

# Design and Implementation of Monitoring Studies to Evaluate the Success of Ecological Restoration on Wildlife

William M. Block<sup>1,2</sup>

Alan B. Franklin<sup>3</sup>

James P. Ward, Jr.<sup>1</sup>

Joseph L. Ganey<sup>1</sup>

Gary C. White<sup>4</sup>

## Abstract

Restoration projects are often developed with little consideration for understanding their effects on wildlife. We contend, however, that monitoring treatment effects on wildlife should be an integral component of the design and execution of any management activity, including restoration. Thus, we provide a conceptual framework for the design and implementation of monitoring studies to understand the effects of restoration on wildlife. Our underlying premise is that effective monitoring hinges on an appropriate study design for unbiased and precise estimates of the response variables. We advocate using measures of population dynamics for response variables given that they provide the most direct measures of wildlife status and trends. The species to be monitored should be those

constituting an assemblage of umbrella species that represent the range of spatial and functional requirements of wildlife in a restored ecological system. Selection of umbrella species should be based on strong empirical evidence that justifies their usage. We also advocate that monitoring be designed as true experiments or quasi-experiments rather than as observational studies to allow for stronger inferences regarding the effects of restoration on wildlife. Our primary message is that if monitoring is to be done, it must be scientifically based.

**Key words:** wildlife, monitoring, status and trends, restoration effects, experiments, quasi-experiments, scale.

## Introduction

The ultimate goal of many ecological restoration projects is to return ecosystem structures, functions, and processes to "natural" or reference conditions. This is typically accomplished by manipulating vegetation and/or the physical environment to move the system towards pre-defined reference conditions that presumably existed at some point in the past. These manipulations alter habitat conditions in various ways for numerous species of wildlife, potentially affecting their population dynamics. A key and implicit assumption is that successful restoration will provide favorable conditions for the native biota. This assumption is rarely tested, but it should be.

The process of testing responses of wildlife to restoration falls under the umbrella of monitoring studies (cf., Thompson et al. 1998). Monitoring treatment effects on wildlife should be an integral component of the design and execution of any management activity, including restoration. This is especially true in the case of adaptive management, which relies heavily on feedback obtained through monitoring results to assess the success of management activities (Walters 1986; Gibbs et al. 1999). Unfortunately, monitoring is rarely done, and when it is done, it often suffers from poor design and lack of statistical rigor.

A complicating factor in designing monitoring programs is defining the appropriate variable(s) to measure wildlife response to restoration. This complication is not trivial. The term "wildlife" includes numerous species that represent diverse life histories. Should restoration focus on responses of the population dynamics of selected species, groups of functionally similar species such as guilds (*sensu* Root 1967), broad taxonomic groupings such as birds or mammals, or communities of organisms? Further, what benchmark should be used to measure the success or failure of restoration? Should this benchmark be a measure of habitats or populations,

<sup>1</sup>USDA Forest Service, Rocky Mountain Research Station, 2500 S. Pine Knoll Dr., Flagstaff, AZ 86001, U.S.A.

<sup>2</sup>Address correspondence to W. M. Block, Rocky Mountain Research Station, 2500 S. Pine Knoll Dr., Flagstaff, AZ 86001, U.S.A., email wblock@fs.fed.us

<sup>3</sup>Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, CO 80523, U.S.A.

<sup>4</sup>Department of Fishery and Wildlife, Colorado State University, Fort Collins, CO 80523, U.S.A.

or some surrogate measure such as indicator species, keystone species, or stressor variables (Noon et al. 1999)? Given that information on wildlife populations and community structure is generally unavailable for target restoration conditions, investigators are charged with developing objective and scientifically defensible criteria by which the success of restoration can be assessed. We know of no established methodology for establishing such benchmarks.

Here we outline some primary considerations for designing monitoring studies that will evaluate effects of restoration on wildlife. We begin by reviewing basic concepts of monitoring and then provide a conceptual framework to be considered for monitoring restoration effects on wildlife. As a frame of reference, we draw upon some ecological restoration projects occurring within southwestern *Pinus ponderosa* (ponderosa pine) forests to emphasize relevant points (USDA Forest Service 1998; USDI Grand Canyon National Park, Science Center, unpublished draft environmental assessment 1999). Our underlying premise is that sampling designs must be appropriate for unbiased and precise inferences about the target population. Failure to conduct monitoring correctly leads to erroneous conclusions and wasted resources.

### Basic Aspects of Monitoring

Monitoring is typically done to assess the change or trend in one or more resources. As such, it assesses the dynamics of the resource, not just its state. Thus, monitoring requires repeated sampling of the variable(s) of interest to measure change or trend. As with any ecological study, monitoring must be scaled to the variable and question being addressed (White & Walker 1997). If one is assessing the change in a species' habitat resulting from restoration treatments, monitoring must be scaled temporally to address how vegetation or other environmental features related to that habitat respond to the treatments over time. Also, because short-term wildlife population responses may differ from long-term responses, monitoring must be conducted over a sufficiently long time to ensure that the population has time to adjust to time-dependent changes resulting from restoration treatments. Monitoring should also be done over a long enough time to incorporate the range of environmental conditions allowing for valid estimates of process variation (USDI Fish and Wildlife Service 1995; White & Walker 1997; Seamans et al. 1999). For example, population trends measured during favorable weather conditions may not represent those occurring during unfavorable conditions, or those occurring within a longer period that includes both favorable and unfavorable weather.

In addition to the time scale, it is equally important to consider spatial scales when monitoring wildlife re-

sponse to restoration. The size of a restoration treatment must be large enough to provide space for placement of enough samples to detect any changes. For example, 8-ha treatment units such as those implemented in southwestern *Pinus ponderosa* restoration (USDA Forest Service 1998; USDI Grand Canyon National Park Science Center, unpublished draft environmental assessment 1999) are probably too small to include enough individuals of most wildlife species to understand effects of the treatment. Thus, both the treatment and the monitoring study design must be scaled to the species or community monitored.

