1

Research and Experimental Design

EDWARD O. GARTON,
JOCELYN L. AYCRIGG,
COURTNEY CONWAY,
AND JON S. HORNE

INTRODUCTION

RELIABLE KNOWLEDGE IS essential for effective wildlife management. It is the basis for implementing management actions to achieve desired outcomes. Thus, wildlife management must be based on sound scientific investigations that produce objective, accurate, and defensible information. Such information is obtained by conducting rigorous research that adheres to the scientific method. This chapter provides an overview of the fundamental concepts of wildlife research and study design. It is a revision of Ratti and Garton (1994), Garton et al. (2005), and Garton et al. (2012).

The **scientific method** (Chamberlin 1890, Dewey 1938, Platt 1964, Bunge 1967, Newton-Smith 1981, Ford 2000, Gauch 2003) provides the most efficient approach to expand our knowledge of wildlife science and is accomplished by following a structured series of steps. These steps guide researchers through the process of devising an explicit research **question**, generating several research **hypotheses** that would answer the question, and deriving **predictions** from those research hypotheses. Next steps are conducting the research to **test** the predictions and using the results to **inform** management or policy decisions while also contributing to our understanding of **ecological processes** (Fig. 1.1, Box 1.1). Conducting research to test the predictions can be accomplished by gathering new or existing observational data to obtain unbiased estimates of important patterns or characteristics (e.g., hunter harvest is additive to mountain lion mortality in adult elk; Robinson et al. 2014) or by performing manipulative experiments (e.g., reducing coyote densities results in higher recruitment of mule deer fawns in years of low lagomorph abundance; Hurley et al. 2011). Another option is to develop a model (e.g., compare effectiveness of removing disease-infected individuals in controlling an epidemic to removing and vaccinating individuals; Potapov et al. 2012). An **integrated approach** combines **estimation**, **experiments**, and **models** to test research hypotheses. We outline a series of steps to describe the research process, although we believe the research process is iterative in that knowledge gained at 1 step can be applied to previous steps as one regularly returns to the original research question and briefly reconsiders earlier steps. This iterative approach is more reflective of how research is conducted, and including these steps in your research will increase the likelihood of your research results informing wildlife management and science.

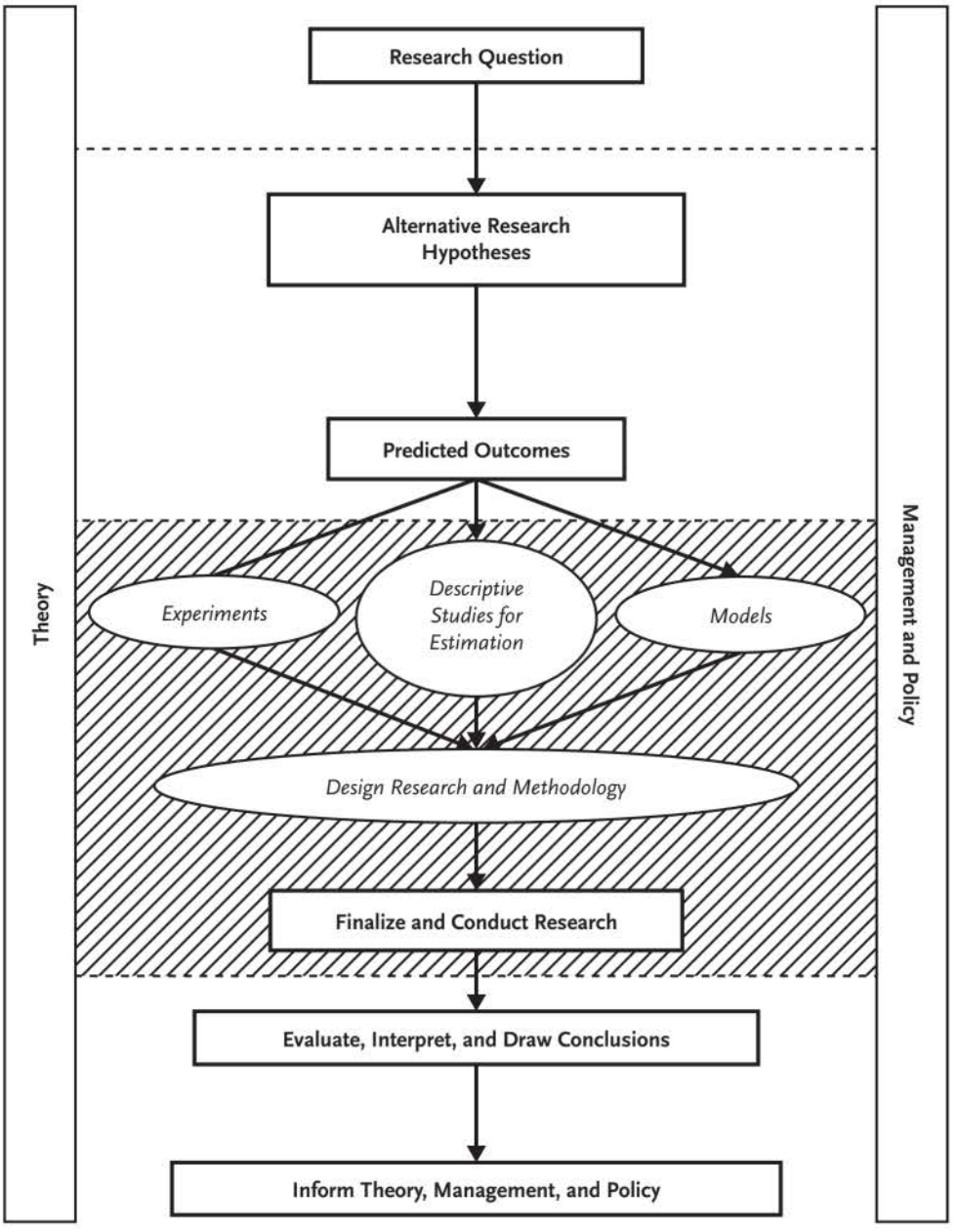

Fig. 1.1. Framework for the research process. The main steps are in boxes with bold outlines. Ovals represent steps or alternative methodologies within the formal testing step (i.e., the shaded section below). Theory, management, and policy are connected throughout the research process. See Box 1.1 for description of recommended steps at each point within this framework. See text for detailed descriptions of each step.

STEP 1: RESEARCH QUESTION

Developing a clear and testable research question is the initial and most important step in the wildlife research process and can be one of the most exciting and rewarding exercises for a researcher. A well-defined, clear, and concise research question is the foundation for all subsequent steps (Fig. 1.1, Box 1.1). Wildlife science is impeded when there is ambiguity in the stated research question because it can lead to research design and methodology that do not address the intended research question. Thus, all researchers and managers should carefully and thoughtfully consider their research question, ensuring it clearly and accurately portrays the intended question as a basis for study design. It is critical to explicitly state the research question in writing in all documents: study design, research proposal, field or lab protocols, and manuscripts and other final products. A good research question will have been well vetted with peers and

Box 1.1 Recommended steps in the research process with an elk population example

We recommend the following steps to increase the likelihood that scientific discoveries will result in reliable knowledge that leads to informed management and policy actions while contributing to the advancement of ecological theory. Ecological theories, management needs, and agency policies should inform these steps (and vice versa) throughout the research process (Fig. 1.1). See text for detailed description of each step in this process.

Step	Description and example
1. Identify research question	Research questions are identified via numerous approaches that are often synergistic. These include brainstorming with peers, exploring prior experiences, soliciting expert input, exploring limitations or assumptions of existing ecological theory, responding to management and policy concerns, identifying knowledge gaps in the published literature, noticing patterns during prior field observations, and performing exploratory data analysis on historic or preliminary data. When possible, it is helpful to try to frame the research question as a "why" question. *While elk populations across the western US are generally stable or increasing, some elk populations have declined in recent years. If managers can understand the causes of the population declines, then management actions could mitigate the causes. This leads to the following research question: why are elk populations declining?*
2. Develop alternative research hypotheses	Theory and previous research on the topic of the research question as well as creativity will lead to alternative research hypotheses. Research hypotheses should be causal statements that are potential answers to the research question and generate 1 or more predictions (Step 3) that can be evaluated via formal testing (Step 4). *Research Hypothesis 1: Elk are declining because the amount of quality forage has declined, leading to a decline in the carrying capacity of elk.* *Research Hypothesis 2: Elk are declining because an increased predation rate by reestablished gray wolves has led to negative population growth rates for elk.* *Research Hypothesis 3: Elk are declining because winter severity (i.e., as measured by snow accumulation and temperature) has increased, leading to a decrease in winter survival of elk calves.*
3. Formulate predictions	One or more predictions should be deduced from each research hypothesis. Predictions should be sufficiently explicit to allow researchers to determine the most suitable approach for evaluating all of the alternative research hypotheses and to develop a methodology for formally testing each prediction (Step 4). *For Research Hypothesis 1:* *Prediction 1: If the amount of quality forage for elk increases, then the population growth rate of elk will increase.* *Prediction 2: If the amount of quality forage for elk increases, then the reproductive potential of adult females will increase because of improved physical condition.*

continued

Prediction 3: If the amount of quality forage for elk increases, then the survival rate of calves will increase because of greater maternal investment.

4. Design formal testing

This is the core of the research process where the research design and methodology are decided upon for each explicit prediction developed. Prior to performing formal testing, the researcher should identify the appropriate population and identify the anticipated methodological approach most suitable for addressing each prediction with assistance from a statistician. Below is a brief description of 3 methodological approaches, the integrated approach, and important considerations for a research design and methodology.

a. Experiments

Experiments (i.e., field, laboratory, or natural) often provide the strongest way to test a research hypothesis. Experiments consist of manipulating conditions either directly or indirectly.

To evaluate Prediction 2 of Research Hypothesis 1, we could design an experiment to increase the amount of quality forage available to a captive population of female elk and measure pregnancy and parturition rates. These rates would be compared to a control population of female elk for which there was no increase in the amount of quality forage. A test would be made of the explicitly deduced prediction that captive female elk that receive an experimentally increased amount of quality forage have higher pregnancy and parturition rates than a control population that received no increase in quality forage.

b. Descriptive studies for estimation

This approach uses descriptive studies to estimate 1 or more traits or variables that will allow researchers to assess whether empirical observations are consistent with a prediction generated by a research hypothesis.

To evaluate Prediction 1 of Research Hypothesis 1, a correlation coefficient estimate could be made to evaluate the strength of the relationship between population growth rate of elk and the amount of quality forage available. The correlation would test the explicitly deduced prediction the population growth rate of elk is positively correlated (i.e., correlation coefficient is significantly >0) with the amount of quality forage available.

c. Models

Predictions also can be assessed with quantitative models, including theoretical and simulation models. Based on a set of assumptions, a model can predict and assess the magnitude of phenomena that can evaluate explicit predictions.

To evaluate Research Hypothesis 2, a prediction could be made that elk population growth rates will respond to changes in gray wolf abundance. This question may need to be answered prior to wolves becoming established in an area (i.e., no empirical data exist to do an experiment or descriptive study). Thus, a model could be developed that predicts elk population dynamics as a function of wolf abundance based on previously described processes, such as density dependence, predator-prey numerical and functional responses, and

	harvest management strategies. Then the explicitly deduced prediction that elk population growth rate and carrying capacity is a declining function of wolf abundance could be tested.
d. Integrated approach	This approach involves testing predictions with a combination of experiments, descriptive studies for estimation, and models. The integrated approach is optimal for addressing any research question because it can leverage the strengths and mitigate the individual weaknesses of the 3 methodological approaches (e.g., experiments, descriptive studies for estimation, and models).
e. Research design and methodology	At this point in the research process, the researcher should have a plan regarding whether the predictions will be tested via experiments, descriptive studies for estimation, models, or an integrated approach and have a set of explicit and testable predictions. The research design and methodology include deciding on an approach for making statistical inference, formulating a robust sampling design, and determining an appropriate sample size. A pilot study should be considered to identify logistical constraints, determine likely sample sizes, and estimate variation in parameters. Additionally, a research proposal should be developed that describes all of the above steps in the research process, including the research question, research hypotheses, predictions, and research design and methodology. Researchers should solicit a critical peer review of the proposal from experts and revise, if necessary.
5. Finalize and conduct research	As stated above, the researcher will have, at this point, a set of explicitly stated and testable predictions (via deduction). A matrix of research hypotheses and predictions should be developed that shows the variables or relationships to be tested, their specific or relative values, and whether each prediction supports or refutes each research hypothesis (see Box 1.11 for example) in order to select the optimal test(s) to conduct.
6. Evaluate, interpret, and draw conclusions from results	The results of the research project should provide a basis for supporting or refuting 1 or more research hypotheses and, hence, informing the research question. Depending on the strength of inference gained, you might recommend additional predictions, increased sample sizes, or new alternative research hypotheses to be tested. This step also includes submitting a manuscript(s) describing the research for peer-reviewed publication, agency report, and presentation at a professional meeting(s) and incorporation into adaptive management plans.
7. Inform theory, management, and policy	Research results should connect back to the ecological theory, management issue, or policy concern that generated the research question. This is a vital step in the research framework, which is required for adaptive management and provides a basis for advancing wildlife science. *The research results would help inform elk harvest policies of state wildlife agencies and also would identify the most appropriate management actions if state and federal agencies decide to take action to increase elk*

continued

populations. It is more likely that effective management actions are identified and implemented if the ultimate cause(s) of the past population decline is identified through a well-designed and executed research project. The research results also could identify the relative contribution of ecological processes in regulating the growth of ungulate populations and inform ecological theory of wildlife population dynamics.

re-examined throughout the research process to ensure it has not shifted or the research design and methodology selected have not caused the research project to drift away from the central research question (e.g., why are mule deer populations declining in southeast Idaho; Hurley et al. 2011). Often too little time is spent explicitly articulating and re-vetting a research question. Instead, a research question should be regularly scrutinized, dissected, probed, and challenged to ensure it is the best and most appropriate question.

Research questions can emanate from a variety of sources. These include brainstorming with peers, exploring prior observations or experiences, soliciting expert input, exploring limitations or assumptions of existing ecological theory, responding to management and policy concerns, identifying knowledge gaps in the published literature, or performing exploratory data analysis on extant field data. Prior field investigations provide the prime foundation for generating some research questions while other research questions are initiated by a management agency interested in a particular wildlife issue. Exploratory data analysis (i.e., examining **patterns** in data from prior studies or from the literature) also can be helpful in generating interesting **causal** or **mechanistic** questions (i.e., "why" questions). Sometimes the initial question is a piece of a larger or broader question. Regardless of the genesis of the research question, it should be explicitly stated, written, and thoroughly scrutinized (via review of the published literature and discussion with peers) until it is unambiguous and the answer will clearly advance wildlife science (Abelson 1997, Johnson 1999).

Research questions will provide the most insight and be most amendable to the scientific method when posed as why questions. Furthermore, **why questions** provide the most insight into the mechanisms and processes responsible for the observed patterns. Unfortunately, studies that address why questions are rare in wildlife science (Gavin 1989, 1991), whereas yes-no questions (i.e., a do/does or is/are question) are more common. When possible, researchers should scrutinize their initial yes-no questions (those often posed at the outset of a study) that highlight or suggest a pattern to uncover a more important and more informative why or causal question that is at the core of the initial yes-no question. Identifying a pattern is often the beginning of good research, but it is sometimes helpful to consider the initial yes-no question as the specific question and the underlying why question as the "broad question" or "causal question." For example, a state wildlife manager might ask, "Can we reduce crop depredations by white-tailed deer?" While this is a perfectly legitimate research question, it addresses a rather narrow topic. However, with some thought, we can broaden this question by asking, "Why do white-tailed deer use crops as a food source?" By broadening the scope of the research question, we can answer the specific management question of how to reduce crop damage, but also provide the opportunity to address research hypotheses related to the nutritional value of native forage versus crops, behavioral responses of deer to actual and perceived risks, and the trade-offs among behaviors for acquiring food, avoiding predation (by humans and nonhumans), and successfully reproducing. An additional benefit of broadening the research question is that it allows us to use and contribute to knowledge of ecological theory. Indeed, the hypotheses generated by our broader research question are directly related to optimal foraging theory, a conceptual treatment of animal foraging behavior that has been studied for decades (Stephens et al. 2007*a*). Thus, our research would benefit enormously by broadening it and evaluating ecological theory to gain insight into the ecological mechanisms responsible for deer depredating on crops. Incorporating mechanisms into the question is invaluable and can usually be accomplished without losing focus on the priorities of the funding organization.

STEP 2: ALTERNATIVE RESEARCH HYPOTHESES

After clearly stating the research question, the next step is to propose alternative research hypotheses—that is, alternative answers to your research question. Chamberlin's (1890; republished in 1965) classic paper on the scientific method emphasized the need to examine **multiple working hypotheses** to address a research question. This approach is especially applicable to wildlife science because ecological systems are inherently complex and difficult to study. Generating research hypotheses (i.e., **retroduction**; Romesburg 1981) is often the most fun step. Each alternative research

hypothesis should include a clear ecological process(es) upon which it is based. The number of alternative research hypotheses only is limited by their plausibility. For example, 20 alternative research hypotheses have been proposed to explain the function of hatching asynchrony in birds (Stoleson and Beissinger 1995). An individual study may be able to only explicitly test 1 or a few research hypotheses, but a good study should at least acknowledge the other plausible research hypotheses.

