RESEARCH AND EXPERIMENTAL DESIGN

Edward O. Garton, John T. Ratti, and John H. Giudice

INTRODUCTION

        Management programs must be based on quality scientific investigations that produce objective, relevant information – and quality science is dependent upon carefully designed experiments, comparisons, and models. This chapter provides an overview of the fundamental concepts of wildlife research and study design, and is a revision of Ratti and Garton (1994).

Emergence of Rigor in Wildlife Science

        Wildlife science is a term the wildlife profession has only recently nurtured. Our profession of wildlife conservation and management was built on natural-history observations and conclusions from associations of wildlife population changes with environmental factors such as weather, habitat loss, or harvest. Thus, we have a long tradition of wildlife management based on "laws of association" rather than on experimental tests of specific hypotheses (Romesburg 1981). Although Romesburg (1981) and others have been critical of wildlife research and resulting management, the wildlife biologist is confronted with tremendous natural variation that might confound results and conclusions of an investigation. Scientists conducting experiments in physics and chemistry have the ability to control variables associated with an experiment, isolating the key components, and can repeat these experiments under the exact same conditions to confirm their results. They also have the ability to systematically alter the nature or level of specific variables

to examine cause and effect.

The wildlife scientist often conducts investigations in natural environments over large geographic areas. It is usually difficult to control potentially causal factors. Responses, such as density of the species in question, are simultaneously subject to the influences of factors such as weather, habitat, predators, and competition, which change spatially and temporally. Thus, rigorous scientific investigation in wildlife ecology is challenging and requires careful design. An early step that moved wildlife investigators from descriptive natural-history studies toward rigorous examination of processes was application of statistics to wildlife data. In the past 30 years the profession has evolved from testing differences among sample means with simple *t-tests* to complex computer models and multivariate analyses. This transition continues as important, new methods emerge.

Scientific inquiry involves a systematic series of steps; wildlife research prior to the last 2 decades had taken only the initial steps in the process. Realization of this problem became more prominent after Romesburg (1981:293) recommended that we move from "unreliable knowledge" to management based on a series of carefully tested hypotheses and conclusions based on sound scientific inquiry.

Experimental vs. Descriptive Research

Most wildlife research prior to 1985 was descriptive. Experimental research is the most powerful tool for identifying causes and should be used more in wildlife studies. However, descriptive natural-history studies, field studies, and carefully designed comparisons based on probability sampling continue to be useful. Descriptive research is an essential, initial phase of wildlife science, and it can produce answers to important questions.

Descriptive research often involves broad objectives rather than tests of specific hypotheses. For example, we might have a goal to describe and analyze gray partridge (*Perdix perdix*) breeding ecology. Thus, we might measure characteristics of nesting habitat, clutch size, hatching success, brood use of habitat, food habits of chicks and adult hens, and mortality due to weather events and predators. From this information, we can learn details of partridge biology that will help us understand and manage the species. If we observe that 90% of gray partridge nests are in habitat "A," 10% in habitat "B," with none in "C" and "D," we are tempted to manage for habitat "A" to increase nesting density. However, many alternatives must be investigated. Possibly habitat "A" is the best available habitat, but partridge experience high nest mortality in this type. Maybe habitat "X" would be the best habitat for nesting, but it is not available on our study area. What habitat types do gray partridge in other regions use? How does nest success and predation differ among regions and habitats? With answers to these questions we can begin to see that defining quality nesting habitat is complex. Nest success may be related not only

to a specific habitat type, but also to the spatial and proportional distribution of habitat types, species of predators present, partridge density, and climatic conditions.

Combining descriptive studies with other studies published in the scientific literature should provide sufficient information to develop a research hypothesis (theory or conceptual model, Fig. 1) that attempts to explain the relationship between habitat and nesting success of partridge. Such descriptive models are general, but can help define specific predictions to be tested to examine validity of the model. These predictions can be stated as hypotheses. We can test hypotheses by gathering more descriptive observations or by conducting an experiment (Fig. 1) in which manipulated treatments are compared with controls (no treatment) to measure magnitude of change (+ or -) resulting from experimental treatments. An alternative approach based on building predictive models is often easier and more feasible than an experimental approach to hypothesis testing. However, it does not provide as firm a basis for drawing conclusions concerning causes of observed differences. Random assignment of plots to treatment and control groups dramatically increases our certainty that measured differences are due to treatment effects rather than some ancillary factor.

Consider again the partridge study and assume we developed a theory (Fig. 1) that partridge adapted to be most successful at nesting in areas resembling their native habitat in Eurasia with its natural complement of predators, food sources, and vegetation cover. From this theory we predict that partridge nesting success in grasslands in North America would be highest in undisturbed native prairie, resembling native, Eurasian partridge habitat and least successful in highly modified agricultural monocultures of corn, wheat, etc. We then formulate the hypothesis that gray partridge nesting density and nest
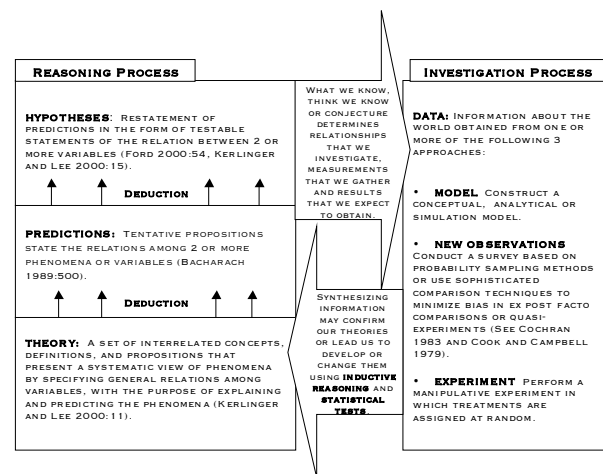

Fig. 1. Circular nature of the scientific method where data are synthesized inductively to develop theories which form the basis for deductively derived predictions and hypotheses that can be tested empirically by gathering new data with experiments, new observations or models (modified from Ford 2000:6).

success are higher in areas dominated (e.g., >75% of the available habitat) by pasture than in areas dominated by cultivated fields. The strongest test of this hypothesis we could perform would involve an experiment (Fig. 1) for which we must establish a series of control and experimental study plots. Our study plots would be randomly chosen from large blocks of land where agricultural practices have not changed in recent years, which contain the types of agricultural practices common to the region where we want to apply our findings. Some of these study plots (commonly half of them) will be randomly selected to act as control plots and will not change throughout the duration of the study. On the experimental plots (the remaining randomly selected plots within the same region as our control plots), cultivated fields will be planted to pasture grass to test the validity of our hypothesis and predictions regarding the effect of habitat on partridge nesting.

This process is difficult, because it requires large blocks of habitat, cooperation from landowners, several years to establish pasture grass on the experimental plots, and additional years of study to measure the response of birds to vegetative changes. The comparison between control and experimental plots will provide a basis to reject the null hypothesis of no effect so we can draw a firm conclusion that increasing cover of pasture or conservation reserve program (CRP) habitat in agricultural areas will increase nesting density and success of gray partridge. In addition, we should be able to estimate the magnitude of effects from management efforts directed at partridge production. If we fail to reject the null hypothesis we cannot draw a firm conclusion since it might be due to insufficient sample size.

Some questions concerning wildlife biology are not amenable to experimentation (e.g., effects of weather on populations, or differences in survival rates between gender or age classes). Other potential treatment effects are too expensive or difficult to accomplish. Some treatments may require substantial efforts to convince interested publics of the value of applying them in any single treatment area. Finally, the need to evaluate effects of many habitat or population factors simultaneously may preclude experimentation. In such cases, an alternative to hypothesis testing is modeling. Information theoretic methods provide powerful tools to evaluate potential models (Burnham and Anderson 2002). Incorporating modeling into the management process is a powerful strategy for predicting consequences of management actions while simultaneously learning about key processes affecting wildlife populations and their habitats (Walters 1986). A key requirement for this process to be successful is the need to monitor consequences of management actions through an adaptive management process (Walters 1986). This adaptive learning process can be facilitated by application of Bayesian statistics, which use additional observations to improve estimates of key relationships assumed prior to the management action (Gelman et al. 1995).

SCIENTIFIC METHOD

One of the early papers published on the scientific method in *Science* in 1890 (Chamberlin 1965) emphasized the need to examine "multiple working hypotheses" to explain an observation. This method is commonly referred to as the hypothetico-deductive method and was formalized in classic contributions by Popper (1959, 1968). The method is a circular process in which previous information is synthesized into a theory, predictions are deduced from the theory, the predictions are stated explicitly in the form of hypotheses, hypotheses are tested through an investigation involving experimentation, observation, or quantitative models, the theory is supported, modified, or expanded on the basis of the results of these tests, and the process starts again (Fig. 1). Platt (1964) re-emphasized the importance of multiple competing hypotheses and proposed a systematic pattern of inquiry, referred to as strong inference, in which the investigator devises alternate hypotheses, develops an experimental design to reject as many hypotheses as possible, conducts the experiment to achieve unambiguous results, and repeats the procedure on the remaining hypotheses. Other major works that provide detailed discussions of the scientific method include Dewey (1938), Bunge (1967), Newton-Smith (1981), Ford (2000), and Gauch (2003).

The most successful applications of the hypothetico-deductive method have been in physics, chemistry, and molecular biology where experiments can isolate the results from all but a small number of potentially causal factors. The classic methods of natural history observation in wildlife biology and other natural sciences have expanded to include experimentation, hypothesis testing, and quantitative modeling. James and McCulloch (1985:1) described this transition for avian biologists: "traditional ornithologists accumulated facts but did not make generalizations or formulate causal hypotheses . . .modern ornithologists formulate hypotheses, make predictions, check the predictions with new data sets, perform experiments, and do statistical tests." This statement is equally applicable to wildlife research. In addition to James and McCulloch (1985), other excellent reviews include Romesburg (1981), Quinn and Dunham (1983), Diamond (1986), Eberhardt and Thomas (1991), Murphy and Noon (1991), Sinclair (1991), Boitani and Fuller (2000), Morrison et al. (2001), and Williams et al. (2001).

The first step in the scientific method is a clear statement of the problem (**Box 1**), which includes a careful review of literature on the topic and preliminary observations or data collection. Published studies and preliminary data should be evaluated and synthesized by exploratory data analysis (Tukey 1977) to develop a conceptual model (theoretical framework or general research hypothesis). This conceptual model is essentially a broad theory (Fig. 1) that offers explanations and possible solutions, and places the problem in a broader

Box 1. Systematic outline of sequential events in scientific research.
1. Identify the research problem.
2. Conduct literature review of relevant topics.
3. Identify broad and basic research objectives.
4. Collect preliminary observations and data as necessary.
5. Conduct exploratory data analysis.
6. Formulate a theory (conceptual model or research hypothesis).
7. Formulate predictions from conceptual model as testable hypotheses (Fig. 1).
8. Design research and methodology for each hypothesis with assistance from a statistical consultant to estimate required sample sizes and anticipate analysis procedures.
9. Prepare written research proposal that reviews the problem, objectives, hypotheses, methodology, and procedures for data analysis.
10. Obtain peer review of the research proposal from experts on the research topic and revise if necessary.
11. Perform experiments, collect observational data, or construct a model.
12. Conduct data analysis.
13. Evaluate, interpret, and draw conclusions from the data.
14. Speculate on results and formulate new hypotheses.
15. Submit manuscript describing the research for peer-reviewed journal publication, agency publication, and/or presentation at scientific meetings.
16. Repeat the process with new hypotheses (starting at step 6 or 7).

context (**Box 1**). The next step is to develop predictions from the conceptual model, i.e., statements that would be true if the conceptual model were true. These predictions are then stated as multiple testable hypotheses. Research should be designed to test these hypotheses; ideally experimentation should be used whenever possible. Included in the design is calculation of sample sizes required to detect the hypothesized effects. Peers and a statistician should review the proposed design before data collection begins. This is also the appropriate time to decide how the data will be analyzed. Data analysis with appropriate statistical procedures leads to rejection of, or failure to reject, hypotheses. Final conclusions usually result in further speculation, modification of the original conceptual model and hypotheses, and formulation of new hypotheses. The publication process is the last, but essential, step, and peer-review comments should be considered carefully before research on new hypotheses is designed.

Problem Identification

The initial step in most wildlife research is problem identification. Most research is either applied or basic. Applied research usually is related to a management problem, e.g., proper habitat management or identification of critical habitat for a declining species. Some problems that require research stem from political controversy or

public demand. For example, we may study specific populations because the hunting public has demanded greater hunting success. Other applied studies may be politically supported due to projected loss of habitat by development or concerns over environmental problems such as contamination from agricultural chemicals.

Unfortunately, few wildlife research projects are basic. We rarely have the luxury of studying wildlife populations to gain knowledge for the sake of knowledge and a more complete understanding of factors that affect behavior, reproduction, density, competition, mortality, habitat use, and population fluctuations. However, once there is political support for research on species of concern (e.g., declining raptor populations), there may be a period when research funding is relatively abundant and allows for basic studies. Research on management questions can often be designed so basic research on underlying principles can be conducted for minimal extra cost as data are gathered to solve the immediate management problem.

Once objectives are identified, research should begin with a thorough literature review followed by a descriptive phase. This phase can be omitted if descriptive work has been completed on the specific problem. Descriptive phases cumulate natural history observations, previous data, and information relevant to the specific study objectives. An important aspect of this phase is exploratory data analysis (Tukey 1977, James and McCulloch 1985). During this process data are quantitatively analyzed in terms of means, medians, modes, standard deviations, and frequency distributions for important groups as well as scatter plots of potential relationships. Exploration of the data should be as complete and biologically meaningful as possible, which may include comparison of data categories (e.g., mean values, proportions, ratios), multivariate analysis, correlation analysis, and regression. The "basic aim of exploratory data analysis is to look at patterns to see what the data indicate" (James and McCulloch 1985:21). If the research topic has received extensive previous investigation, the exploratory phase might even take the form of a meta-analysis of previous data gathered on the question (Osenberg et al. 1999). This phase often involves extensive discussions with other investigators with field or experimental experience on the topic.

Theory, Predictions, and Hypotheses

Exploratory data analysis and perceived associations should lead to development of a theoretical framework (conceptual model) of the problem. Kerlinger and Lee (2000:11) defined theory as "a set of interrelated constructs (concepts), definitions, and propositions that present a systematic view of phenomena by specifying general relations among variables, with the purpose of explaining and predicting the phenomena." Ford (2000:43) identifies 2 parts of a theory consisting of a working part providing information and logical basis for making generalizations and a motivational or speculative component that defines a general direction for

investigation. Predictions or deductive consequences of theory form the basis for hypotheses, which are assertions subject to verification (Fig. 1) (Dolby 1982, James and McCulloch 1985). Normally, the primary research hypothesis is what we initially consider to be the most likely explanation but, if the question has been placed into the proper theoretical framework, several alternate hypotheses are presented as possible explanations for facts observed.

We take an important step from descriptive natural history when we formulate research hypotheses. Interpretation of exploratory data analysis, creation of a theoretical framework, deducing predicted consequences, and formulation of testable hypotheses are difficult aspects of science that require creativity and careful reasoning but they are essential to the future of wildlife science.

Study Design and Statistical Inference

Many different research options are available for answering questions about the biology and management of wildlife species (Fig. 2) (Eberhardt and Thomas 1991, Morrison et al. 2001). These options differ dramatically in terms of 2 criteria: how certain are the conclusions reached and how widely applicable are the conclusions? No single option is perfect. The biologist must weigh the available options carefully to find the best choice that fits within constraints of time and resources.
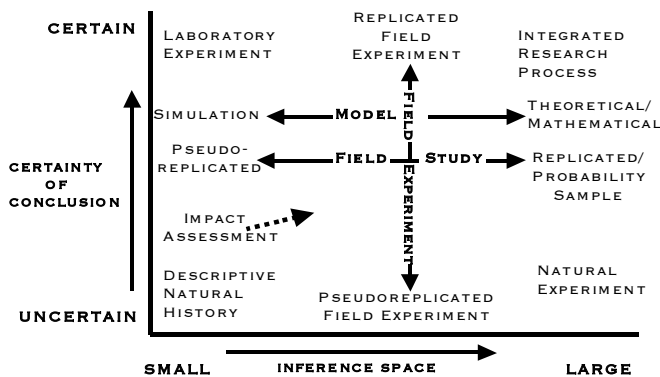

Fig. 2. The potential for wildlife study designs to produce conclusions with high certainty (few alternative hypotheses likely) and widespread applicability (a diversity of research populations where inferences apply).

Experiments consisting of manipulative trials are under-used in wildlife science. Laboratory experiments, in which most extraneous factors are controlled, provide the cleanest results with the most certainty, but results generally have only narrow application to free-ranging wildlife populations. Their demonstrated power and applicability to studies of basic processes in animal physiology, biochemistry, and molecular biology do not easily translate to studies of wildlife habitat and populations. Natural experiments (Diamond 1986), in which large-scale perturbations such as wildfires, disease outbreaks, and hurricanes affect populations and

landscapes naturally, provide only weak conclusions because of lack of replication and inability to control extrinsic factors through random assignment of treatments. Results from studies of natural events (experiments) are applicable to a variety of populations, especially where multiple similar natural events are analyzed. Field experiments (Hurlbert 1984), in which manipulative treatments are applied in the field, combine some of the advantages of laboratory and natural experiments (Fig. 2). Field experiments span a range from pseudoreplicated field experiments (Hurlbert 1984) in which no true replication is used (or possible) and conclusions are not certain, to replicated field experiments in which conclusions are relatively certain (Johnson 2002). Replicated field experiments provide conclusions that are broadly applicable to free-ranging wildlife populations.

Case studies consisting of unreplicated, natural history descriptions (Fig. 2) are most useful at early stages in development of the research process. Pseudoreplicated field studies are only slightly better than descriptive natural history studies. At the other extreme are replicated field studies wherein no manipulation or randomization of treatments occurs, but true replication occurs within a probability sampling framework and information is gathered to evaluate alternate hypotheses. Conclusions from replicated field studies are broadly applicable, but are less certain than those from replicated field experiments. Some questions of importance in wildlife biology and management are not appropriate for experimentation. For example, we may be interested in the effects of weather on a particular animal population, but we cannot manipulate weather. In addition, we may be interested in the relative importance of factors such as predation, habitat, and food limitations on population changes (Quinn and Dunham 1983). In these cases we should formulate primary and alternate hypotheses in the form of models that can be tested statistically or evaluated using likelihood methods to estimate parameters and likelihood ratios or information criteria for comparisons (Burnham and Anderson 2002).

