6 Formulating and testing hypotheses

> "*Construction of a hypothesis implies a belief that there exists a degree of order or regularity that can be identified and measured despite fluctuations in response*"
>
> (Skalski and Robson 1992)

6.1 Hypotheses

The term hypothesis has been mentioned several times in the preceding chapters. Hypothesis has many meanings, ranging from any speculative thought to "*concrete, specific conjectures on the process that lead to an outcome*" (Guthery et al. 2004). The definition I will use is that a hypothesis is a proposition set forth as an explanation for the occurrence of a specified phenomenon. The basis of scientific investigation is the collection of information that is used either to formulate or to test hypotheses. One assesses the important variables and tries to build a model or hypothesis that explains the observed phenomenon. In general, a hypothesis is formulated by rephrasing the objective of a study as a statement, e.g., if the objective of an investigation is to determine if a pesticide is safe, the resulting hypotheses might be that 'the pesticide is not safe' or that 'the pesticide is safe'. A hypothesis is a statistical hypothesis if it is stated in terms related to the distribution of populations. The general hypothesis above might be refined to: 'this pesticide, when used as directed, has no effect on the average number of robins in an area', which is a testable hypothesis. The hypothesis to be tested is called the **null hypothesis** (H_0). The **alternative hypothesis** (H_1) for the above example would be 'this pesticide, when used as directed, has an effect on the average number of robins in an area'. In testing a hypothesis, H_0 is considered to be true, unless the sample data indicate otherwise, (i.e., that the pesticide is innocent, unless proven guilty). Testing cannot prove H_0 to be true but the results can cause it to be rejected. Failing to reject the hypothesis does not mean that it is true. In accepting or rejecting H_0, two types of error may be made. If H_0 is rejected when, in fact, it is true a **type 1** error has been committed. If H_0 is not true and the test fails to reject it, a **type 2** error has been made.

The decision to accept or reject H_0 is made based on some estimated risk of being wrong in that decision, and usually the probability of making a

type 1 error (rejecting a true hypothesis) is of greatest concern. The probability of this error is called the level of significance of the test and the acceptable level of significance should be established prior to, rather than after, testing. The level of significance chosen in any situation is a subjective decision. In most areas of science, this is commonly set at 0.05, i.e., one accepts a 1 in 20 chance of being in error. The choice of a less rigorous test invites criticism and, in many instances, more rigor may be appropriate. However, in many situations dealing with wild populations, the investigator should ask himself or herself, quietly, if the methods available for counting animals and measuring other variables are really reliable enough to justify such confidence.

The choice of which of two alternative hypotheses to use as H_0 and which as H_1 is an important decision because, in statistical tests, H_0 is not rejected unless the evidence against it is overwhelming. In making the choice, one must consider which type of error (type 1 or type 2) is more critical in a biological or real world sense. If you were asked to test the safety of a pesticide, with licensure for widespread use depending on your results, the errors that might be made would (i) result in use of an unsafe pesticide that would risk animal and human health, or (ii) not allow use of a safe chemical resulting in higher costs of production for a crop. Most people would consider it to be far more costly to allow the use of an unsafe chemical than to disallow the use of a safe one. In this case, the appropriate decision should be that H_0 = the chemical is unsafe and H_1 = the chemical is safe, because, in this way, there is a smaller possibility of erring by allowing use of an unsafe chemical.

In other situations, where the risks are less well defined, the hypothesis that there is no effect is usually taken as H_0. For instance, if we were evaluating the efficacy of a new drug for potential use in the control of lungworms in bighorn sheep, we would likely choose that H_0 = the drug has no effect on the number of lungworms. This assumes that there is no effect and places the burden of proof on the pharmacologist (and the worms) to demonstrate an effect by disproving the hypothesis.

Hypotheses are tested by comparing them to observed data. When a hypothesis fails to meet or explain the data, one first checks the data, and then one tries to improve the hypothesis. This process is a continuous one of refining and retesting. In some instances, several competing hypotheses may be proposed and examined to see which one best explains some phenomenon. For instance, Caley and Hone (2002) developed a set of hypotheses or models that might explain how tuberculosis is transmitted to wild ferrets in New Zealand. They fitted these models to age-specific prevalence data collected in the field as a test of the competing hypotheses to determine which model best approximated the field situation. They found that consumption of tuberculous carrion or prey was the most strongly supported model for transmission to ferrets.

While I have presented the traditional approach of using statistical hypothesis testing, the reader should be aware that this approach has limitations and has been criticized for use in wildlife management (Johnson 1999;

Anderson et al. 2000; Guthery et al. 2001). Johnson (1999) raised serious concerns about the usefulness of statistical tests of hypothesis for ecological studies, and the need to be clear about the difference between statistical and biological significance. Alternative approaches, termed "*hypothesis-free science*" by Guthery et al. (2004), include purely descriptive studies, measures of magnitude of effect, and information-theoretic methods that provide strength information on multiple working hypotheses (models), all of which are plausible. Anderson and Burnham (2002) suggested that the need for modeling expertise in the latter of these is "*an excellent reason to seek the help of a statistician*".

The remainder of this chapter will be devoted to methods for collecting information about disease in populations, i.e., epizootiologic data, and very little will be said about collection of clinical information from individual animals.

6.2 Collecting information

Table 6.1 presents a schematic classification of the various methods used for investigating disease conditions. It should be recognized at the outset that there is considerable overlap among the various types and that individual investigations may involve elements of several types. However, each technique has inherent strengths and weaknesses that suit it for particular problems.

