Designing a Nutrient Study

Lisa M. Ganio
Forest Science Department, Oregon State University

Abstract

Good study design is essential for obtaining useful information. Well-designed studies include the appropriate kind of replication and account for the different types of conclusions that are obtained from observational studies and designed experiments. A study that compares any set of groups can be partitioned into 5 steps.

1. Define the question of interest.
2. Plan or design the study.
3. Carry out the plan and collect relevant information.
4. Compile the data into a form to answer the question of interest.
5. Draw conclusions based on the data that were collected.

Considerations for each step and the interactions among the steps are discussed. An example is presented to illustrate the concepts.

Introduction

The information and conclusions that are gleaned from a study are due in part to the design of the study. Good design contributes to the efficiency and validity of the study and therefore to its usefulness. Sometimes the design is deliberate and carefully planned. In other cases, the 'design' is more the result of happenstance. Cox (1958) discussed both cases providing investigators with the tools to understand their own designs. Recent publications have put the concepts into more current settings and applied them to ecological studies (Manly 1992, Skalski and Robson 1992). Hurlbert (1983) discussed the relationship between the temporal and spatial replication in a study and the conclusions that were drawn. His review of ecological literature indicated that appropriate replication was often missing. Ecological study designs have improved over time but there is evidence that the problem still exists (Heffner et al. 1996).

All of the aforementioned authors discuss the dichotomy between *observational studies* and *designed experiments.* In observational studies, data are collected by observing individuals in groups of interest and the data are used to either compare or describe the groups. But the investigators have no control over which subjects belong to which groups, nor over the definitions of

the groups. In a designed experiment, the investigators can manipulate most aspects of the study. The most important manipulation is the control of the assignment of individuals to the groups. Investigators might also control various aspects of the groups, for example, the average age or the amount of fertilizer.

Why is this distinction important? It is true that the analyses of the data from either observational studies or designed experiments can be very similar. But the conclusions that can be drawn differ between the two types of studies. When investigators can randomly assign the subjects to treatments, the randomization process insures that, other than the treatment, there are no systematic differences between the individuals in one group compared to the other. Thus they can conclude that it is the treatment that is ***causing*** the observed result in the experiment. In an observational study, effects outside of the scope of the study can influence the outcome so cause-and-effect relationships cannot be determined. Investigators cannot conclude that the treatment caused the effect; only that the treatment is ***associated with*** the observed effect.

For example, an observational study found greater growth in heavily fertilized seedlings compared to poor growth in poorly fertilized seedlings. It appears as though the growth may be due to fertilizer. .A deeper look into the study revealed that the heavily fertilized seedlings were all observed at one nursery and the poorly fertilized seedlings all were observed at another nursery. Growth differences could be due to factors other than fertilizer including different seedlots, different watering regimes, different growth conditions etc.

In addition to confounding factors, observational studies can result in biased estimates of effects if the subjects are not chosen randomly from the set of all subjects. A ***Literary Digest*** poll in 1936 was used to predict the outcome of the U.S. presidential race (Manly 1992). A survey was mailed to eligible voters listed in the telephone directories and about 2.3 million people responded. Landon was the clear favorite in the survey. But Roosevelt swept the election. Why was the survey wrong? One likely explanation is that presidential choice was associated with economic status and low income subjects without telephones were not sent surveys (Manly 1992). Notice that having a small sample size is not the issue. It is important to have a representative sample of all voting people in the U.S.

Another important dichotomy is whether or not a study contains replication (Skalski 1992). Hurlbert (1983) introduced the term ***pseudoreplication*** to describe types of replication that don't allow studies to produce the conclusion that is desired. If the study's goal is to compare groups, then having replication seems not only logical but essential to concluding that there are consistent differences. But as Hurlbert noted, it's important to replicate the appropriate unit.

But often there are circumstances in which replication is not possible yet investigation is warranted. For example, when one large-scale forest fire occurs, we might want to know its impact on surrounding forests. Or, we might want to study the effects of leaking radiation from a single nuclear power plant. Studies without replication have been termed assessment studies (Skalski 1992). These studies lack replication but provide vital information. Without replication, effects can only be reported as having occurred in this one instance and inference beyond the single case is not possible.