Some research hypotheses undoubtedly have already been proposed and tested for each research question. However, others may have been proposed, but not tested, and still others may be novel (not previously proposed by others). At this step, one should attempt to identify and discuss all plausible research hypotheses. Do not assume all possible research hypotheses have already been proposed and only a thorough review of the literature is needed. Brainstorm in a quiet place or engage in a rigorous discussion with colleagues because a yet-to-be-proposed research hypothesis could be formulated. This is the most creative part of the research process. Numerous methods can be used, but one of the best methods in generating new alternative research hypotheses is to ask what ecological processes (Box 1.2) are relevant to the research question. For example, could some aspect of food limitation explain the research question? What about predation or disease or physiological tolerance to climate? Another useful method is to generate a **conceptual model** (flow diagram or a conceptual framework; see below in "Conceptual Model") that illustrates the various intrinsic and extrinsic factors influencing the trait or pattern of interest. Remember to consult ecological theory related to the research topic, including related theory from other taxa, because that will help ensure all existing research hypotheses have been considered.

Box 1.2 Scientific method terminology and definitions

Assumptions: Conditions that must be true for a hypothesis, theory, or model to be correct (modified from Pickett et al. 2007:63).

Concept: A mental image of an ecological idea or a suite of ecological relationships (modified from Morris 1992).

Data: Two or more pieces of information (modified from Morris 1992).

Deduction: A reasoning process in which a conclusion follows from the evidence presented, so the conclusion cannot be false if the evidence is true; a formal reasoning process in which a prediction is proposed as a logical result of a research hypothesis through a chain of if-then statements.

Domain: The scope in space, time, and phenomena addressed by a theory (Pickett et al. 2007:63); an area or region that is defined or identified in some way (Morris 1992).

Effect size: The magnitude or size of an *effect*, such as the difference between treatments (e.g., the number of animals seen) relative to the noise in the measurements; expresses the magnitude of difference between 2 sample means and therefore is the logical complement to the *P*-values generated from statistical hypothesis tests (Pierce et al. 2012).

Ecological mechanism: An explicit description of causally interacting parts and processes of an ecological system that produces 1 or more effects and explains how an ecological process provides the answer for a research question.

Ecological process: Fundamental or ultimate cause of an ecological relationship; an agent of natural selection responsible for patterns in nature that serve as the causal driver inherent in a research hypothesis.

Induction: A process of reasoning in which a general conclusion is reached from specific data, especially when the conclusion does not necessarily or directly follow from those data (Morris 1992).

Power: The probability that a treatment effect will be detected when it occurs and the probability of rejecting the null hypothesis when it is false (Pierce et al. 2012); relies on the magnitude of the effect, variation in the characteristic being measured, efficiency of the estimator, significance level, and sample size.

Prediction: A testable statement that would be true if a research hypothesis is true.

Pseudoreplication: Use of inferential statistics to test for treatment effects with data from experiments where either treatments are not replicated (though samples may be) or replicates are not statistically independent (Hurlbert 1984).

Research hypothesis: Statement about the ultimate cause or explanation of a pattern, observation, or phenomenon (Romesburg 1981).

Research population: A collection of elements (individuals, spatial units, etc.) about which to make inference in a research project (Scheaffer et al. 2005).

continued

Retroduction: A creative exercise or process of reasoning to propose potential research hypotheses to explain a pattern or observation (modified from Romesburg 1981).

Statistical hypothesis: Restatement of a prediction in the form of a testable statement that can be evaluated via statistical inference from observed data (modified from Ford 200:54 and Kerlinger and Lee 2000:15).

Strong inference: An approach for making progress in scientific inquiry that includes formulating multiple alternative research hypotheses, devising 1 or more predicted outcomes that each would exclude 1 or more of the research hypotheses, conducting a crucial experiment to explicitly test each predicted outcome, and then, based on what was disproved, recycling the whole process by making sub-hypotheses or sequential hypotheses to refine the possibilities that were not disproved (i.e., the scientific method; Platt 1964).

Theory: A set of interrelated 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 (Kerlinger and Lee 2000:11).

Reviewing Literature

The process of generating alternative research hypotheses includes a thorough literature review of past studies that have addressed the research question with an eye for both solid research conclusions and critical gaps in knowledge. Attempt to obtain published and unpublished management agency studies with data related to the research question. Searching Google Scholar and other free online databases provides instant access to titles with links to abstracts and full text of published peer-reviewed literature. Membership in The Wildlife Society and other professional organizations (Ecological Society of America, Society for Conservation Biology, American Fisheries Society) and many public libraries provide access to full-text databases of every paper published in societies' journals. Broad-scale internet searches on Google and other search engines may provide unpublished information of value from public agencies and institutions, but information posted by individuals or unknown organizations should be treated with skepticism.

Conceptual Model

A useful approach to summarizing ideas, suggestions, ecological theory, potential management actions, and published literature related to the research question is to develop a conceptual model (flow diagram) that summarizes, with a simple diagram, the intrinsic and extrinsic factors that influence the research question (Ogden et al. 2005, Margoluis et al. 2009). This exercise will help generate or clarify alternative research hypotheses by illustrating causal pathways, important interactions, and relationships among factors (Fig. 1.2). The goal of creating a conceptual model is to convey the fundamental principles and basic functionality of the system in which the research question occurs. Key characteristics, usually portrayed as variables describing the state of the system, ecological processes (e.g., food limitation, predation, disease), relationships among characteristics, and boundaries identifying internal and external factors influencing the state of the system, are diagrammed graphically for clarity and for communication to

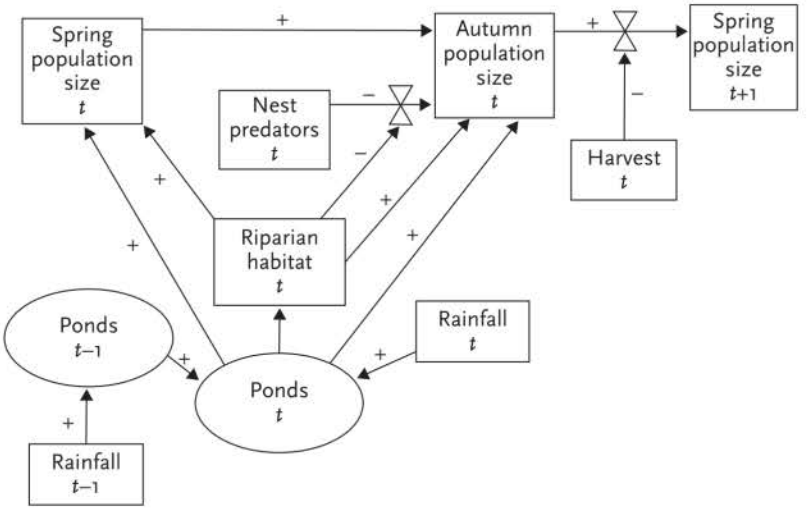

Fig. 1.2. Conceptual model of waterfowl population dynamics: *t* represents time *t*, *t*–1 represents one time step prior to time *t*, and *t*+1 represents one time step after time *t*, "+" represents a positive population influence and "–" represents a negative population influence.

others. This conceptual diagram should be continually scrutinized and revised when necessary to further explore the research question and the research hypotheses.

Let us explore a conceptual model that might have guided Conroy et al. (2002, 2005) in addressing the research question of why American black duck populations fluctuate from year to year (Fig. 1.2). These researchers could have formed a conceptual model by expressing patterns of fall waterfowl populations being positively influenced by spring breeding population size, quantity of ponds, and quantity and quality of riparian habitat and negatively affected by nest predators whose influence likely interacts with quality and quantity of riparian habitat. Hunter harvest influences spring population sizes the following year, but the removals may be associated with compensatory and/or additive effects. Using this conceptual model to depict the dynamics of any particular waterfowl population requires stating a **domain** of interest (Box 1.2) and inference (e.g., American black duck metapopulation, Box 1.3; breeding in 3 regions and harvesting in 6 regions in Canada and the US; Conroy et al. 2005). Some important variables (e.g., abundance of nestling predators) are extremely difficult to estimate, and nest predation rates are a result of a complex suite of factors, including abundance and foraging activities of the main nest predator species, alternative prey availability, annual weather patterns, habitat structure and composition, and parental behaviors. However, their effects might be inferred through changes in nest success and fledging rates resulting from experimental manipulations (e.g., predator removals or manipulation of nesting cover). Generating a conceptual model that adequately depicts key relationships that could influence the research question requires careful thinking and a thorough review of the current literature. We are more likely to contribute to theory and advance wildlife science if we can articulate how our research hypothesis(es) fits within an explicit conceptual model and the associated ecological science.

STEP 3: PREDICTIONS

Formulating predictions (Box 1.1) is a key step in stating explicit statements to be tested formally. Creativity and critical thinking are needed to determine which predictions are likely true if a research hypothesis is true. How could the merit of each individual research hypothesis be tested? What patterns or relationships might we expect or studies might we conduct to test each research hypothesis? This is another fun part of the process where brainstorming and unbridled creativity are useful. After exhausting your creativity (and those of your friends, family, and peers!), consult the literature to examine which predicted outcomes were proposed in past studies, although a healthy dose of skepticism aids in evaluating past research and the degree to which predictions have been adequately tested. This step is the beginning of the **deductive** phase of the research process (Box 1.1). It typically involves iterative thinking about formal tests of the predictions, which often leads to revisions of predictions and formal tests, and so on. The most valuable predictions are those that, if falsified, can differentiate among competing hypotheses.

STEP 4: DESIGN FORMAL TESTING

Once you have a research question, your set of alternative research hypotheses, and some predictions (Box 1.1), you can begin careful planning of the actual research design and methodology. You must choose a research approach that is logistically feasible and will provide the most efficient course for evaluating your research hypotheses. During this stage, researchers need to identify exactly which data will be collected to test explicit predictions and distinguish among alternative hypotheses as well as when, how, how much, and for how long. Furthermore, how will these data be treated statistically? Formal consultation with a statistician is important, and having completed Steps 1 to 3 will be invaluable during this consultation.

Box 1.3 Hierarchy of *spatial population units* (modified from Garton 2002)

Each level is represented by a red-winged blackbird hierarchy of spatial population units at Columbia National Wildlife Refuge, Washington (Fig. 1.3).

Deme — The smallest grouping of individuals approximating random breeding within the constraints of the breeding system where it is reasonable to estimate birth, death, immigration, and emigration rates. Animals in this grouping are ideally distributed continuously across 1 patch of habitat, and their movements are restricted to home ranges for breeders during the breeding season. The size of this patch ideally would be related to the dispersal distance of juveniles or perhaps equal an area 20 to 50 times the size of a female breeding home range (e.g., red-winged blackbirds). Note: for some species, demes are not feasible to delineate because of complex mating patterns and movements (e.g., in

continued

	mallards, males and females form pair bonds on wintering areas and males follow females to nesting areas the following spring; Bellrose 1976:236).
Population	A collection of demes or individuals that are closely related demographically (very high correlations in vital rates), geographically (close proximity), genetically (Manel et al. 2005), and through frequent dispersal. The population occupies a collection of habitat patches closely spaced relative to dispersal distance without large areas of nonhabitat intervening. The area is typically less than 100 times the size of an average home range and not larger than the dispersal distance of 95% of natal dispersers, but it may be much larger if habitat patches are linear in shape and widely dispersed (e.g., all red-winged blackbirds occupying Columbia National Wildlife Refuge during the breeding season might be reasonably treated as a population).
Metapopulation	A collection of populations sufficiently close together that dispersing individuals from source populations occasionally colonize empty habitat resulting from local population extinction. Populations within a single metapopulation may show low or high correlations in demographic rates, but the low rates of dispersal are sufficient to maintain substantial genetic similarity (e.g., red-winged blackbird populations distributed among the 7 national wildlife refuges along 200 km of the Columbia River in the south-central part of the state of Washington constitute a metapopulation; for another example, see Aycrigg and Garton 2014 for elk). A variety of types of metapopulations has been described (Harrison and Taylor 1997, Aycrigg and Garton 2014).
Subspecies	A collection of populations as well as metapopulations, if present, in a geographic region where a very low dispersal rate maintains genetic, morphological, and behavioral similarity, but populations and metapopulations occupy habitat patches that may be separated by large areas of nonhabitat resulting in substantial demographic independence among populations or metapopulations (Mayr 1970).
Species	The collection of interbreeding populations as well as metapopulations and subspecies, if present, encompassing the entire distribution and geographic range of the species. The individuals of the species may show substantial differences in phenotypes (habitat association, physiology, behavior) and genotypes, but the key requirement is separately evolving metapopulation lineages (de Queiroz 2005*b*).

Identify Appropriate Populations

Before starting formal testing of a research question, identify a target population for which we want our results to apply. In general, wildlife professionals work with 3 types of **populations** that affect research design: biological, political, and research. Mayr (1970:424) defined a biological population as a group of potentially interbreeding individuals at a given locality. He defined species as a reproductively isolated aggregate of interbreeding populations, or what de Queiroz (2005*b*) referred to as separately evolving **metapopulations**. Thus, a biological population is an aggregation of individuals of the same species that occupy a specific locality, and where often the boundaries can be described with accuracy. For example, the dusky Canada goose 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.

For other populations, such as red-winged blackbirds, groupings of individuals into a hierarchical organization of **demes**, populations, and metapopulations within the species may require careful consideration of 5 facets (Box 1.3): (1) geographical distribution of individuals, (2) geographical distribution of habitats (resources), (3) correlations in demographic rates (Palsboll et al. 2006), (4) genetic relationships (Manel et al. 2005), and (5) patterns of movement. Identifying the appropriate level within this hierarchy to sample, assign treatments, or build models is critical to obtaining precise estimates and performing valid, powerful tests of predictions. However, processes operating at the appropriate level are influenced by processes occurring at both lower levels in the hierarchy (i.e., mechanisms) and at higher levels (i.e., context). Choosing the level in the biological hierarchy to study (Box 1.3) defines the research population or domain (Box 1.2; Pickett et al. 2007) to which inferences and conclu-

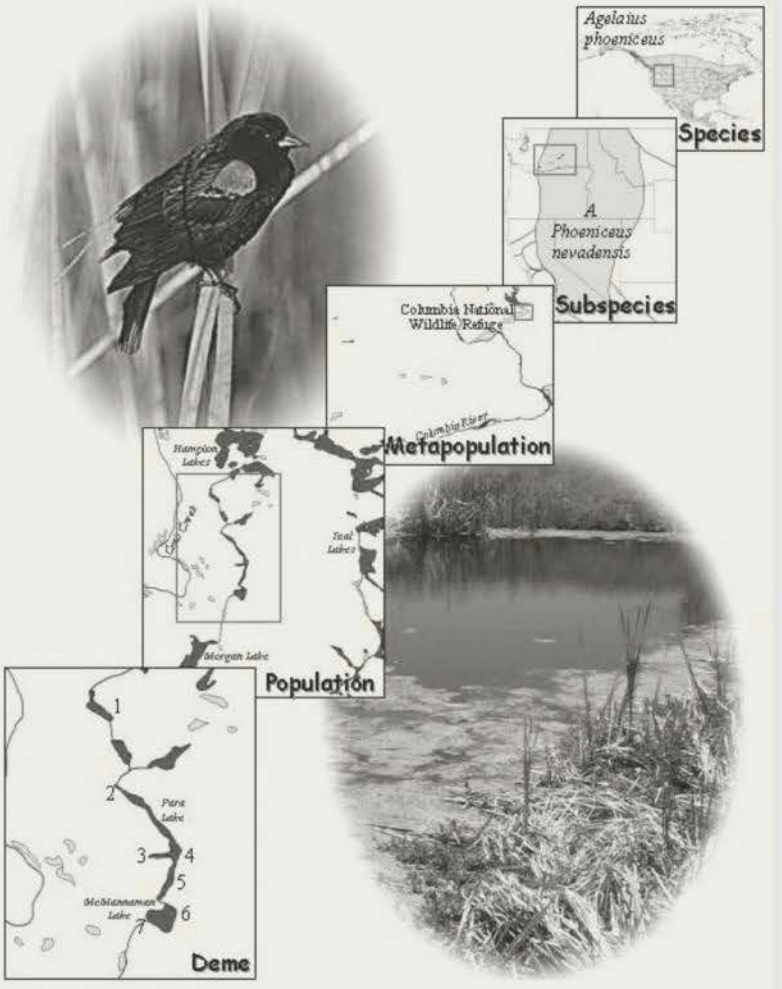

Fig. 1.3. Red-winged blackbird hierarchy of spatial population units from demes to species at Columbia National Wildlife Refuge. *From Garton (2002).*

sions apply. Beletsky and Orians (1996:152) demonstrated with 20 years of banding red-winged blackbirds on the Columbia River National Wildlife Refuge that territorial males and associated females occupying discrete patches of marsh vegetation associated with ponds or streams constituted a deme (Box 1.3). High correlations in demographic rates among demes and genetic similarity due to dispersal among demes make the entire red-winged blackbird population at the refuge an appropriate biological population on which to focus management, but cooperation with other refuges would facilitate management of regional metapopulations (Beletsky 1996, Garton 2002:665). Surrounding irrigated farmlands isolate the populations at refuges from one another to some degree, but populations at refuges throughout the Columbia Basin could be treated as a metapopulation within the subspecies (*Agelaius phoeniceus nevadensis;* Box 1.3). Biological populations for other species may not be as geographically distinct as Canada geese and red-winged blackbirds. Therefore, careful consideration of what constitutes a biological population to which findings will apply is important.

Because wildlife management must consider the practical aspects of implementation, researchers need to consider the political population that often splits biological populations. The political population is artificially constrained by political boundaries, such as county, state, or international entities. For example, a white-tailed deer population within an intensively farmed agricultural region might be closely associated with a river drainage system that provides food resources 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 inadvertently split into 2 political populations, which are subjected to different management strategies and harvest regulations. This is a common wildlife management problem. When biological populations have a political split, it is best to initiate cooperative studies (and a cooperative management program)

so the biological population is maintained and research personnel and funding resources can be pooled to benefit both agencies.