The most basic distinction is between **observational** and **experimental** studies. Observational studies are those in which information is collected about naturally occurring events and in which the investigator does not play an active part in what happens. In contrast, experimental studies measure the effect of manipulations caused by the investigator. To illustrate the difference, consider methods that might be used to study pneumonia in wild sheep. One method might consist of identifying and cataloguing the nasal microflora in bighorn sheep before, during and after a spontaneous outbreak of pneumonia. This is an observational study because the investigator is trying to study events as they occur, without manipulation. A second method might be to study the nasal

Table 6.1 Relationship among various forms of investigative methods that may be used in the study of disease

Investigation		
Experimental	Observational	
	Descriptive	**Analytical**
– laboratory experiment		– cross-sectional study
– field trial		– case-control study
– community trial		– cohort study

microflora before and after the sheep were treated with an antibiotic. This is an experimental study in which the object is to determine the effects of a manipulation. Both studies might be valuable in understanding pneumonia in sheep, and a combination of observational and experimental methods may provide the best information about a disease. As an example, Caley et al. (2001b) used both methods in a study of the relationship between the occurrence of tuberculosis in ferrets and the abundance of brushtail possums. In the observational portion of the study, the prevalence of tuberculosis in ferrets was found to be significantly related to the abundance of possums at a number of sites. When the abundance of possums was experimentally reduced, there was an 80% reduction in the odds of tuberculosis in ferrets in the years immediately after possum depopulation. The conclusion was that the transmission from possum to ferret accounted for most of the tuberculosis in ferrets.

Because scientists are not invisible observers, a problem in all observational studies is the need to minimize the unintentional manipulation that may occur during the investigation because of the presence of the investigator and any handling that may be required to mark animals. This was alluded to earlier in Chaps. 2 and 4, and will be mentioned periodically elsewhere. Whenever possible, the effects of manipulations on factors such as behavior and survival should be measured as part of the study and not assumed to have no effect. Examples of studies that measured the effect of some procedure involved in marking or handling animals are given in Table 6.2.

Table 6.2 Examples of studies that have measured the effect of procedures used for sampling, handling, or marking on wild animals

Species	Handling or marking procedure	Effect
Mallard	Radio-transmitter	Negative effect on reproduction and survival[1]
Wild turkey	Radio-transmitter	Negative effect on wing growth[2]
Moose	Radio-transmitter	No measurable on survival[3]
Grey partridge	Radio-transmitter	Adverse effect on survival, reproduction and body mass in some years[4]
White-winged dove	Radio-transmitter	No effect on blood parameters[5]
Big brown bat	Anesthesia, blood sampling	No measurable effect on survival[6]

[1] Paquette et al. (1997)
[2] Hubbard et al. (1998)
[3] Swenson et al. (1999)
[4] Bro et al. (1999)
[5] Small et al. (2005)
[6] Wimsatt et al. (2005)

Most of the emphasis in the biological sciences and, particularly in post-graduate training, is on experimental methods, so that I will assume that most readers are well acquainted with these techniques. Consequently, the emphasis here will be on observational methods, but many of the general features of sampling, data collection, and analysis apply to both types and individual studies often involve a mixture of observational and experimental elements.

6.2.1 Experimental methods

Before discussing observational methods, a few comments should be made about the various experimental methods. In Table 6.1, three such methods are indicated. In all of these, the investigator alters or manipulates one variable and then measures the resulting change in some other variable. The three methods differ in the way that subjects are chosen for inclusion in the trial, in the degree of control that the investigator has over other variables, and in the method that is used to assess change in the other variables.

It is easiest to explain these differences through the use of an example. Assume that we are interested in determining the efficacy of a vaccine for preventing disease caused by agent X. We could test this by any of the three experimental methods. In both laboratory and field experiments (the latter are usually referred to as clinical trials in human medicine), the experimenter controls the allocation of individuals to the principal and control groups. So, for a study using either of these methods, we might select 100 suitable animals and assign them randomly into two equal-sized groups. The 50 animals in the principal group would be vaccinated while the 50 animals in the control group would not be immunized. To this point the methods are the same but they differ in the technique that is used to test or **challenge** the vaccine. We want to test the efficacy of the immunization. Using the **laboratory experiment** method, each of the 100 animals would be administered a standard challenge dose of agent X and we would determine the effectiveness of the vaccine by comparing the results of this experimental infection in the principal and control groups. The investigator in a laboratory experiment controls all aspects of the challenge (dose, route, timing, etc.). In contrast, if we were to use the **field trial** method we would mark and release all 100 animals back into the wild after having immunized the 50 animals in the principal group. Challenge would occur through natural exposure to agent X and we would have no control over which, or how many, of the animals were exposed. Nor could we control the dose, route or timing of exposure. We would determine the effectiveness of the immunization by measuring and comparing parameters such as the survival time and rate of animals in the two groups, using some of the techniques discussed in Chap. 4 (this assumes that we would be able to find the animals again after release!). In both laboratory and field experiments, the effect of the manipulation is measured by the response in

individual animals. A study of the survival of raccoons immunized against rabies and released into the wild (Brown et al. 1990) is an example of a field trial.

If we were to use a **community trial**, the vaccine would be made available to animals in the area or community, perhaps in the form of an oral bait. We, as investigators, have no control over which or how many animals will consume the bait or become immunized. Challenge of the animals occurs through natural exposure to agent X, as in the field trial. Assessment of the results is done by measuring some indicator of disease occurrence in the **population**, such as the incidence rate, following application of the vaccine to the community. Comparisons might be made to the incidence rate in the population prior to attempted immunization or to the incidence in areas or communities where vaccine was not supplied. Brochier et al. (1988) used this method to study the efficacy of oral rabies vaccination of foxes in Belgium. In an experiment referred to earlier, Caley et al. (2001) reduced possum populations and measured the effect by monitoring the prevalence of tuberculosis in ferrets, but the investigators had no control over which ferrets were exposed to the disease. The important differences that distinguish community trials from other experimental methods are that: (i) the investigator does not choose or allocate which individuals will participate in the trial and, (ii) the effect of the manipulation is measured in the population rather than in the individual.

