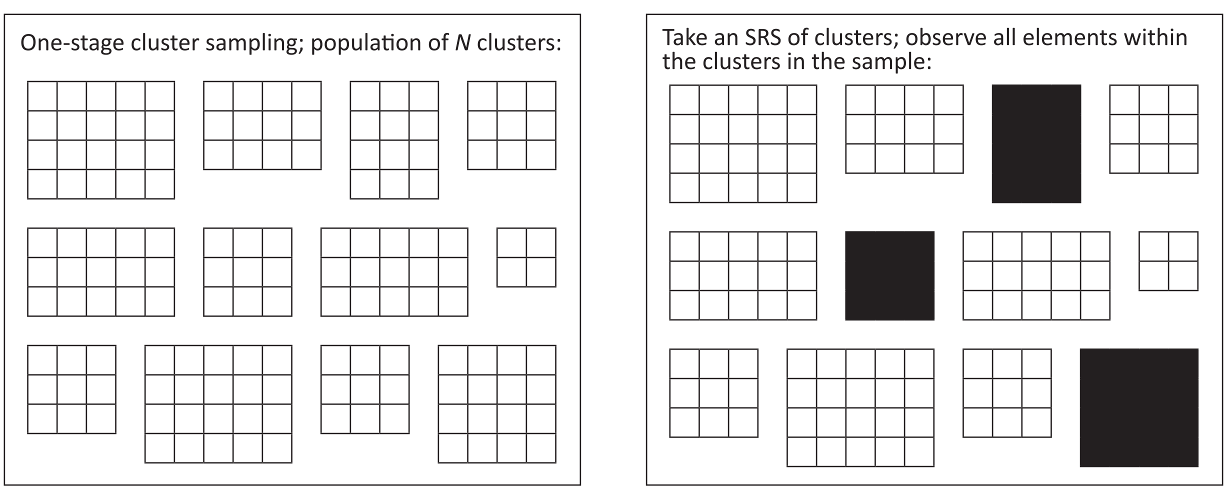


Figure A2.1. ***A cluster sample (source: Lohr 1999:133).***

Adaptive Sampling

What: In adaptive sampling, the sample units are not fully described before the study begins. This strategy begins like any other sampling design with a random selection of sampling units, but additional sampling units are added based on the observed values in the initial sample (Figure A2.2). This is particularly useful for clustered populations in which the presence in a given sampling unit may increase the likelihood of presence in neighbouring populations. Additional samples are added based on a pre-determined rule (i.e., if at least one whale is present in the sampling unit, then look in the four adjacent sampling units). The same (or a different) rule is then applied to the new sampling units. For rare clustered populations, this method increases efficiency as more individuals are likely encountered for the same overall number of samples, resulting in more precise estimates (Thompson and Seber 1996). As this strategy is more difficult to implement than a traditional sampling design, a statistician should be consulted for the design and analysis. MacDonald (2004) obtained mixed feedback on this approach in a survey of statisticians and biologists, finding that logistical difficulties with this approach may arise, particularly in surveys with large spatial scale.

When: Use for rare and clustered populations; examples include (Thompson and Seber 1996):

- Oceanic whale populations estimates, where whales tend to group in pods and cover a small fraction of the possible search area.
- Contagious diseases, such as a person tests positive for a rare disease, then test friends and family of people who have had recent contact with the person.
- Rare plant and animal species.
- Cases in which pollution levels are generally light but where a few hotspots exist.
- Shrimp trawl surveys, where the shrimp tend to cluster, but the cluster locations are not fixed because the shrimp are mobile.
- Deep sea fish (e.g., orange roughy) form large spawning aggregates, making it difficult to use traditional sampling methods to estimate abundance.

Ratio Estimation

What: When two quantities are measured on each sample unit. For example, if you measure yi = bushels of grain harvested from field i and xi = acreage of field i, and you are interested in B = average yield in bushels per acre, or the ratio of y:x. (Lohr 1999).

When/why: Ratio estimation is used in the following cases (Lohr 1999).

- If the ratio itself is of interest, which is a straightforward application.
- If *N* is unknown but is related to some other metric such as weight (i.e., total weight of fish in a net and average weight per fish).