The research (statistical) population is defined as a collection of elements (individuals, spatial units, etc.) about which we wish to make an inference (Scheaffer et al. 2005). In rare cases, the research population may represent all individuals of a species (e.g., endangered species with few individuals, such as whooping cranes). In other instances, a research population might represent an entire biological population, such as one of the bison herds in Yellowstone National Park (Olexa and Gogan 2005). More typically, the research population only is a portion of a much larger biological population. Specifying a research population is essential in the research process and may require a review of existing literature on the species to determine breeding biology and dispersal patterns, geographic sampling to identify distribution of individuals and resources, and a review of literature on biological aggregations (Mayr 1970, Wells and Richmond 1995, Garton 2002, Cronin 2006).

Conclusions from research are only directly applicable to the research population. However, biologists usually want to obtain knowledge and solve problems regarding biological populations and species. To ensure the research conclusion applies to the intended biological population, (1) the sample (see "Sampling Methodology" below) should yield an unbiased representation of the research population, (2) the research population should be an unbiased representation of the biological population, and (3) the biological population should be representative of the species. Because traits (e.g., reproductive rates) often differ among segments of biological populations (e.g., among females) and among populations of a species, we must avoid broad conclusions or inferences relative to a research hypothesis until several projects from different populations and geographic locations provide similar results.

Experiments, Estimation, or Models

After identifying an appropriate research population, we must decide on the best approach for addressing our research question. Wildlife conservation and management were built initially on natural history observations, but numerous more rigorous research designs are available for answering questions about the biology and management of wildlife species (Fig. 1.4; Eberhardt and Thomas 1991, Morrison et al. 2008). These options differ dramatically in terms of 2 criteria highlighted in the figure: how certain are the conclusions reached, and how widely applicable are the conclusions? No single option is perfect in terms of these criteria and others to consider (e.g., feasibility). The biologist must weigh the available options carefully to find the best choice that fits within constraints of time and resources.

At one end of the spectrum are **experiments**, which provide the most certainty in our conclusions, but are often limited in their scope of inference (Fig. 1.4). Experiments consist of manipulating factors hypothesized to influence a particular ecological process (e.g., manipulating the amount of forage to determine if reproductive success is influenced). By randomizing which members of the research population receive the experimental treatments, we can reduce the chance of extraneous factors causing the observed effect and be more confident the factor we manipulated caused the observed effect or the absence of an effect.

Some research hypotheses and predicted outcomes are difficult to test using experiments, such as the effect of weather on a wildlife population. Thus, we may need to consider alternative research approaches. **Replicated field studies**, wherein no manipulation or randomization of treatments occurs, but true replication occurs (e.g., variation among replicates is similar in magnitude to variation among units drawn at random from the entire population), offer somewhat less certainty than true experiments. However, these studies may be more practical or allow for greater scope of inference

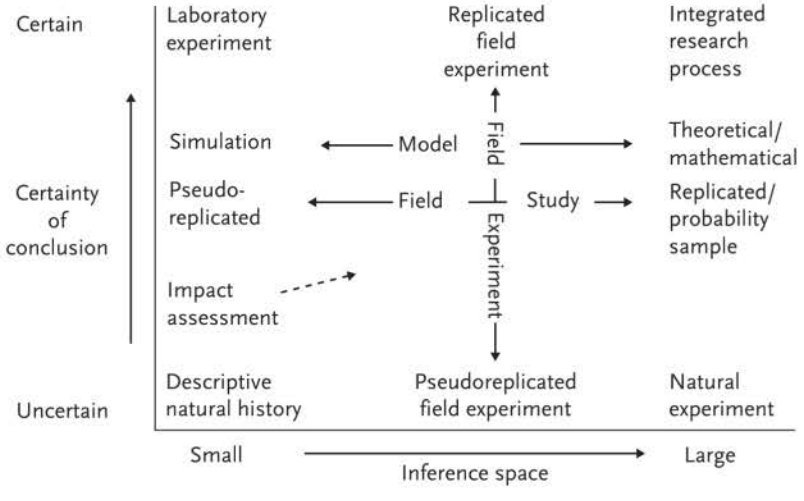

Fig. 1.4. The potential for wildlife study designs to produce conclusions with high certainty (few alternative hypotheses likely) and widespread applicability (a diversity of research populations where inferences apply).

(Fig. 1.4), especially when they are based on probability sampling or model-based designs. Nonreplicated **natural history descriptions** and **pseudoreplicated field studies** (e.g., replicates are more similar to each other than among units drawn randomly from the entire population; Fig. 1.4) typically provide weak inference for the predictions, research hypotheses, or research questions, but may be beneficial in providing observations to develop alternative research hypotheses and predictions (Steps 2 and 3 of the research process). Case studies consisting of unreplicated **natural history descriptions** (Fig. 1.4) are most useful at early stages in the research process. Impact assessments and **pseudoreplicated field studies** (Fig. 1.4) are slightly better than descriptive natural history studies but typically provide weak inference to the validity of research hypotheses (and hence the answer to any important research question).

Models can test research hypotheses, and the inference space can span a continuum from large and general for typical **theoretical models** to small and more realistic for standard **simulation models**, although each can span the full range from small to large inference space (Fig. 1.4). The certainty of conclusions based on these types of models is in part influenced by the measurement of model variables and estimation of model parameters. Theoretical and simulation models provide an important framework from which to understand the interplay among ecological processes that explain many pivotal questions in wildlife science. For example, models can be used to assess the relative extent to which juvenile dispersal influences local population growth. By holding other variables that influence population growth constant and varying juvenile dispersal rates within the model, we can estimate the change in population growth related to small or large variations in juvenile dispersal rates compared to other changes in other demographic rates. Such a **sensitivity analysis** can be very instructive, but such models only are accurate and informative with sufficient existing data with which to populate them. Building a quantitative model from the conceptual model of the question often highlights the most critical measurements and relationships prior to selecting the most important experiment to conduct or estimate(s) to obtain by sampling.

An integrated approach combines numerous approaches to research design into a single overall program for testing the predictions, such as natural history observations, models, natural experiments, and laboratory experiments (**integrated research process**; Fig. 1.4). Using an integrated approach makes the research inference space large and increases the certainty of conclusions by combining multiple formal testing approaches. The complexity of the ecosystems within which wildlife science resides is best addressed with an integrated approach because it enables the wildlife scientist to test predictions across more of the natural variability inherent within ecosystems while using methods ensuring more certainty of conclusions whenever possible (Clark and Stankey 2006, Morrison et al. 2008).

Experiments

Experiments consisting of manipulative trials are underused in wildlife science (Fig. 1.4). Diamond (1986) provided examples of 3 types of experiments (laboratory, field, and natural) and excellent suggestions for improving each type. The literature on experimentation is enormous, but other examples and discussion of experiments useful in wildlife science and ecology were provided by Cook and Campbell (1979), Milliken and Johnson (1984), Kamil (1988), Hairston (1989), Underwood (1997), Tilman et al. (2006), Chalfoun and Martin (2009), and Robinson et al. (2014).

Experiments should be set up to test the predictions of a research hypothesis. In other words, a specific experimental outcome will be predicted if a research hypothesis explains the question (e.g., a captive elk population that is provided with higher availability of quality forage will show higher pregnancy and parturition rates compared to a control population's pregnancy and parturition rates; Box 1.1, Step 4.a).

Laboratory Experiments

Laboratory experiments can be true experiments if they test a hypothesis via randomized assignment of treatments. Drawing inferences from laboratory experiments is easy because of the high level of control, yet this advantage must be weighed against their disadvantages (Table 1.1): (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 larger animals, although less so for plants and smaller animals; and (4) generality—some laboratory results cannot be extrapolated to natural communities. If our research question concerned why Canada geese consume certain plants more than others, laboratory experiments could be designed to test a prediction that Canada geese can select the most nutritious forage when given several alternatives in a cafeteria-feeding trial. Laboratory experiments have provided valuable information on emerging issues such as

Table 1.1. Strengths and weaknesses of different types of experiments.

	Experiment		
	Laboratory	Field	Natural
Control of independent variables[a]	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

[a]Active regulation and/or site matching. *Modified from Diamond (1986).*

wildlife diseases (e.g., Cooke and Berman 2000, Woodhams et al. 2008), efficacy of fertility control (Chambers et al. 1999, Hardy et al. 2006), and interactions between exotic and native species (e.g., Komak and Crossland 2000).

Field Experiments

Field experiments also can be true experiments (i.e., random assignment of treatments) and have a larger scope of inference than laboratory experiments because the treatments typically occur at a larger spatial scale in a more realistic set of circumstances with wild free-ranging animals (Fig. 1.4; Robinson et al. 2014). In field experiments, manipulations are conducted, but other factors are not subject to control (e.g., weather). Thus, field experiments are advantageous because they provide a relatively high degree of certainty in the conclusions as well as greater scope and realism compared to laboratory experiments.

In our Canada goose example, another prediction might involve a field experiment where the researcher selects random pairs of plots in known foraging areas. One member of each paired plot could be randomly assigned to be fertilized to test the expected outcome that geese select fertilized plots more than nonfertilized control plots. If the prediction is realized, a stronger inference about selection of more nutritious foods could be made. 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. Furthermore, support from 2 experimental predictions, 1 in the lab and 1 on the field as part of an integrated research process, would provide very compelling support for the research hypothesis.

Field experiments can be incorporated into **adaptive management** (see concluding section on connecting research and management) by breaking management zones into replicates assigned various treatment levels for comparison to a standard management action (Connelly et al. 2003*a*). Particular advantages of field experiments include random assignment of treatments to units and interspersion of treatment units among those to which the conclusions will apply, which protects against reaching invalid conclusions due to extrinsic factors.

Experimental Design

A well-designed experiment offers great advantages in terms of certainty of conclusions, and a variety of experimental designs are available to researchers (see Underwood 1997). Good experimental designs are based on 3 basic tenets: use of a **control** as an alternative to a **treatment**, **randomization** of assignment of treatments and controls, and **replication** of experimental units within treatments and controls. Powerful experimental designs include nonindependent experimental-unit treatments in which experimental units are **paired** or **blocked** and treatments are randomly assigned to units within pairs or blocks, **crossover** experiments in which treatments randomly assigned to pairs are reversed after an initial period of treatment by switching treatments between members of each pair or block, **multifactor designs** in which levels of 2 or more treatment factors are randomly assigned simultaneously to increase power in detecting treatment effects as well as interactions between treatment factors, and **treatment effects** specified as **fixed** (inference limited to only a small number of treatment levels), **random** (inference across a wide range of treatment levels), **mixed** (both fixed and random effects in multifactor designs), and **nested** effects (limited inference within nested sets of treatment factors).

Replication. In experiments, sample size is the number of replicates to which a treatment or control is assigned. Sample size refers to the number of independent, random sample units drawn from the research population. For logistical reasons, we may measure numerous subsamples closely spaced within a single sample unit. However, we must be careful to distinguish these subsamples in the analysis stage from independent random samples or we will be committing **pseudoreplication** (Hurlbert 1984, Johnson 2002). Subsamples are not independent random sample units because they typically are more similar to each other than independent random samples. Similarly, subsamples in experiments are not true replicates if they cannot be independently assigned to a treatment category. If subsamples are mistakenly treated as sample units or replicates, sample variance will underestimate actual amount of variation in the population, and sample size and precision of the value of interest (i.e., estimate) will be overestimated.

For example, suppose we wanted to address a research question regarding why northern bobwhites have declined and to test the prediction that prescribed fire would improve bobwhite habitat and increase their population. We could do a habitat improvement project over 1,000 km^2 that involves burning 1 km^2 of grassland and brush (e.g., Wilson and Crawford 1979). We could place 20 permanent plots within an area to be burned and 20 in an adjacent unburned area. Measurements of bobwhite abundance before and after the fire on the burned and unburned plots could be compared to test the prediction. However, the 20 plots on the burned area are not replicates, but merely subsamples or pseudoreplicates (Hurlbert 1984). In fact, we only have 1 observation because only a single fire (the treatment) occurred over 1,000 km^2. If we redesign the study, would we expect to see more variation among 20 burns on 20 randomly chosen areas scattered throughout the 1,000 km^2 than among 20 plots in a single burned area? A statistical test would show the fallacy of the first study design because it would only evaluate 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 randomly select 40 1-km^2 sites from throughout the 1,000 km^2 and randomly assign 20 replicates to be burned (i.e., treatments) and 20 replicates to be unburned (i.e.,

control) sites. Each burned and unburned 1-km^2 site would be sampled with some number of (e.g., 5) plots to measure bobwhite habitat before and after the treatment. Data would be analyzed by analysis of variance (ANOVA) with the 40 1-km^2 sites as samples and the 5 plots per site as subsamples. Treating the 100 burned and 100 unburned plots as experimental replicates would be an example of pseudoreplication. Pseudoreplication is a common problem, and investigators must understand the concept of replication and its importance in ecological research (Hurlbert 1984, Johnson 2002).

Controls. In experimental research, controls may be defined as parallel units observed to contrast with units receiving 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 influence conclusions or results. Experimental studies in wildlife that involve repeated measurements through time must include controls because of weather and other factors that vary through time (Morrison et al. 2008). 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. Grass production could increase in the year following burning because of higher rainfall that year. Without control sites, the reason(s) for the increased grass production could not be determined from increased rainfall, burning, or a combination of both factors (or from other nuisance factors), and we would be unable to evaluate the relative importance of each factor.

Determining Sample Size for Experiments. Use a **prospective power analysis** to set goals for sample size prior to starting data collection. The **power** of any statistical hypothesis test is defined as the probability of rejecting the null hypothesis when, in fact, it is false. Power depends on the magnitude of the effect (e.g., magnitude of difference between treatment and control), variation in the characteristic of interest, prescribed significance level (α), and sample size. Zar (1999) provides formulas to calculate power and sample size for statistical hypothesis tests, but a statistician should be consulted for complicated experimental designs and analyses. Many statistical packages (e.g., Statistical Analysis System by SAS Institute 2013) or specialized analysis software (e.g., MARK; White and Burnham 1999) can generate sample data for analysis to determine in advance how large the sample size should be to detect different effect sizes.

Power and sample size calculations should be based on a biologically meaningful **effect size** (magnitude of effect). Identifying a biologically significant effect usually involves expressing a conceptual model as a quantitative model plus value judgments about the importance of a biological response. The significance of a research project prior to beginning fieldwork can be evaluated by estimating the power of a test and calculating sample sizes. Estimates of variability required to estimate sample size and power of tests will only be available after completing a pilot study, which reinforces their value. Sample size analysis may lead to substantial revision of the predictions or perhaps even the research question (see determining sample size below).

Checklist for Experimental Design. The design of any experiment to test predicted outcomes from a research hypothesis must be developed carefully; otherwise, conclusions reached will be doubted. In addition to replication, use of controls and random assignment of treatments discussed above to identify the research population are essential. An experimental-design checklist provides a series of questions to address these critical elements. Many of the questions also are helpful for data collection in studies involving nonexperimental statistical hypothesis testing. Some experimental designs may address several statistical hypotheses (or even several research hypotheses) simultaneously; in other designs, each statistical hypothesis may require independent experimental testing.

1. **What research hypothesis(es) is being investigated and what are the predictions?** As mentioned previously, the research question and the research hypothesis(es) must be stated clearly before any experiment can be designed. If we want to ask why nest success of forest songbirds is lower adjacent to forest edges created during timber harvests, we could test the research hypothesis that nest predators are more successful at finding nests near harvested forest edges compared to intact forest or "natural" edges associated with beetle-killed patches of timber. We could design a study to test the predicted outcome that predation rates are higher at sharp edges (i.e., edges created by forest clearcuts) than at feathered edges (i.e., edges created by beetle kills) or intact forest without apparent edges (Ratti and Reese 1988, Chalfoun et al. 2002, Stephens et al. 2003).

2. **What is the response (dependent) variable(s) and how will it be measured?** The **response** variable should be clear from the research hypothesis and prediction (e.g., nest predation), but selecting the best technique to measure the response variable is more difficult. We must consider all possible methods and identify one that simultaneously maximizes precision and minimizes cost and bias. It is helpful to contact others who have used the techniques, examine the assumptions of the techniques and whether they will satisfy assumptions of anticipated statistical tests, and conduct a pilot study to test the potential techniques (see "Pilot Study" below). In our example, we could search for nests occurring along forest edges (i.e., both managed and beetle caused) and in continuous forest and use a generalized Mayfield estimator of mortality rate (Johnson 1979*b*, Bart and Robson 1982, Klett and Johnson 1982, Jehle et al. 2004, Shaffer 2004). This response variable is continuous, and we could apply any of a variety of designs termed general linear models (GLMs; e.g., ANOVA, linear

regression, analysis of covariance) or an information-theoretic analytical framework.

3. **What is the explanatory (treatment) variable(s) and what levels of it will be tested?** The **explanatory** variable(s) should be clear from the predicted outcome (i.e., sharp and feathered forest edges in our example), but selecting levels to test will depend on the population to which we want to make inferences. To test the effects of our explanatory variable at any level (termed a **random effect**), we must select the levels to test at random (random effects; Zar 1999). If we are only interested in a few levels within our explanatory variable, we only use those levels in our experiment and only make inferences to the levels tested (**fixed effects** or Model I; Zar 1999). For example, to evaluate effects of any type of forest edges on nest predation rates, we would select types of forest edge at random from all types that occur, randomly assign areas to receive cuts with those forest edges, and apply a random-effects model to analyze the data. In our example, we had sharp and feathered forest edges (i.e., 2 types). Clearly, our explanatory variable must be identified and classified correctly or measured precisely. Finally, a control would sharpen our comparison between treatment levels and provide a standard for evaluating the magnitude of predation. For example, comparing nest predation in undisturbed forests to nest predation at 2 types of edges could test the research hypothesis better, and the data could be analyzed with fixed-effects models. Our final conclusions (if we used edge type as a fixed effect) would not apply to nest predation rates on forest birds in all types of forest edges, but rather only to the 2 types that we compared to undisturbed forest.