Designing good field studies is more difficult than designing good experiments because of the potential for extraneous factors to invalidate tests or comparisons. One key step for both experiments and field studies is designing a sampling procedure to draw observations (experimental units or sample units) from the populations of interest. Only if this is done properly can conclusions of the tests be applied to these populations. Survey sampling (Cochran 1963) provides methods that are helpful in designing such sampling procedures. These methods are particularly important for field studies, but are also useful in field experiments for drawing experimental units and subsamples (samples within one experimental unit).

Careful planning of the actual testing process can proceed once a research option has been chosen for each hypothesis. For each hypothesis we must identify exactly what data will be collected and when, how, how much, and for how long. Further, how will these data be treated statistically? Will the data meet assumptions of the statistical test? Is the sample size adequate? Will the

statistical hypothesis provide information directly related to the theory or model? Do biases exist in data collection, research design, or data analysis that might lead to a spurious conclusion? These questions must be considered carefully for each hypothesis before fieldwork begins. Consulting a statistician is important, and the statistician should understand the basic biological problem, the overall objectives, and the research hypotheses.

Peer review of the proposed research should be obtained from several people with expertise and experience with the research topic. Peer review will usually improve a research design, and may disclose serious problems that can be solved during the planning stage. Unfortunately, most peer reviews occur too late, after data collection when the final report or publication manuscript is written.

Sample Size and Power

One of the more challenging steps prior to starting actual data collection is to set goals for sample size using a prospective power analysis. The power of any hypothesis test is defined as the probability of rejecting the null hypothesis when, in fact, it is false. Power depends upon the magnitude of the effect (e.g., magnitude of difference between treatment and control or bound on estimate), variation in the characteristic, significance level (α), and sample size. Zar (1999) provides formulas to calculate power and sample size for hypothesis tests but a statistician should be consulted for complicated experimental designs and analyses. Many statistical packages (e.g., SAS by Statistical Analysis Systems) or specialized analysis software (e.g., CAPTURE and MARK) provide capability to generate sample data for analysis to determine in advance how large the sample size should be to detect effects expected.

Effect size (magnitude of effect) is an important factor influencing sample-size requirements and power of a test. However, power and sample-size calculations should be based on a biologically meaningful effect size. Identifying a biologically significant effect usually involves expressing the conceptual model as a quantitative model plus value judgments about the importance of a biological response. Estimating power of the test and calculating sample size requirements forces the investigator to evaluate the potential significance of the research prior to beginning fieldwork. Sample size analysis may lead to substantial revision of the goals and objectives of the research.

Data Collection

All data should be recorded on preprinted data sheets or entered directly into a hand-held data logger, computer, or personal digital assistant. This ensures that each field person collects exactly the same data as consistent collection of data simplifies analysis. Data sheets should be duplicated after each field day (e.g., computer entry, photocopies, or transcribed) and stored in a separate location from the original data set. Data entered electronically in the field should be downloaded daily and backed up for storage at another location. Transcription of data (including computer data entry) must be followed by careful proofreading, which is greatly facilitated by use of error-checking programs. All field personnel should receive careful instructions regarding data collection, and the principal researcher must check periodically to see that each person has similar skills and uses the same methods for observation, measurement, and recording (Kepler and Scott 1981). The principal researcher has responsibility for quality control and the validity of research results depends upon quality of research design and data collection.

Most novice research biologists are anxious to initiate data collection because of the attractiveness of working out-of-doors and the pleasure derived from observing wildlife-related phenomena. The design phase should not be rushed to initiate fieldwork more quickly. Successful research biologists often spend about 40% of their time in design and planning phases, 20% in actual fieldwork, and 40% in data analysis and writing publications. Data collection can be physically difficult and highly repetitive. Often the most enjoyable and rewarding portion of research comes during the data-analysis phase when the biologist begins to see results from several years of planning and fieldwork.

Data Analysis

Analysis of data should be an enjoyable phase. Frequently this stage takes longer than anticipated because substantial effort must be expended preparing data for analysis. An excellent way to accelerate this part of the process is to begin analyzing the data once a moderately sized portion of it, such as one third, has been gathered. This also facilitates finding major errors in measurement or recording methods before a large amount of effort is wasted. Performing exploratory data analyses to summarize distributions, calculate basic descriptive statistics, and plot preliminary relationships enables the researcher to provide positive feedback to field workers that their work is making a solid contribution while it helps the researcher identify any problems. Analysis methods are covered by Bart and Notz (2004) and in standard statistical references such as Milliken and Johnson (1984), Johnson and Wichern (1988), and Zar (1999). The researcher is cautioned not to lose track of the magnitude of effects or differences in a rush to test their statistical significance. Statistical tests are not the results, but facilitate evaluating the meaning of effects or differences measured. If adequate effort was devoted to planning analysis in cooperation with a statistician during the design phase and quality control was exercised during data collection, data analysis will proceed rapidly.

Evaluation and Interpretation

Evaluation and interpretation is a creative phase,

similar to hypothesis formulation. The quality of conclusions drawn is dependent upon the biologist's past educational and professional experience as well as willingness to consider standard and less-traditional interpretations. One great danger in wildlife science (and other fields) is that researchers often have a conscious or unconscious expectation of results. This bias might begin with development of the overall research objective and carry through to the interpretation phase. This danger is so great that in some fields, such as medicine, experiments are performed with a double-blind approach in which neither researcher nor subjects know membership of treatment and non-treatment groups. A scientist must not design research or interpret data in a way that is more likely to support preconceived explanations of biological systems. Biologists who are consciously aware of their own biases and strive to keep an open mind to new ideas are most likely to make revolutionary discoveries.

The first objective is to organize, clearly and concisely, the results of data collection, exploratory data analysis, and specific statistical analyses. These results must be transformed from a collection of specific information into a synthesis explaining the biological system. Do specific statistical tests support one or more of the theories and hypotheses and clearly reject others? Do the results provide a reasonable explanation of the biological system? Are there alternative explanations of the data and statistical tests? Are there specific problems with the data that should be identified, such as inadequate sample sizes or unusual variation in specific variables measured? What could have introduced bias into the estimates? Are additional data required? These questions must be considered carefully and, if concerns are identified, they must be noted in reports and publications.

During this phase, the biologist usually reaches some conclusions based on the data and results of statistical tests. If the data support the hypothesis, we cannot conclude the theory (model) is true, but only that it has not been rejected (James and McCulloch 1985). The central issue is that we do not prove a research hypothesis or theory to be correct. The credibility of the hypothesis increases as more of its predictions are supported and alternative hypotheses are rejected. The goal of science is knowledge and, in wildlife science, we attempt to explain processes within biological systems and to predict how changes will affect specific wildlife populations.

We can assist other biologists by carefully considering how broadly our conclusions can be generalized to other areas or populations. A relatively common problem is that conclusions often go beyond the data. Interpretation of research data must clearly separate conclusions and inferences based on data from speculation. For example, if we demonstrate that droppings from spruce grouse (*Falcipennis canadensis*) are most abundant under lodgepole pine (*Pinus contorta*) and Engelmann spruce (*Picea engelmannii*), we can conclude that grouse use both tree species for some behaviors, but the type of behavior (e.g., roosting or feeding) is speculation without additional data (e.g., observations of feeding activity, and crop or fecal analyses). Likewise, replication of studies across space and time "provides us greater confidence that certain relationships are general and not specific to the circumstances that prevailed during a single study" (Johnson 2002: 930).

Speculation and New Hypotheses

Rarely does a single research project provide the last word on any problem (Johnson 2002). More commonly, research will generate more questions than it answers. Speculation, based on inconclusive or incomplete evidence, is one of the most important aspects of science. Speculation must be identified and should not be confused with conclusions based on data. But speculation is the fuel for future research. Many facts of nature have been discovered by accident–an unexpected result from some associated research effort. However, most research is directional, i.e., it attempts to support or falsify a theory reached by speculating from facts.

New hypotheses can be considered a form of speculation, which is verbalized in a more formal fashion and has a specific, testable format. For example, considering spruce grouse, we can formulate a basically untestable hypothesis that "spruce grouse have evolved a preference for use of lodgepole pine and Engelmann spruce trees." This statement is too vague and requires historical data that cannot be collected. However, we can hypothesize that spruce grouse use lodgepole pine and Engelmann spruce trees for (1) feeding or (2) roosting. Testing these hypotheses we might learn that 80% of the spruce grouse diet is lodgepole pine even though Engelmann spruce is more abundant. We may then hypothesize that needles from lodgepole pine have higher nutritional quality than needles from Engelmann spruce.

Publication

The final step of the scientific method is publication of the research. Unfortunately, many research dollars are wasted because the knowledge gained was not published, and the information is buried in file cabinets or boxes of data sheets. The publication process is the most difficult phase for many biologists. Clear, concise scientific writing is difficult because most biologists have little formal training and inclination in this activity. Peer review may also be damaging to a person's ego, because we must subject our work to anonymous critiques used by editors to judge whether the manuscript is acceptable for publication.

Agency administrators often do not encourage or reward employees for publishing their work and discourage publication in some instances. Administrators are pressured with calls for immediate answers to management problems; thus, they devalue the long-term benefits of the publication process. Effective administrators recognize that peer review and publication will (1) correct errors and possibly lead to a better analysis, (2) help authors reach the most sound conclusions from their data, (3) make it easier to defend controversial

policies, (4) help their personnel grow as scientists by responding to critical comments and careful consideration of past errors (which may have been overlooked without peer review), and (5) make a permanent contribution to wildlife management by placing results in a literature format available to other agencies, researchers, and students.

Publication is essential to science. Peer reviews normally improve the quality of a manuscript, but some research may not be suitable for publication. This emphasizes the importance of careful planning, design, data collection, etc. Rarely would any research effort that is properly planned, designed, and executed (including a well-written manuscript) be unpublishable. However, the revision process (i.e., responding to criticisms from the editor and referees) may be painful and frustrating to authors. Overall, the system is necessary to insure quality publications, and authors should not be discouraged by the necessity to defend their work and revise manuscripts. Research is not complete and does not make a contribution to knowledge and sound management of wildlife resources until results are published in a way that effectively communicates to the scientific community and user groups (e.g., wildlife managers). In addition to publication in peer-reviewed journals, research findings will improve wildlife management immediately if they are communicated in other forums such as professional meetings, workshops, seminars, general technical reports, informational reports, and articles in the popular press.

MAJOR COMPONENTS OF RESEARCH

Biological and Statistical Populations

The wildlife profession works with 3 types of populations: biological, political, and research populations. Mayr (1970:424) defined a population as a group "of potentially interbreeding individuals at a given locality," and species as "a reproductively isolated aggregate of interbreeding populations." This biological population is an aggregation of individuals of the same species that occupy a specific locality, and often the boundaries can be described with accuracy. For example, the dusky Canada goose (*Branta canadensis*) population breeds within a relatively small area on the Copper River delta of Alaska and winters in the Willamette Valley near Corvallis, Oregon (Chapman et al. 1969). Between the breeding and wintering grounds of the dusky Canada goose is the more-restricted range of the relatively nonmigratory Vancouver Canada goose (Ratti and Timm 1979). Although these 2 populations are contiguous with no physical barriers between their boundaries, they remain reproductively isolated and independent. The Yellowstone National Park elk (*Cervus elaphus*) herds are additional examples of biological populations with separate boundaries (Houston 1982). Biological populations for other species may not be so geographically distinct as Canada geese and Yellowstone elk, in which case the researcher will have to carefully consider from which biological aggregation their

samples are selected and to which their findings will apply. Carefully specifying this biological population is essential in the planning phase of an investigation and may require thorough investigation of existing literature on the species, geographic resources, and reviews of literature on biological aggregations (Mayr 1970, Selander 1971, Stebbins 1971, Ratti 1980, Wells and Richmond 1995, Garton 2002).

The political population has artificial constraints of political boundaries, such as county, state, or international entities. For example, a white-tailed deer (*Odocoileus virginianus*) population within an intensively farmed agricultural region in the Midwest might be closely associated with a river drainage system due to permanent riparian cover and food critical for winter survival. The biological population may extend the entire length of the river drainage but, if the river flows through 2 states, the biological population is often split into 2 political populations that are subjected to different management strategies and harvest regulations. Traditionally, this has been a common wildlife management problem. When biological populations have a political split, it is best to initiate cooperative studies, in which research personnel and funding resources can be pooled to benefit both agencies.

Research populations are usually only a segment of a biological population. From this segment we take a sample resulting in the research population commonly being referenced as the sample frame (Scheaffer et al. 1996). In rare instances, a population may be studied that represents all individuals of a species, e.g., endangered species with few individuals, such as whooping cranes (*Grus americana*). Or, our research population might represent an entire biological population, such as one of the elk herds in Yellowstone National Park. However, the research population usually is only a portion of the biological population and a small segment of the species. Thus, sampling methodology is critical, for it provides the only link between our samples and the research population.

Conclusions from research are directly applicable only to the population from which the sample was drawn, the research population. However, biologists usually have goals to obtain knowledge and solve problems regarding biological populations and species. The key questions are: (1) is the sample an unbiased representation of the research population, (2) is the research population an unbiased representation of the biological population, and (3) is the biological population representative of the species? Because traits among segments of biological populations (and among populations of a species) often differ, broad conclusions relative to a research hypothesis should be avoided until several projects from different populations and geographic locations provide similar results. Combining and synthesizing replicate studies across large spatial extents should be a long-term goal, but may require use of new techniques such as meta-analysis (Osenberg et al.1999).

Approaches to Evaluating Hypotheses

Descriptive natural-history studies are important for gathering basic information and formulating questions but we must go further by placing these questions within the context of theory and using the theory to turn our questions into testable hypotheses (Fig. 1). Three approaches are available to investigate hypotheses: perform a manipulative experiment, gather new observations, or construct a model (Ford 2000:487). We illustrate the application of these approaches with published studies of wildlife questions.

Managers, biologists, hunters, and conservationists have expressed concern about the long-term effects of hunting, predation, and habitat conversion on elk populations from the Lochsa River drainage in northern Idaho to the northern range of Yellowstone National Park. Large scale wildfires such as occurred in the Yellowstone Ecosystem in 1988 raise concerns of park visitors for their potential negative impacts on elk populations, but they probably are a primary cause of some exceptional elk populations such as the Lochsa elk herd which developed following an enormous fire burning across the northern Rocky Mountains in 1910 (Mohler et al. 1958). Do stand replacing wildfires and similar management activities such as clear-cutting harm or benefit elk populations?

Lyon and Christensen (2002) suggest that fire and logging effects are complex. An observational approach to evaluating effects of fire and logging on elk would be the simplest to conduct. For example we might measure changes in forage resources for elk following wildfires, clearcuts, or prescribed burns following Leege and Hickey (1971). This habitat work shows potential effects but responses of elk must be documented also. DelGiudice et al. (1991) used chemical analysis (N and P to creatinine ratios) of elk urine in snow to provide evidence of nutritional stress in elk on the northern range of Yellowstone National Park following the fires in 1988. Better evidence comes from Singer et al. (1997) who estimated survival rates of radio-collared elk calves before and after the large fires. Their work demonstrated the complexity of effects of fire on large mammals because "density-dependent mortality of calves during winter due to malnutrition and summer mortality of calves due to predation were partially compensatory, but severe environmental conditions produced largely additive components of both summer (increased predation) and winter (increased malnutrition) mortality" (Singer et al. 1997: 12).

DelGiudice et al. (2001) used simulation modeling for elk in Yellowstone National Park in combination with snow-urine nitrogen to creatinine ratio analyses to predict more severe nutritional conditions for animals on the Madison-Firehole range than on the northern range immediately following the fires in 1988. Turner et al. (1994) developed a spatially explicit individual-based simulation model to explore the effects of fire scale and pattern on the winter foraging dynamics and survival of free-ranging elk on the northern range. Their simulated elk survival for the winters before, during, and after the fires agreed with observed data and provided the basis for a factorial simulation experiment to "explore effects on ungulate survival of fire size, fire pattern, and winter severity during an initial post fire winter (when no forage was available in burned areas) and a later post fire winter (when forage was enhanced in burned areas)" (Turner et al. 1994:1).

Field experiments to evaluate effects of fires and logging are the most difficult and expensive to conduct but their power to provide insightful results is demonstrated by the work of Cook et al. (1998) who tested the widely-held assumption that thermal cover enhances winter survival and reproduction of elk. They placed small groups of young cow elk in 12 pens at the center of 2.3-ha treatment units assigned one of 4 levels of thermal cover from dense overstory to clearcuts. They found no significant positive effect of thermal cover on condition of elk during 4 winter and 2 summer experiments. "Dense cover provided a costly energetic environment, resulting in significantly greater over winter mass loss, fat catabolism, and (in one winter) mortality" (Cook et al. 1998:1). Experimental investigations often provide clean results but the value of observational studies and models should be also considered.

Pilot Study

A pilot study is a preliminary, short-term trial through all phases of a research project. Pilot studies are an important but often neglected step in the research process. Information can be obtained that will help the researcher avoid potentially disastrous problems during or after the formal research phase. Pilot studies often will disclose hidden costs or identify costs that were over- or underestimated. Optimal sample allocation (Scheaffer et al. 1996) incorporates cost estimates to maximize benefit from limited research budgets. Use of a pilot study should reveal basic logistical problems, e.g., travel time among study plots might have been underestimated or expectations for overall sample sizes might not be feasible without additional personnel and funding. Statistical procedures for estimating needed sample sizes require variance estimates of variables that will be measured, and these variance estimates often are available only from data gathered in a pilot study. These preliminary data might disclose the variance of the population is so large that obtaining adequate sample sizes will be difficult. It is far better to discover these problems before time, energy, personnel, and critical research dollars are committed to a formal research project. If the research is part of an ongoing project, or if much research on the topic has been published, costs, methodology, and variance estimates may already be firmly established.