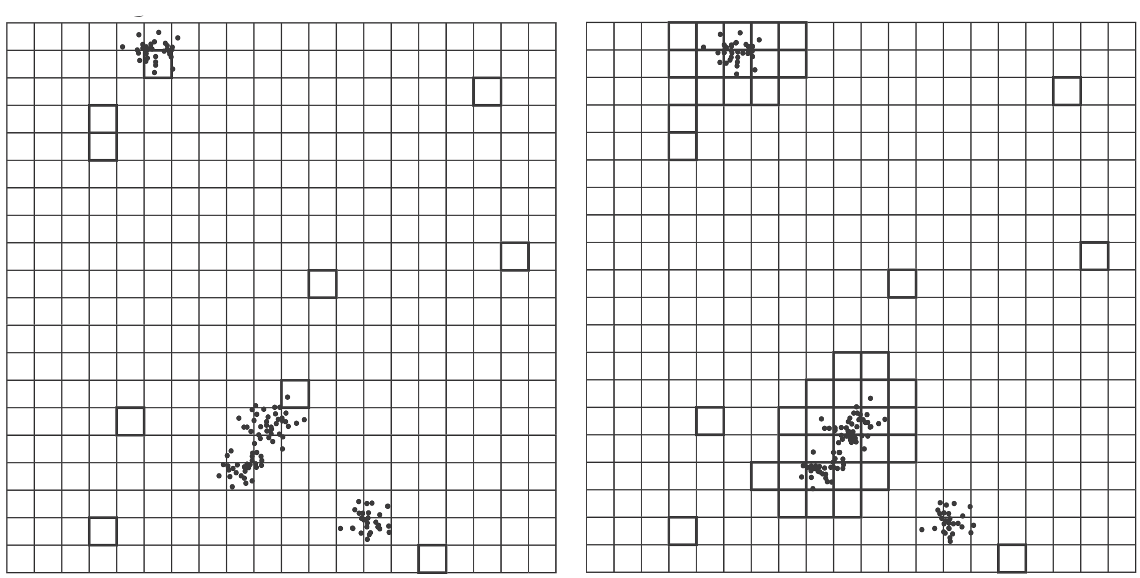


Figure A2.2. ***An example of an initial random sample of units on the left and the resulting adaptive cluster sample on the right (source: Thompson 1990).***

- To improve the precision of an estimate of the total or mean. For example, if you can collect two types of information (x and y) on the same sampling unit and x and y are correlated, then you may possibly improve the precision of the estimate using the relationship between x and y. Although such an estimate may contain bias, it still may reduce variance to a sufficient level. Lohr (1999) provided a good description of the conditions in which this method is suitable, showing how the precision and bias changes with sample size and the strength of the correlation between x and y. This method is essentially a special case of a regression model-based estimate.
- Adjust estimates from the sample based on ratios observed from different groups (such as gender) using "post-stratification" (see discussion below).
- Adjust for non-response (e.g., as sometimes occurs in a survey of businesses); in an environmental study, non-response may occur when access to a sampling unit is denied by a land owner or is not accessible for safety reasons.

Double Sampling Framework

What: In ratio estimation, we measure multiple variables on every observation; however, it is possibly easier or cheaper to collect auxiliary information than the actual variable of interest (Thompson 2002). The term "double sampling" has not been consistently used. Our definition follows Thompson (2002) and Pollock et al. (2002): an initial sample is taken and only the auxiliary information is observed, but then a smaller, second sample (often a subset) is taken where both the variable of interest and the auxiliary information are observed. The observed relationship between the two variables is used to improve the overall estimate for a given cost.

When: This strategy is suitable when the cost of measuring the variable of interest for a sufficient number of sites is prohibitive. This approach is also used when multiple approaches for estimating the variable of interest differ in cost and accuracy. Some examples include:

- Aerial versus ground counts
- "Eyeball" estimates of volume of trees in a stand versus felling trees and measuring
- Time-constrained versus area-constrained searches for tailed frogs
- Number of ponds versus waterfowl counts

Sequential Sampling

The general idea is that sampling continues until a pre-specified number of events are observed (Rao et al. 1981). It is also implemented by sampling until a particular precision level is reached (Elzinga et al. 2001). A first sample is used to determine how many more samples to take (Lohr 1999), and variance estimates need to account for this.

Network Sampling and Snowball Sampling

These strategies all follow a link-tracing design in which information about the links between sampling units is used (Thompson 2002). In network sampling, if one person is selected then everyone related to them (i.e., all siblings) is also sampled. These strategies are useful in hard-to-access and elusive populations, such as drug users or commercial sex workers. The idea is that members from a rare population know one another, so if you interview one homeless person you can then ask them to identify additional homeless people. These ideas are also used in Internet search engines.