Laboratory experiments have been used extensively in the study of disease in wild animals but neither field trials nor community trials have been used widely. A few specific examples will be discussed later in this chapter because they contain elements of both experimental and observational techniques. Extensive guidelines for the conduct of clinical trials are available in epidemiology texts such as Martin et al. (1987) and Thrusfield (2005).

Intervention trials, involving experimental treatment of one segment of a free-ranging population to remove or reduce the effect of a disease agent are a very promising form of field trial for collecting information on the impact of disease on individuals and on a population. Good examples are studies in which selected groups of free-ranging red grouse (Hudson 1986; Hudson et al. 1992) and snowshoe hares (Murray et al. 1997) were treated with an anthelmintic to control parasites. Reproduction in the treated grouse was shown to be superior to that of untreated groups, and treated hares survived at a higher rate that untreated hares.

6.2.2 Observational methods

Observational studies can be **descriptive** or **analytical**. Descriptive studies, as the name implies, involve the description of disease-related events in a population, as well as the identification of those characteristics that define the particular disease. Descriptive studies usually dominate the early stages of an investigation and provide the preliminary data upon which hypotheses may be formulated. For example, during an outbreak or outbreaks of a disease, the

species, sex and age composition of the affected individuals might be defined, and the pathologic features and presence of potential causative agents could be described. For instance, the first step in defining the nature and cause of avian vacuolar myelinopathy was a detailed description of the pathology in affected birds (Thomas et al. 1998) that allowed identification of birds with this disease. If suitable population parameters are known, various rates (morbidity, mortality, prevalence) may be calculated. Associations between factors may be observed or described but the strength of these associations is not tested in purely descriptive studies.

In reviewing literature available on disease in wild animals, it is apparent that the overwhelming bulk of the information is descriptive in nature, reflecting the comparative youthfulness of the science. Descriptive studies are necessary (Herman 2002) and provide the basis for formulating hypotheses about disease that can then be tested. Thus, the stage is set for more widespread use of analytical methods in the study of many diseases of wild animals.

Analytical studies are based upon **comparison** between or among groups that differ in one or more variables. These investigations attempt to **explain** the relationship between disease-related variables and to **measure** the strength of observed associations. Three sub-types of analytic investigation are recognized, based primarily on the manner in which groups are chosen for comparison.

6.2.2.1 Prevalence surveys

The first of these is the **prevalence survey** or **cross-sectional** study, in which data are collected from a broad sample or cross-section of individuals from the population at large. This sample is then sub-divided into two or more sub-groups, based on the presence or absence of some variable. The most common variable used, in our context, is the evidence of disease. Prevalence surveys are concerned with existing disease, i.e., disease present at the time of the survey. Animals that have the disease are designated as **cases** and individuals within the sample that are free of the disease at the time of sampling are included in the **non-cases** or **control** group. The various categories, such as diseased, must be defined in unequivocal terms prior to data collection, so that individuals can be placed into the proper category.

As an example, consider a situation related to lead poisoning in ducks. Descriptive studies have noted the common occurrence of anemia among lead-poisoned ducks and it is thought that this aspect of the disease (anemia) and the risk factor (lead) are associated. A working hypothesis might be that lead causes anemia and a null hypothesis might be that: 'the number of circulating red blood cells in ducks with and without a toxic concentration of lead in their blood is not different'. This implies that lead, at levels causing other signs of intoxication, has no effect on the number of circulating red blood cells in ducks. One approach would be to examine a sample or **cross section** of

ducks from an area where spontaneous lead-poisoning occurs. For this study, anemia, which is the dependent variable, is defined by the number of red blood cells in circulation, measured by determining the packed cell volume (PCV) of a centrifuged blood sample. The independent variable is exposure to lead, measured by analyzing the concentration of lead in whole blood. We must establish unequivocal criteria for each category in advance of the study. On the basis of published literature we might decide that the diagnostic level for lead poisoning will be a concentration 10 ppm of lead in blood and that any duck with a PCV ≤.320 L/L will be considered to be anemic (diseased).

Among a sample of 200 ducks trapped at a lead poisoning hot-spot, we find 40 birds that meet the criteria for lead poisoning and 38 birds that are anemic. Of the 38 anemic birds, 32 also fit the definition for diagnosis of lead poisoning. One method for analysis of this type of data is through the use of a 2×2 contingency table:

	Cases (anemic)	Non-cases (not anemic)	
Exposed (lead- poisoned)	32 (a)	8 (b)	40 (a + b)
Not exposed (not lead-poisoned	6 (c)	154 (d)	160 (c + d)
	38 (a + c)	162 (b + d)	200

Once the data are arranged in this format, there are several methods by which the strength of association between lead and anemia can be measured. One measure used commonly in cross-sectional studies is calculation of **relative risk** (RR). This is the ratio of the rate of occurrence of disease in those exposed to the risk factor to the rate of occurrence of disease in those not exposed. If there is no association between the factor and the disease, RR should = 1. If RR = >1, the size of the value of RR is directly related to the strength of association between the two variables. If RR = <1, there is a negative association between the factors, i.e., the factor may reduce the occurrence of the disease. In this example:

$$\text{RR} = \frac{\text{prevalence of anemia in ducks with lead poisoning}}{\text{prevalence of anemia in ducks without lead poisoning}}$$

$$= \frac{a/a+b}{c/c+d} = \frac{32/40}{6/160} = 20.$$

The risk of being anemic is 20 times greater among lead poisoned birds than in non-lead poisoned birds, indicating a strong association between lead and anemia.