Components of a Comparative Study

In this discussion I would like to focus on studies that compare groups of individuals. Sometimes, in nursery settings, seedlings grow extraordinarily well and we want to identify what produced good growth. Or poor growth prompts us to identify what to avoid in the future. We are ultimately led to compare materials or methods and make decisions for future use. These comparisons take resources away from other aspects of the process, most notably personnel, time and supplies. The best use of resources provides results that have as broad an application as possible for the least investment. We like to optimize the information we obtain while keeping the costs of the comparisons reasonable.

A study that compares any set of groups can be partitioned into 5 steps.

1. Define the question of interest.
2. Plan or design the study.
3. Carry out the plan and collect relevant information.
4. Compile the data into a form to answer the question of interest.
5. Draw conclusions based on the data that were collected.

To conduct the study, it appears that these five steps should follow in sequence. However, in the planning stages, these steps are interconnected, not sequential. Changes at one step can affect every other step. I will discuss the planning (or design) phase of a study and show that each step must be planned and examined in light of all the other steps. Each step is thought of as connected to every other step so that when a change is made at any level, the other steps will be adjusted accordingly.

1. Defining the Question of Interest

Every study has broad goals and objectives that guide the proposed work. These objectives are usually stated explicitly in the planning phase since they provide perspective for the study. The general objectives are refined

into more explicit quantitative statements that define the specific questions to answer or hypotheses to test. From the general objective of "Examine effects of boron in fertilizer", the specific question of interest,

> *"Does the foliar boron concentration of Douglas-fir seedlings differ among the nursery grown seedlings that receive one of 4 different fertilizer regimes, the standard fertilizer with 0 lb/ac of boron, 1 lb/ac of boron, 2 lb/ac of boron, and 4 lb/ac of boron?"*

is generated.

A specific question of interest should explicitly define the response and the groups to compare explicitly. In the statement above, the subjects, Douglas-fir seedlings and the levels to compare, differing fertilizer regimes, are clearly outlined. In addition, foliar boron concentration is defined as the measurement that will be taken and compared among the levels. The specific question of interest will guide the study design. Or, practical considerations may cause the question of interest to be modified. Using the general objective as the question of interest is not helpful in the planning stage since it is too general and it can be misinterpreted. It cannot provide the necessary details about subjects or factors to compare.

2. Designing the Study

After drafting the question of interest, the investigators provide the study design, i.e., the specific plan of action for conducting the study and list of required materials. At this point, modifications, such as the addition of other research questions may occur. But this is expected. The question guides the design by defining the factor(s) and the levels of the factor to be compared. In the example, the factor is "boron enhanced fertilizer" and the levels are 0, 1, 2, and 4 lb/ac. In statistical textbooks, the levels of the factor, or the combinations of the levels of the factors that are used are referred to as the 'treatments'.

Sometimes the levels of the factors are less like treatments whose application is under the investigators' control and more like conditions under which a response is observed. There can be some blurring of the distinction between observational study and designed experiment. Sometimes the groups to compare are environmental conditions such as 'valley soils' and 'upland soils', where subjects were not randomly assigned but data are collected from observed individuals. Investigators must recognize that the inability to assign the individual units to the levels of interest has defined this part of the work as an observational study.

When planning the physical layout of the study there are important concepts to consider in light of available resources. These are *representation, blocking and lack of confounding, replication, randomization, and the anticipated level of precision.*

Representation

Within a study, all conditions of interest to the investigators should be represented. For example, to determine the difference between fertilizer with no boron and fertilizer with 2 lb/ac boron both conditions should be a part of the study. If these conditions were present, then the results would not apply to boron applications of 4 lb/ac. Or, if the results are to apply to both the drier eastern Washington climate and the coastal Washington climate then both climatic conditions need to be represented in the study. A lack of adequate representation in a study will not affect the analysis of the data but it restricts conclusions that can be drawn to the conditions represented in the study.

Blocking and Lack of Confounding

When a factor is identified that can alter the outcome of the study, but there is no interest in comparing the levels of this secondary factor, it is best to control for its effects. It might seem reasonable to choose one level of the secondary factor and carry out the whole study using this level. But usually, representation over a wider environment is desired. A better method selects a variety of levels of the secondary factor. Then all groups of interest are compared at each of the levels of the secondary factor. Thus broader representation is attained while accounting for the extra variation due to the secondary factor. This method is referred to as blocking in statistical textbooks.

Blocking will be effective in removing extraneous variation from the comparisons between the levels of interest when units within blocks are more similar than the blocks themselves. A blocked study is efficient if the blocks are very different. When blocks don't differ more than the units within the blocks, a blocked study will not be able to detect effects as powerfully as it could.