Monitoring effects of restoration can be fundamentally different than monitoring typical applications. In typical applications, the monitoring design is structured to determine when a null hypothesis of no significant change has been rejected. Often the hypothesis is cast as a one-tailed test with some measure of population decline as the alternative hypothesis. With restoration monitoring, however, we might define the system as "restored" when the specified level of population decline is neither met or exceeded. In this case the goal of restoration monitoring is to fail to reject the null hypothesis. Here the statistical power of the monitoring design is probably more important than for typical monitoring because a poorly designed study will often fail to reject the hypothesis when in fact it should be rejected (that is, a type II error). For this reason, valid monitoring of a restoration project requires development of and adherence to a properly designed study.

### Types of Monitoring

Monitoring can be classified into four overlapping categories: implementation, effectiveness, validation, and compliance monitoring (Noss & Cooperrider 1994; Morrison & Marcot 1995). The two types particularly relevant to restoration are implementation and effectiveness monitoring. *Implementation monitoring* is used to assess whether or not a directed management action has been carried out as designed. For example, tree thinning is done to reduce stem densities and achieve a desired size-class distribution of trees as part of restoration of ponderosa pine forests. Implementation monitoring would evaluate whether the target densities and tree distributions were met immediately following thinning. In the context of restoration, implementation monitoring quantifies changes immediately after treatments, and evaluates whether treatments were done as prescribed. *Effectiveness monitoring* is used to determine whether the action achieved the ultimate objective. For example, was the ecosystem restored to reference conditions? Effectiveness monitoring requires response variables to be clearly articulated so that they can be measured accurately and precisely. Typical response variables

for wildlife are related to species' habitats or populations. A difficulty arises, however, because we rarely have accurate or detailed information on populations or habitats of individual species, or on wildlife community structure that existed during reference periods. As a result, we cannot conclude with certainty that species' populations or wildlife community structure have been restored. This uncertainty probably can never be overcome, so alternative approaches that measure population dynamics of selected species are needed as an index of successful restoration. These species could be a group of umbrellas species that collectively represent the spatial, ecological, and functional needs of species likely to occur in a restored ecological system (Lambeck 1997).

### Monitoring Steps

Monitoring involves a series of steps and feedback loops designed to answer some basic questions, namely: Why, what, when, where, and how to monitor? These steps include (1) setting monitoring goals, (2) identifying the resource(s) to monitor, (3) establishing a threshold or trigger point, (4) developing a sampling design, (5) collecting data, (6) analyzing the data, and (7) evaluating the results (Fig. 1). Once monitoring results are obtained and evaluated, the first question should be: Were monitoring goals met? That is, does monitoring demonstrate that restoration was effective with respect to wildlife, or should the restoration prescription be modified to strive for more acceptable results?

### Setting Monitoring Goals

The first step in designing monitoring for a restoration project is to clearly state the goals. In clarifying monitoring goals, a clear description of reference conditions should be stated, preferably in quantitative terms. Reference conditions can be defined along a gradient from pre-settlement conditions (for example, conditions existing prior to European settlement in North America of the area under consideration) to existing conditions. In terms of restoration, reference conditions are usually established between these two ends of the spectrum. The primary goal or objective of many restoration projects is to provide appropriate conditions for the native biota. Implied in this objective is that the wildlife community following restoration will resemble that occurring during reference conditions. This goal is broad and probably impossible to study, however, whereas monitoring goals should be feasible. These goals are perhaps best cast as testable research hypotheses, or as monitoring objectives to estimate appropriate parameters with specified precision (see Establishing Thresholds and Trigger Points).

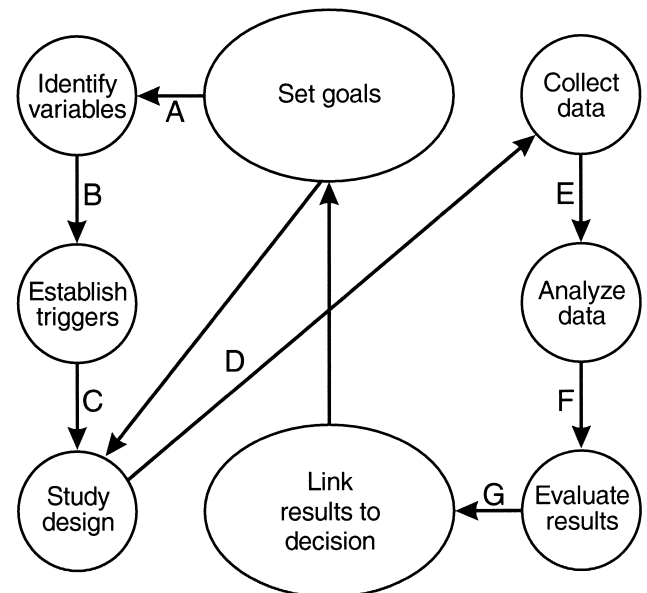


Figure 1. Flow diagram of steps involved with monitoring. Letters A through F signify feedback points when monitoring methods and results are evaluated.

Monitoring goals can be general or project specific. By project specific, we mean the effects on wildlife of a restoration treatment done on a particular piece of land during a specified time. Here a study would be designed to adequately sample that area, and the results would apply only to the place and time of study (cf., Green 1979; Hurlbert 1984). In contrast, were the objective to understand more broadly the effects of a specific treatment on wildlife across a wider area, a different sampling design would be needed. Of the two, project-specific restoration is probably most prevalent.