Precision, Bias, and Accuracy

One measure of quality of estimates is their precision. Precision refers to the closeness to each other of

repeated measurements of the same quantity (Cochran 1963, Krebs 1999, Zar 1999). Precision of an estimate depends on variation in the population and size of the sample. Indicators of precision of an estimator are standard error and confidence intervals. Larger variation in the population leads to lower precision in an estimate, whereas a larger sample size produces higher precision in the estimator. Another measure of quality of an estimator is termed bias. Bias describes how far the average value of the estimator is from the true population value. An unbiased estimator centers on the true value for the population. If an estimate is both unbiased and precise, we say that it is accurate (defined as an estimator with small mean-squared error, Cochran 1963). Accuracy is the ultimate measure of the quality of an estimate (Fig. 3.) and refers to the small size of deviations of the estimator from the true population value (Cochran 1963).
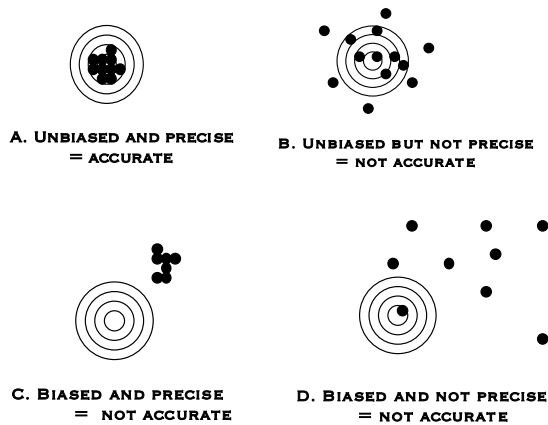

Fig. 3. The concepts of bias, precision, and accuracy are illustrated with targets and a shot pattern (modified from Overton and Davis 1969 and White et al. 1982).

Let us illustrate these concepts with a typical population survey. Suppose we were interested in estimating density of elk on a large winter range. One approach might be to divide the area into a large number of count units of equal size and draw a sample of units to survey from a helicopter. This would define our research population in terms of a geographic area rather than animals. The elements of our target population are count units, and we select a sample of these units using an objective sampling design (a probability sample). Using the helicopter we search each of the sampled units, attempting to count all elk present in each unit. We divide the number of elk counted in each unit by the size of that unit to obtain a density estimate for each unit (Fig. 4A). The histogram suggests little variation in density on this winter range, as most spatial-units (80%) have densities between 1.5 and 2.3 elk/km^2. We need a single value that is representative of the entire winter range, and we choose the mean from our sample as the best estimate of the mean for the winter range. The variation from one unit to the next is small, thus the mean from our sample is a fairly

precise estimate. But suppose we had obtained different results (Fig. 4B). Now the variation from one unit to the next is great, and the sample mean is less precise and not as reliable as the previous estimate. Thus, for a given sample size, the former estimate is more precise because of less variation in the population.

Would the mean from the sample in Area A (Fig. 4A) be an accurate estimate of the mean density of elk on this winter range? To answer this question, we must evaluate the bias in the estimate. If the winter range was partially forested or had tall brush capable of hiding elk from view, aerial counts in each unit would underestimate the
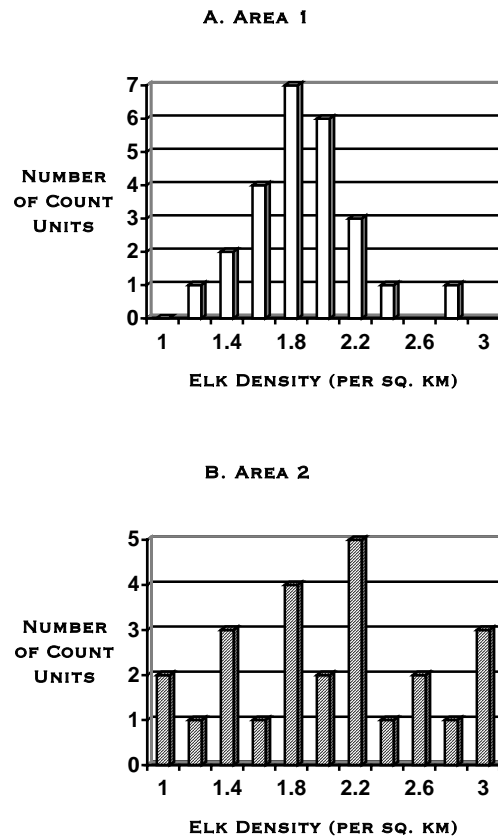
Fig. 4. Hypothetical example of elk counts and density estimates in Areas 1 and 2.

true number of elk present (Samuel et al. 1987). In this example the mean density from the sample would be a biased estimate of elk density on the winter range and, therefore, not highly accurate. If the winter range was a mixture of open brush fields and grasslands where all animals would be visible, mean density from the sample could be an accurate estimate of elk density on the entire winter range. We strive for accuracy in our estimates by selecting the approach with the least bias and most precision, applying a valid sampling or experimental design, and obtaining a sufficiently large sample size to provide precise estimates.

Evaluating bias in an estimate is difficult and, in the past, has been based on the researcher's biological

knowledge and intuition. If bias is constant, the estimate can be used to make relative comparisons and detect changes (Caughley 1977). Usually bias is not constant (Anderson 2001), but its magnitude often can be measured so that a procedure to correct estimates can be developed (Rosenstock et al. 2002, Thompson 2002). For example, Samuel et al. (1987) measured visibility bias in aerial surveys of elk from helicopters, and Steinhorst and Samuel (1989) developed a procedure to correct aerial surveys for this bias.

Replication

Sample size refers to the number of independent, random sample units drawn from the research population. In experiments, sample size is the number of replicates to which a treatment is assigned. For logistical reasons, we may measure numerous subsamples closely spaced in a single sample unit. However, we must be careful to distinguish these subsamples from independent, random samples. Subsamples are not independent random sample units because they typically are more similar to each other than are widely spaced samples. Similarly, subsamples in experiments are not true replicates if they cannot be independently assigned to a treatment category. The precision of a statistic is measured by its standard error. Standard error is calculated from the variation among the true sample units or replicates and the number of samples. If subsamples are mistakenly treated as true sample units or replicates, sample variance will underestimate actual amount of variation in the populations, sample size will overestimate true sample size, and precision of the estimate will be overestimated.

To illustrate this point, suppose we wanted to evaluate the effect of prescribed fire on northern bobwhite (*Colinus virginianus*) habitat in a large valley (1,000 km^2). We might conduct research on a habitat improvement project that involves burning 1 km^2 of grassland and brush (e.g., Wilson and Crawford 1979). We could place 20 permanent plots within the area to be burned and 20 in an adjacent unburned area. Measurements before and after the fire on the burned and unburned plots could be compared to examine effects of fire on bobwhite habitat. However, the 20 plots on the burned area are not really replicates but merely subsamples or pseudoreplicates (Hurlbert 1984). In fact, we have only one observation because we have only one fire in a 1-km^2 plot within the 1,000-km² valley. What would happen if we were to redesign the study to conduct 20 burns on 20 randomly chosen areas scattered throughout the valley. We would expect to see more variation among these plots than among 20 plots in a single burned area. The fallacy of the first design is obvious. A statistical test would evaluate only whether the burned 1-km^2 area differed from the unburned 1-km^2 area and could lead to false conclusions about effects of burning on bobwhite habitat in this area. A more appropriate design would require randomly selecting 40 sites from throughout the entire valley and randomly

assigning 20 to be burned (treatments) and 20 to be control (unburned) sites. Each burned and control site would be sampled with 5 plots to measure bobwhite habitat before and after the treatment, and data would be analyzed by analysis of variance; the 40 sites are samples and the 5 plots per site are subsamples. Thus, the 20 sites of each type would be true replicates. Treating the 100 burned and 100 unburned plots as experimental replicates would be an example of pseudoreplication. Psuedoreplication is a common problem and investigators must understand the concept of replication and its importance in ecological research (Hurlbert 1984, Johnson 2002).

Sample Size and Power of a Test

In descriptive studies, sample size required to obtain an estimate of desired precision can be calculated after an estimate of population variance is obtained from previous studies or a pilot study. Formulas for sample size are available for standard survey designs (Scheaffer et al. 1996, Thompson et al. 1998) and for typical hypothesis tests (Zar 1999).

In studies involving experiments or other types of comparisons, sample size is increased to improve the power of the test (defined as probability of detecting a real difference) and to prevent erroneous conclusions. To illustrate power of a test, consider the following example. Suppose we were using fawn: doe ratio as an indicator of production for a mule deer herd (*Odocoileus hemionus*) (i.e., our research population). We want to know if the fawn:
doe ratio has declined. There are 4 possible outcomes from sampling the herd and testing for a decline in the fawn: doe ratio (the null hypothesis is that there is no change, Table 1). We evaluate whether the fawn: doe ratio has declined by comparing the test statistic we calculate from our data to a value for this statistic at our chosen level of significance (α). The level of significance represents the chance of concluding the ratio changed when in fact it did not. A α of 0.05 indicates that we would make this error only 5 times if the population really did not decline and we tested it by drawing a sample 100 times. This is referred to as Type I error. But we could make another error. We could conclude the ratio had not changed when in fact it had declined. For the situation (Table 1) where we count 500 deer, we would fail to detect the decline in the fawn: doe ratio 50% of the time. This type of error is referred to as Type II error, and its likelihood is measured by β. When we perform a test, we typically set α low to minimize Type I errors. But Type II errors might be as important (Alldredge and Ratti 1986, 1992) or even more important than Type I errors. Obviously we want to detect a change when it occurs; the probability of detecting a change is called the power of the test. The power of the test is calculated as the probability of not making the Type II error (1- β).

Table 1. Possible outcomes (4) of a statistical test for declining production in a deer herd. Counts of 500 antlerless deer (adult does and fawns) were obtained each year, and tests of the null hypothesis of no change in the fawn: doe ratio were performed at the 5% level of significance (**"** = 0.05).

	Fawns per 100 does							
	Actual herd values			Count values				
Case	1988	1989	Change	1988	1989	Conclusion from test	Result of test	Likelihood of this result
1	60	60	None	61	59	No change	No error	95% (1 - **"**)
2	60	60	None	65	50	Declined	Type I error	5% (**"**)
3	65	50	Declined	65	50	Declined	No error	50% (1 - **$**)
4	65	50	Declined	62	57	No change	Type II error	50% (**$**)

The power of the test depends upon several factors including sample size, level of significance (α), variance in the populations, effect size (the true change that occurred), and efficiency of the test or design (Steidl et al. 1997). We cannot control natural variation within the population or the actual change that occurred, but we can control the other 3 factors. Parametric tests (based on standard distributions such as normal distribution, e.g., *t-tests*, *F-tests*, *Z-tests*) have the highest efficiency for normally distributed populations and for large samples. Nonparametric tests (based on ranks of values rather than their actual numerical value, e.g., Mann-Whitney, Wilcoxon signed- ranks test) are superior when sample sizes are small (<30) and populations are not normally distributed (Johnson 1995, Cherry 1998). The power of a test declines as the level of significance is made more stringent (decreasing α). In the example (Table 1), this is a critical problem because the Type II error (failing to detect declining production) is the more serious error. It would be preferable to increase α so that power of the test could be increased. In other situations the Type I error will be more serious and α must be kept low. Increasing sample size increases power of the test. Calculating sample size necessary for a desired level of power is essential to designing a high quality study (Toft and Shea 1983, Forbes 1990, Peterman 1990). However, such calculations should be based on meaningful effect sizes, i.e., what constitutes a biologically significant result (Reed and Blaustein 1997, Cherry 1998, Johnson 1999).

Controls

Observations on control sites are especially important in research design. In nonexperimental research, observations from randomly selected control sites can be compared with observations associated with a particular variable. For example, we may wish to know if habitat used by snowshoe hares (*Lepus americanus*) is different than general habitat availability. To examine this question, we can make observations (e.g., measure vegetation) at habitat-use sites and compare those observations with a series of random sites (controls) that we assume represent general habitat availability. If use sites differ from random control sites, we conclude habitat selection occurred (Pietz and Tester 1983).

In experimental research, a control may be defined as parallel observations used to verify effects of experimental treatments. Control units are the same as experimental units except they are not treated; they are used to eliminate effects of confounding factors that could potentially influence conclusions or results. Creative use of controls would improve many wildlife studies. Experimental studies in wildlife that involve repeated measurements through time must include controls because of the importance of weather and other factors that vary through time (Morrison et al. 2001). Without adequate controls, distinguishing treatment effects from other sources of variation is difficult. For example, in the bobwhite study, control sites were required to distinguish the effects of burning from effects of rainfall and other weather characteristics that affect plant productivity. There might be an increase in grass production in the year following burning because the rainfall was higher that year. Without control sites we cannot tell whether increased grass production resulted from increased rainfall, from burning, or from a combination of both factors. Thus, we cannot evaluate the relative importance of each factor.

SAMPLING

Most information gathered by wildlife biologists is used to meet descriptive rather than experimental objectives. Examples include estimates of population size, recruitment, herd composition, annual production of forage species, hunter harvest, and public attitudes. In these efforts biologists attempt to obtain estimates of characteristics that are important for management decisions. We want to obtain the best estimates possible within the constraints of our resources of time and money. A large body of statistical literature exists to help; these types of studies are referred to as surveys and the topic is known as survey sampling (Cochran 1963, 1983; Scheaffer et al. 1996) or finite population sampling.

The research population is typically synonymous with the statistical population but may differ when we define the statistical population geographically in terms of units of space or habitat. Defining our statistical population as drainages, forest stands, individual ponds, or square-km blocks often facilitates estimating total numbers

of animals and composition of a population. Sampling smaller units of habitat is more likely to be logistically feasible. Likewise this redefinition of our research (statistical) population makes it feasible to apply the powerful tools for sampling from finite populations.

Sampling is also a critical part of experimental research and the test of formal statistical hypotheses. All field studies and most field experiments require creative sampling designs to reduce variation between observations in our treatment or comparison categories. For example, stratification and clustering can sharpen comparisons, but data collected using these methods require analysis by more complicated designs (e.g., block or split-plot designs) (Zar 1999). Choice of specific sampling methods is dependent on the objectives or hypotheses being addressed, the nature of the population being sampled, and many other factors such as species, weather conditions, topography, equipment, personnel, time constraints, and desired sample sizes. A variety of sampling designs is available for biologists to use in wildlife surveys and experimental research (Scheaffer et al. 1996, Thompson et al. 1998, Morrison et al. 2001).

Sampling Design

Simple Random.--A simple random sample requires that every sample unit in the population have an equal chance of being drawn in the sample and the procedure for selecting units is truly random. This can be accomplished by assigning each member of the population a number and then picking numbers, to identify members to sample, from a table of random numbers or a random number generator on a computer or calculator. For example, suppose we wanted to estimate the number of successful hunters in a special hunt where a limited number of permits was issued. We might decide to contact a sample of permit buyers by telephone after the season to measure their hunting success. A survey design checklist (Box 2) helps us design such a survey properly. The population that we want to make statements about is all persons who obtained a permit. The list of the members of the population is usually called the sampling frame (Scheaffer et al. 1996). It is used to draw a random sample from the

population. The sampling frame must be developed carefully or the resulting estimates may be biased. For example, if a portion of our permit buyers did not have telephones and we decided to drop them from the list, the results could be biased if such hunters had different hunting success than permit buyers with telephones. To draw a random sample for our survey we could assign each person who purchased a permit a number and select the numbers to be contacted from a random-numbers table or by using a random-number generator.

In other types of surveys, obtaining a truly random sample of the population might be difficult. In such instances another method such as systematic sampling should be used. One approach, when the research population consists of animals that would be difficult to

sample randomly, is to change the design. We do this by making small geographic units, such as plots or stands, our sample units (or experimental units if we are developing a sampling design for an experimental treatment) and making our measurement on each plot a number or density of animals. Thus, we can take a random sample and use it to infer abundance across our entire study area sampled. A valid random sampling procedure must be independent of investigator decisions. For example, an excellent procedure to locate plots randomly in a study area would be to use a Landsat image of the study area stored in a geographic information system program which allows us to select random locations within the boundary of our

Box 2. Survey design checklist.	
Question	Example
1. What is the survey objective?	Estimate the percentage of successful hunters
2. What is the best method?	Telephone survey of permit holders
3. To what population do we want to make inferences?	Everyone who has a permit for this hunting period
4. What will be the sample unit?	Individual permit holders
5. What is the size of population to be sampled (N)?	N = 350 (for special permit hunt)
6. Which sample design is best?	Simple random sample
7. How large should sample be?[a]	$n = \dfrac{Np(1 - p)}{(N - 1) B^2/4 \ + \ p(1 - p)}$ where: N = population size (350) p = proportion of permit holders who harvested deer (from pilot survey = 0.24) B = bound on the estimate = 0.05 (we want estimate with p ± 0.05 confidence) Therefore $n = \dfrac{350(0.24)(1 - 0.24)}{(350-1)(0.05)^2/4 + 0.24(1-0.24)}$ $n = 159$, we should contact approximately 160 permit holders
8. Have you contacted a statistician to review your design?	Yes!