An alternative design is to treat the explanatory (treatment) variable as being continuous and use regression rather than a classified grouping of treatments. Under this design, the treatment would consist of some level of tree removal on 1 side of the forest edge, and we would apply regression forms of GLMs under either statistical hypothesis testing or information-theoretic model comparisons. The response could be measured as the difference in nest predation rates between the 2 sides of the boundary, which would be predicted from the percentage of overstory removed. A critical requirement with this approach is to select treatment levels (i.e., percentage of tree cover removed) across the full range of forest treatments to which we want to apply our conclusions. Note that this alternative analysis has recast the approach and is addressing a different question than the original approach. Revisiting our original question, as we should regularly during the design process, forces us to evaluate the adequacy of the experimental treatment(s) to test the research hypothesis. For example, we might decide to distinguish between natural edges and ones created during timber harvest and apply analysis of covariance, which allows us to test both continuous (percentage of overstory removed) and categorical (beetle-killed vs. harvested) explanatory variables.

4. **To which population will we make inferences?** If our experimental results are to be applied to the real world, our experimental units must be drawn from a definable research population. The dependent and explanatory variables chosen define the relationship(s) examined and place constraints on the research population. The particular benefit of an experiment is that it minimizes the impact of extraneous factors on the conclusions as long as treatments are assigned at random to experimental units. However, if the research population is defined too broadly, then many extraneous factors could affect the results, and the variation could be too large to detect meaningful effects under the research hypothesis (i.e., low internal validity). If the population is defined too narrowly, then we have something akin to a laboratory experiment where application of the results might be severely limited (i.e., low generality or external validity).

Reaching the proper balance between internal and external validity takes thought and insight. For example, it would be challenging to compare nest-predation rates between sharp and feathered forest edges at sites throughout the northern Rocky Mountains due to logistics and costs. However, we could restrict our research population to 1 national forest, and we may wish to restrict our research population to a single forest type to remove extraneous factors that may covary with forest types. In addition, we need to determine the types of sharp and feathered edges that occur to decide which ones are relevant to our research population, as well as which ones are amenable to random assignment of treatments. Sharp edges are commonly produced by power lines and road rights-of-way as well as clearcuts, which differ in size, shape, human access, and disturbance after treatment. Also, designing an experiment involving random assignment of treatments is only feasible for clearcuts. Therefore, we could restrict our research population to sharp edges created by clearcuts and feathered edges created by selective harvests. As the examples above imply, identifying your research population is an essential step that helps determine the experimental design and methodology.

5. **What is the experimental unit?** The experimental unit is the smallest unit that is independent of other units and allows random assignment of a treatment. Not identifying the experimental unit correctly could result in an experiment without replication and could represent a case of pseudoreplication (Hurlbert 1984). For example, we could erroneously decide the experimental unit for our nest-predation study will be an individual nest. The resulting design could select 3 areas randomly assigned to clearcut, control, and selective cuts. We could find and monitor 20 nests along the edge of each area. If we use the 20 nests in each area as replicates, we have pseudoreplication because only a single area was given each treatment and only 1 area was randomly assigned each treatment. The 20 nests in each area are subsamples. Pseudoreplication restricts our potential inferences because we have sampled populations from only 2 logged areas and 1 unlogged area, and our inferences can only be made to those 3 areas, not to clearcuts, selective cuts, or undisturbed forests in general.

Pseudoreplicated designs are sometimes unavoidable, but interpretation of the results is restricted because, without true replication, confounding factors rather than the treatment could have caused the results. This potential problem can be minimized by increasing the sample size of true replicates. For example, in our nest-predation experiment, if 1 area was within the home range of a common raven pair and the other areas were not, this single confounding factor could affect the results regardless of treatment. A more reliable experiment could involve identifying about 30 areas scheduled for logging that were sufficiently separated to be independent of each other and randomly assigned, with 10 to be clearcut, 10 to be selectively cut, and 10 to be controls (i.e., undisturbed). Nests could be located and monitored in each area, and nests in a single area would be correctly treated as subsamples with their overall nest success treated as the observation for that area. This approach removes effects of confounding factors and allows conclusions to be drawn for the populations sampled (i.e., edges created by clearcuts and selective cuts within a specific habitat type in this region). Including control areas without an edge, as well as beetle-killed openings, would provide invaluable information for assessing the biological significance of the difference between the 2 types of edges created by timber harvests with regard to nest predation of forest songbirds.

6. **Which experimental design is best?** We describe a few of the most widely used experimental designs to consider (e.g., paired, multifactor, etc.; see "Experimental Design" above), but we advise you to consult texts on experimental design and a statistician before making the final selection (Scheiner and Gurevitch 2001, Quinn and Keough 2002, Morrison et al. 2008). The choice depends primarily on the type of explanatory and response variables (categorical, discrete, or continuous), number of levels of each, ability to block experimental units together, and type of relationship predicted (additive or with potential interactions such that combined effects of multiple explanatory variables may be larger or more complex than simply adding together their individual effects). For our study of nest predation with 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 could be randomly assigned to treatment and controls, with areas cast as blocks resulting in a randomized complete blocks design (Zar 1999).

7. **What sample size is needed?** Estimating sample size for proper analysis is essential because if the sample size is too costly or difficult to obtain, it would be better to redesign the experiment or evaluate a different research hypothesis. Sample size depends on the magnitude of the effect that you wish to detect, variation in the research population, type of relationship to be predicted, and desired power for the test (see "Determine Sample Size" below). Typically, preliminary data from a pilot study (see "Pilot Study" below) 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 to ensure that you have a high (80–90%) chance of detecting a biologically meaningful effect size (e.g., differences between your treatment and control categories).

8. **Have you consulted a statistician and received peer review on your experimental design?** Consulting a statistician before data collection is essential. A statistician cannot salvage an inadequate experimental design after a study is completed, but can almost always help improve a design prior to collecting data. Peer review by biologists with experience on similar studies can prevent wasted effort by helping assess if measurements or treatments proposed will work on a large scale in the field. Prior to the start of data collection is the time to get these comments!

Descriptive Studies for Estimation

Descriptive studies to estimate important characteristics or relationships provide another of the 3 general types of approaches (along with experiments and models) for evaluating research hypotheses. Most information gathered by wildlife biologists comes from descriptive rather than experimental study designs (i.e., lower left quadrant of Fig. 1.4), in which conclusions are less certain and only applicable locally. Such descriptive studies often provide the raw material in managing wildlife, developing research questions, evaluating local impacts, and initiating a research program. However, one also can design more rigorous field-based observational studies (see below under "Sampling Designs") that yield strong inferences about wildlife populations based on design-based and model-based sampling protocols, which measure key management variables or responses to large-scale management actions. In other words, observational data can be collected to address an explicit a priori prediction of an explicit research hypothesis (rather than merely as part of a descriptive information-gathering endeavor). For example, estimates of population size, recruitment, age-specific survival rates, herd composition, annual production of forage species, hunter harvest, and public attitudes (and comparisons of those estimates across time or space) can be part of an explicit prediction as a way to test a research hypothesis.

Given the constraints of time and money, it is important to obtain reliable, rigorous estimates for use in testing research hypotheses and informing management actions. These studies are referred to as surveys in the statistical literature, and the topic is known as **survey sampling** (Cochran 1977, Scheaffer et al. 2005) or finite population sampling. A modern alternative to this classic **design-based sampling** incorporates **model-based sampling**, which requires building a model to estimate important characteristics of the population based on characteristics measured for every member of the population (see "Design-Based versus Model-Based Inference" below).

Descriptive studies may appear similar to field experiments when randomization is included in the study design, but the

key distinction is that **treatments** are not assigned at random in descriptive studies. For example, a descriptive study addressing why geese feed in particular areas might randomly select plots where flocks of Canada geese have fed and other plots where they have not fed to test the explicit directional prediction that geese feed in areas with more nutritious vegetation. For this study, plots were chosen at random to measure nutrition, but the amount of nutrition was not randomly assigned to each plot, and the design may have a fundamental flaw if geese select individually more nutritious plants that were removed by the foraging geese prior to sampling. If the prediction is supported, a weak inference would be that geese are choosing plots with more nutritious food, but numerous alternative explanations remain untested (e.g., geese may prefer hilltop sites with good visibility, and coincidentally, these were heavily fertilized sites to compensate for wind-soil erosion from years of tillage). Making inferences from descriptive studies is difficult because we make ex post facto comparisons between groups (Kerlinger 1986). These groups also may differ in many other ways.

Predictions based on a descriptive study often take the form of an if-then statement. For example, if Research Hypothesis 1 explains the question, then variable x should be positively correlated with variable y (see Box 1.1, Step 3). The applied nature of wildlife management makes the realism and generality of descriptive studies an important advantage, but their widespread applicably is questionable unless multiple cases are analyzed and the analysis sharpens comparisons by controlling for relevant covariates and nuisance variables.

Modeling

Modeling can be used as a deductive tool to synthesize theoretical understanding with creative ideas about potential solutions to a problem or question. Creating a quantitative model makes the assumptions, accepted facts, or generalizations explicit for making valid and/or testable predictions. Kitching (1991:31) suggested that modeling involves 9 steps that closely align with the steps in the research process (Box 1.1). Starfield and Bleloch (1991) describe the modeling steps in a straightforward manner with lots of wildlife examples created in spreadsheets, while Clark (2007) presents a rigorous account of ecological modeling using free statistical and modeling software such as R (R Development Core Team 2016).

Adaptive management and population viability analysis both use modeling extensively. **Adaptive management** requires building predictive models that summarize what is known or assumed about a management question to examine alternative management actions. Managers choose 1 alternative included in the model. Monitoring also is conducted to (1) confirm the action was accomplished, (2) evaluate whether the consequence predicted by the model occurred, and (3) provide feedback on results of the action to improve understanding of the system, its behavior, key parameters, and relationships incorporated into the model. A more active version of adaptive management intentionally perturbs the system more in the interests of expanding knowledge of the behavior and relationships. **Population viability analysis** uses models and population data to estimate the probability that a population(s) of rare species will persist for a specified time (Mills 2007:254). This type of modeling is valuable in making scientifically defensible decisions concerning the listing, delisting, or management of a rare species or population. These models can be components of the research process (Fig. 1.1) that produce knowledge in the form of forecasts of the future. Modeling also can increase our understanding of interrelationships (e.g., long-term impacts and dynamics of wolf, mountain lion, and coyotes on their prey), which are difficult to manipulate experimentally. Building conceptual and quantitative models is a helpful step in any investigation as it focuses on critical relationships and assumptions.

Modeling also can provide an alternative to experiments because randomly assigning experimental treatments can be difficult in some situations, and sometimes the high cost of treatments precludes adequate replication. Modeling can assist in finding the best solutions to pressing problems (Starfield and Bleloch 1991, Shenk and Franklin 2001), selecting the best of alternative choices (Holling 1978, Walters 1986), and determining the relative magnitude of effects from multiple causes acting simultaneously (Wisdom and Mills 1997, Saltelli et al. 2001). A successful, overall strategy for modeling is to build the simplest model that describes the relationships between causative factors and the resulting effects. Try to select a modeling strategy that clearly incorporates your hypotheses and predictions at the simple empirical end of the continuum in terms of model objectives and complexity (Table 1.2). Long-term monitoring data and extensive measurements of demographic traits and habitat relationships provide the basis for accurate, useful models. Frequently, the goal is to model the responses of wildlife populations or habitats with the fewest predictors necessary to make good predictions where the predictors represent causal/mechanistic factors related to explicit research hypotheses. On the basis of Kitching (1991), we suggest the following 8 steps to build an ecological or wildlife model.

Steps to Building a Model

1. Develop a Research Question. The research question must have been identified earlier as one of the first steps in the research process (see "Step 1: Research Question" above, Fig. 1.1, Box 1.1), and a modeling approach can narrow the focus to concentrate on a limited set of available management options. Now the researcher should answer why a numerical or mathematical model is an appropriate way of tackling the research question (Kitching 1991:31). For example, the research question might be "Why is this species rare?" as a component of a conservation program to prevent further declines in abundance. The research hypotheses would propose the most likely causes, and a model could be developed that would predict how each of several management actions will likely cause changes in abundance of this species. If a par-

Table 1.2. Modeling strategies along gradients of simple to complex for scientific and statistical models.

	Gradient	
	Simple	Complex
Scientific models		
Quantification	Conceptual (verbal) →	Quantitative
Theoretical	General →	Complex simulation
Relationships	Linear →	Nonlinear
Variability	Deterministic →	Stochastic
Time scale	Time specific →	Dynamic
Mathematical formulation	Difference equations →	Differential equations
Factors	Single →	Multifactor
Spatial	Single site →	Multisite
Number of species	Single species →	Multispecies
Statistical models		
Sampling	Simple random →	Stratified, clustered, or multistage
Hypothesis testing	Fixed or random effects →	Mixed fixed and random effects
Independence of observations	Complete independence →	Dependence between observations in space, time, or both
Errors	Single term →	Separate process and observation errors

ticular management option does result in the largest increases of this species modeled population(s), it supports an explicitly deduced prediction and lends support to 1 of the research hypotheses. So, in the process of identifying the management option most likely to recover this species, the study also contributes to the broader question of why this species declined (i.e., puts the whole effort into a conceptual context). It is important to embrace the modeling approach as a pragmatic one. Kitching (1991:31) noted there was no point in building an ecological model that is more complex, more complete, or more time-consuming than is justified by the terms of reference of the problem to which the model is a response. Hence, use the deductive methods above as well as inductive methods applied to literature reviews and expert opinions to identify key predictors in the model and minimize covariates and interactions in the model where justified.

2. System Identification. After settling on a research question, define the system boundary delimiting interacting characteristics and processes internal to the system versus exterior drivers and level of resolution to model within the hierarchy of ecological levels ranging from individual animals with associated spatial extent to a population or metapopulation (Box 1.3; Fig. 1.5). Next select a set of components (characteristics and processes) to model (Fig. 1.2). One strategy is to pursue a parsimonious approach by making the model as simple as possible and only selecting critical components essential to describe the system. This approach is followed in developing conceptual models, simple simulation models, or general, theoretical models taking the form of analytical mathematical models (Table 1.2). Another approach, which takes the form of complex simulation models, is to include all components likely involved in the processes of interest. This approach, however, rarely succeeds because of the myriad unmeasured parameters required to incorporate such complexity. Typically, a simpler model is chosen between these 2 ends of the spectrum of complexity. One should tend toward the parsimonious end while including enough complexity to produce realistic predictions. In addition, interactions between model variables (i.e., positive, negative, feedback loops, and complex combinations) must be defined based on creative thinking and literature. Creating a simple conceptual model clarifies these interactions (e.g., Fig. 1.2) and guides the literature review (see "Conceptual Model" above).

3. Selecting Model Type. Our research question and the research hypotheses we hope to address should guide us in selecting the appropriate model type along the continuum from simple to complex with preference toward the simplest model necessary. Building complex models requires estimating more characteristics with more complex relationships. Great insight for many wildlife management problems can be gained with simple, linear models incorporating deterministic effects of a few independent factors at a single or small number of sites. Even forecasts for population viability analysis requiring stochastic models with time lags can be modeled with simulation models based on estimates from standard linear regression methods (Garton et al. 2011, Garton et al. 2016). In reality, many, if not most, ecological phenomena are synergistic, interactive, bimodal, context dependent, and/or curvilinear, implying the fullest insight from models will result from their incorporation into long-term integrated research programs, which combine iterative modeling, sampling, and experimental efforts. Such efforts are characterized as adaptive management at the agency level or long-term investigations by university-, institute-, or agency-funded programs.

4. Mathematical Formulation. Wildlife models can be formulated as difference equations portrayed with simple spreadsheets (Starfield and Bleloch 1991) because of the seasonal and annual patterns, which makes it difficult to estimate parameters for continuous time models formulated as differential equations. Differential equations have been successful for

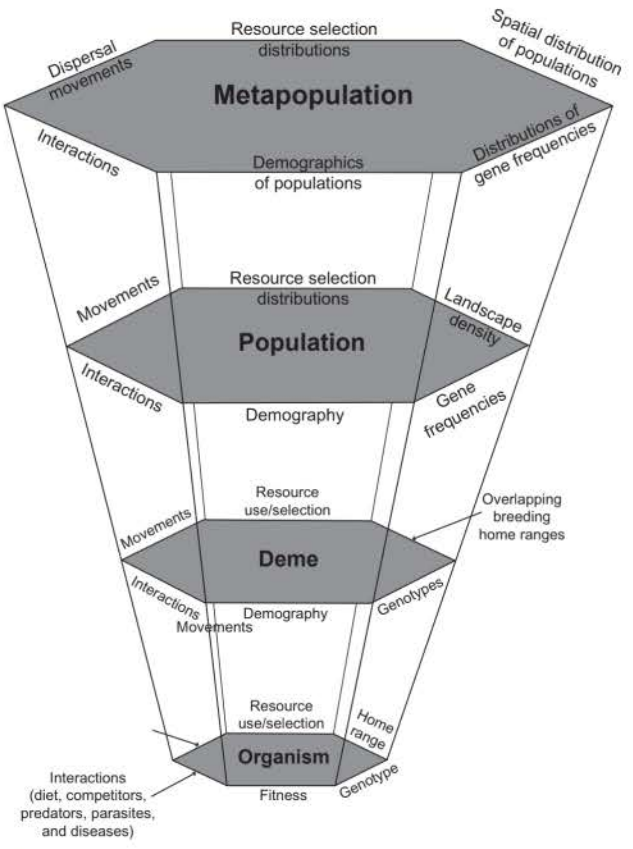

Fig. 1.5. Hierarchical arrangement of individuals from organism to metapopulation illustrating multiple facets to consider in delineating levels to study: demography, genetics, geographical distribution of individuals, distribution and selection of resources, patterns of movement, and interactions (e.g., diet, competitors, predators, parasites, and diseases). Processes operating at 1 level in this ecological hierarchy are influenced by processes and characteristics at lower levels in the hierarchy (i.e., mechanisms) and at higher levels in the ecological hierarchy (i.e., context). *Modified from Pickett et al. (2007:29).*

developing general theoretical models of the ecological theories underlying the principles of wildlife population ecology (Ginzburg 1986, Turchin 2001, Colyvan and Ginzburg 2003, Berryman 2003). However, translating these models into stochastic difference equation models (Garton et al. 2011, Garton et al. 2016) has proven successful for modeling time series of populations with complex dynamics (e.g., population viability analysis for San Joaquin kit fox [*Vulpes macrotis mutica*] incorporating density dependence and a 2-year lag in rainfall effects on plant productivity; Dennis and Otten 2000).