### Identifying Resources to Monitor

The next step is to decide what to monitor. Should the investigator monitor populations or habitats, or both, or perhaps some surrogate measure? Should the focus be single species, guilds, taxonomic assemblages, or entire communities? Measuring population response is perhaps the most direct way to understand effects of restoration, and wildlife science has developed the capability to monitor populations of many species (White et al. 1999). However, costs of monitoring single-species populations are not trivial (Verner 1983; USDI Fish and Wildlife Service 1995) and exceed the total budgets of many projects. Habitat is often monitored in lieu of populations because it costs less to sample. Unfortunately, we have limited knowledge of most species' relationship with their habitat, and especially lack information that links habitat conditions to population status. Without a clear understanding of a species' habitat and how it re-

lates to the population, inferences drawn from monitoring habitat are tenuous and probably inaccurate. In monitoring habitat alone, numerous other factors (such as disease, weather, predation, or events that occur on wintering grounds or along migration routes) can depress populations even when the habitat could support a larger population, further complicating our understanding of species-habitat relationships.

Surrogate measures such as indicator species, umbrella species, or guilds are often monitored rather than the populations or habitats of all individual species. Selection of indicator species involves three key steps. The first is to develop a conceptual model that outlines how the community is organized, including interrelations among system components (Fig. 2), and then to anticipate changes in pathways among components following restoration treatments (Noon et al. 1999). The second step identifies potential indicator species. Parameters monitored for indicator species should be (1) objective and quantitative, (2) accurately and precisely measured, (3) cost-effective, (4) used at the appropriate spatial and temporal scales, and (5) clearly linked to the parameter(s) that they indicate (Landres et al. 1988; Noon et al. 1999). The third step entails completing a pilot study to validate that the indicator meets the five criteria listed above. Caution is warranted when using indicator species, as the literature is replete with examples of their shortcomings (Mannan et al. 1984; Morrison 1986; Block et al. 1987; Landres et al. 1988). Thus, the use of indicators should not be simply a matter of convenience, but must be based on strong empirical evidence that supports their usage.

One possible use of surrogates would be to identify a set of indicator or umbrella species that would constitute a "restoration assemblage" (Lambeck 1997). The spatial and functional requirements of species in this assemblage should include those of all other species expected to be present in the restored ecological system. Populations of these species could be monitored to assess the success of restoration on wildlife. Selection of these species must follow the guidelines outlined above for indicator species.

Monitoring populations, habitats, or even indicators for all species in a community is impossible in most situations and may not provide the appropriate metrics for understanding community responses. Investigators often use various indirect measures to index community structure, such as species richness, composition, similarity, diversity, and evenness (Pielou 1977; Hayek 1994). By and large, these indices are ad hoc and lack theoretically valid sampling distributions. Further, biological interpretation of many of these indices, especially those that index species diversity, is difficult (Hayek 1994; Austin 1999). Community metrics may have some value, however, when one compares an area before and after restoration, or a treated area with an

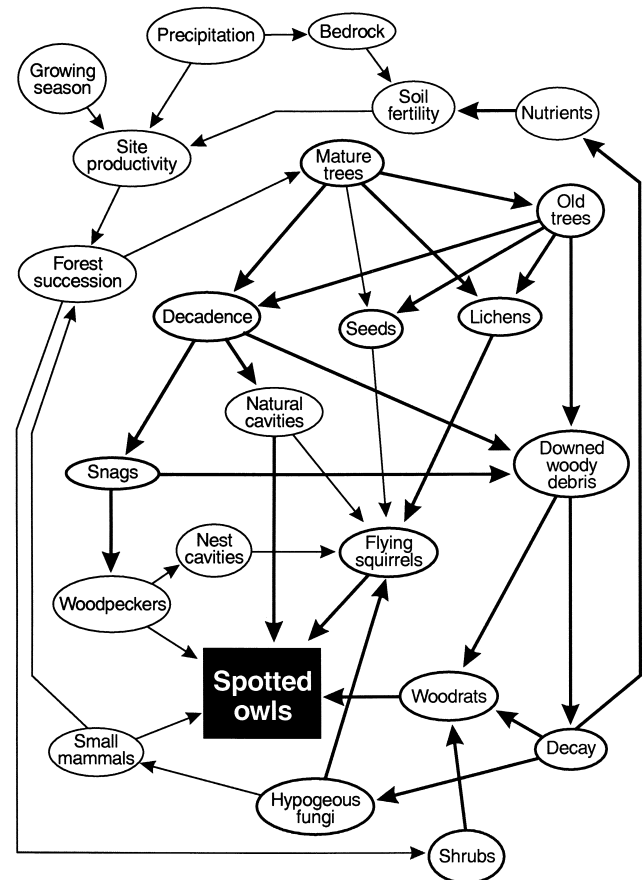


Figure 2. Conceptual model of the components and interrelationships of forested ecosystems inhabited by California Spotted Owls (from Verner et al. 1992).

untreated one. Unfortunately, we lack information for interpreting when these metrics indicate that the system has been restored.

We conclude that monitoring population dynamics is most appropriate for evaluating effects of restoration on wildlife for the following reasons. First, population response by species is the most direct measure. Second, documentation of strong relationships between habitat and population is lacking for most species. Third, the use of surrogate measures such as indicator species is fraught with numerous shortcomings. Finally, we know of no clear measure of wildlife community response to restoration.