[a] Scheaffer et al. (1996:99).

study area using Universal Transverse Mercator (UTM) coordinates (Fig. 5A). The UTM coordinates of these selected plot locations can be entered into a hand-held GPS (Global Positioning System) unit that will guide us to the exact location. Random-like methods, referred to as haphazard or representative, have been used in place of truly random designs, but should be avoided because they are subject to investigator bias. An example
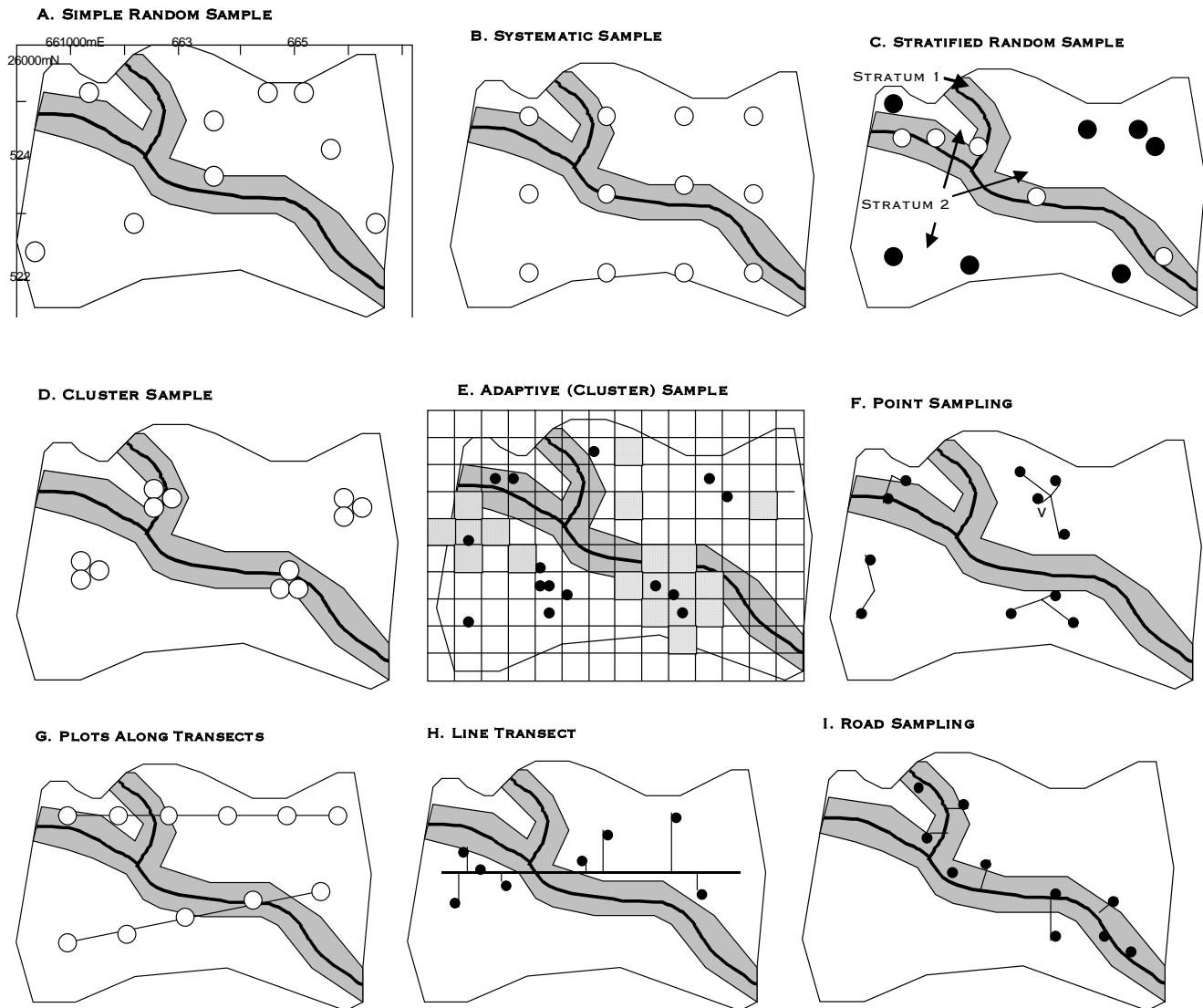
A. Simple Random Sample **B. Systematic Sample** **C. Stratified Random Sample**

D. Cluster Sample **E. Adaptive (Cluster) Sample** **F. Point Sampling**

G. Plots Along Transects **H. Line Transect** **I. Road Sampling**

Fig. 5. Examples of sampling methods: A = simple random sample, B = systematic sample, C = stratified random sample, D = cluster sample, E = adaptive cluster sampling, F = point sampling, G = plots along transects, H = line transect, and I = road sampling.

of these methods is the technique of facing in a random direction and throwing a pin over the shoulder to obtain the center for a vegetation plot. Although this sounds random, the odds of a field crew randomly facing away from a dense stand of thorny shrubs such as multiflora rose (*Rosa multiflora*) and throwing the pin into the middle of such a patch is practically zero. Truly random samples occasionally produce poor estimates by chance due to poor spatial coverage of the area or population of interest (e.g., in an area with a small number of important habitat patches, all of the patches may be missed by a truly random approach) (Hurlbert 1984, Johnson 2002).

Systematic.--A systematic sample is taken by selecting elements (sampling units) at regular intervals as they are encountered. This method is easier to perform and less subject to investigator errors than simple random sampling. For example, if we wanted to sample bird-watchers leaving a wildlife management area it would be difficult to draw a truly random sample. However, it would be easy to draw a systematic sample of 10% of the population by sampling every tenth person leaving the area. Systematic sampling is also used extensively in vegetation measurements because of its ease of use in the field. Systematic sampling is almost exclusively used in geographic sampling because it makes possible evaluation of the spatial pattern of variability (e.g., spatial

autocorrelation), which is used for most modern spatial modeling. A valid application requires random placement of the first plot followed by systematic placement of subsequent plots, usually along a transect or in a grid pattern (Fig. 5B). This approach often provides greater information per unit cost than simple random sampling because the sample is distributed uniformly over the entire population or study area. For random populations (i.e., no serial correlation, cyclic pattern, or long-period trend), systematic samples give estimates with the same variance as simple random samples.

The major danger with systematic samples is that they may give biased estimates with periodic populations (i.e., with regular or repeating cycles). For example, if we were interested in estimating the number of people using a wildlife management area, we might establish a check station and take a systematic sample of days during the season. This procedure could yield extremely biased results if we chose to take a sample of one-seventh of the days. If our sample day fell during the workweek, we could obtain different results than if it was during the weekend. Additionally, our estimate of variance would likely be too small, leading us to conclude that our estimate was much more precise than in reality. In this situation the population sampled obviously is periodic; in other situations the periodicity might be quite subtle. Thus, systematic sampling must be used with caution. The formal procedure is conducted by randomly selecting one of the first k elements to sample and every k^{th} element thereafter. For example, if we wanted to sample 10% of our population, k would equal 10 and we would draw a random number between 1 and 10. Suppose we selected 3, we would then sample the third element and every tenth element thereafter (i.e., 13th, 23rd, 33rd, . . .). At a check station we might use this to sample 10% of the deer hunters or bird watchers that came through the station. When locating plots along a transect, we would randomly locate the starting point of the transect and then place plot centers at fixed intervals along the transect such as every 100 m. Advantages and disadvantages of random and systematic sampling have been reviewed by Thompson et al. (1998), Krebs (1999), and Morrison et al. (2001).

Stratified Random.--In many situations, obvious subpopulations exist within one total population. For example, tourists, bird-watchers, and hunters are readily divided into residents and nonresidents. A study area can be divided into habitats. A population of animals can be divided into age or gender groups. If members of these subpopulations are similar in terms of the characteristics we are estimating and the subpopulations themselves differ from each other in the characteristic of interest, a powerful design to use is stratified random sampling. Subpopulations are referred to as "strata," and we draw a simple random sample of members from each of these strata. Stratified random sampling is also useful if we are particularly interested in the estimates for the subpopulations themselves. The strata are chosen so they contain units of identifiably different sample characteristics, usually with lower variance within each stratum.

For example, if the objective of a study of moose (*Alces alces*) is to estimate moose density, we might define strata on the basis of habitats (e.g., bogs and riparian willow [*Salix* spp.] patches, unburned forests, and burned forest). We then draw a simple random sample from each of these strata (Fig. 5C). If moose density is different among strata, variation in each stratum will be less than the overall variation. Thus, we will obtain a better estimate of moose density for the same or less cost. If strata are not different, stratified estimators may not be as precise as simple random estimators. In some instances, cost of sampling is less for stratified random sampling than for simple random sampling. A final advantage of stratified random sampling is that separate estimates for each stratum (e.g., moose density in willows or in forests) are obtained at no extra cost. The formal procedure for stratified random sampling consists of 3 steps: (1) clearly specify the strata – they must be mutually exclusive and exhaustive, (2) classify all sampling units into their stratum, and (3) draw a simple random sample from each stratum. Formulas are available to calculate the sample size and optimal allocation of effort to strata (Scheaffer et al. 1996, Krebs 1999). A pilot survey can be analyzed using analysis of variance (ANOVA) to learn if stratification is indicated. If cover types define strata, most GIS software will automatically select random coordinates within cover types making stratified random samples easy to select.

Cluster Sampling.--A cluster sample is a simple random sample in which each sample unit is a cluster or collection of observations (Fig. 5D). This approach has wide application in wildlife biology because many birds and mammals occur in groups during all or part of the year. When we draw samples from such populations we draw clusters of observations, i.e., groups of animals. Likewise, many wildlife user groups (e.g., waterfowl hunters, park visitors) occur in clusters (e.g., boats in wetlands, vehicles along highways). Cluster sampling is also useful where cost or time to travel from one sample unit to the next is prohibitive. This is commonly the situation in surveys of animals and habitat. The formal procedure for cluster sampling consists of 3 steps: (1) specify the appropriate clusters and make a list of all clusters, (2) draw a simple random sample of clusters, and (3) measure all elements of interest within each cluster selected.

Making a formal list of clusters is rarely possible or essential. Instead, we emphasize obtaining a random sample of clusters. If the sample units are animals, which naturally occur in groups, the size of the clusters will vary from group to group depending on the social behavior of the species. Cluster sampling of habitat is performed by choosing a random sample of locations and then locating multiple plots in a cluster at each location. In this case, the researcher sets cluster size. The optimal number of plots (cluster size) depends upon the pattern of variability in habitat. If plots in a cluster tend to be similar (little variability within a cluster), cluster size should be small.

If plots in a cluster tend to be heterogeneous (high variability within a cluster), cluster size should be large. For other types of cluster samples such as groups of animals or people in vehicles, cluster size is not under control but is a characteristic of the population. For example, aerial surveys of elk and deer on winter ranges result in samples of animals in clusters. Estimates of herd composition (e.g., fawn: doe or bull: cow ratios) are readily obtained by treating these data as cluster samples (Bowden et al. 1984).

Adaptive Sampling.--Adaptive sampling is a recent development in sampling design. It differs from earlier methods because the sample size is not set at the start of the sampling effort, but rather depends upon the results obtained during sampling. Thompson and Ramsey (1983) pioneered adaptive cluster sampling for gathering information on rare animals and plants, which often are clustered in occurrence. In adaptive cluster sampling, an initial sample of units is drawn by a random or other standard design and neighboring units are also sampled for any unit that satisfies a criterion such as having more than *x* individuals present (Thompson and Seber 1996). The initial sampling unit and neighbors (where sampled) form neighborhoods analogous to clusters and are treated similar to cluster sampling. Size of clusters does not need to be constant nor is it known in advance. For spatially clustered animals or plants, the neighborhood consists of adjacent spatial sample units (Fig. 5E). Smith et al. (1995) showed that adaptive cluster sampling would be relatively more efficient than simple random sampling for estimating densities of some species of wintering waterfowl if the right sample unit size and criterion for further sampling in the neighborhood were chosen. The species for which it would be superior show more highly clustered distributions. For other species, conventional sampling designs with fixed sample sizes were superior. Thompson and Seber (1996) provided numerous examples of applications of adaptive sampling under conventional sampling designs and estimation methods, as well as applications based on maximum likelihood methods and Bayesian approaches. Thompson et al. (1998) and Morrison et al. (2001) also review the basic concept and provide simple examples.

Sequential Sampling.--Another recent development is use of sequential sampling, which differs from the classical statistical approach in that sample size is not fixed in advance. Instead, samples are drawn one at a time and, after each sample is taken, the researcher decides whether a conclusion can be reached. Sampling is continued until either the null hypothesis is rejected or the estimate has adequate precision. This type of sampling is applicable to wildlife studies where sampling is performed serially, i.e., the result of each sample is known before the next sample is drawn (Krebs 1999). The major advantage of this approach is that it usually minimizes sample size saving time and money. After an initial sample of moderately small size is obtained, successive samples are added until the desired precision is met, the null hypothesis can be rejected, or a maximum sample size under a

stopping rule has been reached. This approach typically requires <1/3 the sample size required in a standard design (Krebs 1999:304). For example if we wanted to survey deer on a winter range to insure that harvest had not reduced buck abundance below a management guideline of 5% bucks, we would develop a graph (Fig. 6) and plot the results of successive samples as shown (Krebs 1999:312). We must choose a level of significance for our test (e.g., $\alpha = 0.10$) and a power for the test ($1-\beta = 0.90$) and specify an upper rejection region (> 10% bucks) above, which we assume the population has not been adversely impacted by buck-only harvests. Once an initial sample of 50 deer has been obtained, sequential groups of deer encountered are added and totals plotted on the graph until the line crosses one of the upper or lower lines or the stopping rule is reached. For example, (Fig. 6) the lower rejection line was reached at a sample size of 140. At this point the null hypothesis that bucks were >5% would be rejected and the conclusion would be there were <5% bucks remaining. An important constraint is the sample must be distributed throughout the entire population so that a simple random sample of deer groups is obtained. This would be most feasible using aerial surveys from helicopter or fixed-wing aircraft.
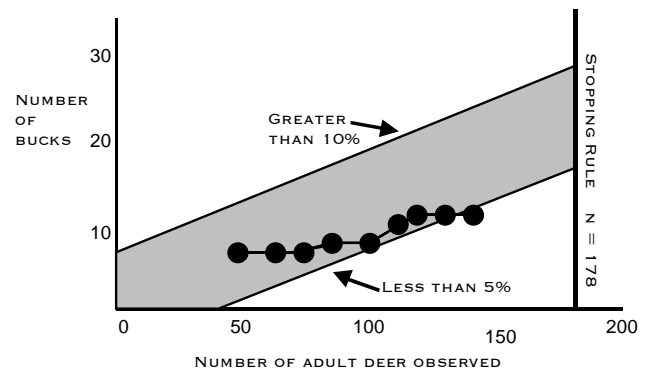

Fig. 6. Sequential sampling for percentage bucks in a deer herd.

Other Sampling Methods.--Many other sampling designs are available. For example, two-stage cluster sampling involves surveying only a portion of the members of each cluster drawn in the sample. This approach is efficient when clusters are large. Cluster sampling is one version of the more general method referred to as ratio estimation (Cochran 1963). Related methods are regression estimation and double sampling (Scheaffer et al. 1996), which have wide potential for application to wildlife research. The interested reader should consult a standard reference on sampling techniques (Scheaffer et al. 1996) and work with a statistician experienced in survey sampling.

Sampling Methodology

Plots.--Plots are used widely to sample habitat characteristics and count animal numbers and sign. Plots represent small geographic areas (circular, square, or rectangular) that are the elements of the geographically defined population. The research population size is the number of these geographic areas (plots) that would cover the entire study area. Sufficient time, money, and personnel to study an entire area are usually not available, and a subset of plots is used with the assumption that it is representative of the area. Any of the survey designs (simple random, systematic, stratified random, cluster, etc.), or more complicated designs such as 2-stage designs may be applied (Cochran 1963). Selecting the best design requires insight into characteristics and patterns of distribution of species across the landscape. One advantage of this approach is that size of the population is known and totals can be estimated (Seber 1982). Selection of plot size and shape, also an important consideration, has been reviewed by Krebs (1999).

Point Sampling.--In point sampling, a set of points is established throughout the population and measurements are taken from each sample point (Fig. 5F*)*. A common measurement is distance from the point to a member of the population (e.g., plant or calling bird). Examples include point quarter and nearest neighbor methods used widely to estimate density of trees and shrubs (Mueller-Dombois and Ellenberg 1974), and the variable circular plot method of estimating songbird density (Reynolds et al. 1980). If observers doing point counts for birds record the distance to each bird detected, as in the variable circular plot approach, transforming distances to areas makes it easy to apply the extensive methods and algorithms developed for line transects (Buckland et al. 1993, Laake et al. 1994). Selection of sample points usually follows a systematic design, but other sample designs can be used as long as points are spaced sufficiently that few members of the population are sampled more than once. Necessary sample size can be estimated from formulas even if population size is assumed to be large or unknown (Zar 1999).

Transects.--A transect is a straight line or series of straight-line segments placed in the area to be sampled. Transects are used to organize or simplify establishment of a series of sample points or plots, and as a sample unit themselves. Transects are used widely to obtain systematic samples of spatially distributed populations (e.g., plants). In these situations, plots along transects are actual sample units (Fig. 5G), and should be treated as described under systematic sampling. Plots can also be placed along transects at random intervals. When transects are used as sample units, they are commonly referred to as line transects (Burnham et al. 1980). Measurements of perpendicular distance, or sighting distance and angle, to the sampled elements (e.g., flushing animals, groups of animals, carcasses, snags, etc.) are recorded (Fig. 5H). These distances are used to estimate effective width of the area sampled by the transect (Seber 1982, Buckland et al. 1993). Each transect is treated as an independent observation, and transects should be non-overlapping according to established sampling designs (e.g., simple random, systematic, stratified random). Transects are often easier to establish in rough terrain than are plots, but they must be established carefully with compass or transit and measuring tape or with a GPS unit. Use of transects is becoming more widespread in aerial survey work because of development of precise navigational systems (Patric et al. 1988, Anthony and Stehn 1994). The critical assumptions for transect methods for sampling mobile objects such as animals (i.e., 100% detection for objects directly on the line, no movement toward or away from the observer before detection) must be examined carefully before this sampling method is selected (Burnham et al. 1980). In certain cases, more-sophisticated methods may be used to adjust counts for less-than-perfect detection on the line (Buckland et al. 1993, Manly et al. 1996, Quang and Becker 1996). A strip transect appears similar, but it is really a long, thin plot, because the method assumes all animals or objects in the strip are counted (Krebs 1999).

Road Sampling.--Sampling from roads is a widely used method for obtaining observations of species sparsely distributed over large areas, or for distributing observations of abundant species over a large geographic area. This sampling method is usually the basis for spotlight surveys of nocturnal species such as white-tailed deer (Boyd et al. 1986), black-tailed jackrabbits (*Lepus californicus*) (Chapman and Willner 1986), brood and call counts of upland game birds (Kozicky et al. 1952), scent-station surveys (Nottingham et al. 1989), and the Breeding Bird Survey (Robbins et al. 1986). This approach involves drawing a sample from a population defined as that population occupying an area within a distance x of a road (Fig. 5I). The distance x is generally unknown and varies with any factor that would affect detection of an animal, such as conspicuousness, density, type of vegetation cover, or background noise for surveys based on aural cues.

Roads rarely provide unbiased estimates for a region because they are generally placed along ridges or valleys and avoid steep or wet areas. Further, roads modify habitat for many species and may attract some wildlife. For example, during snow periods some bird species will come to roads for grit and spilled grain. Thus, sampling along roads rarely provides a representative sample of habitat (e.g., Hanowski and Niemi 1995). Although this bias is well known, it is often ignored in exchange for the cost efficient and easy method. As with all indices, every effort should be made to standardize counting conditions along fixed, permanently located routes (Caughley 1977); however, this alone does not guarantee reliable counts (Anderson 2001, Thompson 2002). Sampling along roads can be an efficient approach if it is designed as a random sample from a stratum adjacent to roads that is one element of a stratified random sample of the entire area including other strata distant from roads (Bate et al. 1999).