Post-stratification

At the outset of an experiment, you often will not know what proportion of the sampling units belong to a particular group or stratum (males vs. females, grassland vs. forest) and, therefore, it is impossible to stratify the sample before implementing the study. Post-stratification occurs when the sample is grouped into strata after the study is complete and stratified estimates are obtained. A major danger to post-stratification involves "data snooping." If you choose your strata after seeing the data, you might obtain arbitrarily small variances (Lohr 1999). Holt and Smith (1979) described post-stratification as "a device for protecting the statistician's inference against those occasions when his randomization gives an unbalanced or unrepresentative sample." They noted, however, that it is possible to stratify to the point where the strata might contain a single individual. In addition, Royall (1968) noted that: "the statistician needs to consciously ignore some of the information in the data so that meaningful inference is possible." A safe compromise to prevent data snooping is to choose the groups that you will stratify on ahead of time but assign the observations to each group after you see the data. To use post-stratification, the relative size of each stratum must be known or estimated (Thompson 2002). The variance estimate is slightly different with post-stratification and approximations vary among texts. Thompson (2002) provided a brief summary and justification for the variance estimate he recommends.

Quota sampling is a variation, where the surveyor continues sampling until a pre-specified number is obtained from each stratum. If the samples are chosen at random, then this is equivalent to a stratified random sample (Cochran 1977); however, in practice this would require significant effort near the end of the study, as many of the individuals drawn would be from strata that already had their quota filled. As a result, quota samples are often drawn in some convenient, non-random manner. As an example, this type of sampling would be useful in estimating sea otter populations that occurred in both large groups and very small groups (Thompson 2004:30). However, if the samples are not obtained using probability sampling, then we cannot draw inference from these samples except in a model-based approach (Lohr 1999).

APPENDIX 3. IMPORTANT STATISTICAL SAMPLING CONCEPTS

Hypothesis Testing

Null hypothesis (H_0) – The null hypothesis refers to a default position or belief—it is the hypothesis that the study is trying to disprove. The null hypothesis is not rejected unless there is substantial evidence to support the alternative hypothesis. By formulating the problem this way, the null hypothesis is initially favoured, similar to the concept in law that the "accused is assumed innocent until proven guilty." The null hypothesis usually refers to the case where "there is no effect" (e.g., no difference between two groups; or the mean is equal to some value); however, this does not have to be the case. A null hypothesis is either "rejected" by statistical testing, or "not rejected"; a null hypothesis is never "accepted," as the failure to reject may simply be a result of insufficient data (see Table A3.1).

Alternative hypothesis (H_A) – This hypothesis represents the alternative claim, and usually refers to the case where "there is an effect" (e.g., group A is bigger than group B, or the mean is less [or not equal, or greater] than some value). The test is formulated so that there must be substantial evidence to support this alternative assertion (i.e., the burden of proof is on the alternative hypothesis).

The terms "alpha level" and "beta level" are often used to describe two possible data analysis errors:

1. **Type I error (α)** refers to the probability of detecting an effect (i.e., rejecting the null hypothesis) when an effect does not exist (false positive). This probability (α) is typically referred to as the significance level of the test.
2. **Type II error (β)** refers to the probability of not detecting an effect (i.e., failing to reject the null hypothesis) when an effect does exist (false negative).

Power ($1-\beta$) – Power is the probability of detecting an effect given that an effect does exist. To calculate the power, you must know what kind of statistical test is planned (e.g., one- vs. two-tailed *t*-test).

Effect size – The desired precision, or effect size of interest, refers to the effect size that is biologically meaningful. For example, you might be interested in a difference of 1 cm when comparing the length of male and female frogs; however, if comparing the length of male and female grizzly bears, then 1 cm is likely irrelevant but a difference of 10 cm might be relevant. If you try to detect a difference of 1 cm, then the required sample size will be far greater than if you try to detect a difference of 10 cm. Therefore, if you didn't really need that level of precision, you will have wasted a great deal of sampling effort.