5. Selecting Computing Methods and Programs. Using spreadsheet programs for developing wildlife models can provide remarkable insight into important wildlife population dynamics (Starfield and Bleloch 1991). Specialized software designed for specific purposes, such as population viability analysis, can be applied to projecting persistence for endangered and rare species, such as RAMAS (Akçakaya 2000), VORTEX (Lacy 1993), and PVAClone R package (Nadeem and Lele 2012; also see Horne et al. 2016). Programs designed for statistical analysis, such as SAS (SAS Institute 2003) and R (R Development Core Team 2006), are equally adaptable to simulating both deterministic and stochastic models as well as estimating model parameters (Bolker 2008, Garton et al. 2011, Horne et al. 2016).

6. Parameter Estimation. Sampling methods, least squares for GLMs, and maximum likelihood methods are all useful for estimating state variables and parameters for alternative models. Information-theoretic approaches based on maximum likelihood methods to evaluate competing models provide excellent tools to determine their relative precision in predicting responses (Burnham and Anderson 2002). Burnham and Anderson (2002) contend that information-theoretic methods, such as using **Akaike's information criterion** (**AIC**) to assess the information content of a model, should be applied when we cannot experimentally manipulate causes or predictors (Akaike 1973). Model-averaged parameter estimates are readily calculated within this framework using Akaike weights (Burnham and Anderson 2002:133), although Cade (2015) demonstrated that model averaging fails to obtain unbiased estimates in the presence of collinearity among predictors.

7. Model Validation. Model validation can take at least 2 forms. First, comparing model predictions to data analyzed to build the model provides a preliminary validation that is always performed as part of constructing the model. This is essential to verify the model is performing as expected. Second, a stronger test of model validity requires comparing the output from the model to independent data not used in the construction of the model (e.g., from partitioning the original data into a first part to estimate parameters and a second part for a validation test). An even better test involves gathering additional data for this purpose or as a component of an adaptive management program. The comparison is usually made using standard statistical tools (see below under "Approaches for Data Analysis"), such as correlation and regression, which may be evaluated under a **frequentist** perspective using either statistical hypothesis tests or likelihood measures or a **cross-validation** perspective for **Bayesian** models (Gelman et al. 2013). If models use all data possible to maximize precision in parameter estimation during the model-building stage, then approaches such as **jackknifing** are applied, in which each individual observation is predicted from a series of leave-one-out models fit to all the rest of the data (Efron and Tibshirani 1993).

8. Model Implementation. Once all the above steps are completed, the model is ready to be implemented to address explicit predictions of the research hypothesis(es) and thereby help address the original research question. Usually, implementation involves a sensitivity analysis that is performed by manipulating key input parameters to assess model response to anticipated alternative management actions and/or potential environmental trends, changes, or variation. A useful model is an invaluable aid to both research and management, but verac-

ity of any prediction rests firmly on assumptions built into the model structure and relationships as well as validity of any parameters estimated from field observations. A useful model should be used interactively with ongoing research and management activities whereby the modeling exercise helps identify critical relationships and parameters that are then investigated in the field by gathering new observations or performing experiments (i.e., an integrated approach—see below). This process is adaptive management, where model predictions guide management actions and continued monitoring provides feedback to validate and improve model assumptions expressed as model parameters and relationships.

Integrated Research Process

The best evaluation of a research hypothesis often involves using some combination of field studies, experimentation, and models as part of an integrated research process (Fig. 1.4). For example, field observations by Ratti et al. (1984) indicated spruce grouse fed exclusively on certain lodgepole pine and Engelmann spruce while ignoring numerous other similar trees of the same species. They developed the following research question: why do spruce grouse select specific individual trees for forage? This led to a research hypothesis that spruce grouse selected specific trees of these 2 species because of their higher nutritional content. The authors also developed a predicted outcome that stated if spruce grouse selected a specific tree on which to forage, then this tree has higher nutritional content than surrounding trees. Hohf et al. (1987) developed a laboratory experiment with captive birds to test the predicted outcome of that research hypothesis. Evaluating other alternative hypotheses, such as lower inhibitory compounds, would be productive in an integrated research process. The integrated research process (Fig. 1.4) is often initiated by natural-history observations. Field observations and conceptual models should lead to experiments, and the results of natural experiments should lead to field and laboratory experiments and quantitative models. For example, Mook (1963) modeled bird predation on western spruce budworm. Takekawa and Garton (1984) also observed birds feeding heavily on western spruce budworms during a budworm outbreak, which led them to suggest a research hypothesis: that avian predators act to control the abundance of budworms. Field experiments were conducted to test and confirm a predicted outcome of their research hypothesis: survival of budworms on trees with netting (to experimentally exclude avian predators) was 3–4 times higher than on control trees exposed to bird predation (Campbell et al. 1983, Takekawa and Garton 1984). The experimental work confirmed the prediction developed from both observations and models. The level of certainty in the answer to the research question increases as many predicted outcomes are supported and alternate hypotheses are rejected in successively more rigorous tests that use a combination of field sampling, experiments, and models. After such findings are repeated over broad geographic areas or throughout the range of the species, the research hypothesis that receives repeated support may become a **principle** of wildlife science (Johnson 2002). The integrated research process should be the goal of wildlife science (Clark and Stankey 2006, Morrison et al. 2008).

Outstanding examples of integrated research programs include long-term research on the causes of population cycles in red grouse in Scotland (Jenkins et al. 1963, Watson and Moss 1972, Moss et al. 1984, Watson et al. 1994, Kerlin et al. 2007, New et al. 2009, Martinez-Padilla et al. 2014) and snowshoe hares in North America (Keith 1963, 1974; Windberg and Keith 1976; Keith and Windberg 1978; Keith et al. 1984; Krebs et al. 2001; Krebs et al. 2014). Other examples include ecological processes responsible for population regulation in red deer in Scotland (Lowe 1969, Guinness et al. 1978, Clutton-Brock et al. 1985, Coulson et al. 1997, McLoughlin et al. 2008, Stopher et al. 2008, Pérez-Barberia et al. 2013). The integrated research process used in the research program on snowshoe hares has evaluated the role of predators as well as alternative proposed causes of the classic 10-year cycle in snowshoe hare and Canada lynx numbers. In all 3 examples, descriptive studies and field observations formed the basis for subsequent research that included a series of innovative field studies and experiments (natural, field, and laboratory).

For example, preliminary studies of red grouse in Scotland (Jenkins et al. 1963) provided information on fundamental population processes: births, deaths, immigration, and emigration. This information was used to form alternative research hypotheses about causes of population fluctuations. Postulated causes initially included food quality, breeding success, spacing behavior, and genetics (Watson and Moss 1972, Kerlin et al. 2007). 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 shoots and flowers) affected egg quality, breeding success (viability of young), and spacing behavior of males and females. Territory size also affected recruitment and population density (but see Bergerud [1988] for a critique of the self-regulation hypothesis and inferences based on red grouse research). Watson et al. (1984*a*) then used innovative field experiments to test predictions of a nutritional quality hypothesis in which (1) fertilized fields were used to assess grouse response to increased nutritional quality of the heather and (2) implanted males with time-release hormones were used 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 did not support research hypotheses that nutrition, genetics, and parasitism were causal factors (although Dodson and Hudson [1992] make a counterargument 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*b*, Moss et al. 1990, New et al. 2009). These findings led to more research because the mechanisms underlying density-dependent relationships, including summer and winter emigration, were unclear. Recent research has focused

Box 1.4 Concepts of *precision*, *bias*, and *accuracy*

Precision is 1 measure of quality of an estimate. Precision refers to the closeness to each other of repeated estimates of the same quantity (Cochran 1977, Krebs 1999, Zar 1999). Precision of an estimate depends on form of the estimate, variation in the population, and size of the sample. Indicators of precision of an estimator are standard error and confidence intervals. Greater variation in the population leads to lower precision in an estimate, whereas a larger sample size produces higher precision in an estimate. Bias is another measure of quality of an estimate. Bias describes how far the average value of an estimate is from the true value. An unbiased estimate centers on the true value. If an estimate is both unbiased and precise, we say it is accurate (defined as an unbiased estimate with small mean-squared error [Cochran 1977]). Accuracy is the ultimate measure of the quality of an estimate and refers to the small deviations of the estimator from the true value (Cochran 1977).

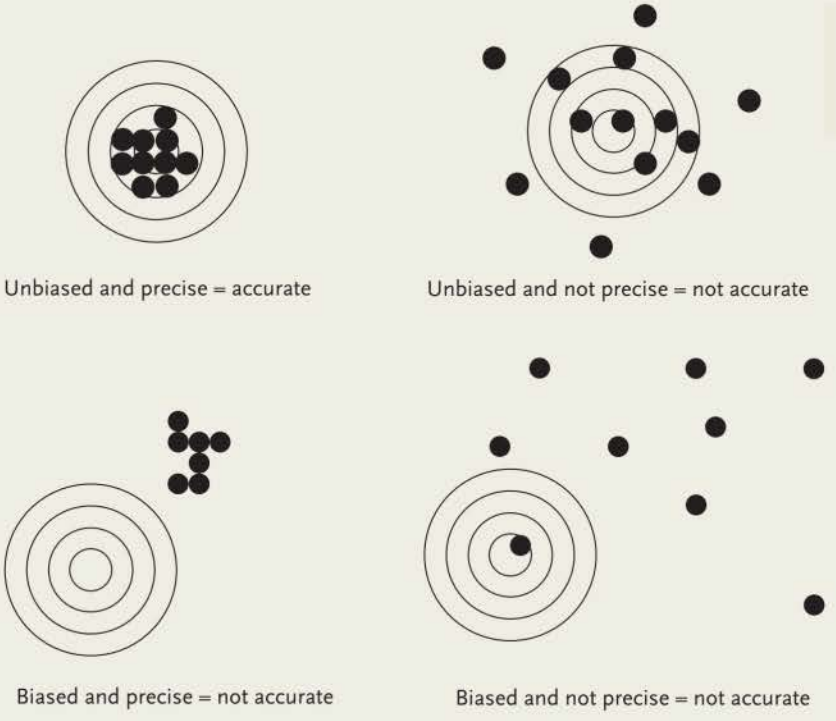

Fig. 1.6. Concepts of bias, precision, and accuracy are illustrated with targets and shot patterns below. *(Modified from Overton and Davis 1969 and White et al. 1982).*

on 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, Kerlin et al. 2007). Thus, the integrated approach to understanding this ecosystem continues, with research findings to date informing management of these species and populations as well as the ecological theory on population cycles and regulation.

Research Design and Methodology

At this point, researchers should have a research question, a set of research hypotheses and their predictions, and a general idea of whether the predictions should be tested via experimentation, a descriptive study based on estimation, or with the use of models (or a combination of these approaches via the integrated research process). Now begins the phase of determining the research design and methodology that includes deciding on an approach that will be used for making statistical inference, formulating a robust sampling design, and determining appropriate sample sizes based on a practical understanding of the key concepts of **precision**, **bias**, and **accuracy** for estimates (Boxes 1.4 and 1.5; Fig. 1.6). After understanding these key concepts, a good initial step in making these decisions is to conduct a pilot study.

Pilot Study

A pilot study is a preliminary, short-term trial designed to provide some useful information to design a more thorough study. Pilot studies are an important, but often neglected step in the research process, and they might occur at (or concurrent with) numerous stages in the process. Information obtained from a pilot study can help the researcher avoid potentially disastrous problems during or after the full-blown implementation of a larger research study. Pilot studies often will evaluate whether proposed field methods are valid, help estimate sample size (both the number needed to test predictions and the number likely given the proposed methods/resources), and disclose hidden costs or identify costs that were over- or underestimated. Optimal sample allocation incorporates cost estimates to maximize benefit from limited research budgets (Scheaffer et al. 2005). A pilot study can help reveal basic logistical problems, such as travel time among study plots being underestimated or sample sizes being too small, 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 only available 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 for an estimator will be difficult (see Fig. 1.7B Area 2). It is far better to discover such problems be-

Box 1.5 An example of a *population survey*

If there is an interest in estimating elk density on a large winter range, the area could be divided into numerous spatial count units and a sample of units drawn to survey from a helicopter. This would define the research population in terms of a geographic area rather than animals. The elements of the target population are count units, and a sample of these units is selected using a sampling design (see Sampling Design section in text). Using a helicopter, all elk present are counted in each sample unit, and density estimates are obtained by dividing the number of elk counted in each unit by the unit area. The data displayed as a histogram (Fig. 1.7A) suggest little variation in density on this winter range, as 80% of spatial units have densities between 1.6 and 2.2 elk/km^2. Because a single value is needed, the mean from our sample is used as representative of the entire winter range and the best estimate of the mean for the winter range. The variation from 1 unit to the next is small, and thus the mean from our sample is a precise estimate (see Box 1.4). If different results had been obtained that indicated the variation from 1 unit to the next is large, then the sample mean is less precise and not as reliable as the previous estimate (Fig. 1.7B). 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 1 be an accurate estimate of the mean density of elk on this winter range? To answer this question, the bias in the estimate must be evaluated. If the winter range was partially forested or had tall brush capable of hiding elk from view, aerial counts in each sample unit would underestimate the true number of elk present (Samuel et al. 1987). Therefore, the mean density from the sample would be a biased estimate of elk density on the winter range and 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. Accuracy in the estimates is made by selecting the approach with the least bias and most precision by applying a valid sampling or experimental design and obtaining a sufficiently large sample size to provide precise estimates.

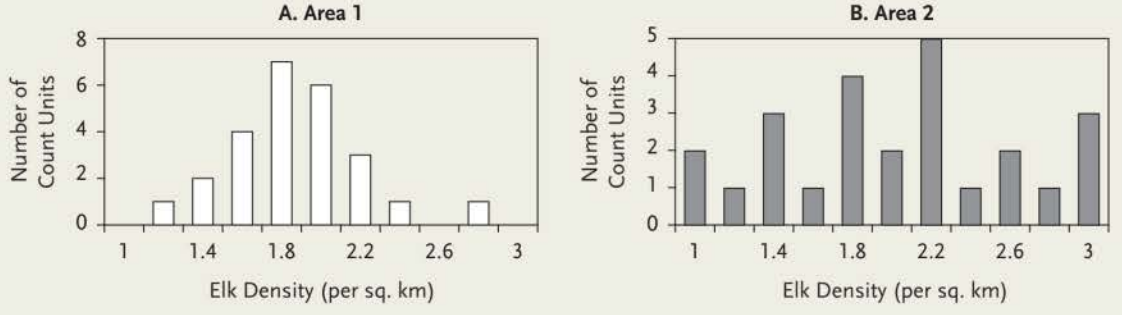

Fig. 1.7. Hypothetical example of elk counts and density estimates in Areas 1 and 2.

continued

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 to correct biased estimates (Rosenstock et al. 2002, Thompson 2002*b*). 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 adjust aerial surveys for this bias.

fore time, energy, personnel, and extensive research dollars are committed to a research project doomed to fail. Costs, methodology, and variance estimates may already be firmly established if the research is part of an ongoing project or if much research on the topic has been conducted.

Design-Based versus Model-Based Inference

To test our research hypotheses, we use some type of statistical inference. In other words, we must extrapolate from the limited data (our sample and/or experimental units) that our time and resources allow us to collect. One key step for both experiments and descriptive studies is designing a sampling procedure to draw observations (experimental units or sample units) from the research population(s) of interest. Only if this is done properly can conclusions of the tests be applied to these populations.

In general, 2 approaches (design based and model based) are commonly used for statistical inference (Koch and Gillings 2006). **Design-based** inference relies on **random selection** of observational units from a population of units that have an **a priori** probability of being included in the sample. Various quantities of interest (e.g., total abundance, clutch size) can be obtained by weighting observed values in the sampled units by the inverse of their probability of being included in the sample.

Model-based inference uses assumptions external to the study design to extend inferences to the target population. Under this approach, the observed data are assumed to be a realization of some predictable pattern or process. The goal of model-based inference is to devise a model that is an appropriate description of this process and then use the model to predict the values of unobserved units. Provided the model is a good **approximation** (i.e., all models are simplifications and subject to some degree of misspecification) of the process, model-based inference can be more efficient than design-based inference (i.e., provides more accurate estimates with limited data).

The main advantage of design-based inference is that it provides unbiased estimates without the need for additional assumptions. However, this strength of design-based inference also is a weakness. With limited data, design-based estimation may be significantly less efficient (i.e., higher variance in the estimates, resulting in low accuracy) when predictable relationships can be accurately portrayed with a model (Ver Hoef 2008). Understanding the difference between these 2 approaches not only helps in choosing the most efficient approach, but also is important because of the implications for data collection and analyses. For example, if researchers intend to use design-based inference, then some type of random sampling for data collection is required. Conversely, collecting a random sample for model-based inference is inefficient because researchers should collect data that will provide the most precise estimates of the model's parameters. Pilot study data help achieve this by ensuring that observations attempt to cover the full range of values for the response variable as well as any covariates that are being used to model the response variable.

Sampling Designs

If researchers decide to use design-based inference, survey sampling (Cochran 1977) is helpful in designing such sampling procedures (Box 1.6). Survey sampling also is useful in field experiments for drawing experimental units and subsamples (i.e., samples within 1 experimental unit). Model-based methods such as regression estimation and double sampling potentially have wide application in wildlife research (Scheaffer et al. 2005). To learn more about other sampling designs, consult a reference on sampling techniques (e.g., Scheaffer et al. 2005) and work with a statistician experienced in sampling. The survey design checklist (Box 1.7) helps to identify key decisions essential for a successful sampling study.