Another ratio that may be calculated is the **odds ratio** (referred to briefly in Chap. 2). Odds ratio is the probability (or odds) that a case (an anemic bird) has been exposed to lead, divided by the probability that a control (non-anemic) bird has been exposed:

$$\frac{\frac{a/(a+c)}{c/(a+c)}}{\frac{b/(b+d)}{d/(b+d)}} = \frac{a/c}{b/d} = \frac{ad}{bc} = \frac{32 \times 154}{6 \times 8} = 102.7$$

In this sample of birds, the association between exposure to lead and anemia is obviously strong. When the prevalence of the disease within the population is low (<5%), odds ratio is similar to RR. Comparison could also be made using a more conventional method, such as chi-square (X^2), in which case:

$$X^2 = \frac{n([ad\text{-}bd] - n/2)^2}{(a+b)(c+d)(a+c)(b+d)}$$ with one degree of freedom.

The value for significance at the 5% level is 3.84. The calculated value in this example is 116.4 and since this exceeds 3.84, there is less than a 5% probability that a difference as large as observed would occur due to sampling error. We can reject our hypothesis and infer that there is an association between lead poisoning and anemia. Because both variables were measured on a continuous scale, one could also use regression analysis in this instance.

An advantage of a prevalence study is that one is comparing samples drawn from a single population and all diseased and non-diseased individuals in the population should theoretically have an equal chance of being included in the sample. In real life, as has been pointed out elsewhere, this assumption is probably seldom valid. For instance, many of the most severely lead poisoned birds are probably unavailable for capture, while some birds that have been exposed to smaller amounts of lead might be unusually susceptible to the method of capture. A disadvantage of prevalence studies is the large total sample size that may have to be examined. The sample size required is inversely related to the prevalence of the disease, or other factor under consideration, in the population. If the prevalence rate is very low, a large number of individuals must be included in the sample to ensure that the sample contains sufficient diseased individuals for comparison. As in every type of study, selection and collection of an appropriate sample is important; this subject will be discussed more in the following chapter.

6.2.2.2 Case-control method

In many instances, a second type of analytical study, the **case-control** technique, is more efficient than the prevalence survey, because individuals with a special characteristic such as the presence of disease are specifically chosen for inclusion in the study. The basic method in such studies is to identify the association to be measured, e.g., the relationship between lead and anemia,

and then to identify an appropriate number of individuals that have one of the features to be studied. These individuals are the **cases**, and often these are chosen on the basis of presence of the disease. Another group of individuals that do not have this factor are then identified and used as **controls** for comparison.

We can apply this technique to the example of anemia and lead poisoning. Measurement of the concentration of lead in blood is relatively costly, whereas anemia can be detected in the field by centrifuging a small volume of blood to determine PCV using a simple microhematocrit centrifuge. The prevalence survey was wasteful because blood from 200 ducks was analyzed for lead content, of which only 38 of the ducks were anemic. An alternative would be to **screen** a large group of birds, using the inexpensive PCV measurement, to select a sub-sample of birds with anemia (the cases) from within this group for study. An appropriate number of birds without anemia (controls) could also be chosen and lead analysis would then be done only on blood from the ducks in these two groups. In the prevalence study, 200 lead analyses were done, including 38 anemic birds. A case-control study that included 38 anemic birds and an equal number of non-anemic birds would require analysis of only 76 samples, for a substantial financial saving. This relative advantage of case-control studies over prevalence studies becomes progressively greater as the prevalence of the disease in the population declines.

The most difficult part of a case-control study lies in choosing appropriate controls. Ideally, controls should differ from cases only in the single factor under consideration but it is seldom possible to match cases and controls this completely. In choosing controls, three basic decisions must be made: (i) source of controls, (ii) selection of controls from within the source, and (iii) number of controls. The source of controls is obvious in some situations, e.g., if interested in the effect of a water-borne pollutant on animals using river water, one might sample downstream from the source of contamination for cases and above the source for controls. Alternatively, one might sample from two similar watersheds, one of which was contaminated. In other situations the choice is more difficult. Assume that we are interested in the association between renal lesions and antibodies to *Leptospira* spp. in skunks. One source of specimens might be skunks submitted for necropsy to a diagnostic laboratory. These animals would be submitted for many reasons, but primarily to determine the nature of some observed illness. Cases, i.e., animals with renal disease, could be selected from among the animals submitted to the laboratory. The advantage of this source of specimens is that little cost would be incurred in collecting the animals. Several sources might be considered for control animals, including animals without renal disease from among those submitted to the laboratory. However, this sample should be questioned, as the animals have already been selected from the population because of the presence of illness. Hence, they are not likely to be representative of the population. Other sources of controls might be nuisance animals collected by pest-control operators, or from a sample of skunks collected specifically for

the study by trapping. Each of these sources is subject to bias and arguments could be mounted in favor or against the suitability of each. Identification of the biases, and their probable effect on the data, is the most important consideration in choosing the source. In some circumstances, one might choose to use more than one source for controls. If similar results are obtained using control groups chosen from different sources, this is evidence that the observed association is true, whereas if the estimates of risk are different, one should suspect that one or both of the control groups is biased, and the source of bias should be investigated (this should not be taken as condoning trying various control groups until one is found that yields the desired result and then reporting only this result!).

One source of controls that may be appropriate for certain investigations is specimens collected at a time different from that of the cases. Reference collections and various types of specimen banks are particularly valuable in this regard. For example, much of our knowledge of the effects of chlorinated hydrocarbons on the thickness of eggshells is the result of case-control type comparisons between eggs collected from contemporary birds that had been exposed to these agents, and eggs collected in the pre-insecticide era held as museum specimens. Similarly, the concentration of mercury in the feathers of contemporary birds has been compared with that in feathers from museum specimens collected prior to industrialization and to the use of mercurial seed-dressing agents. The latter comparison clearly documented temporal changes associated with this risk factor (Berg et al. 1966).