#### Establishing Thresholds and Trigger Points

Monitoring is a critical step in adaptive management whereby decisions are made to continue, modify, or discontinue management actions depending on monitoring results. This is particularly relevant to restoration, which represents a relatively new management approach. Because restoration ecology is a developing

field, information detailing whether and how well treatments move ecological systems towards restored conditions is often unavailable. Consequently, threshold values or trigger points would provide periodic benchmarks to evaluate whether treatments are effective (points F and G in Fig. 1). Examples of benchmarks could be that the population size of a selected species exceeds or falls below a specified level, that certain species are present or presumed absent in an area following restoration treatments, that a certain percentage of a project area possesses specified habitat types (such as old-growth forest or gallery riparian forest), or that primary productivity meets or exceeds a minimal level.

Essentially, thresholds and trigger points are values towards which the restoration activity is striving to achieve or to avoid. When a threshold value is achieved (such as reaching a population size of 1,000 for a given species), it signals that conditions are appropriate to initiate the next phase of restoration or to continue on the present course. In contrast, when a trigger point is reached (such as falling below a population size of 1,000 for a given species), it signals the need to cease or alter restoration activities and possibly to mitigate their effects. Different targets or threshold values can be established for various stages during and after restoration. However, the parameter being estimated (for example, population size) during monitoring will always have some level of precision associated with its estimation. Therefore, establishing values for trigger points also requires that the parameter linked with the trigger point be measured at some predefined level of precision. This requires an a priori study design that considers the precision of the estimated parameters that will be used as the response variable.

### Monitoring Study Design

Accurate and precise information on wildlife responses to restoration requires that the study be designed to incorporate the effects of the treatment on selected wildlife populations as a primary objective. In the following, we briefly touch on some key points to consider when monitoring effects of restoration on wildlife.

Many of the considerations involved with proper design of any research study apply to monitoring studies. The foundation of any proper study design is application of the appropriate sampling design. The choice of sampling design depends on numerous factors including the biology of the monitored species, the species' spatial distribution, sampling variances, logistics, costs, and efficacy of field methodologies. Study designs can be broadly classed as either *experimental*, where one wants to test ideas about how things function (cause and effect), or *observational*, where one wants to accurately and precisely measure patterns (Manly 1992).

The key difference between the two types is that experimental studies, if properly designed, allow for inferences about cause and effect (e.g., Did the restoration result in a population increase for the focal species?) (James & McCullough 1995). Observational studies, including those based on probability (survey) sampling or just observations of some phenomena, do not allow for any inferences with respect to cause and effect. In the case of restoration, we want to know whether restoration (the treatment) had an effect on whatever response we are monitoring. Therefore, we discuss here only experimental study designs.

Experiments can be categorized as either *true (controlled) experiments* or *quasi-experiments* (Eberhardt & Thomas 1991; Manly 1992; James & McCullough 1995). The key differences between the two categories are outlined in Table 1. The ideal is the true experiment where treatments and controls are randomly assigned to experimental units and experimental units are replicated, because strong inference depends on a controlled experiment (Eberhardt & Thomas 1991). However, conducting a true experiment usually is not feasible in natural settings or under specific management conditions (Michener 1997). Therefore, an alternative is the quasi-experiment, which can have some of the properties of a true experiment though it lacks the randomization of treatments and controls (Table 1). By and large, restoration projects are planned. This allows restoration to be considered a treatment.

**Design of True Experiments.** The purpose of any experimental design is to provide the maximum amount of information relevant to the problem under investigation (Ostle 1983). A true experiment provides this in terms of monitoring restoration activities because it assesses not only whether change has occurred, but more importantly, whether that change is due to the restoration activity. There are three basic principles in designing an experiment: replication, randomization, and local control (Ostle 1983). Replication is the ability to repeat a treatment and is not merely repeated observation of the same units (Hurlbert 1984; Eberhardt & Thomas 1991); replication provides the estimate of experimental error. Randomization ensures that replicates are independent (assuming adequate spacing among units), thus making statistical comparisons valid by meeting assumptions of independence among experimental units. More importantly, random assignment of treatments and controls to experimental units helps ensure that treatments will not be affected by extraneous sources of variation over which there is no control by "averaging" out effects of those factors (Ostle 1983). Both randomization and replication provide a framework for estimating variability in experimental units that are treated alike (Eberhardt & Thomas 1991). Finally, local control refers to refine-

**Table 1.** Differences in the design aspects of true experiments and quasi-experiments.

Design Aspect	Type of Experimental Design	
	True Experiment	Quasi-experiment
Treatments	randomly allocated to experimental units	self-assigned to experimental units
Controls	randomly allocated to experimental units	randomly allocated to experimental units, self-assigned, or lacking
Confounding factors	controlled by design	not controlled by design
Cause and effect	directly inferable	not directly inferable
Inference	strong	weak
Potential design	randomized complete block; completely randomized; factorial treatments; split plot	nonequivalent controls; interrupted time series; before-after-control-impact

ments in experimental design, such as balancing blocking and grouping of experimental units to make the experimental design efficient and reduce the magnitude of the experimental error (Ostle 1983). Because many textbooks cover the design of true experiments, we will not attempt to cover this topic in detail here. Some potential designs for restoration monitoring experiments are shown in Table 1.

A key point is that inferences are strongest with the use of true experiments (Table 1). The further a study design deviates from a true experiment, the weaker its inferences become. However, a strictly experimental approach in large-scale studies is often difficult and expensive to achieve (Eberhardt & Thomas 1991; Michener 1997). Thus, quasi-experiments, observational studies, and modeling represent alternatives in restoration monitoring designs. Because the literature is more scattered in the design of quasi-experiments, we consider some potential designs in more detail here.

**Design of Quasi-Experiments.** Given prior knowledge of a restoration prescription, one could collect before implementing the treatment and an unrestored "control" site. This quasi-experimental design comes under the auspices of the general "Before-After-Control-Impact" (BACI) design commonly used for impact assessments (Green 1979; Stewart-Oaten et al. 1992). The control site should be as similar as possible to the restored site and should be as distant enough from the treated site to be regarded as an independent sample (that is, the chance is small that individuals sampled on the treated unit are also sampled on the control unit). Typically, subsamples are taken at both the treated and control sites before and after treatment. The two sites are regarded as the population to which inferences can be extended based on the subsamples.