In general, the less variability in the data, the more precise a difference you can detect. The power to detect an effect of a given size depends on the variability in the estimates (resulting from sampling/measurement error and

Table A3.1. ***Matrix showing four possible scenarios resulting from testing a hypothesis, where the null hypothesis is that there is "no effect." The conclusion is correct (green) in two cases and also incorrect (red) in two cases.***

		Conclusion after testing the hypothesis	
		Reject H_0 (conclude there is an effect)	**Fail to reject H_0 (conclude there is no evidence of an effect)**
Truth, which we are trying to evaluate	H_0 true (no effect)	Type I error (α)	$1-\alpha$
	H_0 false (there is an effect)	Power ($1-\beta$)	Type II error (β)

true differences in the population). Increasing sample size is one way to reduce the variability and produce a more precise estimate (e.g., a narrower confidence interval, for a given significance level, α). Here are some examples of effect sizes of interest for different questions:

- a confidence interval for stubble height of +/– 5 cm
- a change in occupancy of 50% between two periods
- a decline in abundance of 10% over 10 years

Determining Acceptable Error Rates

Determining acceptable error rates depends on the question of interest and the way in which the hypotheses are framed. The consequences of making each type of error must be evaluated for each question and study. Reducing Type I error, results in an increase in Type II error, and vice versa. The only way to improve both simultaneously is to increase the sample size.

The standard approach to designing your study is to determine the maximum Type I error (α) you are willing to accept, and then to make the Type II error as small as possible within this constraint (Devore 1995). Common α values are: 0.01, 0.05, and sometimes 0.1; whereas a minimum of 80% power is usually suggested. The more serious the consequence of Type I and II errors on humans or the environment, the smaller the α and β values you should choose.

Traditionally, Type I errors (or false positives) have been considered the most serious. For example, in medicine when testing a new drug, the cautious assumption is that it is not acceptable unless there is conclusive evidence otherwise. However, in environmental studies the consequences of making a Type II error may actually be more serious. For example, if trying to determine whether there has been a decline in a population over the past 10 years:

- H_0: There has been no decline in population *X* over the past 10 years
- H_A: There has been a decline in population *X* over the past 10 years
- Type I error: Conclude that there has been a decline when in fact there has not.
- Type II error: Conclude that there has not been a decline when in fact there has been.

You must consider the potential consequences of each error in the context of your hypotheses. For this example, the potential consequences of a Type II error are that: there are no changes in the status assessment, no habitat protections implemented, no change to direct any harvest restrictions, all of which may lead to an increased rate of decline or a failure to stop the rate of decline. In the long run, recovery may be more costly if the declines are not caught soon enough. Alternatively, if you make a Type I error, and implement management activities to protect species when they aren't actually declining, the consequences are temporary lost development costs or misplaced expenditure of protection funds. However, as more information is gained, these costs can be adjusted; whereas in the former case, the costs continue to increase as the population declines.

Alternatively, the hypothesis could be reworded to be more conservative (e.g., H_0: there has been a decline of 10% in population *X* over 10 years; H_A: there has been a decline of less than 10% in population *X* over 10 years), putting the burden of proof on those trying to prove there has not been a decline.

The consequences of both Type I and Type II errors should be considered and explicitly documented as part of the study design phase.

REFERENCES

Box, G. E. P., W. G. Hunter, and J. S. Hunter. 1978. Statistics for experimenters: an introduction to design, data analysis, and model building. John Wiley & Sons, Toronto, Ont.

Buckland, S. T., D. R. Anderson, K. P. Burnham, and J. L. Laake. 1993. Distance sampling: estimating abundance of biological populations. Chapman and Hall, London, U.K. http://www.colostate.edu/Dept/coopunit/download.html (Accessed January 2014).

Bunnell, F. 2005. Biodiversity monitoring guideline: terrestrial biological and physical monitoring and aquatic biological and physical monitoring. Forest Investment Account Activity Guidance Document. http://www.env.gov.bc.ca/fia/documents/monitoring_guidelines.pdf (Accessed January 2014).

Cochran, W. G. 1977. Sampling techniques. John Wiley & Sons, New York, N.Y.

Devore, J. L. 1995. Probability and statistics for engineering and the sciences. 4th edition. Brooks/Cole, Pacific Grove, Calif.

Doran, G. T. 1981. There's a S.M.A.R.T. way to write management's goals and objectives. Management Review 70(11):35–36.