Selection of individual controls from within the source usually involves sampling and, in most instances, also involves some degree of matching between case and control samples. Careful selection and matching of cases with controls maximizes the information available from a comparison, because it reduces differences between groups in variables other than the one being considered. Some variables, such as sex, age and species, are so obvious that researchers should not need to be reminded of the need for their consideration in matching. In some studies, it may be advantageous to pair individual cases and controls, e.g., a 5-year-old female deer from aspen habitat (the case) would be matched with a control animal of the same species, age and sex collected from a similar habitat (providing that the relationship between these variables and the disease is not under consideration). Overmatching, in which case and control are matched for some determinant that is important in the disease may occur and result in a falsely low estimate of relative risk.

The number of controls required in a study depends on the ease of collection, cost of analysis, and the statistical methods used. In general, at least as many controls as cases should be examined. Many statistical methods benefit from having equal-sized samples and this is a requirement for some techniques. In some circumstances, it may be desirable to analyze more controls than cases to reduce variation within the sample. The same types of statistical analyses used for prevalence studies are applicable to case-control studies and RR and the odds ratio may be calculated.

6.2.2.3 Cohort studies

Prevalence surveys and case-control studies deal with disease existing at the time of the study. The third type of analytical study, **incidence** or **cohort** studies is concerned with **development** of disease in a group of animals. Often the animals studied are free of the disease at the initiation of the study. The term cohort describes a group of individuals who have something in common at the time they are assembled as a group and who are then followed for a period of time to see what happens to them. Cohort studies are potentially useful because they are a more direct method of measuring the risk associated with a disease factor or agent. These studies can be done in two ways. A group of animals can be assembled in the present and followed into the future (a **concurrent cohort study**) or a group can be identified from past records and followed to the present (an **historical cohort study**). In general, cohort studies require the ability to monitor both the occurrence of disease and exposure to one or more risk factors in individual animals over time. Exposure to the risk factor may occur prior to, at the time of, or after the beginning of the study. The occurrence of disease can be monitored in many ways such as through periodic observation or examination of the animals, or through collection and analysis of blood, feces or other specimens at specified intervals.

Bird banding and other mark/recapture techniques that are used to measure mortality or survival represent a form of cohort study. A cohort of hatch-year mallards that is banded in one year and has their subsequent fate monitored through band returns represents a type of concurrent cohort study. A historical cohort study might involve examining band returns to date from all blue-winged teal banded in 1980. In such studies, death, monitored remotely through band returns, is the only measure of disease and the risk factor under study is the summation of all causes of mortality. Through such studies, comparisons can be made among cohorts. For example, the survival rate of birds of the same species banded in the same year in different flyways and, presumably, exposed to different risk factors could be compared. This might be a technique for monitoring the effect of replacement of lead shot by non-toxic steel shot, assuming that the level of usage of steel shot is different among flyways, and can be quantified (however, it would be difficult to separate the effects of lead from those of all other causes of mortality).

The basic requirement for any cohort study is the ability to follow the animals through time. The longer it takes for disease to develop following exposure to the risk factor, the longer the cohort must be followed. Cohort studies have received relatively little use in the study of diseases of wildlife to date because of the difficulty in finding and following individuals. Studies of neonatal mortality of ungulates (Ballard et al. 1981; Nelson and Woolf 1987) and of mortality in a variety of other species (Schultz 1980; Sargeant et al. 1982; Nicholson and Hill 1984; Evelsizer 2002) using radiotelemetry have many characteristics of cohort studies. Burns et al. (2005) used a cohort study design to assess the effects of bot flies on white-footed mice.

Cohort study design can also be used to measure the effect of experimentally applied risk factors. Studies of the effects of ingested lead pellets on duck survival are examples of a form of cohort study (Bellrose 1959; Deuel 1985). In these studies, the risk factor (lead pellets) was artificially applied, so the studies are, in reality, experimental rather than observational; however, these studies are useful for explanation of methodology and for explaining some limitations of this type of study. In both studies a large number of wild ducks was trapped and banded. Lead shot were administered orally to approximately half of the birds before they were released into the wild. The fate of the birds was then monitored through band returns (note that this study has features of a field trial). The cohorts for comparison were the group of ducks that was exposed to the risk factor (lead) and the group that was not exposed. The assumption in both studies was that band returns accurately measured mortality.

A problem in both of these studies was related to having an unequivocal definition of the groups. Bellrose (1959) examined birds with a fluoroscope to detect previously ingested lead pellets prior to the onset of the study. This technique does not identify all birds exposed to lead but was the most acceptable method for measuring lead exposure at the time. No attempt was made to ensure that the birds in the California study were free of lead at the onset of the trial (Deuel 1985). Thus, some birds in both the non-exposed and the exposed group in each study may have been exposed to lead, and the proportion of such birds in each group was unknown. The inherent assumption was that any such exposure was the same in the two groups and that any effect was associated with the administered dose of lead.

Data in Table 6.3 were taken from Bellrose (1959) to demonstrate calculation of RR of mortality occurring in association with exposure to the administered dose of lead. Bellrose used the term "*relative hunting vulnerability*" but it was calculated in the same way as RR.