Commonly, restoration projects are done on a site-specific basis with no replication. In this case, a restoration effect would be indicated by a significant interaction between the difference between the control and restored site before the treatment and that difference af-

ter the treatment (Underwood 1994, 1997). A problem with this design is that the changes may result from natural variation over time and not from the treatment. To partially account for this possibility, Underwood (1991) proposed taking measurements at multiple times before and after the treatment. Each measurement is assumed to be an independent, temporally replicated sample, and each replicate is regarded as an estimate of true change. The timing of taking repeated samples must be based on previous information that documents that they are not serially correlated and can be considered independent, or based on time-series data that are analyzed using a mixed-model repeated measures design that allows correlations to be partitioned from other sources of variation (Littel et al. 1996).

Another major shortcoming of the basic BACI design is that the lack of replication limits the inferences that can be drawn from the results (Green 1979; Hurlbert 1984). Results derived from a BACI study with no replication would apply only to the place and time of study. In the case of only one treated area, one way to strengthen inferences drawn from treatment effects would be to have multiple controls. Using multiple controls allows for a clearer interpretation of temporal variation, because they expand the scope of the sampled population and thus the scope of inference. That is, multiple controls allow one to distinguish temporal from spatial variability within the population, providing a perspective for evaluating variation on the treated site. If the variation among controls in both space and time is less than the variation within the restored site, then it is deduced that the change detected in the restored site is greater than expected based on just natural variability and results from the effects of the restoration treatment. Even with replicated controls, extrapolation of results to other locations may be limited. If an objective was to acquire results with broader application, it would be desirable to have multiple restored sites with multiple controls. This general design provides a certain amount of flexibility in the exact experimental design used. For example, treated and control sites could be paired, or

they could be randomly selected or stratified across the landscape. If this level of replication is feasible, then a true experimental design should be considered, as long as treatments and controls can be randomly allocated to sites.

Beyond the basic designs discussed above, investigators may choose or be relegated to using less optimal approaches. For example, if no control is available, the comparison is simply one of before and after treatment at the treated site(s), which is essentially a time-series design (Green 1979). However, under this design, natural temporal change cannot be separated from treatment effects because changes detected following treatment may be of the result of factors extraneous to the treatment.

**Other Design Considerations.** Once a general study design is established, numerous sampling aspects must be addressed. These include defining the target population, selecting sampling units or plots, and determining plot size and shape, field methodologies to use, and timing and frequency of data collection.

A major step in designing a monitoring study is to clearly establish the target population and sampling frame. Generally, defining the target population defines the sampling universe and the extent to which inferences can be extrapolated. In rare cases, the sampling frame is small enough to permit a complete sampling of the entire area. More typically, however, the entire sampling frame cannot be measured, so a sample of population units is required. Some primary considerations for establishing sampling units are (1) what constitutes an independent member of the population, (2) their size and shape, (3) the number needed, and (4) how to place them within the sampling frame.

The size and shape of sampling units depends on the methods used to collect data, biological edge effects, spatial distribution of the species studied, biology of the species, and logistics of collecting data. Thompson et al (1998:44–48) summarize the primary considerations and tradeoffs in choosing a plot design; for example, long and narrow plots may allow for more precise estimates, but square plots will have less edge effect. They conclude that no single design is optimal for all situations, and they suggest trying several in a pilot study. Plot size depends largely on the biology and distribution of the species under study. Larger plot sizes are needed for species with larger home-range sizes (such as *Strix occidentalis* [Spotted Owl]) and for species with clumped distributions (such as *Oreortyx pictus* [Mountain Quail]).

The number of sample units and allocation of those units depends on sampling variances, which are largely influenced by species' distribution and abundances. Sample size is largely guided by the number of plots needed to provide precise estimates of the parameter(s) of inter-

est. Sample plots should be allocated to minimize sampling variances. Species with clumped distributions generally require more samples to achieve a desired precision than species with regular distributions. This can be accomplished in various ways, depending on characteristics of the population to be sampled. Some basic sampling designs include simple random, systematic random, stratified, one-stage and two-stage cluster, and ratio estimators (Cochran 1977; Thompson 1992).

Other key parts of any study design include the timing of data collection and the length of time over which data should be collected. The choice of timing and length is influenced by the biology of the organism, the objectives of monitoring, intrinsic and extrinsic factors that influence the parameter(s) to be estimated, and the resources available to conduct monitoring. Obviously, studies of breeding animals should be conducted during the breeding season, studies of migrating animals during the migration period, and so on. Within a season, however, timing can be critical. For example, many passerines are more detectable during the early part of the breeding cycle when males are displaying as part of courtship and territory display. Thus, detection probabilities for many species will be greater during this period than during others. Another consideration is that the very population being monitored can change within a season. For example, age-class structures and numbers of individuals change during the breeding season as juveniles fledge from nests and are recruited to the population. This is typical of many temperate small mammal species, whose populations are low in late winter and early spring and peak in late summer and fall. As a result, population estimates for a given species may change depending on when data were collected. However, once it is decided when to collect data, it is crucial that data be collected during the same phase in the phenology of the species during subsequent years to control within-season variation.