Determine Sample Size

One of the more challenging steps prior to starting actual data collection is to set goals for sample size. In experimental studies, a **power analysis** provides sample size requirements for testing a hypothesis(es), while in descriptive studies, it provides sample size requirements for obtaining an estimate of desired precision to evaluate the validity of our predictions. Many statistical packages (e.g., SAS Statistical Analysis System; SAS Institute 2013) or specialized analysis software, such as CAPTURE (White et al. 1982) and MARK (White et al. 1999), can generate sample data for analysis to determine the sample size needed to detect expected effects. In experiments or comparative studies, sample size is increased to improve the power of a statistical hypothesis test and to prevent erroneous conclusions. The statistical power of a test depends on sample size, level of significance (α), variance in the populations, effect size (i.e., the true change that occurred), and efficiency of the test or design (Steidl et al. 1997). Zar (1999) provides formulas to calculate power and sample size for sta-

Box 1.6 Sampling designs

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. To identify members to sample, each member of the population is assigned a number, and then numbers are picked from a table of random numbers or a random number generator on a computer.

When the research population consists of animals difficult to sample randomly, a potential solution is to change the sampling design such that small geographic units, such as plots or stands, are the sample units (or experimental units if we are developing a sampling design for an experimental treatment) and make a measurement on each plot. Thus, a random sample can be taken and used to infer abundance across the entire study area sampled. A valid random sampling procedure must be independent of researcher decisions. For example, to locate plots randomly in a study area, a Landsat image of the study area could be used. Random locations could be selected within the boundary of the study area using Universal Transverse Mercator (UTM) coordinates, and the UTM coordinates of these selected plot locations can be stored in a Global Positioning System unit that gives the exact location (Fig. 1.8A).

Random-like methods, referred to as haphazard or representative samples, have been used in place of truly random designs, but should be avoided because they are subject to researcher bias (e.g., the technique of facing in a random direction and throwing a pin over the shoulder to obtain the center of a vegetation sampling 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, and throwing the pin into the middle of such a patch is practically 0. 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 the patches may be missed by a truly random approach; Hurlbert 1984, Johnson 2002).

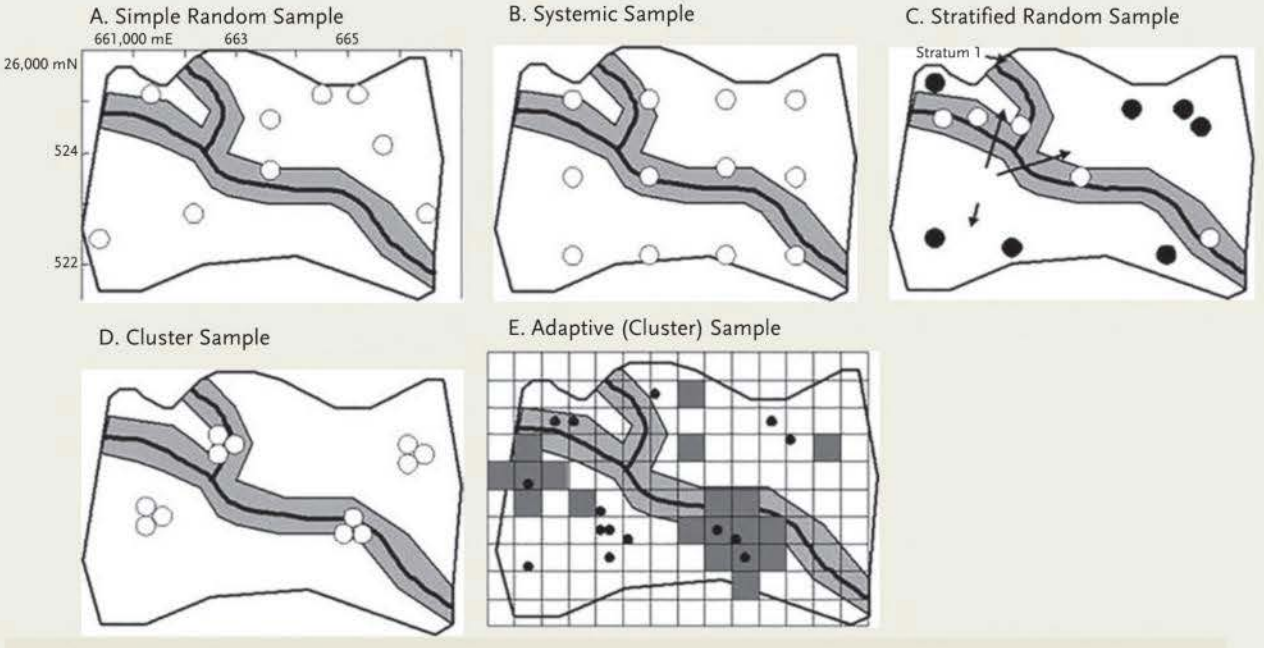

Fig. 1.8. Examples of sampling designs: (A) simple random sample selected in a GIS based on UTM coordinates, (B) systematic sample, (C) stratified random sample, (D) cluster sample, (E) adaptive cluster sampling.

In a **systematic sample**, sampling units are selected at regular intervals as they are encountered (Fig. 1.8B). This method is easier to perform and less subject to researcher errors than simple random sampling. For example, it would be difficult to draw a truly random sample of birdwatchers leaving a wildlife management area. It would be easier to systematically sample

continued

10% of the population by sampling every 10th person leaving the area. Systematic sampling is used extensively in vegetation measurements because of its ease of use in the field. Systematic sampling is frequently used in geographic sampling because it evaluates the spatial pattern of variability (e.g., spatial autocorrelation), which is used in 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. 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.

Systematic sampling may give biased estimates with periodic populations (i.e., with regular or repeating cycles). For example, to estimate the number of people using a wildlife management area, a check station could be established to systematically sample days during the season. This procedure could yield biased results if a sample was taken on once a week. If the sample day fell on a workday, different results could be obtained than if it fell on a weekend. Additionally, the estimate of variance could be too small, leading to the conclusion the estimate was more precise than it truly was. In this situation, the population sampled obviously is periodic; in other situations, the periodicity could be quite subtle. Thus, systematic sampling must be used with caution. The formal procedure is conducted by randomly selecting 1 of the first k elements to sample and every kth element thereafter. For example, if 10% of the population was to be sampled, k would equal 10 and a random number between 1 and 10 would be drawn. Suppose 3 was drawn, then the third element would be sampled and every 10th element thereafter (i.e., 13th, 23rd, 33rd, etc.). At a check station, this approach could be used to sample 10% of deer hunters that came through the station. To locate plots along a transect, randomly locate the starting point of the transect and then place plot centers at fixed intervals along the transect (i.e., every 100 m). Advantages and disadvantages of simple random and systematic sampling have been reviewed by Thompson et al. (1998), Krebs (1999), and Morrison et al. (2008).

Stratified random sampling can be use if members of subpopulations exist within 1 total population that is similar in the characteristics, but the subpopulations differ from each other in the same characteristic (see Fig. 1.8C). For example, tourists, birdwatchers, and hunters are readily divided into residents and nonresidents or a study area can be divided into habitat types. A wildlife population can be divided by age or sex. Subpopulations are referred to as strata, and a simple random sample of members is drawn from each stratum. Stratified random sampling also is useful if there is interest in the estimates for the subpopulations. The strata are chosen to contain units of identifiably different sample characteristics, usually with lower variance within each stratum.

For example, to test a prediction that requires an estimate of moose density, the strata could be defined by vegetation types (e.g., bogs and riparian willow patches, unburned forests, and burned forest) and then by drawing a simple random sample from each of these strata. If moose density is different among strata, variation in each stratum will be less than the overall variation. Thus, a more precise estimate of moose density is obtained for the same or less cost. If strata are not different, stratified estimates may be less precise than simple random estimates. Occasionally, the cost of stratified random sampling is less than simple random sampling. Also, stratified random sampling provides separate estimates for each stratum (e.g., moose density in willows or in forests) at little or no extra cost. 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 strata, 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 (Krebs 1999, Scheaffer et al. 2005). A pilot survey can be analyzed using analysis of variance (ANOVA) to learn if stratification would be beneficial. If cover types define strata, most Geographic Information System (GIS) software will automatically select random coordinates within cover types making stratified random samples easy to select.

A **cluster sample** is a simple random sample in which each sample unit is a cluster or collection of observations (see Fig. 1.8D). This approach has wide application because many birds and mammals occur in groups during all or part of the year. When samples are drawn from such populations, clusters of observations are drawn (i.e., groups of animals). Furthermore, many wildlife user groups (e.g., waterfowl hunters, park visitors) occur in clusters (e.g., boats in wetlands, vehicles along highways). Cluster sampling is useful when cost or time to travel from 1 sample unit to the next is prohibitive, which commonly occurs in surveys of animals and habitat.

Cluster sampling consists of 3 steps: (1) specify the appropriate clusters and make a list of all clusters, if feasible; (2) draw a simple random sample of clusters; and (3) measure all elements of interest within each cluster selected. Making

a list of clusters is straightforward for spatial units using a GIS, although rarely possible or essential for clusters of organisms, but we can obtain 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 with the researcher setting the cluster size. The optimal number of plots (i.e., cluster size) depends on the pattern of variability within the habitat. If plots in a cluster tend to be similar (i.e., little variability within a cluster), then cluster size should be small. If plots in a cluster tend to be dissimilar (i.e., high variability within a cluster), then cluster size should be large. For other types of cluster samples, such as groups of animals or people in vehicles, cluster size cannot be controlled, but is a characteristic of the population. For example, aerial surveys of elk and mule 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 differs from the previous 4 methods because sample size is not set at the start of the sampling effort, but rather depends on 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 randomly or by other sampling designs (e.g., systematic), and neighboring units also are sampled if the initial sample meets a selected criterion, such as having more than X individuals present (Thompson and Seber 1996, Seber and Mohammad 2005). The initial sampling unit and its neighbors (where sampled) form neighborhoods analogous to clusters in a cluster sampling design. Size of clusters does not need to be constant or known in advance. For spatially clustered animals or plants, the neighborhood consists of adjacent spatial sample units. Smith et al. (1995*a*, *b*) showed that adaptive cluster sampling could be more efficient than simple random sampling for estimating densities of wintering waterfowl species if the right sample unit size and criterion for further sampling in the neighborhood were chosen. Thompson and Seber (1996) provided numerous examples of adaptive sampling under conventional sampling designs and estimation methods, as well as examples based on maximum likelihood methods and Bayesian approaches. Thompson et al. (1998), Morrison et al. (2008), and Seber and Mohammad (2005) also review the basic concept and provide simple examples.

Sequential sampling involves drawing samples 1 at a time, and after each sample is taken, the researcher decides whether a conclusion can be reached concerning the prediction under a hypothesis-testing inferential framework. Sampling continues until either the study prediction is supported (rejection of the statistical null hypothesis) or the estimate has adequate precision to detect the expected pattern. Therefore, the sample size is not fixed in advance. Sequential 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). This approach usually minimizes sample size, prevents oversampling or undersampling, and thereby saves time and money. After an initial sample of moderately small size is obtained, successive samples are added until the desired precision is met or a maximum sample size under a stopping rule is reached. Sequential sampling typically requires <1/3 the sample size required for other sampling designs (Krebs 1999:304). For example, to survey mule deer on a winter range to ensure that harvest had not reduced male abundance below a management guideline of 5% males, we could graph and plot the results of successive samples (Krebs 1999:312). A level of significance must be chosen for the test (e.g., α). This approach involves drawing samples 1 at a time, and after each sample is taken, the researcher decides whether a conclusion can be reached concerning whether the population has not been adversely affected by male-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 1 of the upper or lower lines or the stopping rule is reached (e.g., the lower rejection line was reached at a sample size of 140). The conclusion would be there were <5% males remaining on the winter range. It is important the sample be distributed throughout the entire population so that a simple random sample of deer groups is obtained regardless of when you stop sampling. This would be most feasible using aerial surveys from helicopter or fixed-wing aircraft.

In addition to the 6 sampling designs discussed above, many other sampling designs are available, such as 2-stage cluster sampling, which involves surveying only a portion of the members of each cluster drawn in the sample, an efficient design when clusters are large. Cluster sampling is 1 version of the more general method referred to as ratio estimation (Cochran 1977).

continued

Box 1.7 A simple survey design checklist of key questions to answer in designing a survey to estimate characteristics or parameters of interest with an example from a survey of hunters to estimate percentage hunter success

Question	Example
1. What is the survey objective?	Estimate the percentage of successful hunters
2. What is the best technique to sample the population?	Telephone survey of permit holders
3. To which 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 the population to be sampled (N)?	$N = 350$ (for special permit hunt)
6. Which sample design is best?	Simple random sample
7. How large should the sample be?[a]	$n = \frac{Np(1-p)}{(N-1)\ B^2/4 + p(1-p)}$, 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 an estimate with $p \pm 0.05$ confidence) Therefore, $n = \frac{350(0.24)(1-0.24)}{(350-1)((0.05)^2/4) + 0.24(1-0.24)}$ $n = 159$ (i.e., we should contact approximately 159 permit holders)
8. Have you contacted a statistician to review design?	Yes!

[a]Scheaffer et al. (2005).

tistical hypothesis tests, but a statistician should be consulted for complicated experimental designs and analyses.

In contrast to conducting a prospective power analysis during the design phase, performing a retrospective power analysis after the data are collected or during the analysis phase is controversial or contraindicated (Thomas 1997, Steidl et al. 1997). Retrospective power analysis is uninformative unless effect sizes are set independently of the observed effect (Steidl et al. 1997).

For example, we want to test the prediction the fawn/doe ratio in a mule deer herd (i.e., our biological population is our research population) has declined. There are 4 possible results from sampling the herd and testing the prediction the fawn/doe ratio has declined (i.e., the null statistical hypothesis is there is no change; Table 1.3). 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), which represents the chance of concluding the ratio changed when in fact it did not. An $\alpha = 0.05$ indicates we would only make this error 5 times if the population ratio did not decline, and we tested it by drawing a sample 100 times. This is referred to as **Type I error**. Alternatively, we could conclude the ratio had not changed when in fact it had declined. For the situation where we count 500 deer, we would fail to detect the decline in the fawn/doe ratio 50% of the time (Table 1.3). This type of error is referred to as **Type II error**, which is measured by β. Typically, we set α low to minimize Type I errors, but Type II errors are equally 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 ($1 - \beta$). Managers wanting to detect declines in productivity would probably strive for a larger sample size in this case as well as accepting a 10% or 20% chance for a Type I error, which could jointly increase power to 90% or better.

Effect size (i.e., magnitude of effect) is an important metric influencing sample size requirements and power of a statistical hypothesis test. Power and sample size calculations should be based on a biologically meaningful effect size. Identifying a biologically significant effect size usually involves expressing the

Table 1.3. 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 ($\alpha = 0.05$).

	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% ($1-\alpha$)
2	60	60	None	65	50	Declined	Type I error	5% (α)
3	65	50	Declined	65	50	Declined	No error	50% ($1-\alpha$)
4	65	50	Declined	62	57	No change	Type II error	50% (α)

conceptual diagram as a quantitative model plus value judgments about the importance of a biological response. Estimating power of the test and calculating sample size requirements force the investigator to evaluate the potential significance of the research prior to beginning fieldwork. Sample size analysis may lead to substantial revision of the sampling and experimental design or the funds requested in a research proposal.

We cannot control natural variation within the population or the actual change that occurred (effect size), but we can control sample size, efficiency, and significance level (Steidl et al. 2000). For large sample sizes, **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 estimators and for large samples. When sample sizes are small (<30) and populations are not normally distributed, **nonparametric tests** that are based on ranks of values rather than their actual numerical value (e.g., Mann-Whitney, Wilcoxon signed-ranks test) are superior (Johnson 1995, Cherry 1998). The power of a test increases as sample size increases as long as samples or experimental units are truly random samples from the entire population to which the estimates or hypotheses will apply. Increasing variation with increased sample size indicates serial correlation in samples, which can lead to erroneous confidence intervals or conclusions from hypothesis tests due to pseudoreplication. 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).

Sampling Methodology

Many types of sampling methods exist, including plots, points, transects, road sampling, and paired sampling. These are briefly discussed in Box 1.8.

Model Experimentation and Validation

The validation and experimental phases of the modeling process described earlier (see "Steps to Building a Model" above) constitute an effort to confront theories with data. The models express our theoretical understanding of the system along with its characteristics and processes. The validation and experimental manipulation of a model confronts theory with data. Both model validation and experimentation are useful, especially when we conduct these activities within an adaptive management process where future management actions are accompanied by monitoring to simultaneously validate the predictions of the models (our theory or understanding of the system) and probe the behavior of the system (Walters 1986:250). Comparing model predictions to data completes a feedback loop to improve our understanding of ecological systems. Frequently, managers and researchers break the loop by ignoring inconsistencies detected because of the effort expended in developing the models and trade-offs in the selection of management actions. This leads to passive adaptation rather than a probing through experimental management actions. Walters (1986:251) noted that conservative, risk-averse decision making created a particularly difficult situation for learning, especially when effects of management actions are compounded with environmental changes and there are lags inherent in the responses. For example, if the desired outcome is a harvestable surplus of a game species, a manager can face substantial social, economic, and political pressure to find the right answer (see "Adaptive Management: Connecting Research with Management" below). Models are useful in efforts to ensure that management and ecological understanding are based on valid estimates and relationships rather than wishful thinking.

Problems to Avoid

Examine your methods carefully to avoid 3 common problems: procedural inconsistency, nonuniform treatments, and pseudoreplication (Box 1.9).