Table 6.3 Band recovery within the season of banding from wild mallards exposed to different numbers of lead pellets. The number of birds banded in the 0- and 1-pellet groups is a total for 3 years, whereas all of the birds in the 2- and 4-pellet groups were banded during a single year. Data are from Bellrose (1959)

Number of pellets administered	Number banded	Number recovered
0 (non-exposed)	1,456	116
1	1,455	161
2	392	95
4	504	99

If the data in Table 6.3 for ducks receiving either 0 or 1 pellet are arranged in a 2×2 table:

	Recovered	Not recovered	
Exposed (1 pellet)	161 (a)	1,294 (b)	1,455 (a + b)
Not exposed	116 (c)	1,340 (d)	1,456 (c + d)

RR associated with 1 pellet is $\frac{a/a+b}{c/c+d} = \frac{161/1445}{116/1456} = 1.38$

The RR of being killed by a hunter was 1.38 times greater for ducks receiving one pellet than for ducks not exposed to additional lead. The RR for ducks given two and four pellets was 1.89, and 2.12, respectively, compared to ducks not exposed to additional lead. This indicates a dose/effect interaction. Deuel (1985) used a different approach and monitored band returns over the 5 years following experimental exposure to lead. No significant difference was found in the rate of band returns between birds given two lead pellets and those not given any lead. The calculated RR, using these data, was 1.

A study of bovine tuberculosis in European badgers (Cheeseman et al. 1988) illustrates the value of a cohort study for determining the evolution of a disease within a population and in individual animals. The spatial distribution of individual groups of badgers was determined and fecal samples were collected from each group biweekly to monitor occurrence and spread of infection among groups within the population. Individual badgers within groups with fecal samples positive for *Mycobacterium bovis* were captured and examined clinically at 3-month intervals. During the initial 5 years of the study, the spread of infection among groups was slow and restricted, and mortality related to *M. bovis* was low, with some infected badgers surviving ≥ 22 months. There was evidence of both horizontal and vertical transmission within groups and no relationship was apparent between population density and prevalence of infection.

Weigler et al. (1988) followed individual koalas naturally infected with *Chlamydophila psittaci* for 24 weeks and observed the development and/or resolution of clinical disease in the animals. This provided an understanding of the course and significance of this infection that could not have been attained by other methods, such as cross-sectional sampling.

Brown et al. (1990) used a cohort design to study the effect of vaccination for rabies on the survival of adult raccoons in an area where rabies was enzootic. Equal numbers of vaccinated and unvaccinated wild-caught raccoons were fitted with radios and released (note that this study is in reality a field trial). The animals were monitored for several months but no difference in survival was detected between the two groups.

A disadvantage of cohort studies for the study of disease in wild animals is the large number of animals that may be required, because of the difficulty in following subjects. For instance, a minimum of 7,946 female pintails would be required to provide an 80% chance of detecting a 20% difference in recovery between lead-dosed and non-dosed birds, because of the low rate of band

returns (Deuel 1985). It is probably not surprising that no difference in survival between lead-dosed and non-exposed birds was detected in that study. Despite the limitations, cohort studies deserve consideration in situations where animals can be monitored regularly and the development of disease can be measured. The technique is particularly suitable for situations in which animals are predictably available for periodic reassessment. A prime example are colony nesting birds, where many individuals are available, and a cohort can be followed through the nestling period and into subsequent years, because of their nest site fidelity. For instance, Hannsen et al. (2004) measured the effect of vaccination with non-pathogenic antigens on survival of nesting common eider females and Wimsatt et al. (2005) used a cohort study design to measure the effects of anesthesia and blood sampling on the survival of big brown bats. Use of radiotelemetry to relocate animals may extend the use of cohort studies to a wide variety of other disease situations. Evelsizer (2002) used a cohort design to compare survival of radio-marked mallards on lakes where bird carcasses were and were not removed during botulism outbreaks.

Observational studies of disease may be either **retrospective** or **prospective** in nature. The major difference between the two types relates to the timing of data collection. Retrospective studies use data recorded in the past, i.e., before the start of the study, while prospective studies involve the active collection of information for the specific purpose of the study. Retrospective analysis is dependent upon the quality of data collected in the past. A common problem is that because the information was not collected specifically for the study, portions of data may be missing or recorded in a manner inappropriate for the desired review. The lack of detailed records of disease in wild animals has limited the use of retrospective analysis; however, such analyses may be an efficient method for gathering information, particularly on diseases that occur infrequently. For example, about once each year, a pronghorn antelope found dead or dying with severe skin lesions has been submitted to our diagnostic laboratory. These cases have been handled routinely and the bacterium *Arcanobacterium pyogenes* has been isolated from the lesions in almost all instances. Each of these cases was an interesting (but seemingly unrelated) curiosity at the time it was examined. However, when records of disease conditions recognized in pronghorns were reviewed, it was obvious that these cases fit together to form a distinct pattern. This pattern or syndrome was characterized by a distinct sexual prevalence (all cases were in males), seasonality (autumn–early winter), distinctive pathologic lesions (necrotizing purulent dermatitis confined to, or most severe on, the head and neck), and presence of *A. pyogenes*. The collected data allow description of this syndrome and formulation of a hypothesis that the disease is associated with wounds suffered by males during the rut, and that pronghorns may have poor resistance to this common bacterium. This retrospective review of the available records could provide a basis for further analytic study. Davidson et al. (1990) used a similar method to study brain abscesses in white-tailed deer. As data collections become more available in future, retrospective studies will become increasingly useful for the study of disease in wild animals. Use of

historical materials, such as museum specimens of eggs and bird skins, as controls for retrospective studies was mentioned earlier. A remarkable data set based on > 2,000 clutches of eggs collected from British sparrowhawks between 1870 and 1990 was used to demonstrate that (i) eggshell thinning coincided with the introduction of DDT, and (ii) shell thickness increased as use of the pesticide was restricted and then banned (Newton 1998).