Length of study refers to how long a study must be continued to estimate the parameter of interest. Effects of restoration treatments may occur at different temporal scales, from immediate wildlife responses to those that occur decades or even centuries post-treatment. Given time-varying effects, monitoring should occur over corresponding periods. A primary consideration for monitoring studies should be temporal qualities of the ecological parameter being measured. Temporal qualities include frequency, magnitude, and regularity, which are influenced by biotic and abiotic factors acting stochastically and deterministically (Franklin 1989). For example, proposed population monitoring for *Strix occidentalis lucida* (Mexican Spotted Owl) should occur for 15 years or longer, a period that would include a full generation and encompass varying environmental conditions (USDI Fish and Wildlife Service 1995).

Alternatives exist for conducting long-term studies (Michener 1997) such as retrospective studies (Davis 1989), substituting space for time (Pickett 1989), using systems with fast dynamics as analogies for those with slow dynamics (Strayer et al. 1986), and modeling (Shugart 1989). Each of these alternatives has merits and limitations that must be understood before they are used as an alternative to long-term data collection.

In summary optimal monitoring of restoration effects requires a rigorous study design. To ensure that reliable information can be obtained on wildlife responses, the restoration project should be designed and implemented with that in mind. Failure to do so may greatly reduce the ability to monitor and understand effects on wildlife.

### Analyzing the Data

If monitoring is designed correctly at the inception of planning a restoration project, the analysis is determined beforehand by the experimental design. The experimental design defines the sources of variation or factors to consider. Thus, specific analytical methods and/or statistical tests should be determined prior to data collection and should be appropriate to the experimental design.

We must note that the analysis of ecological data is moving away from traditional hypothesis testing and towards parameter estimation with associated measures of precision (Johnson 1999). Often, a "significant" hypothesis test indicates that a treatment effect exists, but not that a key threshold was met or exceeded, which is an important consideration for monitoring restoration effects. The magnitude of the response parameters, their levels of precision, and their relationship to established threshold levels are more important than simply whether a statistically significant effect exists (Fig. 3). To address this information need, hypothesis tests can be designed using the threshold value and its confidence interval as the reference point. An alternative to hypothesis testing is the use of model selection procedures for a given analysis, such as regression (Burnham & Anderson 1998). For example, linear and nonlinear regression in an analysis of covariance framework could be used to model the parameters in Figure 3, using model-selection criteria to determine which regression model best explains the data. Using this approach does not alter the experimental design, just the interpretation of the results stemming from the analysis dictated by that design.

Another consideration in analysis of the data is incorporation of detection probabilities. A detection probability is the chance of confirming the presence of an animal at a given place and time (Thompson et al. 1998). Numerous factors, including vegetation structure, be-

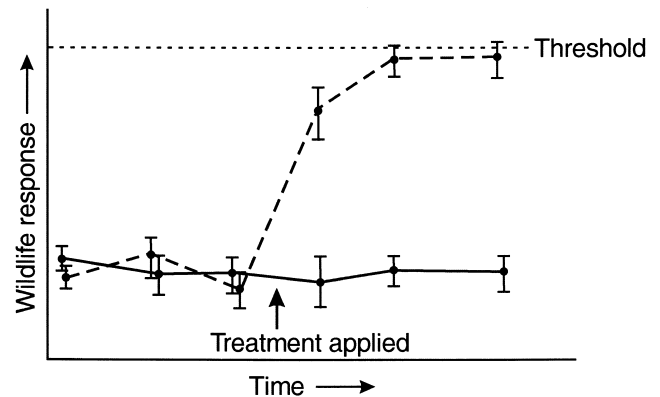


Figure 3. Example of data from a hypothetical BACI design with a single control site (solid line) and a single treatment site (dashed line) to illustrate use of parameter estimates and their confidence intervals to examine whether a pre-established threshold (dotted line) has been reached after a restoration treatment.

havior of individuals, and sampling methods can influence the probability of detecting an animal. The probability of detecting a bird, for example, might be greater in an open forest than in a dense one. If more birds are detected in an open than in a dense forest, one could mistakenly conclude that the open forest contained more birds when the difference between forests was in detectabilities and not population numbers. This is particularly relevant to ecological restoration where treatments often entail modifications to vegetation structure. Failure to adjust for heterogeneous detection probabilities adds a source of bias that may render the results of monitoring invalid.

### Evaluating the Results

If monitoring is conducted properly, results provide information that can be used to evaluate and adjust restoration practices. If results indicate that restoration treatments meet or exceed wildlife objectives, such treatments can be duplicated in future restoration projects. If effects are deleterious to the species or habitat(s) considered, then prescriptions should be modified to ameliorate or mitigate those impacts. These scenarios fall under the auspices of adaptive management (Gibbs et al. 1999). The concept of adaptive management rests largely on monitoring effects of land-management activities, in this case restoration, on key resources, and then using monitoring results as a basis for modifying those activities when warranted (Walters 1986; Gibbs et al. 1999). Adaptive management is an iterative process whereby management practices are initiated and their effects monitored at regular intervals. Effectively, restoration treatments are conducted incrementally and desired out-



comes are evaluated at each step. If outcomes are consistent with predictions, the project continues as designed. If outcomes deviate negatively from predictions, then restoration can continue, end, or be modified. These feedback loops assume that monitoring was correctly designed and that both treatments and monitoring efforts were implemented as designed. If not, the problem rests not with restoration and monitoring but with how they were implemented.

## Conclusions

### Current Situation

Restoration of degraded ecosystems is a relatively new management approach, and effects of most restoration treatments on various system processes and components are poorly understood. How restoration affects native wildlife is a primary question or concern in many areas. This question is addressed most efficiently when monitoring wildlife response is considered during the design phase of restoration. In reality, most restoration projects are not designed to consider wildlife primarily, so options for addressing these concerns are limited. Consequently, investigators often settle for suboptimal study designs for monitoring restoration effects on wildlife. These designs include observational studies, unreplicated comparisons of restored and natural sites, small study areas, spatial dependence between sampling units, and studies of short duration (Patten 1997; Kus 1998; Brown 1999). Certainly, past wildlife-restoration studies have provided useful information for the species studied, and they represent a solid foundation for the development of more rigorous monitoring approaches.