Dependent (Paired) And Independent Observations.--If we wish to make population comparisons, pairing observations is a powerful tool for

detecting differences. If there is a correlation between members of a pair, treating them as dependent or paired observations can improve the power of tests for differences. For example, to compare diets of adult female mountain sheep (*Ovis canadensis*) and lambs, we might treat a ewe with a lamb as a pair and measure the diet of each animal by counting the number of bites of each plant they eat while foraging together. Treating these observations as pairs would sharpen comparison between age classes because it would compare animals foraging together and experiencing the same availability of plants. Pairing is a powerful technique in other contexts in which there is dependency between the observations. Pairing should be used only if an association really exists, otherwise power of comparison will be decreased.

Pairing also can be used to help answer a different question. For example, studies of habitat selection are often made by locating areas used by a species (i.e., nest sites or radio locations) and measuring habitat characteristics at these use sites with sample plots. Available habitats are measured from random sample plots throughout the study area (Fig. 7A). A comparison of use and random plots is used to identify characteristics of areas selected by the species. An alternative approach involves pairing use and random plots by selecting a random plot within a certain distance of the use plot (Fig. 7B). For analysis, use and random plots are paired (i.e., random plot locations are dependent on use sites). This comparison could produce different results from the unpaired comparison because it would be testing for habitat differences within areas used by the species (microhabitat selection), whereas the unpaired comparison (e.g., independent plots) would be testing for habitat differences between areas used by the species and typical habitat available within the general study area (macrohabitat selection). Choosing a paired or unpaired design will depend on the objectives of the study but both may be useful in applying a hierarchical approach to studying habitat selection (Wiens 1973, Johnson 1980).

STUDY DESIGN

Hypothesis testing incorporates one or more of 4 basic research options: field studies, natural experiments, field experiments (Hurlbert 1984), and laboratory experiments (Fig. 2). Wildlife field studies are common, but interpretation of the results has severe limitations. Experiments span a continuum from natural experiments over which we have no control to completely controlled laboratory experiments (Table 2).

Field Studies

Field studies, sometimes called mensurative or observational experiments, are similar to experiments in that they are conducted to test hypotheses, but they differ because treatments are not assigned at random. Making inferences from field studies is difficult because we make *ex post facto* (Kerlinger 1973) or after-the-fact
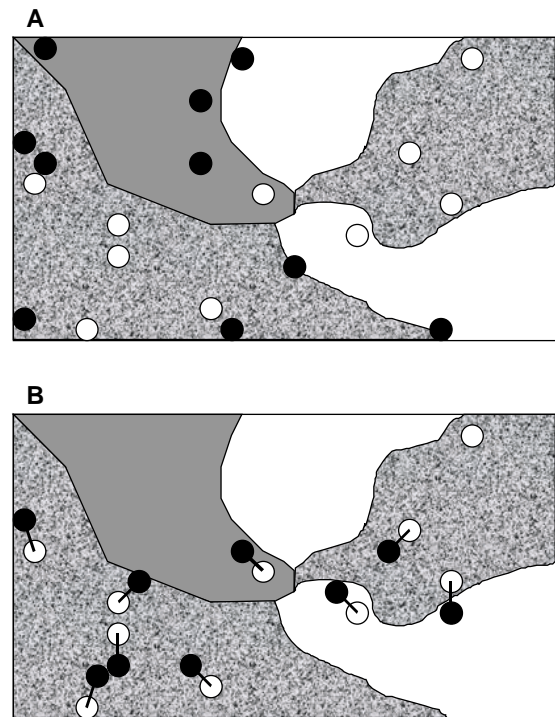

Fig. 7. Illustrative examples of (A) use (◯) and random plots (●), and (B) use plots paired with random plots.

comparisons between groups. Drawing firm conclusions is difficult, because these groups also differ in many other aspects. For example, in a field study of dietary selection by Canada geese we might randomly select plots where flocks of geese have fed and those where they have not fed to examine if geese chose areas with vegetation that is more nutritious. If they did, a weak inference would be that geese are choosing nutritious food, but numerous alternative explanations remain untested (e.g., maybe geese preferred hill-top sites where visibility was good, and coincidentally these were also sites farmers fertilized most heavily to compensate for wind-soil erosion from previous years of tillage). The important aspect of a field study is that we have comparison groups (e.g., fed upon vs. non-fed upon plots), but we have no treatments. Well-designed field studies can make important contributions to wildlife science and management (e.g., Paltridge and Southgate 2001), but their limitations must not be overlooked. Many statisticians strongly object to performing hypothesis tests on observational data and recommend alternative approaches for evaluating the data (Cherry 1998, Johnson 1999, Anderson et al. 2000).

Natural Experiments

Natural experiments are similar to field studies except that we study the effects of **uncontrolled** treatments such as wildfires, hurricanes, mass mortality from diseases, agricultural practices, and range expansions by animals or plants. A key problem in evaluating natural experiments is that we cannot assign treatments randomly.

Table 2. Strengths and weaknesses of different types of experiments (modified from Diamond 1986).

	Laboratory Experiment	Field Experiment	Natural Experiment
Control of independent variables[1]	Highest	Medium	Low
Ease of inference	High	Medium	Low
Potential scale (time and space)	Lowest	Medium	Highest
Scope (range of manipulations	Lowest	Medium	High
Realism	Low	High	Highest
Generality	Low	Medium	High

[1]Active regulation and /or site matching.

In natural experiments the treatment precedes the hypothesis and most comparisons must be made after the fact. In laboratory and field experiments the treatment follows the hypothesis. Many hypotheses of interest to wildlife biologists can be tested only with natural experiments, yet it is difficult to draw inferences from such experiments. The applied nature of wildlife management makes the realism and generality of natural experiments an important advantage. With our Canada goose example, a natural experiment might be to survey farmers in the region to locate pastures that have been fertilized and those that have not been fertilized in recent years. If our observations of feeding geese show more use of pastures that had been fertilized, we have more evidence indicating they select more-nutritious forage. However, many alternative explanations remain. For example, perhaps those pastures that were fertilized were grazed later in the summer, and geese preferred fields with the shortest grass where ability to detect approaching predators was greatest.

Field Experiments

Field experiments offer advantages over natural experiments in terms of ease of inference and control but disadvantages of restricted scale and lower generality (Table 2). Compared to laboratory experiments, field experiments have greater scope and realism. Their main advantage is that we can randomly assign treatments. In field experiments, manipulations are conducted, but other factors are not subject to control (e.g., weather). In many situations in wildlife science, field experiments offer the best compromise between the limitations of laboratory experiments and natural experiments (Wiens 1989). In our Canada goose example, a subsequent field experiment would be to select random pairs of plots in known foraging areas. One member of each pair would be randomly assigned to be fertilized to learn if geese select the fertilized plots more than nonfertilized control plots. If they did, a stronger inference about selection of nutritious foods could be made, because random assignment of a large number of plots to fertilization and control groups should have canceled the effects of extraneous confounding factors. Interspersion of treatment and control plots (Hurlbert 1984, Johnson 2002) in fields

naturally used by geese strengthens our belief that our conclusion would apply in systems where geese typically forage.

Impact Assessment

The most basic form of impact assessment compares measurements of wildlife and other characteristics at a site potentially affected by development to similar measurements at an unaffected reference site (Fig. 2). This most simple form of impact assessment provides almost no basis for inference because the reference site may differ for a multitude of reasons besides absence of the development. Green (1979) noted the potential improvement in this design that results from making measurements before and after development at both reference and development sites. The basic before-after/control-impact (BACI) design has become standard in impact assessment studies (Morrison et al. 2001) and has also been used in predator-removal studies (e.g., Risbey et al. 2000). However, differences from before to after at reference (control) and impacted (treatment) sites is confounded by natural variation and may not be produced by the impact itself (Hurlbert 1984, Underwood 1994). In contrast to a well-designed field experiment, neither reference nor impacted site is chosen randomly over space and the treatments are not assigned randomly. These severely limit certainty of conclusions and inferences to other areas. The goal is not to make inferences to all possible sites (Stewart-Oaten et al. 1986) for a power plant, for example, but to the particular power plant site developed. For larger impact studies where the goal is to make inferences with more certainty and to more sites (Fig. 2), the basic BACI design must be improved through addition of replication and randomization (Skalski and Robson 1992, Underwood 1994). Stewart-Oaten et al. (1986) emphasized the value of expanding the BACI design to include temporal replication and noted the advantage of taking samples at nonregular time intervals rather than on a fixed schedule. Hurlbert (1984) emphasized that comparing abundances of wildlife from repeated surveys at one impact and one reference site constitutes pseudoreplication that is only eliminated by having several replicated impact and reference sites. Replicated reference sites with environmental

characteristics similar to the impact site are quite possible and highly desirable; however, replicated impact sites are only feasible in large-scale impact studies, typically involving meta-analysis of many single impact-site studies.

Laboratory Experiments

Drawing inferences from laboratory experiments is easy because of the high level of control, yet this advantage must be weighed against their disadvantages (Table 2) in terms of (1) scale --laboratory experiments are restricted to small spatial scales and short time periods, (2) scope --only a restricted set of potential manipulations is possible in the laboratory, (3) realism --the laboratory environment places many unnatural stresses and constraints on animals, and (4) generality --some laboratory results cannot be extrapolated to natural communities. Laboratory experiments in biology have been most useful in studying basic molecular or biochemical processes common to all organisms of a class. Laboratory experiments have also provided valuable information on emerging issues such as wildlife diseases (e.g., Cooke and Berman 2000), efficacy of fertility control (Chambers et al. 1999), and interactions between exotic and native species (e.g., Komak and Crossland 2000). In a continuation of our example, laboratory experiments could be designed to examine if geese really can select the most nutritious forage when given several alternatives in a cafeteria-feeding trial.

Identifying one research option as best for all situations is not possible. All options should be considered as possibilities when a hypothesis test is designed. Sometimes the best evaluation of a hypothesis involves using a combination of field studies and several types of experiments. For example, Takekawa and Garton (1984) obtained field observations of birds feeding heavily on western spruce budworms (*Choristoneura occidentalis*) during a budworm outbreak, which suggested that birds were a major source of budworm mortality. Field experiments were conducted to test this hypothesis by placing netting over trees to exclude birds. Survival of budworms on trees with netting was 3-4 times higher than on the control trees exposed to bird predation (Takekawa and Garton 1984). Field observations by Ratti et al. (1984) indicated spruce grouse fed exclusively on certain trees while ignoring numerous other similar trees of the same species. This led to a laboratory experiment with captive birds by Hohf et al. (1987) that tested the hypothesis that trees selected for feeding had higher nutritional content than random trees. Diamond (1986) provided examples of the 3 types of experiments and excellent suggestions for improving each type. Other examples and discussion of experiments were provided by Cook and Campbell (1979), Milliken and Johnson (1984), Kamil (1988), Hairston (1989), and Underwood (1997).

Integrated Research Process

The integrated research process (Fig. 2) builds on a solid base of natural-history observations. Field observations should lead to experiments, and the results of natural experiments should lead to field and laboratory experiments. The level of certainty increases as many predictions from the research hypothesis are supported and alternate hypotheses are rejected in successively more rigorous tests that use replicated research options. After such findings are repeated over broad geographic areas or throughout the range of the species, the research hypothesis may become a principle of wildlife science (Johnson 2002). The integrated research process should be the goal of wildlife science.

Outstanding examples of integrated research programs include long-term research on red grouse (*Lagopus lagopus scoticus*) in Scotland (Jenkins et al. 1963, Watson and Moss 1972, Moss et al. 1984, Watson et al. 1994), red deer on the Isle of Rhum (Scotland) (Lowe 1969, Guinness et al. 1978, Clutton-Brock et al. 1985, Coulson et al. 1997), and snowshoe hares in North America (Keith 1963, 1974; Windberg and Keith 1976; Keith and Windberg 1978; Keith et al. 1984). Research on red grouse and snowshoe hares has focused on hypothesized causes of population cycles, while research on red deer has focused on population regulation and density-dependent effects on survival, fecundity, reproductive success, spacing behavior, and emigration. In all 3 examples, descriptive studies and field observations formed the groundwork for subsequent research that included a series of innovative field studies and experiments (natural, field, and laboratory).

Preliminary studies of red grouse in Scotland (Jenkins et al. 1963) provided information on fundamental population parameters: births, deaths, immigration, and emigration. This information was used to form research hypotheses about causes of population fluctuations. Postulated causes initially included food quality, breeding success, spacing behavior, and genetics (Watson and Moss 1972). Using data from long-term field studies coupled with field and laboratory experiments, Watson and Moss (1972) concluded that quality of spring and summer foods [heather (*Calluna vulgaris*) shoots and flowers] affected egg quality, breeding success (viability of young), and spacing behavior of males and females, but territory size ultimately affected recruitment and population density [but see Bergerud (1988) for a critique of the self-regulation hypothesis and inferences based on red grouse research].

The level of certainty increases as predictions from a research hypothesis are supported and alternate hypotheses are rejected via replicated research and successively more rigorous observational, experimental, and modeling tests. Watson, Moss and co workers took these findings into innovative field experiments in which they 1) fertilized fields to assess grouse response to increased nutritional quality of the heather (Watson et al. 1984*b*) and 2) implanted males with time-release hormones to monitor changes in territory size associated with aggressiveness induced by higher or lower levels of androgens and estrogens (Watson 1967). Additional and more rigorous research rejected hypotheses that nutrition,

genetics, and parasitism were causal factors (although Dodson and Hudson [1992] make a counter argument for the role of the parasite *Trichostrongylus tenuis*), and instead focused on emigration as the key factor in population declines (Moss et al. 1984, Watson et al. 1984*a*, Moss et al. 1990). These findings led to more research because the mechanisms underlying density-dependent relationships, including summer and winter emigration, were unclear. Recent research has focused on the hypothesis of kin selection and differential aggression between kin and non-kin to explain cyclic changes in red grouse (Moss and Watson 1991, Watson et al. 1994) and synchronization of cycles across large regions by weather (Watson et al. 2000). Thus, the integrated research process continues.

A Checklist for Experimental Design

The design of any experiment must be developed carefully or the conclusions reached will be subject to doubt. Four particularly critical elements in the design of a manipulative experiment are: (1) specification of the research population, (2) replication, (3) proper use of controls, and (4) random assignment of treatments to experimental units. An experimental-design checklist is useful to provide a series of questions to assist in addressing these critical elements. Many of the questions will be helpful with design of data gathering for studies involving nonexperimental hypothesis testing. Some experimental designs may address several hypotheses simultaneously; in other designs, each hypothesis may require independent experimental testing.

1. **What is the hypothesis to be tested?** The hypothesis developed from the conceptual model must be stated clearly before any experiment can be designed. For example, we could test the hypothesis that nest predation on forest songbirds is higher at sharp edges, such as occur at typical forest clearcuts, than at feathered edges (partial timber removal), such as occur at the boundary of selectively-logged areas (Ratti and Reese 1988).

2. **What is the response or dependent variable(s) and how should it be measured?** The response variable should be clear from the hypothesis (e.g., nest predation), but selecting the best technique to measure it might be more difficult. We must consider all possible methods and identify one that will simultaneously maximize precision and minimize cost and bias. It is often helpful to contact others who have used the techniques, examine the assumptions of the techniques, and conduct a pilot study to test the potential techniques. In our example, we might search for naturally occurring nests along forest edges and use a generalized Mayfield estimator (Heisey and Fuller 1985) of mortality rate. This response variable is continuous and we could apply any of a variety of designs termed general linear models (GLM) (e.g., ANOVA, linear regression, analysis of covariance). Alternately, we could measure our response for each nest as successful (at least one young fledged) or unsuccessful and use appropriate analysis methods such as chi-squared statistics applied to

contingency tables or log-linear models (Fienberg 1970, 1980).

3. **What is the independent or treatment variable(s) and what levels of the variable(s) will we test?** The independent variable(s) should be clear from the hypothesis (sharp and feathered forest edges in our example), but selecting levels to test will depend upon the population to which we want to make inferences. If we want to test the effects of our independent variable at any level, we must select the levels to test at random (random effects or Model II ANOVA, Zar 1999). If we are interested in only a few levels that our independent variable could take, we use only those levels in our experiment and make inferences only to the levels tested (fixed effects or Model I ANOVA, Zar 1999). For example, if we wanted to evaluate effects of forest edges of any type on predation rates, we would select types of forest edges at random from all types that occur and apply a random effects or Model II ANOVA to analyze the data. In our example we are interested only in the 2 types categorized as sharp and feathered. Additionally, our independent variable must be identified and classified clearly or measured precisely. Finally, how can we use controls to expand our understanding? In our example, comparing nest predation in undisturbed forests to predation at the 2 types of edges might be enlightening and we would analyze the data with fixed-effects or Model I ANOVA. Our final conclusions would not apply to predation rates in all types of forest edges but only to the 2 types that we compared to undisturbed forest. An alternative approach to the design would be to treat the independent (treatment) variable as being continuous and use regression for the analysis rather than ANOVA. Under this design we might specify the treatment would consist of some level of overstory removal on one side of the forest edge and we would apply regression forms of GLM. The response could be measured as the difference in predation rates between the 2 sides of the boundary, which would be regressed on percent of overstory removed. Here it becomes critical to select treatment levels (percent overstory removed) across the full range of forest treatments to which we want to apply our conclusions.

4. **To which population do we want to make inferences?** If the results of the experiment are to be applied to the real world, our experimental units must be drawn from some definable portion of that world, the research population. The dependent and independent variables chosen should define the relationship(s) examined and place constraints on the definition of this population. Finally, we must consider the impact of potential extraneous factors in selecting the population of interest. If the population is defined so broadly that many extraneous factors impact the results, the variation might be so large that we cannot test the hypothesis (low internal validity). If the population is defined so narrowly that we have essentially a laboratory experiment, application of the results might be severely limited (low generality or external validity).