Cohort studies may be historical in nature. These represent a form of retrospective analysis in which individuals with a particular disease are traced back in time to examine their exposure to various risk factors in the past. This type of study has proven particularly valuable for the study of rare diseases in humans but requires an accurate historical record on the individual, something that is seldom available for wild animals. However, this type of study can be used in wildlife for investigating diseases that leave recognizable traces in the animal. For example, antibody in serum is evidence of past exposure to a disease agent and lead accumulated in bone or mercury in plumage are evidence of exposure to these heavy metals. Similarly, analysis of elements in tissue, such as copper in hair, may reveal the availability of this nutrient to the animal during the period that the hair was growing. Thus, if one was interested in a neurologic disease in birds, it might be possible to select individual birds affected with the disease as a cohort and measure their past exposure to certain viruses by looking for antibody in their serum, and to lead and mercury by analysis of bone and feathers. Findings in these birds could be compared to those from a group of similar birds that did not have the disease. This example blends the characteristics of cohort and case-control studies, illustrating the overlap that may occur among methods.

In prospective studies, the process of information collection can be planned carefully to fulfill specific objectives of the study and, in most instances, the period of data collection is relatively short. This often requires an intense effort but should result in the collection of information of uniform quality. Some diseases occur so infrequently that it is impossible to amass sufficient data over a short period of time and as the period of data collection lengthens problems of non-uniformity of data become more severe. Information collected over a period of years may suffer from many of the same shortfalls described for retrospective studies. This is a problem particularly for investigators working in diagnostic laboratories or disease investigation units. These individuals have a unique opportunity to see and handle diseased animals but their primary responsibility is to investigate each new problem as it arises, rather than to do in-depth research on any one problem. Information collected from the routine activities of such laboratories and individuals may be valuable for retrospective analysis but often suffers from the deficiencies mentioned earlier. There is the risk in any extended study that short-term trends related to a disease may become obscured by long-term trends in population density or abundance unrelated to the disease under study.

One method of combining the benefits of planned data collection with the intermittent availability of specimens and information is an **opportunistic**

prospective study. As an example, our diagnostic laboratory receives a small number of beaver each year for necropsy. Among these animals there have been several with severe degenerative joint disease. A retrospective review of records on these cases indicated that in most instances, the joint lesions were considered to be the major disease process, although the ultimate cause of death was often starvation or misadventure. The animals were usually described as 'aged' in the records, but their actual age had not been determined and, in a few instances, the sex had not been recorded. Based on the observations, one might suspect that debilitating degenerative joint disease is an age-related phenomenon of unknown prevalence in beaver. To investigate this phenomenon further would require additional beaver for examination. One way to proceed might be to collect a large sample of beaver, perhaps from trappers, and do a cross-sectional survey to determine the frequency of occurrence of the disease in various age groups. However, the prevalence of the condition in the population is probably quite low, so that a very large sample would be required, (minimum sample size will be discussed in Chap. 7), and this would require a major research effort.

An alternative would be to do a prospective study using all beaver submitted to the laboratory in the future as a sample, which would be available at little cost, and to collect uniform information related to joint disease from each beaver (the obvious disadvantage, as mentioned earlier, is that such animals may not be representative of the population). For this type of study, we have found that a specific protocol form (usually 1–2 pages) should be designed. The protocol contains a brief statement of the rationale and objectives of the study, a detailed definition of the disease under consideration, together with specific instructions on the information and specimens to be collected. The latter information is arranged in checklist format so that omissions are obvious. Thus, in a study of the association between age and the occurrence of joint disease in beaver, we might provide space on the form for recording weight, sex, and certain body measurements of each animal. The protocol would also specify that a specific tooth be removed and sectioned for aging by cementum annuli examination; that specific bones and joints be examined with lesions being described in a specific manner and photographed; and that certain specimens, perhaps synovial membranes, would be collected for histology.

An advantage of this system of data collection is that different individuals, who may be working in the laboratory, can follow the protocol and collect data in a uniform manner. We have found that several small research projects of this type can be done simultaneously without disrupting the normal diagnostic function of the laboratory unduly. Thus, presently in the Canadian Cooperative Wildlife Health Centre laboratories we have separate protocols for collecting tissues from raptors for lead and anticoagulant analysis, for collecting tissues from some piscivorous birds for mercury analysis, for examining the spinal column of raptors for fractures, as well as examining all wild ungulates for chronic wasting disease. Each study is activated only when an appropriate specimen became available.

The sequence in the investigation of a particular disease is usually, first, recognition of its occurrence, followed by descriptive studies that define the disease and provide the information needed for formulation of hypotheses. Once a hypothesis has been developed, the investigator can then choose among the experimental and observational techniques available for testing it. In general, experimental studies are more rigorous and may be subject to less bias than are observational studies. Experimental studies can be replicated, if necessary, whereas it is impossible to replicate observational studies exactly. However, the results of observational studies may be more directly applicable to a field situation, since they measure naturally occurring, rather than contrived, disease occurrences. Observational studies also may be the only method feasible for situations where the conditions that prevail during a disease occurrence cannot be reproduced experimentally or where experimental studies are impossible, such as in some parks or when dealing with endangered species. The basic techniques described in this chapter can be modified to fit almost any situation. Even elaborate techniques, such as discriminant or multivariate analysis, in which a myriad of environmental factors are measured in relation to disease occurrence, are extensions of simple observational methods.