We are encouraged by some of the designs of recent restoration projects in northern Arizona in ponderosa pine forests within Grand Canyon National Park (USDI Park Service 1999) and Coconino National Forest (USDA Forest Service 1998). The general experimental designs consist of three blocks, each consisting of three treated units and an untreated control. Although the application of an experiment is a step forward, limitations still exist for understanding treatment effects on wildlife. For example, experimental units are relatively small (about 8 ha) and are placed in close proximity to each other. Also, blocks differ with respect to management histories, vegetation, topography, and edaphic features, and therefore cannot be considered valid replicates. The small size of treatment units, edge effects of adjoining treatments, and lack of replication will greatly limit inferences related to effects of the treatments on wildlife. Even so, understanding these limitations will provide a basis for improving the design of future projects.

### Future Directions

We contend that monitoring restoration activities should always involve experiments because researchers want to know not only if there was an effect after restoration, but also whether that effect was due to restoration. Given this, we also argue that whenever feasible restoration monitoring should be designed as a true experiment rather than a quasi-experiment.

Given that restoration is a relatively recent management approach, we have an opportunity to establish a monitoring framework that results in a logical flow of information. This approach should be standardized to enable broad-based questions to be addressed, yet remain flexible to allow for answers to more local, site-specific questions. Replicated projects could be implemented across an ecotype (such as ponderosa pine forest, valley oak woodland, shortgrass prairie) to understand effects on the appropriate restoration assemblage. Once general relationships are established, then more project-specific studies can be done to understand restoration effects on selected species such as threatened or endangered species, species of high societal value, or ecological keystone species, or on selected groups of species. Core study designs and field methods for collecting data on a limited set of variables could be standardized to permit pooling data from different projects across large geographic areas for future meta-analyses. Investigators would then have latitude to collect beyond the standardized base design by adding to the sampling design or collecting data on additional variables to address project-specific questions.

Still unresolved is deciding exactly what species to monitor. Some choices are politically driven, such as the mandate to monitor populations of threatened and endangered species or species that are highly valued by the public. These species may or may not be the best measures of the effectiveness of restoration for wildlife. Clearly a procedure is needed for selecting appropriate umbrella species to represent the range of ecological needs and population responses by species expected to be present in restored ecological systems. For some systems, these choices are readily apparent based on previous work; in other systems, research is needed to determine the appropriate species. Until we tackle this immediate and pressing need, we will lose opportunities to gain reliable information and move restoration ecology forward.

### Acknowledgments

We thank Michael Morrison for involving us in the wildlife session at the SER meetings in San Francisco. Elizabeth Ammon, Rudy King, and an anonymous reviewer provided comments on an earlier draft that greatly improved this paper.