Downes, B. J., L. A. Barmuta, P. G. Fairweather, D. P. Faith, M. J. Keough, P. S. Lake, B. D. Mapstone, and G. P. Quinn. 2002. Monitoring ecological impacts: concepts and practice in flowing waters. Cambridge University Press, Cambridge, U.K.

Eberhardt, L. L. and J. M. Thomas. 1991. Designing environmental field studies. Ecological Monographs 61(1):53–73. http://www.esajournals.org/doi/abs/10.2307/1942999 (Accessed January 2014).

Edwards, D. 1998. Issues and themes for natural resources trend and change detection. Ecological Applications 8(2):323–325. http://www.jstor.org/stable/2641071 (Accessed January 2014).

Elzinga, C. L., D. W. Salzer, J. W. Willoughby, and J. P. Gibbs. 2001. Monitoring plant and animal populations. Blackwell Science, Inc., Malden, Mass.

Environmental Protection Agency. 2006. Guidance on systematic planning using the data quality objectives process. Office of Environmental Information, Washington, D.C. EPA QA/G-4. http://www.epa.gov/quality/qs-docs/g4-final.pdf (Accessed January 2014).

Gordon, D. R., J. D. Parrish, D. W. Salzer, T. H. Tear, and B. Pace-Aldana. 2005. The Nature Conservancy's approach to measuring biodiversity status and the effectiveness of conservation strategies. In: Principles of conservation biology. 3rd Ed. M. J. Groom, G. K. Meffe, and C. R. Carroll (eds.). Sinauer Press, Sunderland, Mass. https://www.conservationgateway.org/Documents/Principles-in-Conservation-Chapter-03.pdf (Accessed January 2014).

Green, R. H. 1979. Sampling design and statistical methods for environmental biologist. Wiley Interscience, New York, N.Y.

Hankin, D. G. 1984. Multistage sampling designs in fisheries research: applications in small streams. Canadian Journal of Fisheries and Aquatic Science 41:1575–1591. http://www.nrcresearchpress.com/doi/pdf/10.1139/f84-196 (Accessed January 2014).

Holt, D. and T. M. F. Smith. 1979. Post stratification. Journal of the Royal Statistical Society: Series A (General) 142(1):33–46. http://www.jstor.org/stable/2344652 (Accessed January 2014).

Houde, I., F. L. Bunnell, and S. Leech. Assessing success at achieving biodiversity objectives in managed forests. BC Journal of Ecosystems and Management 6(2):17–28. http://jem.forrex.org/index.php/jem/article/download/311/230 (Accessed January 2014).

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54(2):187–211. http://www.jstor.org/stable/1942661 (Accessed January 2014).

Jessen, R. J. 1942. Statistical investigation of a sample survey for obtaining farm facts. Iowa Agricultural Experiment Station, Iowa State College, Ames, Iowa. Research Bulletin No. 304, pp. 54–59.

Keough, M. J. and B. D. Mapstone. 1995. Protocols for designing marine ecological monitoring programs associated with BEK mills. Commonwealth Scientific and Industrial Research Organization, Canberra, Australia. Technical Report No. 11.

Kinley, T. A. 2009. Effectiveness monitoring of badger wildlife habitat areas: summary of current areas and recommendations for developing and applying protocols. B.C. Ministry of Forests and Range, and B.C. Ministry of Environment, Forest and Range Evaluation Program, Victoria, B.C. FREP Report No. 25. http://www.for.gov.bc.ca/ftp/hfp/external/!publish/FREP/reports/FREP_Report_25.pdf (Accessed January 2014).

Korman, J. and P. S. Higgins. 1997. Utility of escapement time series data for monitoring the response of salmon populations to habitat alteration. Canadian Journal of Fisheries and Aquatic Science 54:2058–2097. http://www.nrcresearchpress.com/doi/pdf/10.1139/f97-115 (Accessed January 2014).

Krebs, C. J. 1989. Ecological methodology. Harper Collins Publishers Inc., New York, N.Y.

Lint, J., B. Noon, R. Anthony, E. Forsman, M. Raphael, M. Collopy, and E. Starkey. 1999. Northern spotted owl effectiveness monitoring plan for the Northwest Forest Plan. U.S. Department of Agriculture Forest Service, Pacific Northwest Research Station, Portland, Oreg. General Technical Report No. PNW-GTR-440. http://www.fs.fed.us/pnw/pubs/pnw_gtr440.pdf (Accessed January 2014).