Reaching the proper balance between internal and

external validity takes thought and insight. For example, we might want to compare nest-predation rates in sharp and feathered forest edges throughout the northern Rocky Mountains, but the logistics and cost would make the study difficult. Thus, we might restrict our population to one national forest in this region. Next we need to consider the types of forests. We might want to test the hypothesis for the major forest types, but we know the species of birds nesting in these forests and their nest predators differ among forest types. Thus, we may need to restrict our population to one important type of forest to remove extraneous factors that could impact our results if we sampled a large variety of forest types. We need to ask what types of sharp and feathered edges occur to decide which ones we will sample. Sharp edges are commonly produced by clearcuts, power line rights-of-way, and road rights-of-way. These 3 types differ dramatically in factors such as size, shape, human access, and disturbance after treatment. Additionally, our ability to design a true experiment involving random assignment of treatments is severely limited for all but the clearcuts. Therefore, we might restrict our populations to sharp edges created by clearcuts and feathered edges created by selective harvests.

5. **What will be our experimental unit?** What is the smallest unit that is independent of other units, which will allow random assignment of a treatment? This must be identified correctly or the resulting experiment might not have true replication, but represent a case of pseudoreplication (Hurlbert 1984). For example, we might erroneously decide the experimental unit for our nest-predation study will be an individual nest. The resulting design might entail selecting 3 areas and randomly assigning them to be clearcut, control, and the other to be selectively logged. By intensive searching we find 20 nests along the edge of each area and monitor them for predation. The resulting data would suggest 20 replicates of each treatment but, in fact, only a single area was given each treatment. Only one area was randomly assigned each treatment and the 20 nests are subsamples. Thus, pseudoreplication restricts our potential inferences. In effect, we have sampled from populations consisting only of 2 logged areas and one unlogged area, and our inferences can be made only to those 3 areas, not to clearcuts, selective cuts, or undisturbed forests in general.

In some situations, pseudoreplicated designs are unavoidable, but interpretation of their results is severely restricted because, without replication, confounding factors rather than the treatment could have caused the results. For example, in our nest predation experiment if one of the areas was within the home range of a pair of common ravens (*Corvus corax*) and the other areas were not, this single confounding factor could affect the results regardless of which treatment was in which area. A more reliable experiment would require that we identify several areas with potential to be logged, perhaps 15, sufficiently far apart to be independent of each other, and that we randomly assign 5 to be clearcut, 5 to be selectively harvested, and 5 to be controls. We would locate and monitor several nests in each area. The nests in a single

area would be correctly treated as subsamples and their overall success treated as the observation for that area. This approach attempts to remove effects of confounding factors and allow development of a conclusion with general application to the populations sampled, i.e., edges created by clearcuts and selective cuts within this habitat type in this region. Including control stands without an edge provides invaluable information for assessing the biological significance of the difference between the 2 types of edges.

6. **Which experimental design is best?** A few of the most widely used designs are described, but we advise consulting texts on experimental design and a statistician before making the final selection. The choice depends primarily upon the type of independent and dependent variables (categorical, discrete, or continuous), number of levels of each, ability to block experimental units together, and type of relationship hypothesized (additive or with interactions). For our study of nest predation along 2 types of forest edges, a single-factor design would be appropriate, but Hurlbert's (1984) argument for interspersion of treatments and controls could be incorporated using a more sophisticated design. For example, 3 adjacent stands in 5 different areas might be randomly assigned to treatment and controls, with areas cast as blocks resulting in a randomized complete blocks design (Zar 1999).

7. **How large should the sample size be?** Estimating sample size needed for proper analysis is essential. If the necessary sample size were too costly or difficult to obtain, it would be better to redesign the project or work on a different question that can be answered. Sample size depends upon the magnitude of the effect to be detected, variation in the populations, type of relationship that is hypothesized, and desired power for the test. Typically some preliminary data from a pilot test or from the literature are required to estimate variances. These estimates are used in the appropriate formulas available in statistical texts (e.g., Zar 1999) and incorporate a prospective power analysis through which we are assured that we have a high (80-90%) chance of detecting biologically meaningful differences between our treatment and control categories.

8. **Have you consulted a statistician and received peer review on your design?** Obtaining review by a statistician before the data are gathered is essential. The statistician will not be able to help salvage an inadequate design after a study is completed. Peer review by other biologists with experience on similar studies could also prevent wasted effort if measurements or treatments are proposed that will not work on a large scale in the field. Now is the time to get these comments!

Single-factor vs. Multifactor Designs

Single-factor analyses are the simplest because they involve only comparisons between 2 or more levels of one factor. Evaluating the simultaneous effect of 2 or more independent variables (factors) at once requires use of

complicated statistical methods, which should be discussed with a statistician. Under many conditions we can test 2 factors at once without expending more effort than would be required to test either of the factors alone. A complicating issue is the potential for interaction between factors (Steel and Torrie 1980). An interaction occurs if the effects of one factor on the response variable are not the same at different levels of the second factor. For example, if we are interested in the effect of snowmelt date on nest success by arctic nesting, polymorphic snow geese (*Chen caerulescens*), we might discover an interaction between color phase and the onset of spring snow melt. Thus, darker blue-phase birds would have higher nesting success during early snowmelt years because they are more cryptically colored once snow has melted and experience less nest predation. During late snowmelt years white-phase birds are more cryptically colored and experience less nest predation. Many observations might be required to clarify possible relationships in these situations.

Dependent Experimental Units

Special designs have been developed to handle many types of dependency in experimental units. A common design involves pairing. In a paired design we match experimental units in pairs that are as similar as possible. The treatment is then applied to one member of each pair at random. If there is a confounding factor, which we succeed in matching in the pairs, this approach will lead to a more powerful test than if pairing is not performed. For example, if we were studying the effects of spring burning on bobwhite habitat, we could place pairs of plots throughout our study area, being careful to place each pair in a homogeneous stand of vegetation. We would then randomly assign one member of each pair to be burned in spring. The analysis would then examine the differences between the members of a pair and test for a consistent improvement or decline in the burned portion of the pair. Pairing would remove the effects of vegetation difference from one part of the study area to another and result in a more sensitive experiment. If members of pairs are not more similar than members of the general population, the test will be less powerful because of the pairing.

When more than 2 levels of a factor are compared, pairing is referred to as blocking. A block is a set of similar experimental units. Treatments are randomly assigned to units within each block, and the effectiveness of blocking can be tested during the analysis. For example, if we expanded our study of burning to include spring and autumn burning as treatments, a block design would be appropriate. Three adjacent plots would be placed in homogeneous vegetation stands, and spring and autumn burning would be applied randomly to 2 of the 3 plots. The analysis would entail a randomized block ANOVA.

Another common form of dependency occurs when repeated measurements are taken on the same experimental unit through time. This is common in wildlife research wherein the effects of treatments may change over time and must be monitored over a series of years. For example, in our study of spring and autumn burning the effects may be different in the first, second, and third growing seasons after treatment. The plots should be monitored over several years to measure these effects. The measurements are repeated on the same plots, so they are not independent. This must be treated correctly in the analysis by using repeated measures or multivariate analysis of variance (Milliken and Johnson 1984, Johnson and Wichern 1988). Dependency is also common in count data, especially when animals occur in groups (Eberhardt 1970). This lack of independence is often referred to as over dispersion. To properly cope with significant over dispersion the dependency should be modeled. Unless the biologist has extensive training in this topic, close cooperation with a consulting statistician is essential in designing and analyzing experiments involving such complicated designs.

Crossover Experiments

Crossover experiments provide a powerful tool to evaluate treatments that do not produce a long-lasting effect. Selecting pairs of experimental units and randomly assigning one member of each pair to be treated during the first treatment period initiates a crossover experiment. The second member of each pair serves as the control during this treatment period. In the second treatment period, the control unit becomes the treatment and the former treatment becomes the control. In this way the effects of any underlying characteristics of experimental units are prevented from influencing the results. This technique is valid only if treatment effects do not persist into the second treatment period.

Consider the following example. Suppose we wanted to test the hypothesis that mowing hay before 4 July decreases ring-necked pheasant (*Phasianus colchicus*) nest success. We could test this by dividing our study area into 5 homogeneous hayfield regions and then dividing each region into 2 portions. In one randomly selected portion of each region we could pay farmers not to mow their hay fields until after 4 July (treatments). In the other portion of each region, hay mowing would proceed as in most years, with the first cutting during mid-June; these portions will serve as controls. To monitor nest success, we locate nests by systematic field searches, being sure to search treatment and control areas with identical methodology, e.g., search intensity and seasonal timing. Nest success will be measured with standard techniques. After one year, we might measure significantly higher nesting success in the treatment portions, i.e., those areas with delayed hay mowing. However, the number of treatments is small and we are not able to conclude with confidence if higher nest success resulted from the treatment or from some undetected, inherent differences in treated portions of each region, such as nest predators. We implement the crossover experiment by switching in the

second year so the original control portions of the study regions now have mowing delayed until after 4 July (new treatments), and the original treatment portions revert to the standard practice of first cutting in mid-June (new controls). If the portions with late cutting treatments again have higher nest success, we have better evidence that delayed mowing is responsible for higher nest success than we had at the end of the first year (i.e., we have better evidence for a cause-and-effect relationship). If even stronger support for the hypothesis is desired, the crossover experiment might be repeated in the same region and in other farming regions.

Fixed, Random, Mixed, and Nested Effects

One of the most critical decisions we must make in design concerns choosing the population to which we want to make inferences. If only a few levels of a treatment factor are relevant or would occur in the future, we set a limited number of values at which the treatment would be applied and the factor is termed a fixed effect or Model I ANOVA. If we want the conclusion to apply to any level of a treatment factor, we must select the treatment levels as a random sample from the population of potential values so a conclusion drawn about the effect of this factor applies across all levels at which it occurs. This design is termed a random effect or Model II ANOVA. A mixed model (Model III) includes both fixed and random effects. In simple 2-factor or multi-factor designs all levels of each factor are applied to all levels of other factors and the design is considered to be a crossed design. When this is not possible, the design must use approaches in which one factor is nested within another factor. A nested design can be described as a hierarchical design, which occurs most commonly where certain levels of one factor only occur within some of the levels of another factor. For example, a study evaluating the effect of vegetation treatment on bird communities might have 3 plant communities with treatments of clearcut, burn, partial-cut, and controls. These factors would need to be nested if one of the plant communities was a shrub community where timber harvest does not occur. Decisions about design of experiments must be reflected correctly in the analysis, as different measures of variance are appropriate for fixed, random, mixed, or nested effects.

ALTERNATIVES TO HYPOTHESIS TESTING

Many of the questions that wildlife biologists and managers wish to answer are not amenable to experimental manipulation and testing. Answers to these questions rely on field studies and natural experiments (Fig. 2). Applying hypothesis testing to such comparisons is invalid under a frequentist philosophy (Fig. 8, Efron 1998), but treating them as quasi-experiments (Cook and Campbell 1979) is done widely in the social sciences. Unfortunately the strength of causal inferences is weak because of high potential for confounding other factors with the measured "treatment" or comparison factor. Alternatives to

hypothesis testing are available and have recently become more accepted and feasible through development of powerful computer algorithms.
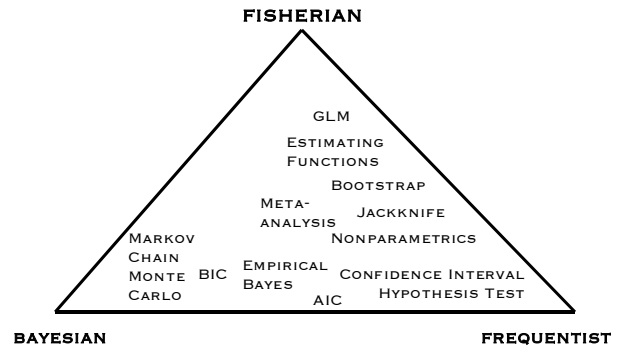

Fig. 8. Selecting analysis methods from 3 dominant statistical philosophies (GLM = General Linear Model, BIC = Bayesian Information Criterion, AIC = Akaike's Information Criterion) (modified from Efron 1998).

Regression and General Linear Models

One of the most flexible approaches to identifying predictive and potentially causal relationships between wildlife and environmental or management characteristics involves use of ordinary least squares to estimate parameters of regression or general linear models (GLM, Fig. 8). Experimental manipulations that produce different levels of predictor variables are more readily analyzed by ANOVA, regression, or analysis of covariance versions of general linear models under a Fisherian philosophy (Fig. 8), named after R. A. Fisher who pioneered a "spirit of reasonable compromise, cautious but not overly concerned with pathological situations" (Efron 1998:99) in the analysis of experiments. Designing a study to gather data on a variety of potential causal variables rather than manipulating those variables through a designed experiment is an appealing alternative but yields inferences of much lower certainty (Fig. 2). Performing hypothesis tests on such data (e.g., testing point null hypotheses) is easily performed with modern regression programs. However, it may not be justified as an inferential approach and may readily lead into a "fishing-expedition" doomed to failure due to high Type I errors. Many statisticians refuse to analyze such data using hypothesis tests and instead encourage biologists to apply maximum likelihood and information-theoretical approaches under a modeling perspective, i.e., identifying the most parsimonious model with good predictive ability (Milliken and Johnson 1984, Anderson et al. 2000, Burnham and Anderson 2002).

It is essential in designing manipulative or observational studies to estimate linear models to strive to obtain observations throughout the full range of the predictive variables. It is especially important to obtain observations at both low and high values of the predictive variable because these set limits for the range of values that can be used later for prediction. The values at the

ends of this range have the most leverage on slope estimates. If too narrow a range is measured, a significant relationship may not be detected among the variability. However, a relationship may be linear only through a portion of its range such that beyond a certain level an increasing effect may turn into a negative effect at progressively higher levels. Such situations should be apparent from exploratory data analyses (Anderson 2001, Johnson 2002).

Model Building vs. Hypothesis Testing

Rigorously evaluating ideas concerning wildlife habitats and populations using experimental manipulations may be difficult because we cannot randomly assign treatments and the high cost of treatments precludes adequate replication. However, modeling methods (Shenk and Franklin 2001) and information-theoretic approaches to evaluating competing models can be useful (Burnham and Anderson 2002). A biologist's goal should be to build the simplest model that describes the relationships between causative factors and the effects they produce. The biologist or manager is most likely to select a modeling strategy at the simple, empirical ends of the continua in terms of model objectives and complexity (Table 3). Long-term monitoring data provide the basis for more complex models. In most cases the goal is to model the responses of wildlife populations or habitats with the smallest number of predictors necessary to make good predictions. Likelihood methods are used to compare how well a number of potential models perform against each other. Akaike (1973) developed one of the most popular measures, Akaike's Information Criterion (AIC). This criterion penalizes the likelihood of the model based on how many parameters are required to estimate from the observed data. The model with the lowest AIC (Burnham and Anderson 2002) is the most parsimonious model that gives the best predictions. Burnham and Anderson (2002) contend that information theoretic methods, such as using AIC to assess the information content of a model, should be applied where we cannot experimentally manipulate causes or predictors. Unfortunately, information theoretic approaches do not provide guidance for key issues during the design phase such as necessary sample sizes, which are best addressed by a prospective power analysis. After data are gathered, AIC does not identify how well any model performs but does provide an objective basis for choosing the best model among the models evaluated. One of the limitations of using AIC is that the results are dependent upon the models chosen for the analysis, which may differ between biologists.

Bayesian Statistics

Bayesian data analysis is described as "practical methods for making inferences from data using probability models for quantities we observe and for quantities about which we wish to learn" (Gelman et al. 1995:3). One of

the primary appeals of Bayesian statistics is that after sampling a population and calculating statistics such as the mean, variance, and confidence interval for the mean, Bayesian analysis allows us to state we are 95% certain the true mean for the population is within this 95% confidence interval. Johnson (1999) provided an easily understood description of the conceptual differences between the frequentist and Bayesian approaches (Fig. 8). A Bayesian analysis requires performing 3 basic steps (Gelman et al. 1995).

1. **Specify a probability distribution for all quantities** [i.e. use prior studies and creative thinking to specify a particular "**prior**" probability for the parameter(s)]. We begin by stating what is the range of all possible values for the characteristics that we are attempting to measure and make our best guess, from earlier studies and clear thinking, if any of the values are more likely than others. This step is controversial because it introduces subjective decisions into the process and has potential for misuse if one's goal is to "cook the books" to produce a particular result (Dennis 1996). However, well-designed research should gather historical data so that knowledge is available on the probability distribution of the parameter(s) (Box 1).

2. **Use the observed data to calculate a posterior distribution for the parameter of interest as a conditional probability distribution.** This second step in Bayesian analysis follows data collection. We improve our prior guess of the value of the characteristic by combining it with the new data we gathered to state conclusively our best "posterior" guess of the value of the characteristic. This step is performed using Bayes' rule and this Bayesian estimate might be considered as a weighted average estimate based on the sample data and the assumed prior value where weights are proportional to the precision of the observed and prior values (Gelman 1995:43). As sample size increases, the Bayesian value approaches the maximum likelihood estimate and any influence of the prior probability vanishes. Markov chain Monte Carlo methods are used widely for these calculations (Fig. 8).

3**. Evaluate the fit of the model and the implications of the resulting posterior distribution.** The third step in Bayesian analysis (Gelman et al. 1995:3) consists of "evaluating the fit of the model and the implications of the resulting posterior distribution: does the model fit the data, are the substantive conclusions reasonable, and how sensitive are the results to the modeling assumptions?"

COMMON PROBLEMS

Sample Size and Power

The importance of sample size cannot be overemphasized. Sample size and experimental design are the major factors under the control of the biologist that strongly influence power of the test, i.e., the likelihood of

Table 3. Modeling strategy based on objectives (A) and complexity (B).

A. Objectives

Empirical---**Theoretical**

------------------------------Statistical estimation-------------------------- ---Forecasting --------Projection ------General

Sampling   Fixed effects models   Random effects models   Adaptive Management   Principles

B. Complexity

Simple---**Complex**

Conceptual (verbal) --Quantification---------------------------------------Quantitative

 Linear --Relationships--Non-linear

Deterministic --Variability--- Stochastic

Time-specific --Time Scale--Dynamic

Single ---Factors---Multifactor

Single Site --Spatial --Multi-site

Single Species --Number of Species ---------------------------------Multi-species

detecting a significant difference when one really occurs. Inadequate sample size usually results from (1) inadequate consideration of population variance, (2) inability to collect data (e.g., observe a rare species), or (3) insufficient funding, time, or personnel. Often a sample-size problem is overlooked initially because of failure to consider sample-size reduction throughout the study, i.e., we focus mostly on the initial sample size and not on the final sample size that represents the most important data for consideration of a hypothesis. For example, in a study of mallard (*Anas platyrhynchos*) brood movements almost 10 times as many nests were required to be found as the sample size of broods indicated because of an 89% sample-size reduction from nests located to actual brood data (Rotella and Ratti 1992*a*, *b*).