## LITERATURE CITED

- Austin, M. P. 1999. A silent clash of paradigms: some inconsistencies in community ecology. *Oikos* **86**:170–178.
- Block, W. M., L. A. Brennan, and R. J. Gutiérrez. 1987. Evaluation of guild-indicator species for use in resource management. *Environmental Management* **11**:265–269.
- Brown, S. C. 1999. Vegetation similarity and avifaunal food value of restored and natural marshes in northern New York. *Restoration Ecology* **7**:56–68.
- Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: a practical information-theoretical approach. Springer-Verlag, New York.
- Cochran, W. G. 1977. Sampling techniques. 3rd edition. John Wiley and Sons, New York.
- Davis, M. B. 1989. Retrospective studies. Pages 71–89 in G. E. Likens, editor. Long-term studies in ecology: approaches and alternatives. Springer-Verlag, New York.
- Eberhardt, L. L., and J. M. Thomas. 1991. Designing environmental field studies. *Ecological Monographs* **61**:53–73.
- Franklin, J. F. 1989. Importance and justification of long-term studies in ecology. Pages 3–19 in G. E. Likens, editor. Long-term studies in ecology: approaches and alternatives. Springer-Verlag, New York.
- Gibbs, J. P., H. L. Snell, and C. E. Causton. 1999. Effective monitoring for adaptive wildlife management: lessons from the Galapagos Islands. *Journal of Wildlife Management* **63**:1055–1065.
- Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. John Wiley and Sons, New York.
- Hayek, L. C. 1994. Analysis of amphibian biodiversity data. Pages 207–269 in W. R. Heyer, A. Donnelly, R. W. McDiarmid, L. C. Hayek, and M. S. Foster, editors. Measuring and monitoring biological diversity: standard methods for amphibians. Smithsonian Press, Washington, D. C.
- Hurlbert, S. A. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**:187–211.
- James, F. C. and C. E. McCullough. 1995. The strength of inferences about causes of trends in populations. Pages 40–51 in T. E. Martin and D. M. Finch, editors. Ecology and management of neotropical migratory birds: a synthesis of critical issues. Oxford University Press, New York.
- Johnson, D. H. 1999. The insignificance of statistical significance testing. *Journal of Wildlife Management* **63**:763–772.
- Kus, B. E. 1998. Use of restored habitat by the endangered Least Bell's Vireo (*Vireo belli pusillus*). *Restoration Ecology* **6**:75–82.
- Lambeck, R. J. 1997. Focal species: a multi-species umbrella for nature conservation. *Conservation Biology* **11**:849–856.
- Landres, P. B., J. Verner, and J. W. Thomas. 1988. Ecological uses of vertebrate indicator species: a critique. *Conservation Biology* **2**:316–328.
- Littel, R. C., G. A. Milliken, W. W. Stroup, and R. D. Wolfinger. 1996. SAS system for mixed models. SAS Institute, Cary, North Carolina.
- Manly, B. J. F. 1992. The design and analysis of research studies. Cambridge University Press, Cambridge, United Kingdom.
- Mannan, R. W., M. L. Morrison, and E. C. Meslow. 1984. Comment: the use of guilds in forest bird management. *Wildlife Society Bulletin* **12**:426–430.
- Michener, W. K. 1997. Quantitatively evaluating restoration experiments: research design, statistical analysis, and data management considerations. *Restoration Ecology* **5**:324–377.
- Morrison, M. L. 1986. Birds as indicators of environmental change. *Current Ornithology* **3**:429–451.
- Morrison, M. L., and B. G. Marcot. 1995. An evaluation of resource inventory and monitoring program used in national forest planning. *Environmental Management* **19**:147–156.
- Noon, B. R., T. A. Spies, and M. G. Raphael. 1999. Conceptual basis for designing an effectiveness monitoring program. Pages 21–48 in B. S. Mulder, B. R. Noon, T. A. Spies, M. G. Raphael, C. J. Palmer, A. R. Olsen, G. H. Reeves, and H. H. Welsh, technical coordinators. The strategy and design of the effectiveness monitoring program in the Northwest Forest Plan. General Technical Report PNW-437. U.S. Forest Service, Portland, Oregon.
- Noss, R. F., and A. Y. Cooperrider. 1994. Saving nature's legacy: protecting and restoring biodiversity. Island Press, Washington, D.C.
- Ostle, B. 1983. Statistics in research. Iowa State University Press, Ames.
- Patten, M. A. 1997. Reestablishment of a rodent community in restored desert scrub. *Restoration Ecology* **5**:156–161.
- Pickett, S. T. A. 1989. Space-for-time substitutions as an alternative to long-term studies. Pages 110–135 in G. E. Likens, editor. Long-term studies in ecology: approaches and alternatives. Springer-Verlag, New York.
- Pielou, E. C. 1977. Mathematical ecology. John Wiley and Sons, New York.
- Root, R. B. 1967. The niche exploitation pattern of the Blue-gray Gnatcatcher. *Ecological Monographs* **37**:317–350.
- Seamans, M. E., R. J. Gutiérrez, C. A. May, and M. Z. Peery. 1999. Demography of two Mexican Spotted Owl populations. *Conservation Biology* **13**:744–754.
- Shugart, H. H. 1989. The role of ecological models in long-term ecological studies. Pages 90–109 in G. E. Likens, editor. Long-term studies in ecology: approaches and alternatives. Springer-Verlag, New York.
- Stewart-Oaten, A., J. R. Bence, and C. W. Osenberg. 1992. Assessing effects of unreplicated perturbations: no simple solutions. *Ecology* **73**:1396–1404.
- Strayer, D., J. S. Glitzenstein, C. G. Jones, J. Kolas, G. Likens, M. J. McDonnell, G. G. Parker, and S. T. A. Pickett. 1986. Long-term ecological studies: an illustrated account of their design, operation, and importance to ecology. Occasional Paper 2. Institute for Ecosystem Studies, Millbrook, New York.
- Thompson, S. K. 1992. Sampling. John Wiley and Sons, New York.
- Thompson, W. L., G. C. White, and C. Gowan. 1998. Monitoring vertebrate populations. Academic Press, San Diego, California.
- Underwood, A. J. 1991. Beyond BACI: experimental designs for detecting environmental impacts on temporal variations in natural populations. *Australian Journal of Marine and Freshwater Research* **52**:569–587.
- Underwood, A. J. 1994. On beyond BACI: sampling designs that might reliably detect environmental disturbances. *Ecological Applications* **4**:3–15.
- Underwood, A. J. 1997. Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge University Press, Cambridge, United Kingdom.
- USDA Forest Service. 1998. Fort Valley environmental assessment. Unpublished report. Peaks Ranger District, Coconino National Forest, Flagstaff, Arizona.
- USDI Fish and Wildlife Service. 1995. Recovery plan for the Mexican Spotted Owl (*Strix occidentalis lucida*). Vol. 1. U.S. Fish and Wildlife Service, Albuquerque, New Mexico.
- USDI Park Service. 1999. Draft environmental assessment Grand Canyon forest restoration research. Unpublished report. Grand Canyon National Park, Science Center, Grand Canyon, Arizona.
- Verner, J. 1983. An integrated system for monitoring wildlife on the Sierra Nevada Forest. *Transactions of the North American Wildlife and Natural Resources Conference* **48**:355–366.
- Verner, J., R. J. Gutiérrez, and G. I. Gould, Jr. 1992. The California

- Spotted Owl: general biology and ecological relations. Pages 55–77 in J. Verner, K. S. McKelvey, B. R. Noon, R. J. Gutiérrez, G. I. Gould, Jr., and T. W. Beck, technical coordinators. The California Spotted Owl: a technical assessment of its current status. USDA Forest Service, General Technical Report, PSW-GTR-133. Pacific Southwest Research Station, Albany, California.
- Walters, C. J. 1986. Adaptive management of renewable resources. Macmillan, New York.
- White, G. C., W. M. Block, J. L. Ganey, W. H. Moir, J. P. Ward, Jr., A. B. Franklin, S. L. Spangle, S. E. Rinkevich, J. R. Vahle, F. P. Howe, and J. L. Dick, Jr. 1999. Science versus political reality in delisting criteria for a threatened species: the Mexican Spotted Owl experience. Transactions of the North American Wildlife and Natural Resources Conference **64**:292–306.
- White, P. S., and J. L. Walker. 1997. Approximating nature's variation: selecting and using reference information in restoration ecology. Restoration Ecology **5**:338–349.