Approaches for Data Analysis

At this point, researchers have developed a well-thought-out and biologically meaningful research question, alternative research hypotheses, predictions, and statistical hypotheses; decisions have been made regarding study, experimental, and sam-

Box 1.8 Examples of sampling methods

Plots are used often to sample habitat characteristics and count animal numbers or animal signs. Plots represent small geographic areas (circular, square, rectangular, or hexagonal) that are the elements of the geographically defined population. The size of the research population is the number of these geographic areas (i.e., plots) that would cover the entire study area. Plots have the advantage that size of the research population (the domain) is known, and totals can be estimated (Seber 1982). However, sufficient time, money, and personnel to sample an entire study area are usually not available; therefore, a subset of plots is used to represent the entire study area. Any of several sampling designs (see Box 1.6), including more complicated designs such as the 2-stage designs, can be applied (Cochran 1977). Review of different sampling designs and consideration of the characteristics and patterns of distribution of species across the landscape are important prior to choosing a sampling design (Box 1.6). Selection of plot size and shape, also an important consideration, has been reviewed by Krebs (1999).

In **point sampling**, a set of points is established throughout the research population, and measurements are taken from each sample point (Fig. 1.9A). A common measurement is distance from the point to a member of the population (e.g., plant, calling bird). Examples include point quarter and nearest neighbor methods used 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 count surveys for birds record the distance to each bird detected, as in the variable circular plot approach, then transforming distances to areas makes it easy to apply the extensive methods and algorithms developed for line transect or distance sampling (Buckland et al. 1993, Laake et al. 1994, Buckland et al. 2004) and to correct for variability in detection probability (Kissling and Garton 2006). Selection of sample points usually follows a systematic design, but other sampling designs can be used if points are spaced sufficiently apart that few members of the population are sampled more than once. Necessary sample sizes can be estimated from formulas even if the population size is assumed to be large or unknown (Zar 1999).

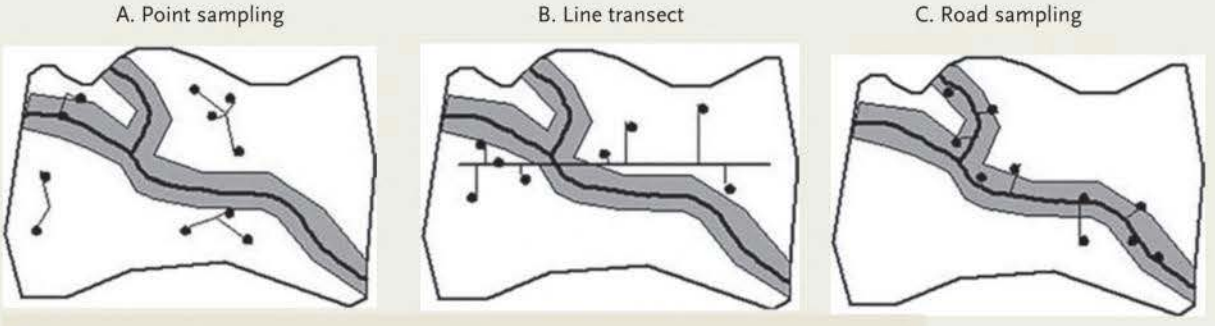

Fig. 1.9. Examples of sampling methods: (A) point sampling, (B) line transect, (C) road sampling.

A **transect** is a straight line or series of straight-line segments placed in the area to be sampled (Fig. 1.9B). Transects are used to organize or simplify establishment of a series of sample points or plots and as sample units. Transects are often used to obtain systematic samples of spatially distributed populations (e.g., plants or territorial animals). Plots along transects are actual sample units and should be treated as described under systematic sampling. Plots also can 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) are recorded. These distances are used to estimate effective width of the area sampled by the transect (Seber 1982, Buckland et al. 1993, 2004) and probability of detection at each distance. 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 plots, but should be established carefully with compass or transit and measuring tape or with a GPS unit. The critical assumptions of 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, other methods could 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, Buckland et al. 2004). A strip transect is not a line transect, but a long thin plot in which it is assumed all animals or objects are counted (Krebs 1999).

Sampling from roads is used for obtaining observations of species sparsely distributed over large areas or for distributing observations of abundant species over a large geographic area (see Fig. 1.9C). This sampling method is used for spotlight surveys of nocturnal species such as white-tailed deer (Boyd et al. 1986) and black-tailed jackrabbits (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, Sauer et al. 2015). This approach involves drawing a sample from a population defined as that population occupying an area within a distance x of a road. 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.

Road surveys rarely provide unbiased estimates for a region because they are generally placed along ridges or valleys and avoid steep or wet areas. Furthermore, roads modify habitat for many species and may attract some wildlife. For example, during snowy periods, some bird species are attracted to roads for grit and spilled grain. Thus, sampling along roads rarely provides a representative, unbiased sample of habitat and the research population (e.g., Hanowski and Niemi 1995). Although this bias is well known, it is often ignored in exchange for a 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 unbiased counts (Anderson 2001, Thompson 2002*a*). 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).

If we wish to make population comparisons, **paired 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 test a prediction that adult female bighorn sheep eat more woody plants than do bighorn lambs, we might treat a female with a lamb as a pair and measure the diet of each animal by counting the number of bites of each plant eaten while foraging together. Treating these observations as pairs would strengthen the comparison between age classes because it would compare animals foraging together and experiencing the same availability of plants. Pairing is a powerful technique when there is dependency between observations. Pairing only should be used if an association exists; otherwise, power of the comparison will be decreased.

Paired observations also can be used in studies of habitat selection 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, which are paired with available habitats measured at random sample plots throughout the study area (Fig. 1.10A). The logic behind this pairing is not obvious, but comparison of the use sites and random plots can identify characteristics of areas selected by the species. The more logical approach for Fig. 1.10A would be to treat the sample of use observations independently from the random available locations yielding a sample size for the comparison twice as large as if they were treated as paired observations. An alternative approach of pairing "use" and random plots is selecting a random plot within a certain distance of the use plot (Fig. 1.10B). For analysis, use and random plots are paired (i.e., random plot locations are dependent on use sites), which could produce different results from the unpaired comparison because it is testing for habitat differences within areas used by the species (microhabitat selection). By contrast, 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 research question, the explicit hypotheses of interest, and the spatial autocorrelation among sampling units, but both may be useful in applying a hierarchical approach to studying habitat selection (Wiens 1973, Johnson 1980).

continued

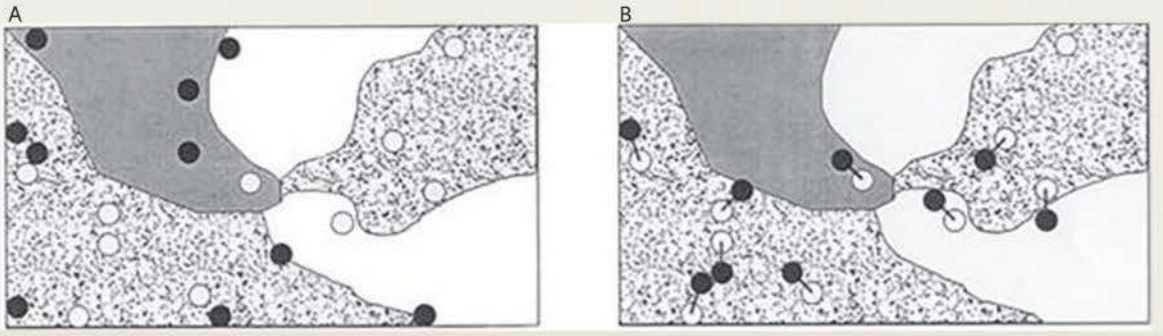

Fig. 1.10. Paired observations with examples of (A) use (○) and random (●) plots drawn at random from the study area and (B) use plots paired with random plots within a short distance of used plots.

Box 1.9 Common problems to avoid

Procedural inconsistency

This is a common problem that can be prevented with proper research design. For example, if a project is dependent on field personnel to accurately identify songs of forest birds, the data could be biased if 1 or more of the field personnel misidentified birds during the surveys (Cyr 1981). In this situation, the magnitude of the bias will depend on 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 in an explicit protocol, and all individuals collecting data should have similar skills and knowledge of the research methods (Kepler and Scott 1981). The design should incorporate double sampling or similar procedures to estimate the extent of observer bias (Farnsworth et al. 2002, Kissling and Garton 2006). This type of bias is often overlooked (or ignored) as a potential problem and seldom reported in research publications.

Nonuniform treatments

This common bias can be illustrated by considering a field experiment to evaluate effects of sharp and feathered forest edges on nest success of forest songbirds. If we used both clear cuts and road ways as sharp edges, we might have confused our treatment results because of differences in attractiveness of sharp edges near roads where carrion is an abundant attractant to generalist predators like common ravens. High variability between replicates in nonuniform treatments substantially reduces our power to detect biologically significant effects.

Pseudoreplication

If sample or experimental units are not independent, but instead are subsamples and are treated incorrectly as independent samples or experimental units, then the units are pseudoreplicated. This is often a problem in field ecology that should be avoided (Hurlbert 1984). In manipulative experiments, experimental units only are independent 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 sampling design should be reconsidered or the lack of independence must be treated correctly in the analysis through use of cluster sampling or adjustments in degrees of freedom for tests. More sophisticated analysis programs commonly test for this lack of independence between successive observations. The sample or experimental unit must have a direct tie to the research population. If the research population consists of 1 meadow in Yellowstone National Park, then 2 or more samples drawn from that meadow

would be replicates and our inferences or conclusions only would apply to that single meadow. If our research population consisted of all meadows in Yellowstone National Park, then 2 plots in the same meadow would constitute subsamples, not true replicate samples. Also, repeated sampling of the same radio-marked animal often constitutes a form of pseudoreplication. For example, if the research population consisted of moose in 1 ecoregion, repeated observations of habitat use by a single animal would not be true replicates. A similar problem would arise if 2 radio-marked animals were traveling together; thus, their habitat selection would not be independent. The data would have to be summarized into a single value, such as the proportion of the observations in a certain habitat or a more complicated design treating repeated observations of the same animal as subsamples for statistical analysis. This would reduce sample size to the number of radio-marked moose. Treating repeated observations of the same animal as replicates only is justified when the individual animal is the research population. In this situation, tests for serial correlation should be conducted to ensure that observations are not repeated so frequently to still be pseudoreplicates (Swihart and Slade 1985).

Box 1.10 Approaches for statistical data analysis

Described are 3 paradigms for analyzing data: frequentist, Bayesian, and Fisherian (modified from Efron 1998).

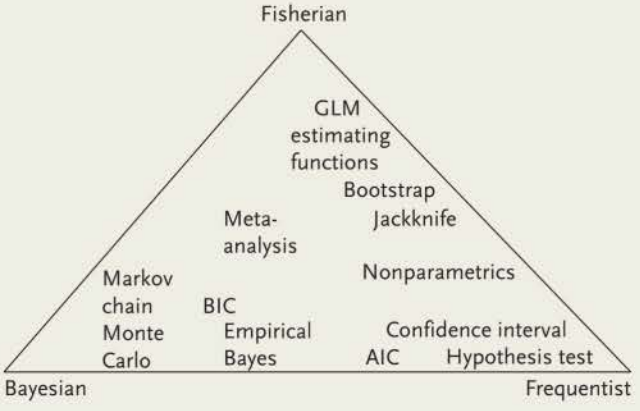

Fig. 1.11. Three paradigms for analyzing data: frequentist, Bayesian, and Fisherian. GLM = generalized linear model, BIC = Bayesian information criterion, and AIC = Akaike's information criterion. *Modified from Efron (1998).*

Frequentist approaches include statistical **hypothesis testing**, which is sometimes referred to as null hypothesis testing and is probably the most familiar and widely used approach for data analysis. Statistical hypothesis testing is rooted in the philosophical idea of falsification, in which an attempt is made to disprove a statistical hypothesis, leaving the alternate statistical hypothesis to be tentatively accepted (Underwood 1997). This is referred to as a frequentist method because it estimates a probability (frequency) that the observed estimate would have occurred solely by chance if the null hypothesis were really true. Johnson (1999) described the basic steps of statistical hypothesis testing that mirror the approach suggested by Underwood (1997). First, the researcher uses predictions from the alternative research hypotheses to develop a statistical hypothesis that reflects an expected pattern or the effects of some treatment. Next, the logical opposite of this statistical hypothesis (e.g., no effect of the treatment or explanatory factor) is usually set as the null hypothesis, and data are collected to assess the validity of the null hypothesis.

Information-theoretic (IT) model selection also is a frequentist approach, but offers a distinct alternative to statistical hypothesis testing. IT model selection has seen widespread growth in wildlife and ecological sciences in recent years (Hilborn and Mangel 1997, Burnham and Anderson 2002, Johnson and Omland 2004, Richards 2005, Cade 2015). In con-

continued

trast to traditional statistical hypothesis testing, IT can be used to evaluate a suite of statistical models devised to represent predictions from alternative research hypotheses and thereby identify the one that is most strongly supported by the data while acknowledging all are wrong or incomplete characterizations of the ecological process. The IT approach can be used to pursue Chamberlin's (1965 [1890]) approach to science by constructing multiple working hypotheses that are subject to repeated confrontation with empirical data (e.g., **AIC**, **BIC**). Those that are supported by the data tend to be retained while those with little support tend to be dropped from consideration (Burnham and Anderson 2002).

Statistical hypothesis testing and IT model selection are only 1 component of statistical inference, while estimation of effect sizes and measures of precision are at least as important (Johnson 1999, Robinson and Wainer 2002, Stephens et al. 2007*b*). Quinn and Dunham (1983:613) suggested the objective of biological research typically is to assess the relative contributions of a number of potential causal agents operating simultaneously. If this is the case, estimation of effect sizes is of primary importance to wildlife science, and these results should be emphasized in data analysis as well as the parallel approach of using IT methods (**AIC** and **BIC**) to compare the likelihood of multiple alternative hypotheses stated as competing models with parameters estimated using maximum likelihood methods. Läärä (2009:152) echoed this sentiment and noted the basic tools for statistical reasoning are based on the strength of associations and the sizes of differences. The author further noted the effects are provided by point estimates, their standard errors, and associated confidence intervals. Reporting effect sizes not only are important for practical interpretation of the research, but are also the critical components for any subsequent **meta-analysis** (Gurevitch et al. 2001, Hobbs and Hilborn 2006). Läärä (2009) contains several practical recommendations for presenting and interpreting effect sizes that are useful.

Efron (1998) described the **Fisherian** approach as follows: "Fisher's philosophy is characterized as a series of shrewd compromises between the Bayesian and frequentist viewpoints, augmented by some unique characteristics that are particularly useful in applied problems." Efron (1998) emphasized Fisher's approach as fundamentally driven by his top priority of providing statistical tools of value to application of statistics to challenging studies in natural resources. He pioneered use of maximum likelihood estimation in analyses of experimental designs and broadened its applicability with inferences based on randomization, which expand the arsenal of statistical tools beyond general linear models (**GLMs**) based on assuming normal distributions of estimators into the widely applicable approaches termed **estimating functions**, **bootstraps,** and **jackknifes**. Wildlife scientists and managers will find these tools extremely valuable for their analyses if they succeed in interacting with forward-thinking applied statisticians.

Bayesian data analyses are 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 (Hilborn and Mangel 1997, Gelman et al. 2003:3). Johnson (1999) provided an easily understood description of the conceptual differences between the frequentist and Bayesian approaches. A Bayesian analysis requires performing 3 basic steps (Gelman et al. 2003, 2013).

1. **Specify a probability distribution for all quantities** or, in other words, use prior studies and creative thinking to specify a prior probability for the parameter(s). To begin, state the range of all possible values for the characteristic(s) to be measured and make the best guess of the values that are most likely based on earlier studies and critical thinking. This is controversial because it involves subjective decisions and could be misused if someone wanted to produce a specific result (Dennis 1996). However, well-designed research should gather historical data to inform the probability distribution of the parameter(s) as part of a literature review and pilot study.

2. Use the observed data to **calculate a posterior distribution for the parameter** of interest as a conditional probability distribution. This step follows data collection because our prior guess of the value for the characteristic we are measuring is improved by combining it with the newly collected data, and we can state conclusively our best **posterior guess** of the value of the characteristic. Bayes's rule is used for this step, and the Bayesian estimate could be considered a weighted average estimate based on the sample data and the assumed prior value in which weights are proportional to the precision of the observed and prior values (Gelman et al. 2003:43). As sample size increases, the Bayesian estimate approaches the maximum likelihood estimate, and any influence of the prior probability vanishes. **Markov chain Monte Carlo** methods on a computer are used widely for these calculations as they make it feasible to estimate probabilities for complex models that cannot be solved analytically.

3. **Evaluate** the fit of the model and the implications of the resulting posterior distribution. This involves addressing questions such as the following: does the model fit the data, are the substantive conclusions reasonable, and how sensitive are the results to the modeling assumptions?

pling designs; and empirical data (e.g., pilot study or previous research) have been collected to shed light on the necessary sample sizes for observations or experiments. Now researchers must decide on a statistical approach. General approaches for data analysis include Bayesian to frequentist paradigms (Box 1.10; Fig. 1.11) with distinct differences in how probability should be interpreted (Cox 2006). We provide an overview of approaches for data analysis (Box 1.10), but a thorough treatment of these paradigms is beyond the scope of this chapter. However, we refer readers to Collier and Schwertner (Chapter 2, this volume) and textbooks on statistical analysis (e.g., Zar 1999). We also refer readers to Ellison (2004), who summarized the main differences between Bayesian and frequentist approaches to statistical inference (also see Ellison 1996, Dennis 1996, Taper and Lele 2004, Hobbs and Hilborn 2006). The Fisherian approach (Box 1.10) provides a compromise between these 2 approaches, with Fisher's GLMs (including linear regression, ANOVA, and analysis of covariance) and more recent variations (estimating functions, bootstraps, and jackknifes as well as fixed-, random-, and mixed-effects linear models based on maximum likelihood estimation) providing the workhorses of analysis for experimental designs in natural resources. Preliminary ideas about statistical analysis should be vetted with biometricians and applied statisticians at this point in the design phase, and their suggestions often will require iterating back to earlier steps in the investigation process because of challenges in meeting statistical assumptions for analyses.