A potentially rewarding method, which has received relatively little attention, is the combination of experimental and cohort techniques. More than 70 years ago, Aldo Leopold (1939) recognized that observational and correlational studies have limitations for understanding disease. More recently others have expressed the need for experimental perturbation or manipulation to extend our knowledge of disease processes in wild populations (Tompkins et al. 2001). The manipulation might consist of either adding or removing a disease agent and then studying the effect on the population. The study by Bellrose (1959) of mortality associated with lead ingestion by ducks was an early example of adding a disease agent. Other examples also deal with the effects of toxicants on birds. Gilman et al. (1978) extracted organochlorine contaminants from gull eggs in the contaminated environment of Lake Ontario and injected this material into uncontaminated eggs in a colony in New Brunswick in an attempt to separate the direct effects of the toxicants from other factors that may have been operating in the contaminated environment. The cohort for study consisted of eggs in the New Brunswick colony, some of which were exposed to the risk factor (the contaminants) and some of which were not. The eggs were incubated and hatched by the natural parents and embryo and chick mortality were monitored and compared. McEwen and Brown (1966) used this method to determine the effect of two pesticides on sharp-tailed grouse. Wild adult male grouse were trapped, fitted with radio-transmitters, given a single oral dose of one of the pesticides or lactose (control birds) and then released. Survival and behavior of the birds was monitored by radiotelemetry and direct observation on the breeding grounds. The lethal dose of pesticide for these birds was found to be similar to that determined in prior experiments using penned birds. However, changes in social hierarchy, breeding behavior, and vulnerability to predators

detected in the free-ranging birds exposed to sublethal doses of pesticide had not been detected in earlier trials with penned birds.

This method may also be appropriate for certain infectious diseases. Samson et al. (1987) exposed some lambs within a free-ranging bighorn sheep herd to a known number of larvae of *Protostrongylus* spp. lungworms and then monitored the health of the artificially exposed (and of unexposed) members of the lamb cohort by measuring larvae in the feces, clinical signs and survival over the subsequent winter. The advantage of this type of study is that exposure to the risk factor is controlled, as in an experiment, while other variables that may be important in the natural disease are allowed to occur in a manner not possible in the laboratory. Conversely, it may be possible to remove or reduce the effect of a disease agent on a cohort within the population. This has been done by using anthelmintics to study the effects of cecal nematodes on red grouse (Hudson et al. 1992), abomasal worms on Soay sheep (Gulland 1992), intestinal nematodes on snowshoe hares (Ives and Murray 1997), gastrointestinal parasites in yellow-necked mice (Ferrari et al. 2004), and fleas on Richardson's ground squirrels (Jardine et al. 2006) and by experimental supplementation with a nutrient (selenium) in mule deer (Flueck 1994). In each of these situations information was discovered that could not have been identified by observation alone.

6.3 Use of indicator or sentinel species

In some situations it may be advantageous to use a species other than the one of direct concern to collect information about disease. One reason for doing this may be in circumstances in which it is impossible to adequately sample the main species, because it is rare or endangered. Northern bobwhites were used as a surrogate to investigate the presence and prevalence of disease agents on range occupied by the endangered Attwater's prairie chicken (Purvis et al. 1998) and black-footed ferret X Siberian polecat hybrids and domestic ferrets were used as a surrogate in developing disease control measures for endangered black-footed ferrets (Williams et al. 1995). Another reason for using a surrogate is that it may be much easier to work with the surrogate than the species of concern, e.g., domestic chickens have been used as sentinel birds for western equine encephalomyelitis virus for many years in Saskatchewan because it is much easier to put out small flocks of chickens around the province that can be bled for serology periodically, than it is to capture an equivalent number of wild birds. A third reason for using a sentinel species occurs in situations in which a scavenging or carnivorous species screens a large number of the species of concern (which is at a lower trophic level). Measuring evidence of disease in the carnivore/scavenger provides an index to the relative frequency of occurrence of disease in the primary species of concern. Wild carnivores (Gage and Montenieri 1994) and domestic dogs and cats (Leighton et al. 2001) have been used to monitor disease, including

plague, in small rodents. A relatively small sample of carnivores yields information comparable to that obtained by trapping a large number of rodents. A further advantage is that carnivores are longer lived and, hence, available for sampling over a more extended time period than the rodents. Nugent et al. (2002) proposed that feral pigs marked with a radio-transmitter prior to release are an efficient and sensitive sentinel for detecting the presence of bovine tuberculosis in brushtail possums in areas of New Zealand. Of 17 pigs released in an area with a low density of possums, 15 were recovered > 2 months later and all had become infected with *M. bovis*.

6.4 Summary

- A hypothesis is a proposition (set forth as an explanation for the occurrence of a phenomenon) that can be tested.
- The basis for scientific investigation is the collection of information to formulate and test hypotheses.
- Experimental methods measure the effect of manipulations caused by the investigator; observational methods collect information about naturally occurring events.
- There are three sub-types of experimental techniques that differ in the way subjects are chosen for inclusion in the study, in the amount of control that the investigator has over variables, and in the method used to assess changes in other variables.
- Descriptive observational studies dominate the early phase of most investigations and involve the description of disease-related events in the population. Associations among factors may be observed but the strength of the associations is not measured.
- Analytical observational techniques are of three basic types: prevalence surveys, case:control studies, and incidence or cohort studies; all attempt to explain the nature of relationships among various factors and to measure the strength of associations.
- Prevalence surveys and case:control studies deal with disease existing at the time of the study; incidence studies are concerned with the development of disease over time.
- Observational studies may be retrospective, using existing data, or prospective with collection of new information.
- The investigation of a disease may require application of several different techniques singly or in combination. The methods that have been used to study wildlife diseases often have elements of several types of technique.
- Experimental manipulation or perturbation, for example by adding or removing a disease agent from some animals in a population, may be very useful for detecting and measuring population-level effects of disease.
- In some situations it may be advantageous to collect data about disease using another species as a surrogate, indicator, or sentinel for disease in the species of primary concern.