Another common problem is fairly large overall data sets that are not sufficiently similar among years (or seasons) to combine, resulting in annual sample sizes that are too small for analysis. At the beginning of a research project we often set our desired sample size based on combining data collected over several continuous years. However, if the characteristic of interest were different among years of the study, combining the data would not be valid. For example, in a study of habitat selection by red fox (*Vulpes vulpes*), habitat use might differ between mild and severe winters with heavy snow cover. In this example, combining the data would not be valid, yet the sample size in each year may be too small to detect selection (Alldredge and Ratti 1986, 1992).

A primary defense against weak tests of hypotheses is to perform a prospective power analysis at the start of the research following a pilot study. Zar (1999:559) provided simple ways to estimate the required sample size for most common sampling and experimental designs. In contrast to this essential prospective power analysis during the design phase, performing a retrospective power analysis after the data are collected, during the analysis phase, is controversial or contraindicated (Thomas 1996, Steidl et al. 1997). Retrospective power analysis is uninformative unless effect sizes are set independently of the observed effect (Steidl et al. 1997).

Procedural Inconsistency

Procedural inconsistency is another common problem that can be prevented with proper research design. Problems of this type occur from seemingly minor variations or alterations in methodology. For example, if a project is dependent upon field personnel to accurately identify songs of forest passerine birds, the data set may be biased by identification errors (Cyr 1981). In this situation, magnitude of the bias will depend upon the rate of errors by individuals, difference in the rate of errors among individuals, and relative proportion of data collected by each individual. Research methodology should be defined with great detail, and all individuals collecting data should have similar skills and knowledge of methods used (Kepler and Scott 1981). If inconsistencies cannot be eliminated through selection and training of field workers, the design must incorporate double-sampling or similar procedures to remove inherent biases (Farnsworth et al. 2002). One unfortunate aspect of biases of this type is that they are often overlooked (or ignored) as potential problems and are seldom reported in research publications.

Non-uniform Treatments

A third common bias is non-uniform treatments. This problem is illustrated by considering 2 previous research examples. In the discussion of crossover experiments, we described a 2-year study in which

mowing on the treatment areas was delayed until after 4 July. Assume that in the first year of this study, all treatment areas were mowed between 4 and 7 July, as planned. But during year 2 of the study, a 3-day rainstorm began on 4 July, and the treatment areas were not cut until 9-12 July. Although this 5-day difference in mowing of the treatment areas may seem insignificant, the impact on the results and interpretation of our experiment is really unknown–and may be serious. Thus, the second year of the experiment should be repeated. Since dates of pheasant nesting and plant growth varies from year to year in response to temperature and rainfall patterns, a better way to set the date for the mowing treatment might be based on the cumulated degree-days widely published in farm journals.

In the second example, consider again our field experiment to evaluate effects of sharp and feathered edges on nest success of forest birds. If we had used both clear cuts and road ways as sharp edges we might have hopelessly confused our treatment results because of differences in attractiveness of sharp edges near roads where carrion is an abundant attractant to generalist predators like ravens. High variability between replicates in non-uniform treatments substantially reduces our power to detect biologically significant effects.

Pseudoreplication

Pseudoreplication occurs when sample or experimental units are not independent, i.e., they are really subsamples rather than replicates, but are treated as though they are independent samples or experimental units. This is a widespread problem in field ecology (Hurlbert 1984) that should be avoided wherever possible. In manipulative experiments, experimental units are independent only if we can randomly assign treatments to each unit. In field studies, a simple test for pseudoreplication is to ask if the values for 2 successive observations are more similar than values for 2 observations drawn completely at random from the research population. If so, the successive observations are probably not true replicates and the research should be redesigned or this lack of independence must be treated correctly in the analysis through use of cluster sampling or adjustments in degrees of freedom for tests.

There must be a direct tie between the sample or experimental unit and the research population. If the research population consists of one meadow in Yellowstone National Park, then 2 or more samples drawn from that meadow would be replicates. In this example, our inferences or conclusions would apply only to that single meadow. If our research population consisted of all meadows in Yellowstone National Park, then 2 plots in the same meadow would not constitute true replicate samples. Also, repeated sampling of the same radio-marked animal often constitutes a form of pseudoreplication, e.g., if our research population consisted of moose in one ecoregion, repeated observations of habitat use by a single animal would not be true replicates (and a similar problem would

arise if 2 radio-marked animals were traveling together, thus, their habitat selection would not be truly independent). The data would have to be summarized into a single value such as the proportion of the observations in a certain habitat for statistical analysis. This would reduce sample size to the number of radio-marked moose. Treating repeated observations as replicates is strictly justified only when the individual animal is the research population. In this situation, tests for serial correlation (Swihart and Slade 1985) should be conducted to assure the observations are not repeated so frequently that they are still pseudoreplicates.

THE RESEARCH-MANAGEMENT CONNECTION

Wildlife management programs should be developed from application of scientific knowledge, i.e., we should apply scientific facts and principles resulting from research on specific topics such as population ecology, habitat selection, or behavior. Initially, this is a sound practice for development of a new management program. The logic behind formulation of a management program is similar to formulation of a research hypothesis; both provide opportunity for predictive statements. Our management prediction is that our plan of action will achieve a desired result. However, a major problem with nearly all wildlife management programs throughout the world is the lack of research on the effectiveness of programs (Macnab 1983, Gill 1985). Seldom is the question "does our management lead to the desired result?" addressed in formal, well designed, long-term research projects. For example, disparate gender ratios are common among North American mallard populations (i.e., more males than females, Bellrose et al. 1961). Our long-term management response (with monogamous species) is to set hunting regulations that direct more harvest pressure on males. Initially this management plan seems appropriate, the assumption being that we shift harvest to the "surplus" segment of the population that adds little to overall recruitment. However, several important questions should be considered. Does reduction of excess males in the mallard population affect overall recruitment? For example, unpaired males often fertilize females attempting to renest. Is there an evolutionary adaptation to disparate gender ratios? With given levels of harvest, mallard population levels may not be maximized by disproportional harvest of the male segment. No research to date has adequately addressed these questions. If these basic biological questions cannot be answered, hunting regulations to increase male harvest may not be justified because of expensive public education and enforcement problems they create.

A second common example is prescribed burning as a management practice to increase deer and elk populations. The effectiveness of this management has not been addressed, and most evaluations have only noted increases in browse forage species and changes in animal distributions. Increased population levels in response to prescribed burning have not been adequately documented

or thoroughly studied (Peek 1989).

A third example is the use of population indices to monitor changes in population levels (e.g., ring-necked pheasant crowing counts). The primary assumption for use of a population index is that the index is directly related to density. Although nearly every wildlife management agency uses trend data from population indices for management decisions, only a few rare examples of index validation exist (e.g., Rotella and Ratti 1986, Crête and Messier 1987). Some studies have disclosed that index values are not related to density (Smith et al. 1984, Rotella and Ratti 1986, Nottingham et al. 1989).

Walters (1986) proposed a systematic solution to these problems, which he called adaptive management. It involves a more formal specification of management goals and responses to management actions through the use of predictive models (Table 3), which can be compared to actual system responses through detailed monitoring (Thompson et al. 1998). Management actions are treated as experiments, which must be monitored carefully to ascertain if goals were met and to identify errors in understanding the dynamics of the natural systems that we manage. Actual responses to management actions are compared to predictions from our models based on current knowledge and assumptions (e.g., adaptive harvest management, Williams and Johnson 1995, Williams et al. 1996). Some management actions are taken to intentionally disturb the system to learn more about system response and its determinants. System disturbances can be problematic if there is considerable public interest in a particular issue and strongly held opinions as to preferred management action.

If wildlife agencies have the responsibility for management of wildlife populations, they also have the responsibility to conduct research on the effectiveness of management programs. Wildlife agency administrators should strive to develop well-designed, long-term management-research programs as a basic component of annual agency operations.

SUMMARY

Carefully designed wildlife research will improve the reliability of knowledge that is the basis of wildlife management. Research biologists must rigorously apply the scientific method and make use of powerful techniques in survey sampling and experimental design. Modeling offers powerful tools to predict consequences of management choices, especially when it is based on carefully designed field studies, long-term monitoring, and management experiments designed to increase understanding. More effort should be dedicated to the design phase of research, including obtaining critiques from other biologists and statisticians, and avoiding common problems such as insufficient sample sizes, procedural inconsistencies, non-uniform treatments, and pseudoreplication. Wherever possible, we must move from observational studies to experimental studies that provide a more reliable basis for interpretation and conclusions; these studies need to be replicated across space and time. Wildlife biologists have a tremendous responsibility associated with management of animal species experiencing increasing environmental-degradation problems, loss of habitat, and declining populations. We must face these problems armed with knowledge from quality scientific investigations.

ACKNOWLEDGMENTS

LITERATURE CITED

AKAIKE, H. 1973. Information theory and an extension of the maximum likelihood principle. pp.267-281. in B. N. Petran and F. Csaki, eds, Second international symposium on information theory, Akadeemiai Kiadi, Budapest, Hungary.

ALLDREDGE, J. R., AND J. T. RATTI. 1986, Comparison of some statistical techniques for analysis of resource selection. Journal of Wildlife Management 50:157-165.

_____, AND _____. 1992. Further comparison of some statistical techniques for analysis of resource selection. Journal of Wildlife Management 56:1-9.

ANDERSON, D. R. 2001. The need to get the basics right in wildlife field studies. Wildlife Society Bulletin 29:1294-1297.

_____, K. P. BURNHAM, AND W. L. THOMPSON. 2000. Null hypothesis testing: problems, prevalence, and an alternative. Journal of Wildlife Management 64:912-923.

ANTHONY, R. M. AND R. A. STEHN. 1994. Navigating aerial transects with a laptop computer map. Wildlife Society Bulletin 22:674-676.

BACHARACH, S. B. 1989. Organizational theories: Some criteria for evaluation. Academy of Management Review 14: 496-515.

BART, J., AND W. NOTZ. 2004. Analysis of data. Pages 000-000 in C. E. Braun, editor. Wildlife investigational techniques. Sixth edition. The Wildlife Society, Bethesda, Maryland, U.S.A.

BATE, L. J., E. O. GARTON, AND M. J. WISDOM. 1999. Estimating snag and large tree densities and distributions on a landscape for wildlife management. U.S. Department of Agriculture, Forest Service General Technical Report PNW-GTR-425.

BELLROSE, F. C., T. G. SCOTT, A. S. HAWKINS, AND J. B. LOW. 1961. Sex ratios and age ratios in North American ducks. Illinois Natural History Survey Bulletin 27:391-474.

BERGERUD, A. T. 1988. Population ecology of grouse. Pages 578-685 in A. T. Bergerud and M. W. Gratson, editors. Adaptive strategies and population ecology of northern grouse. Volume II. Theory and synthesis. University of Minnesota Press, Minneapolis, Minnesota, USA.

BOITANI, L. AND T. K. FULLER, Editors. 2000. Research techniques in animal ecology: controversies and consequences. Columbia University Press, New York, USA.

BOWDEN, D. C., A. E. ANDERSON, AND D. E. MEDIN. 1984. Sampling plans for mule deer sex and age ratios. Journal of Wildlife Management 48:500-509.

BOYD, R. J., A. Y. COOPERRIDER, P. C. LENT, AND J. A. BAILEY. 1986. Ungulates. Pages 519-564 in A. Y. Cooperrider, R. J. Boyd, and H. R. Stuart, eds. Inventory and Monitoring of Wildlife

Habitat. U.S. Department of Interior Bureau of Land Management Service Center, Denver, Colorado, U.S.A.

BUCKLAND, S. T., D. R. ANDERSON, K. P. BURNHAM, AND J. L. LAAKE. 1993. Distance sampling: estimating abundance of biological populations. Chapman and Hall, New York, U.S.A.

BUNGE, M. 1967. Scientific research. I The search for system. Springer-Verlag, New York, U.S.A.

BURNHAM, K. P., AND D. R. ANDERSON. 2002. Model selection and inference: a practical information-theoretic approach. Second edition. Springer-Verlag, New York, U.S.A.

_____, _____, AND J. L. LAAKE. 1980. Estimation of density from line transect sampling of biological populations. Wildlife Monograph 72.

CAUGHLEY, G. 1977. Analysis of vertebrate populations. John Wiley & Sons, New York, U.S.A.

CHAMBERLIN, T. C. 1965. The method of multiple working hypotheses. Science 148:754-759.

CHAPMAN, J. A., C. J. HENNY, AND H. M. WIGHT. 1969. The status, population dynamics, and harvest of the dusky Canada goose. Wildlife Monograph 18.

CHAPMAN, J. A., AND G. R. WILLNER. 1986. Lagomorphs. Pages 453-473 in A.Y. Cooperrider, R. J. Boyd, and H. R. Stuart, editors. Inventory and monitoring of wildlife habitat. U.S. Department of Interior Bureau of Land Management Service Center, Denver, Colorado, U.S.A.

CHERRY, S. 1998. Statistical tests in publications of The Wildlife Society. Wildlife Society Bulletin 26:947-953.

CLUTTON-BROCK, T. H. M. MAJOR, AND F. E. GUINNESS. 1985. Population regulation in male and female red deer. Journal of Animal Ecology 54: 831-846.

COCHRAN, W. G. 1963. Sampling techniques. Second edition John Wiley & Sons, New York, USA.

_____ 1983. Planning and analysis of observational studies. John Wiley & Sons, New York, USA.

COOK, J. G., L. L. IRWIN, L. D. BRYANT, R. A. RIGGS, AND J. W. THOMAS. 1998. Relations of forest cover and condition of elk: a test of the thermal cover hypothesis in summer and winter. Wildlife Monographs 141:1-61.

COOK, T. D., AND D. T. CAMPBELL. 1979. Quasi-experimentation: design and analysis issues for field settings. Houghton Mifflin, Boston, Massachusetts, USA.

COULSON, T., S. ALBON, F. GUINNESS, J. PEMBERTON, AND T. CLUTTON-BROCK. 1997. Population substructure, local density, and calf winter survival in red deer (*Cervus elaphus*). Ecology 78: 852-863.

COOKE, B. D., AND D. BERMAN. 2000. Effect of inoculation rate and ambient temperature on the survival time of rabbits, *Oryctolagus cuniculus* (L.), infected with the Rabbit Haemorrhagic Disease virus. Wildlife Research 27:137-142.

CHAMBERS, L. K., G. R. SINGLETON, AND L. A. HINDS. 1999. Fertility control of wild mouse populations: the effects of hormonal competence and an imposed level sterility. Wildlife Research 26:579-591.

CRETE, M., AND F. MESSIER. 1987. Evaluation of indices of gray wolf, *Canis lupis*, density in hardwood-conifer forests of southwestern Quebec. Cananadian Field-Naturalist 101:147-152.

CYR, A. 1981. Limitation and variability in hearing ability in censusing birds. Pages 327-333 in C. J. Ralph and J. M. Scott, editors. Estimating numbers of terrestrial birds. Studies in Avian Biology 6.

DELGIUDICE, G. D., F. J. SINGER, AND U. S. SEAL. 1991. Physiological assessment of winter nutritional deprivation in elk of Yellowstone National Park. Journal of Wildlife Management. 55:653-664.

DELGIUDICE, G. D., R. A. MOEN, F. J. SINGER, AND M. R. RIGGS. 2001. Winter nutritional restriction and simulated body condition of Yellowstone elk and bison before and after the fires of 1988. Wildlife Monographs. 147:1-60.

DENNIS, B. 1996. Discussion: should ecologists become Bayesians? Ecological Applications 6:1095-1103.

DEWEY, J. 1938. Scientific method: induction and deduction. Pages 419-441 in J. Dewey, editor, Logic–the theory of inquiry. Henry Holt and Co., New York, USA.

DIAMOND, J. 1986. Overview: laboratory experiments, field experiments and natural experiments. Pages 3-22 in J. Diamond and T. J. Case, editors. Community ecology. Harper & Row, New York, USA.

DODSON, A. P., AND P. J. HUDSON. 1992. Regulation and stability of a free-living host-parasite system: *Trichostrongylus tenuis* in red grouse. II. Population models. Journal of Animal Ecology 61:487-498.

DOLBY, G. R. 1982. The role of statistics in the methodology of the life sciences. Biometrics 38:1069-1083.

EBERHARDT, L. L. 1970. Correlation, regression, and density dependence. Ecology 51:306-310.

_____, AND J. M. THOMAS. 1991. Designing environmental field studies. Ecological Monographs 61:53-73.

EFRON, B. 1998. R. A. Fisher in the 21st century. Statistical Science 13:95-122.

FARNSWORTH, G. L., K. H. POLLOCK, J. D. NICHOLS, T. R. SIMONS, J. E. HINES, AND J.R. SAUER. 2002. A removal model for estimating detection probabilities form point-count surveys. Auk 119:414-425.

FIENBERG, S. E. 1970. The analysis of multidimensional contingency tables. Ecology 51:419-433.

_____. 1980. The analysis of cross-classified categorical data. Massachusetts Institute of Technology Press, Cambridge, Massachusetts.

FORBES, L. S. 1990. A note on statistical power. Auk 107:438-439.

FORD, E. D. 2000. Scientific method for ecological research. Cambridge University Press, Cambridge, U.K.

GARTON, E. O. 2002. Mapping a chimera? Pages 663-666 in J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Sampson editors. Predicting species occurrences: issues of accuracy and scale. Island Press, Washington, D. C., USA.

GELMAN, A. B. J. S. CARLIN, H. S. STERN, AND D. B. RUBIN. 1995. Bayesian data analysis. Chapman and Hall, Boca Raton, USA.

GILL, R. B. 1985. Wildlife research–an endangered species. Wildlife Society Bulletin 13:580-587.

GREEN, R. H. 1979. Sampling design and statistical methods for environmental biologists. Wiley, New York, USA.