Research Proposal

At this step in the research process, you have a research question, alternative research hypotheses, and predictions. You also have a general plan about how to test your predictions and make statistical inferences as well as have conducted a pilot study or gathered previously collected data, determined sample design, and estimated sample sizes. All of these components of the research process can now be brought together into a written research proposal. A research proposal is important so that you can solicit peer review (see below) regarding your plan's feasibility for answering your research question and informing management, policy, and theory. It should include detailed methodology and the potential limitations of your project. Writing a research proposal can help ensure that your alternative research hypotheses address the research question and explain how your sampling or experimental design and estimates of sample sizes will inform statistical inferences. A formal research proposal is a vital step in developing a successful project and advancing wildlife science (Pierce et al. 2012).

Peer Review

Peer review (i.e., evaluation by independent, qualified reviewers) of the proposed research should be obtained from several people with expertise and experience with the research topic. Reach out to both local peers in your community and experts who have published work on the research topic and ask them to review your research proposal. Peer review will often improve a research design and could identify serious problems that can be solved prior to proceeding with the research project. Unfortunately, most peer reviews occur too late, after data collection or when the final report or manuscript is written. Avoid the common mistake of thinking that your research proposal is not sufficiently polished to send to others for peer review—if that is truly the case, then you should not go into the field to collect data yet either.

STEP 5: FINALIZE DESIGN AND CONDUCT RESEARCH

The outcome of Step 4 above (i.e., design formal testing) is an explicit list of predictions suitable for testing along with an experiment(s), sampling plan(s), and/or model(s) designed to test each prediction(s). Adding estimates of the statistical power, costs, logistics, and feasibility of each test will help ensure that you obtain a clear answer to the original research question. Predictions are sometimes in the form of statistical null hypothesis tests, which are typically directional (1-tailed) statistical hypotheses because an explicit relationship is being tested that is predicted by a research hypothesis (e.g., H_0: elk density is not related to forage quality vs. the alternative H_1: elk density is positively correlated with forage quality) rather than a point-null hypothesis test (e.g., H_0: elk density is not related to forage quality vs. the alternative H_1: elk density is positively or negatively correlated with forage quality).

Although predictions are often proposed as a good way to test a specific research hypothesis, researchers should consider whether alternative research hypotheses make the same prediction, and if so, the prediction may not be as informative as other predictions. To assess the effectiveness of different predictions and to indicate inferences that can be made from testing the predictions, make a list of specific testable predictions and each of the research hypotheses in a matrix (Box 1.11). In the matrix, include details of the variables or relationships to be tested, their specific or relative values, and whether each test performed supports or refutes each hypothesis (Box 1.11). This matrix organizes the predictions and research hypotheses in an effective way to (1) determine if the outcome of your tests will refute any of the research hypotheses, (2) highlight the potential for an integrated approach to identify interactions and multicausal challenges, and (3) summarize the overall structure of your study and the reason for doing each test. For example, in the elk study outlined in Box 1.11, all of the predictions provide clear tests of research hypotheses except the prediction that severe winters decrease elk calf survival.

Data Collection

Most novice research biologists are anxious to initiate data collection because of the attractiveness of working outdoors and the pleasure derived from observing wildlife-related phenomena. However, the design phase should not be rushed to

Box 1.11 Example of a matrix of research hypotheses and their associated predictions for the research question of why elk populations are declining

This example is based on the research hypotheses and predictions described in Box 1.1. To assess the effectiveness of different predictions and to indicate inferences that can be made from testing the predictions, make a list of specific testable predictions and each of the research hypotheses in a matrix. In the matrix, include details of the variables or relationships to be tested, their specific or relative values, and whether each test performed supports or refutes each research hypothesis. This matrix organizes the predictions and research hypotheses in an effective way to (1) determine if the outcome of your tests will refute any of the research hypotheses, (2) highlight the potential for an integrated approach to identify interactions and multicausal challenges, and (3) summarize the overall structure of the study and the reason for doing each test. For example, in the elk study outlined in Box 1.1, all the predictions provide clear tests of research hypotheses except the prediction that severe winters decrease elk calf survival. To make this prediction stronger, the mechanism for winter severity, such as snow depth, could be addressed.

Predictions	Research Hypothesis 1: Decrease in amount of quality forage	Research Hypothesis 2: Increase in predation rate by wolves	Research Hypothesis 3: Increase in winter severity
Experimental reductions in wolf abundance will increase population growth rates of elk	No	Yes	Yes
Captive female elk that received a higher amount of quality forage will have higher pregnancy and parturition rates than a control population that does not receive supplemental forage	Yes	No	No
Among elk populations, the population growth rate of elk is most strongly correlated with:	Amount of quality forage (positive correlation)	Predation rate (negative correlation)	Winter severity (negative correlation)
An integrated population model will suggest that elk population dynamics are most strongly affected by changes in:	Amount of quality forage	Wolf predation rate	Winter severity
In severe winter, elk calf survival decreases	Yes	Yes	Yes

initiate fieldwork more quickly. Successful research biologists often find their projects are most successful when they spend about 20–40% of their time in design and planning phases, 20–40% in actual fieldwork, and 40% in data analysis and writing publications.

All data should be recorded on preprinted data sheets or entered directly into a handheld data logger, computer, tablet, or mobile device. This ensures that each observer collects the same data, yielding consistent collection of data that simplifies data preparation and 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. An explicit quality assurance/quality check (i.e., QA/QC) protocol should be developed prior to collection of any data, and the protocol should be followed to reduce errors. Transcription of data, including computer data entry, must be followed by careful proofreading, which is greatly facilitated by checking for valid entries using database queries and spreadsheet scripts. All field personnel should receive the QA/QC protocol regarding data collection, and the prin-

cipal 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 the final responsibility for quality control, and the validity of research results depends on quality of research design and data collection.

Data Analysis

Performing statistical analysis during data collection, such as estimating descriptive statistics and testing null hypotheses will facilitate detecting errors in the methods or their application and sharpen interest of field personnel and supervisors. Any problems detected can be corrected before they degrade results of the whole project. Problems and questions that develop may require further contact with experts and consulting statisticians. All of this effort to prepare the data for data analysis often takes as long as you expected the entire statistical analysis to require. Final statistical analysis (Box 1.10) provides an opportunity to use the full data set to formally confront theories and alternative hypotheses with new data to answer the original question. This exciting step provides a major payoff for all the effort expended in designing and gathering the right data to answer the question by evaluating relative support for the alternative hypotheses as well as rejecting any hypotheses that are not supported by tests of their predictions.

STEP 6: EVALUATE, INTERPRET, AND DRAW CONCLUSIONS

Evaluation and interpretation of results also is a creative phase of the research process. The quality of conclusions drawn is dependent on the biologist's past educational and professional experience as well as willingness to consider standard and less-traditional interpretations. A danger in wildlife science (and other fields) is that researchers often have a conscious or unconscious expectation of results. This bias could begin with development of the overall research question, what research hypotheses to consider (Platt 1964), and what predictions to test, which carry through to interpretation of results. To reduce this bias, scientific fields, such as medicine, perform experiments with a double-blind approach in which neither researcher nor subjects know membership of treatment and nontreatment groups. Researchers who are consciously aware of their own biases and strive to keep an open mind to new ideas are most likely to gain reliable knowledge.

The objective of this step is to organize and interpret, clearly and concisely, the results of specific statistical analyses. A matrix (Box 1.11) makes interpretation more straightforward and helps reduce subjectivity at this stage. Research results are transformed from a collection of specific information into a synthesis explaining which research hypotheses were supported and what insights can be gained regarding the answer to the research question. Do statistical evaluations support 1 or more of the research hypotheses and clearly reject others? Do the results provide a reasonable explanation of the cause of the ecological pattern or process responsible for it? Are there alternative explanations of the data and statistical tests? Are there data limitations that should be identified, such as inadequate sample sizes or unusual variation in specific variables measured? What could have introduced bias into the estimates or comparisons? Are additional data required? Are there other research hypotheses that were not considered (see below)? These questions must be considered carefully, and if concerns are identified, they must be noted in the evaluation and interpretation of research results.

If the data and statistical results support or reject a research hypothesis(ses), we can provide some insight (a tentative answer) into the research question. We do not prove a research hypothesis to be correct. Indeed, some would argue that **all** research hypotheses are to some degree incorrect, but instead, the credibility of a research hypothesis increases as more predictions are supported and alternative research hypotheses are not supported. In complex multicausal questions, the success at identifying important predictors only may come from tests evaluating multiple causes simultaneously. The goal of science is knowledge, and in wildlife science, we attempt to explain the ecological processes responsible for ecological patterns and phenomena and thereby help predict how changes (natural or anthropogenic) will affect wildlife.

We need to carefully consider how broadly our conclusions can be generalized to other areas or populations and not allow our conclusions to go beyond the data. Interpretation of research data must clearly separate conclusions and inferences based on data from speculation. For example, if we demonstrate via a statistical hypothesis test that droppings from spruce grouse are most abundant under lodgepole pine and Engelmann spruce trees, we can conclude that grouse use both tree species. However, without conducting additional tests of explicit predictions deduced from research hypotheses to explain why (e.g., predictions might focus on feeding activity or fecal analyses), we do not know the reason for the observed pattern. Likewise, replicated studies across space and time provide greater confidence that certain relationships are general and not specific to the circumstances that prevailed during a single study (Johnson 2002). Hypotheses repeatedly tested and supported become theories widely accepted as the basis for both ecological understanding and successful management.

Publication

A vital step of the research process is publication because publishing the results provides an outlet for sharing research results and adding to the current knowledge on wildlife research and management. One can make the case that if the results of a study are not published in a journal or other rigorously reviewed publication, then the study should not have been conducted because its ability to inform management, policy, or theory will be severely limited. Unfortunately, many research dollars are wasted because knowledge gained was not published, and the information is buried in file cabinets, boxes of

data sheets, or an unpublished graduate thesis or agency report. The publication process takes substantial time and patience, but it is rewarding. The ability to write clearly and concisely is gained with practice and feedback. Therefore, write a draft manuscript and get it peer reviewed. Requesting a friendly, courtesy review from colleagues prior to polishing your manuscript for submission to a journal will assist you enormously in the process of manuscript revision and editing essential to steering your work through the publishing process. Peer reviews by journals can sometimes be daunting, because we must subject our work to anonymous critiques used by editors to judge whether the manuscript is acceptable for publication. Do not be unduly deterred by rejection of your manuscript, but use the critical comments to improve your writing and resubmit to the same or another outlet. The reward of publication is knowing your research is worthy of publication because your peers have approved it and you are adding to the body of knowledge on wildlife science.

Agency administrators should encourage or reward employees for publishing their work even though they are pressured for immediate answers to management problems. Effective administrators recognize that peer review and publication will (1) correct errors and possibly lead to an improved analysis, (2) help authors reach sound conclusions with 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 science and management by making results readily available to other agencies, managers, researchers, and students.

Publication and peer review are essential to science because they promote better-quality research. Authors should not be discouraged by defending their research and revising manuscripts because peer review can improve the quality of a manuscript. Research is not complete and does not contribute to knowledge and sound management of wildlife until research results are published to effectively communicate to the scientific community and 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.

Formulate New Hypotheses

Rarely does a single research project provide the last word on any problem (Johnson 2002). More commonly, research will generate as many questions as answers. Speculation, based on inconclusive or incomplete evidence for or against research hypotheses, is the fuel for future research and one of the most important aspects of science. Speculation (e.g., retroduction) must be identified as such and should not be confused with conclusions based on data (e.g., interpretation of results from hypothesis tests). Many natural phenomena have been discovered by accident—an unexpected result from some associated research effort (Royston 1989).

New research hypotheses are a form of speculation, which are verbalized in a formal fashion to answer a specific research question. For example, while conducting a study to test whether spruce grouse use lodgepole pine and Engelmann spruce trees for (1) feeding or (2) roosting, we might learn that 80% of spruce grouse diet is lodgepole pine even though Engelmann spruce is more abundant. Hence, we could formulate a new research hypothesis, such as lodgepole pine needles have higher nutritional quality than Engelmann spruce needles, to explain why grouse eat more lodgepole pine despite its relative rarity.

Adaptive Management: Connecting Research with Management

Wildlife management programs should be developed from scientific knowledge based on research on topics such as population ecology, habitat selection, and/or behavior. The formulation of a management program is similar to formulation of a research hypothesis; both are based on assumed relationships and predicted outcomes. Our management prediction is that our plan of action will achieve a desired management result. However, nearly all wildlife management programs throughout the world lack research on the effectiveness of their management actions (Macnab 1983, Gill 1985, Westgate et al. 2013). Seldom is the question "Does our management lead to the desired result?" addressed in formal, well-designed, long-term research projects. For example, does prescribed burning as a management practice increase mule deer and elk populations? The effectiveness of this management practice has not been addressed, and most evaluations have noted increases in forage species and changes in animal distributions, but increases in population levels in response to prescribed burning have not been adequately documented (Peek 1989). A second example is the use of population indices to monitor changes in wildlife population densities (e.g., ring-necked pheasant crowing counts). The primary assumption of a population index is 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 population index validation exist (e.g., Rotella and Ratti 1986, Crête and Messier 1987). Some studies have disclosed that population 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 called adaptive management. Adaptive management involves formal specification of management goals and responses to management actions through predictive models (Table 1.2) based on alternative management strategies (akin to alternative research hypotheses), which can be compared to actual system responses through monitoring (Thompson et al. 1998, Sauer and Knutson 2008, Conroy and Peterson 2009). Management actions are treated as experiments, which must be monitored to ascertain if goals were met and to identify misinformation

regarding the dynamics of a managed natural system and effects of management actions. Using monitoring data results, 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, Johnson et al. 2002, Westgate et al. 2013). Adaptive management is an interactive, iterative process in which learning over time improves management if a monitoring program provides feedback to both our understanding of the system and management actions (Conroy and Peterson 2009, Westgate et al. 2013). A less desirable, but sometimes necessary alternative is termed **scenario planning**, an approach applicable to situations with responses occurring at such long time scales that monitoring cannot provide timely feedback in the face of future changes in relationships associated with highly variable external drivers such as climate change (Peterson et al. 2003*a*).

Adaptive management is a specific case of **structured decision making**, which is a process that addresses complexity, uncertainty, multiple objectives, and different perspectives to provide defensible management decisions after accounting for all available knowledge (Clemen 1996). Structured decision making has multiple steps: problem definition, objectives, alternatives, consequences, trade-offs, uncertainty, risk tolerance, and linked decisions (Conroy et al. 2008). The basic strength of structured decision making is that it allows wildlife managers to make effective decisions more consistently and to provide a structured, objective approach for working on hard management decisions (Clemen 1996). Wildlife managers are faced with difficult decisions regarding both management and conservation of wildlife. For example, how can bison be restored to their former range, which could benefit other threatened prairie species, while also considering the economic and social impacts to cattle ranchers if the infectious disease brucellosis can spread from bison to cattle? Both structured decision making and adaptive management are being used more often by wildlife managers (Conroy et al. 2002, Johnson et al. 2002, Dorazio and Johnson 2003, Regan et al. 2005, Moore and Conroy 2006, McCarthy and Possingham 2007, Martin et al. 2009).

Because wildlife management agencies have the responsibility for management of wildlife populations and their habitats, they also have the responsibility to assess and improve the effectiveness of management programs. Administrators and managers in wildlife management agencies that wish to base management decisions on the best available science should strive to develop well-designed, long-term management-research programs as a basic component of annual agency operations (see Westgate et al. 2013 for additional recommendations for successfully applying adaptive management).

STEP 7: INFORM THEORY, MANAGEMENT, AND POLICY

To increase the likelihood that research results inform management, policy and theory once a study is completed, do the following: (1) explicitly state how results support, refute, or change existing theory in the research proposal and in the discussion section of final reports and manuscripts; (2) explicitly state how results inform specific management actions or policies in the research proposal and in the discussion section of final reports and manuscripts; (3) propose new research hypotheses based on results or propose different predictions that were not tested in the discussion section of final reports and manuscripts; and (4) include wildlife managers and policy decision makers in the planning of the project, specifically request their review of the research proposal, and keep them engaged throughout the project. Discussion sections for all reports and manuscripts should strive to do both the first and second steps above, but many only attempt 1 (or none!) of the 2 steps. It is useful and informative to explore both angles. In a well-designed study that follows the research process laid out in this chapter, it will be obvious how the results of the study will inform theory, management, and policy, but only if this is considered early in the design.

SUMMARY

Carefully designed wildlife research will improve the reliability of knowledge that is the basis of wildlife management and conservation. Researchers must rigorously apply the research process (i.e., the scientific method) by stating clear research questions, evaluating alternative research hypotheses, formulating predictions, and testing those specific predictions using powerful techniques in survey sampling, modeling, and experimental design. Modeling offers a powerful tool 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 ecological understanding. More effort should be dedicated to the design phase of research, including obtaining critiques from managers, experts, other biologists, and statisticians; conducting a pilot study; and avoiding common problems, such as insufficient sample sizes, procedural inconsistencies, nonuniform treatments, and pseudoreplication. Wherever possible, an integrated approach should be considered, including a combination of descriptive studies for estimation, experimental studies, and models, because an integrated approach will provide a more reliable basis for conclusions and interpretations. We view the research process as iterative even though for ease of organization, we laid it out here as a linear series of steps. Wildlife biologists have a tremendous responsibility associated with management of wildlife populations, which are experiencing a multitude of environmental stressors, including loss of habitat, habitat fragmentation, climate change, and the spread of invasive species, all of which may lead to declining populations. We must face these changes armed with knowledge based on rigorous scientific research.