Lohr, S. L. 1999. Sampling: design and analysis. Brooks/Cole Publishing Company, Pacific Grove, Calif.

Marcot, B. G., R. S. Holthausen, M. G. Raphael, M. M. Rowland, and M. J. Wisdom. 2001. Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. Forest Ecology and Management 153(1–3):29–42. http://www.sciencedirect.com/science/article/pii/S0378112701004522 (Accessed January 2014).

Marmorek, D. R., M. Porter, and D. Pickard (editors). 2006. Collaborative systemwide monitoring and evaluation project: Year 3, Annual Report. Prepared by ESSA Technologies Ltd., Vancouver, B.C. on behalf of the Columbia Basin Fish and Wildlife Authority, Portland, Oreg. Project No. 2003-036-00.

McDonald, L. L. 2004. Sampling rare populations. In: Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters. W. L. Thompson (ed.). Island Press, Washington, D.C. pp. 11–42.

McDonald, T. L. 2003. Review of Environmental Monitoring Methods: Survey Designs. Environmental Monitoring and Assessment 85(3):277–292.

Mellina, E. and S. G. Hinch. 1995. Overview of large-scale ecological experimental designs and recommendations for the British Columbia Watershed Restoration Program. B.C. Ministry of Environment, Lands and Parks and B.C. Ministry of Forests, Watershed Restoration Program, Victoria, B.C. Watershed Restoration Project Report No. 1. http://www.env.gov.bc.ca/wld/documents/wrp/wrpr_1.pdf (Accessed January 2014).

Meyer, P. J. 2003. What would you do if you knew you couldn't fail? Creating S.M.A.R.T. goals. In: Attitude is everything: if you want to succeed above and beyond. Meyer Resource Group, Inc., Waco, Tex.

Montgomery, D. C. 1997. Design and analysis of experiments. 4th Ed. John Wiley & Sons, New York, N.Y.

National Park Service. 2012. Guidance for designing an integrated monitoring program. U.S. Department of the Interior, National Park Service, Fort Collins, Colo. Natural Resource Report NPS/NRSS/NRR—2012/545. http://science.nature.nps.gov/im/monitor/docs/NPS_Monitoring_Design_Guidance.pdf (Accessed January 2014).

Newhouse, N. J., T. A. Kinley, C. Hoodicoff, and H. Page. 2007. Wildlife Habitat Area effectiveness evaluations: Protocol for monitoring the effectiveness of badger wildlife habitat areas. Version 1.4. B.C. Ministry of Forests and Range, Victoria, B.C. https://www.for.gov.bc.ca/ftp/hfp/external/!publish/frep/values/Wildlife_Badger_WHA_EE_Protocol_Mar2007.pdf (Accessed January 2014).

Noon, B. R. 2003. Conceptual issues in monitoring ecological resources. In: Monitoring ecosystems: interdisciplinary approaches for evaluating ecoregional initiatives. D. E. Busch and J. C. Trexler (eds.). Island Press, Washington, D.C. pp. 27–71.

Nyberg, B. 2010. Conceptual models for wildlife effectiveness evaluations: a recommended approach. B.C. Ministry of Forests and Range, Forest Practices Branch, Victoria, B.C. FREP No. 24. http://www.for.gov.bc.ca/ftp/hfp/external/!publish/FREP/reports/FREP_Report_24.pdf (Accessed January 2014).

Oakley, K. L., L. P. Thomas, and S. G. Fancy. 2003. Guidelines for long-term monitoring protocols. Wildlife Society Bulletin 31(4):1000–1003.

Pollack, K. H., H. Marsh, L. L. Bailey, G. L. Farnsworth, T. R. Simons, and M. W. Alldredge. 2004. Separating components of detection probability in abundance estimation: an overview with diverse examples. In: Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters. W. L. Thompson (ed.). Island Press, Washington, D.C. pp. 43–58.

Pollack, K. H., J. D. Nichols, T. R. Simons, G. L. Farnsworth, L. L. Bailey, and J. R. Sauer. 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. Environmetrics 13(2):105–119. http://onlinelibrary.wiley.com/doi/10.1002/env.514/abstract (Accessed January 2014).

Rao, P. V., K. M. Portier, and J. A. Ondrasik. 1981. Density estimation using line transect sampling. In: Estimating numbers of terrestrial birds. C. J. Ralph and J. M. Scott (editors). Studies in Avian Biology 6:441–445. https://sora.unm.edu/sites/default/files/journals/sab/sab_006.pdf (Accessed January 2014).