GUINNESS, F. E., T. H. CLUTTON-BROCK, AND S. D. ALBON. 1978. Factors affecting calf mortality in red deer (*Cervus elaphus*). Journal of Animal Ecology 47:817-832

HAIRSTON, N.G. 1989. Ecological experiments: purpose, design, and execution. Cambridge studies in ecology. Cambridge University Press, New York, USA.

HANOWSKI, J. M. AND G. J. NIEMI. 1995. A comparison of on- and off-road bird counts: do you need to go off road to count birds accurately? Journal of Field Ornithology 66:469-483.

HEISEY, D. M. AND T. K. FULLER. 1985. Evaluations of survival and cause-specific mortality rates using telemetry data. Journal of Wildlife Management 49:668-674.

HOHF, R. S., J. T. RATTI, AND R. CROTEAU. 1987. Experimental analysis of winter food selection by spruce grouse. Journal of Wildlife Management 51:159-167.

HOUSTON, D. B. 1982. The northern yellowstone elk: ecology and management. Macmillan Publishing Co., New York, USA. 474pp.

HURLBERT, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monograph 54:187-211.

JAMES, F. C., AND C. E. MCCULLOCH. 1985. Data analysis and the design of experiments in ornithology. Pages 1-63 in R. F. Johnston, editor. Current Ornithology. Vol. 2. Plenum Press, New York, USA.

JENKINS, D., A. WATSON, AND G. R. MILLER. 1963. Population studies on red grouse, *Lagopus lagopus scoticus* (Lath.) in north-east Scotland. Journal of Animal Ecology 32:317-376.

JOHNSON, D. H. 1985. Statistical sirens: The allure of nonparametrics. Ecology 76:1998-2000.

_____. 1999. The insignificance of statistical significance testing. Journal of Wildlife Management 63:763-772.

_____. 2002. The importance of replication in wildlife research. Journal of Wildlife Management 66:919-932.

JOHNSON, R. A., AND D. W. WICHERN. 1988. Applied multivariate statistical analysis, Second edition Prentice-hall, Englewood Cliffs, New Jersey, USA.

KAMIL, A. C. 1988. Experimental design in ornithology. Pages 313-

346 *in* R. F. Johnston, editor. Current Ornithology. Vol. 5. Plenum Press, New York, USA.

KEITH, L. B. 1963. Wildlife's ten-year cycle. University of Wisconsin Press, Madison, Wisconsin, USA.

_____. 1974. Some features of population dynamics in mammals. Proceedings of the International Congress of Game Biologists 11:17-58.

KEITH, L. B., AND L. A. WINDBERG. 1978. A demographic analysis of the snowshoe hare cycle. Wildlife Monographs 58.

KEITH, L. B., J. R. CARY, O. J. RONGSTAD, AND M. C. BRITTINGHAM. 1984. Demography and ecology of a declining snowshoe hare population. Wildlife Monographs 90.

KEPLER, C. B., AND J. M. SCOTT. 1981. Reducing bird count variability by training observers. Pages 366-371 *in* C. J. Ralph and J. M. Scott, editors. Estimating numbers of terrestrial birds. Studies in Avian Biology 6.

KERLINGER, F. N. AND H. B. LEE. 2000. Foundations of behavioral research. Fourth edition. Harcourt College Publishers, New York, USA.

KOMAK, S., AND M. R. CROSSLAND. 2000. An assessment of the introduced mosquitofish (*Gambusia affinis holbrooki*) as a predator of eggs, hatchlings and tadpolse of native and non-native anurans. Wildlife Research 27:185-189.

KOZICKY, E. L., G. O. HENDERSON, P. G. HOMEYER, AND E. B. SPEAKER. 1952. The adequacy of the fall roadside pheasant census in Iowa. Transactions of North American Wildlife and Natural Resources Conference 17:293-305.

KREBS, C. J. 1999. Ecological methodology. Second edition. Harper & Row, New York, USA.

LAAKE, J. L., S. T. BUCKLAND, D. R. ANDERSON, AND K. P. BURNHAM. 1994. DISTANCE User's Guide V2.1. Colorado Cooperative Fish and Wildlife Research Unit Colorado State University, Fort Collins, USA.

LEEGE, T. A. AND W. O. HICKEY. 1971. Sprouting of northern Idaho shrubs after prescribed burning. Journal of Wildlife Management. 35:508-515.

LOWE, V. P. W. 1969. Population dynamics of red deer (*Cervus elaphus* L.) on Rhum. Journal of Animal Ecology 38:425-457.

Lyon, L. J., and A. G. Christensen. 2002. Elk and land mangement. Pages 557-581 *in* D. E. Toweill and J. W. Thomas (editors). North American elk: ecology and management. Smithsonian Institution Press, Washington, D.C., USA.

MACNAB, J. 1983. Wildlife management as scientific experimentation. Wildlife Society Bulletin 11-397-401.

MANLY, B. F. J., L. L. MCDONALD, AND G. W. GARNER. 1996. Maximum likelihood estimation for the double-count method with independent observers. Journal of Agricultural and Biological Environmental Statistics. 1:170-189.

MAYR, E. 1970. Populations, species, and evolution: an abridgement of animal species and evolution. Belknap Press, Harvard University Press, Cambridge, Massachusetts, USA. 453pp

MILLIKEN, G. A., AND D. E. JOHNSON. 1984. Analysis of messy data: designed experiments. Vol. I. Van Nostrand Reinhold, New York, USA. 473pp.

MOHLER, L. L., P. D. DALKE, AND W. M. SHAW. 1958. Elk and elk hunting in Idaho. Transactions of North American Wildlife Natural Resource Conference 23:491-501.

MORRISON, M. L., W. M. BLOCK, M. D. STRICKLAND, AND W. L. KENDALL. 2001. Wildlife study design. Springer, New York, USA.

MOSS, R., AND A. WATSON. 1991. Population cycles and kin selection in red grouse *Lagopus lagopus scoticus*. Ibis 133(supplement 1):113-120.

MOSS, R., A. WATSON, AND P. ROTHERY. 1984. Inherent changes in the body size, viability, and behavior of a fluctuating red grouse (*Lagopus lagopus scoticus*) population. Journal of Animal Ecology 53:171-189.

MOSS, R., I. B. TRENHOLM, A. WATSON, AND R. PARR. 1990. Parasitism, predation and survival of hen red grouse *Lagopus lagopus scoticus* in spring. Journal of Animal Ecology 59:631-642.

MUELLER-DOMBOIS, D., AND H. EELLENBERG. 1974. Aims and methods of vegetation ecology. John Wiley & Sons, New York, USA.

MURPHY, D. D., AND B. D. NOON. 1991. Coping with uncertainty in wildlife biology. Journal of Wildlife Management 55:773-782.

NEWTON-SMITH, W. H. 1981. The rationality of science. Routledge and Kegan Paul, Boston, Massachusetts, USA.

NOTTINGHAM, B. G., JR., K. G. JOHNSON, AND M. R. PELTON. 1989. Evaluation of scent-station surveys to monitor raccoon density. Wildlife Society Bulletin 17:29-35.

OSENBERG, C. W., O. SARNELLE, S. D. COOPER, AND R. D. HOLT. 1999. Resolving ecological questions through meta-analysis: Goals, metrics, and models. Ecology 80:1105-1117.

OVERTON, W. S., AND D. E. DAVIS. 1969. Estimating the numbers of animals in wildlife populations. Pages 403-456 *in* R. H. Giles, editor. Wildlife management techniques. Third edition. The Wildlife Society, Washington, D.C., USA.

PALTRIDGE, R., AND R. SOUTHGATE. 2001. The effect of habitat type and seasonal conditions on fauna in two areas of the Tanami Desert. Wildlife Research 28:247-260.

PATRIC, E. F., T. P. HUSBAND, C. G. MCKIEL, AND W. M. SULLIVAN. 1988. Potential of LORAN-C for wildlife research along coastal landscapes. Journal of Wildlife Management 52:162-164.

PEEK, J. M. 1989. Another look at burning shrubs in northern Idaho. Pages 157-159 *in* D. M. Baumgartner, D. W. Breuer, and B. A. Zamora, editors. Proceedings symposium on prescribed fire in the intermountain region: forest site preparation and range improvement. Washington State Univ., Pullman, USA.

PETERMAN, R. M. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Science 47:2-15.

PIETZ, P. J., AND J. R. TESTER. 1983. Habitat selection by snowshoe hares in north central Minnesota. Journal of Wildlife Management 47:686-696.

PLATT, J. R. 1964. Strong inference. Science 146:347-353.

POPPER, K. R. 1959. The logic of scientific discovery. Hutchinson and Co., London, U.K.

_____. 1968. Conjectures and refutations: the growth of scientific knowledge. Second edition. Harper & Row, New York, USA.

QUANG, P. X., AND E. F. BECKER. 1996. Combining line transect and double count sampling techniques for aerial surveys. J. Ag. Biol. Environ. Stat. 2:230-242.

QUINN, J. F. AND A. E. DUNHAM. 1983. On hypothesis testing in ecology and evolution. American Naturalist 122:602-617.

RATTI, J. T. 1980. The classification of avian species and subspecies. American Birds 34:860-866.

_____, D. L. MACKEY, AND J. R. ALLDREDGE. 1984. Analysis of spruce grouse habitat in north-central Washington. Journal of Wildlife Management 48:1188-1196.

_____ AND K. P. REESE. 1988. Preliminary test of the ecological trap hypothesis. Journal of Wildlife Management 52:484-491.

_____, AND D. E. TIMM. 1979. Migratory behavior of Vancouver Canada geese: recovery rate bias. Pages 208-212 *in* R. L. Jarvis and J. C. Bartonek, editors. Proceedings management biology of Pacific Flyway geese: a symposium. Northwest Section, The Wildlife Society, Portland, Oregon, USA.

_____, AND E. O. GARTON. 1994. Research and experimental design. Pages 1-23 *in* T. A. Bookhout, editor. Research and management techniques for wildlife and habitats. Fifth Edition. The Wildlife Society, Bethesda, Maryland, USA.

REYNOLDS, R. T., J. M. SCOTT, AND R. A. NUSSBAUM. 1980. A variable circular-plot method for estimating bird numbers. Condor 82:309-313.

RICKER, W. E. 1954. Stock and recruitment. Journal of The Fisheries Research Board of Canada 11:559-623.

RISBEY, D. A., M. C. CALVER, J. SHORT, J. S. BRADLEY, AND I. W. WRIGHT. 2000. The impact of cats and foxes on the small vertebrate fauna of Heirisson Prong, Western Australia. II. A field experiment. Wildlife Research 27:223-235.

ROBBINS, C. S., D. BYSTRAK, AND P. H. GEISSLER. 1986. The breeding bird survey: its first 15 years, 1965-1979. U.S. Department of Interior, Fish and Wildlife Service Resource Publication 157.

ROMESBURG, H. C. 1981. Wildlife science: gaining reliable knowledge. Journal of Wildlife Management 45:293-313.

ROTELLA, J. J., AND J. T. RATTI. 1986. Test of a critical density index assumption: a case study with gray partridge. Journal of Wildlife

Management 50:532-539.

_____, AND _____. 1992a. Mallard brood survival and wetland habitat conditions in southwestern Manitoba. Journal of Wildlife Management 56:499-507.

_____, AND _____. 1992b. Mallard brood movements and wetland selection in southwestern Manitoba. 1992. Journal of Wildlife Management 56:508-515.

ROSENSTOCK, S. S., D. R. ANDERSON, K. M. GIESEN, T. LEUKERING, AND M. F. CARTER. 2002. Landbird counting techniques: current practices and an alternative. Auk 119:46-53.

SAMUEL, M. D., E. O. GARTON, M. W. SCHLEGEL, AND R. G. CARSON. 1987. Visibility bias during aerial surveys of elk in northcentral Idaho. Journal of Wildlife Management 51:622-630.

SCHEAFFER, R. L., W. MENDENHALL, AND L. OTT. 1996. Elementary survey sampling. Fifth edition. Duxbury Press. Boston, Mass, USA.

SEBER, G. A. F. 1982. The estimation of animal abundance and related parameters. Second edition. Charles Griffin, London, U.K.

SELANDER, R. K. 1971. Systematics and speciation in birds. Pages 57-147 in D. S. Farner and J. R. King, editors. Avian biology. Vol. I. Academic press, New York, USA.

SHENK, T. M., AND A. B. FRANKLIN, editors. 2001. Modeling in natural resource management: development, interpretation, and application. Island Press, Washington, D.C., USA.

SHIPLEY, W. 2000. Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference. Cambridge University Press, Cambridge, UK.

SINCLAIR, A. R. E. 1991. Science and the practice of wildlife management. Journal of Wildlife Management 55:767-773.

SINGER, F. J., A. HARTING, K. K. SYMONDS, AND M. B. COUGHENOUR. 1997. Density dependence, compensation, and environmental effects on elk calf mortality in Yellowstone National Park. Journal of Wildlife Management. 61:12-25.

SKALSKI, J. R., AND D. S. ROBSON. 1992. Techniques for Wildlife Investigations. Academic Press, San Diego, California, USA.

SMITH, D. R., M. J. CONROY, AND D. H. BRAKHAGE. 1995. Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. Biometrics 51:777-788.

SMITH, L. M., I. L. BRISBIN, JR., AND G. C. WHITE. 1984. An evaluation of total trapline captures as estimates of furbearer abundance. Journal of Wildlife Management 48:1452-1455.

STEBBINS, G. L. 1971. Processes of organic evolution. Second edition. Prentice-Hall, Englewood Cliffs, New Jersey. USA.

STEEL R. G., AND J. H. TORRIE. 1980. Principles and procedures of statistics: a biometrical approach. Second edition. McGraw-Hill Book Co., New York, USA.

STEIDL, R. J., J. P. HAYES, AND E. SCHAUBER. 1997. Statistical power analysis in wildlife research. Journal of Wildlife Management 61:270-279.

STEINHORST, R. K., AND M. D. SAMUEL. 1989. Sightability adjustment methods for aerial surveys of wildlife populations. Biometrics 45:415-425.

STEWART-OATEN, A., W. M. MURDOCH, AND K. R. PARKER. 1986. Environmental impact assessment: Pseudoreplication in time? Ecology 67:929-940.

SWIHART, R.K., AND N.A. SLADE 1985. Testing for independence of observations in animal movements. Ecology 66:1176-1184.

TAKEKAWA, J. Y., AND E. O. GARTON. 1984. How much is an evening grosbeak worth? Journal of Forestry 82:426-428.

THOMPSON, S. K., AND F. L. RAMSEY. 1983. Adaptive sampling of populations. Technical Report 82, Department of Statistics, Oregon State University, Corvallis, Oregon, USA.

THOMPSON, S. K., AND G. A. F. SEBER. 1996. Adaptive sampling. John Wiley & Sons, New York, USA.

THOMPSON, W. L. 2002. Towards reliable bird surveys: accounting for individuals present but not detected. Auk 119:18-25.

_____, G. C. WHITE, AND C. GOWAN. 1998. Monitoring vertebrate populations. Academic Press, San Diego, California, USA.

TOFT, C. A., AND P. J. SHEA. 1983. Detecting community-wide patterns: estimating power strengthens statistical inference. American Naturalist 122:618-625.

TUKEY, J. W. 1977. Exploratory data analysis. Addison-Wesley Publishing Co., Reading, Massachusetts, USA.

TURNER, M. G., Y. WU, L. L. WALLACE, W. H. ROMME, AND A. BRENKERT. 1994. Simulating winter interactions among ungulates, vegetation, and fire in northern Yellowstone Park. Ecological Applications 4:472-496.

UNDERWOOD, A. J. 1994. On beyond BACI: Sampling designs that might reliably detect environmental disturbances. Ecological Applications 4:3-15.

_____. 1997. Experiments in Ecology. Cambridge University Press, Cambridge, UK.

WALTERS, C. 1986. Adaptive Management of Renewable Resources. MacMillan Publishing Co., New York, NY, USA.

WATSON, A. 1967. Social status and population regulation in the red grouse (*Lagopus lagopus scotius*). Royal Society Population Study Group 2:22-30.

WATSON, A., AND R. MOSS. 1972. A current model of population dynamics in red grouse. Proceedings of the International Ornithological Congress 15:134-149.

WATSON, A., R. MOSS, AND R. PARR. 1984a. Effects of food enrichment on numbers and spacing behavior of red grouse. Journal of Animal Ecology 53:663-678.

WATSON, A., R. MOSS, R. PARR, M. D. MOUNTFORD, AND P. ROTHERY. 1994. Kin landownership, differential aggression between kin and non-kin, and population fluctuations in red grouse. Journal of Animal Ecology 63:39-50.

WATSON, A., R. MOSS, AND P. ROTHERY. 2000. Weather and synchrony in 10-year population cycles of rock ptarmigan and red grouse in Scotland. Ecology 81:2126-2136.

WATSON, A., R. MOSS, P. ROTHERY, AND R. PARR. 1984b. Demographic causes and predictive models of population fluctuation in red grouse. Journal of Animal Ecology 53:639-662.

WELLS, J. V. AND M. E. RICHMOND. 1995. Populations, metapopulations, and species populations: what are they and who should care? Wildlife Society Bulletin 23:458-462.

WHITE, G. C., D. R. ANDERSON, K. P. BURNHAM, AND D. L. OTIS. 1982. Capture-recapture and removal methods for sampling closed populations. Report LA-8787-NERP, UC-11, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

WIENS, J. A. 1973. Pattern and process in grassland bird communities. Ecological Monographs 43:237-270.

_____. 1989. The ecology of bird communities: foundations and patterns. Vol. I. Cambridge University press, New York, USA.

WILLIAMS, B. K. AND F. A. JOHNSON. 1995. Adaptive management and the regulation of waterfowl harvests. Wildlife Society Bulletin 23:430-436.

_____, _____, AND K. WILKINS. 1996. Uncertainty and the adaptive management of waterfowl harvests. Journal of Wildlife Management 60:223-232.

_____, J. D. NICHOLS AND M. J. CONROY. 2001. Analysis and management of animal populations. Academic Press, San Diego, California, USA.

WILSON, M. M. AND J. A. CRAWFORD. 1979. Response of bobwhites to controlled burning in south Texas. Wildlife Society Bulletin 7:53-56.

WINDBERG, L. A., AND L. B. KEITH. 1976. Snowshoe hare population response to artificial high densities. Journal of Mammalogy 57:523-553.

ZAR, J. H. 1999. Biostatistical analysis. 4th ed. Prentice-Hall. Englewood Cliffs, N.J. 718pp.