Roni, P., T. J. Beechie, R. E. Bilby, F. E. Leonetti, M. M. Pollock, and G. P. Pess. 2002. A review of stream restoration techniques and a hierarchical strategy for prioritizing restoration in the Pacific Northwest watersheds. North American Journal of Fisheries Management 22:1–20.

Royall, R. M. 1968. An old approach to finite population sampling theory. Journal of the American Statistical Association 63(324):1269–1279. http://www.jstor.org/stable/2285883 (Accessed January 2014).

Schueller, S. K., S. L. Yaffee, S. J. Higgs, K. Mogelgaard, and E. A. DeMattia. 2006. Evaluation sourcebook: measures of progress for ecosystem and community-based projects. Ecosystem Management Initiative University of Michigan, Ann Arbor, Mich. http://www.snre.umich.edu/ecomgt/evaluation/pdf/EMI_SOURCEBOOK_August_2006.pdf (Accessed January 2014).

Schwarz, C. J. 2012. Course notes for beginning and intermediate statistics. Simon Fraser University, Department of Statistics and Actuarial Science, Burnaby, B.C. http://people.stat.sfu.ca/~cschwarz/CourseNotes/ (Accessed January 2014).

Seber, G. A. F. 1986. A review of estimating animal abundance. Biometrics 42(2):267–292. http://www.jstor.org/stable/2531049 (Accessed January 2014).

Shaffer, T. L. and D. H. Johnson. 2008. Ways of learning: observational studies versus experiments. Journal of Wildlife Management 72(1):4–13. http://www.jstor.org/stable/25097498 (Accessed January 2014).

Skalski, J. R., K. E. Ryding, and J. J. Millspaugh. 2005. Wildlife demography: analysis of sex, age, and count data. Elsevier Academic Press, San Diego, Calif.

Stewart-Oaten, A., W. W. Murdoch, and K. R. Parker. 1986. Environmental impact assessment: “pseudoreplication” in time? Ecology 67(4):929–940. http://www.jstor.org/stable/1939815 (Accessed January 2014).

Thompson, S. K. 1990. Adaptive cluster sampling. Journal of the American Statistical Association 85(412):1050–1059. http://www.jstor.org/stable/2289601 (Accessed January 2014).

———. 2002. Sampling. 2nd Ed. John Wiley & Sons, New York, N.Y.

Thompson, S. K. and G. A. F. Seber. 1996. Adaptive sampling. John Wiley & Sons, Toronto, Ont.

Thompson, W. L. 2004. Sampling rare or elusive species: concepts, designs and techniques for estimating population parameters. Island Press, Washington D.C.

Underwood, A. J. 1994. On beyond BACI: sampling designs that might reliably detect environmental disturbances. Ecological Applications 4(1):3–15. http://www.jstor.org/stable/1942110 (Accessed January 2014).

Urquhart, N. S. and T. M. Kincaid. 1999. Designs for detecting trend from repeated surveys of ecological resources. Journal of Agricultural, Biological, and Environmental Statistics 4(4):404–414. http://www.jstor.org/stable/1400498 (Accessed January 2014).

Vesely, D., B. C. McComb, C. D. Vojta, L. H. Suring, J. Halaj, R. S. Holthausen, B. Zuckerberg, and P. M. Manley. 2006. Development of protocols to inventory or monitor wildlife, fish, or rare plants. U.S. Department of Agriculture Forest Service, Washington, D.C. General Technical Report WO-72. http://www.fs.fed.us/biology/wildecology/docs/SpProtocolTechGuide.pdf (Accessed January 2014).

Wieckowski, K., D. Pickard, M. Porter, D. Robinson, D. Marmorek, and C. Schwarz. 2008. A conceptual framework for monitoring Fisheries Sensitive Watersheds. Report prepared by ESSA Technologies Ltd. for B.C. Ministry of Environment, Victoria, B.C. http://essa.com/wp-content/uploads/2011/02/FSW-Conceptual-Report-Nov-5-2008.pdf (Accessed January 2014).

Wu, C. F. J. and M. Hamada. 2000. Experiments: planning, analysis, and parameter design optimization. John Wiley & Sons, Toronto, Ont.

Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space and time. Trends in Ecology and Evolution 16(8):446–453.