Correct blocking also prohibits the secondary factor from being confounded with the factor(s) of interest. For example, if seedlings from one seedlot are assigned to a high boron fertilizer and seedlings from a different seedlot are assigned to a low boron rate, then observed differences in foliar boron content between the rates might be due to genetic differences in the seedlots and not due to differences in fertilizer. We say that the application rate is confounded with the seedlot. Any difference we observe between the groups cannot be attributed solely to application rate nor to seedlot. Confounding of any kind is an undesirable condition because there is no way to tell which of the confounded factors (seedlot, boron application rate) are resulting in the observed effect. Blocking avoids confounding because all levels of the factor(s) of interest occur within each level of the secondary factor. So in the previous example, all boron levels are assigned to seedlings from one seedlot and that is repeated for each seedlot. The blocks are the seedlots.

Randomization

Blocking is used to account for *identifiable* components of a study that can affect the outcome. Similarly, there are likely to be non-identifiable components, not of interest to the investigators, affecting the outcome. But blocking cannot be applied to non-identifiable components. Rather than trying to adjust for unknown effects, investigators use random allocation to remove systematic effects of unidentifiable sources. By randomly assigning the experimental units to the levels of interest, systematic differences between the experimental units are evenly distributed across all levels of interest.

In a random assignment every individual replication has a *known* chance of being assigned to any group. In a simple random assignment, every individual replication has the *same* chance of being assigned to any group. For example, if 1000 seeds are to be assigned to treatments, the chance that any single seed is assigned to a treatment is 1/1000 = 0.001. One common method used to produce a random assignment is to number the levels of the treatment and then use a random number table to assign individuals to the treatment levels in the order supplied by the random number table.

For example, suppose that 50 seedlings (25 from each of two seedlots) were to be used to compare 0 lb/ac and 4 lb/ac of boron and that one seedlot was unknowingly infected with a disease. If the infected batch of seedlings was assigned to the high boron application level then it might be concluded that the high level of boron resulted in poor performance, when in fact the unknown disease was the cause. But if a random number table was used to generate a string of O's and 4's and then individual seedlings were assigned to the 0 and 4 lb/ac levels in that random order, the infected seedlings would occur in both application rates and thus reduce the performance in both applications to some extent, but not disproportionately in one rate over the other.

It is important to note that "randomization", or the random allocation of units to groups, is not the same thing as 'haphazard' or a subjective assignment. In the absence of random allocation, even if we don't perceive that we've employed systematic bias in our assignment process, subjective bias is inevitable (Cox 1958). In addition, it is difficult to describe a haphazard assignment in which you 'just picked a unit' and you cannot insure the lack of bias. In the long run, a random assignment process has no bias. And since it is quantifiable in the sense that each individual has a known chance of being assigned to any level, the error associated with each individual can be computed and an estimate of the overall precision of the study can be calculated. Although the assignment of individuals to units will differ if the random allocation method is repeated, the method itself can be repeated exactly.

Replication

Before we accept the existence of an effect (e.g., greater foliar boron concentration) the effect must be observable in replicates that represent the range of variation over which inference is to be made (Hurlbert 1983). Thus the experiment must include replication of units of a particular type and, in

the presence of randomization, replicates provide an estimate of the variation associated with an effect.

The technical question of what 'unit' should be replicated in order to have true replication that allows useful inference should be considered carefully. Hurlbert (1983) pointed out that replication of a particular type, e.g., in time, may not suffice for the desired inference. Cox (1958) developed a method for discerning the definition of the unit requiring replication. First, decide which treatments or levels to be compared. Then find the unit that corresponds to 'the smallest piece of material that receives an *independent* application of the treatment'. This unit is called the experimental unit. This is the unit that requires replication in order to conclude whether or not the treatments differ.

In the boron fertilizer example, if the fertilizer is applied individually to each seedling, then each seedling receives an independent application of the treatment so the seedling is the experimental unit. Replication means having multiple seedlings per application rate. Alternatively, if the fertilizer is aerially applied to many seedlings in one acre plots, then the experimental unit is not the seedling but the group of seedlings on the one-acre-plots; true replication involves having more than one one-acre-plot that receives the application rate. Defining how the treatment is physically applied is crucial to the definition of replication.

In comparative studies an estimate of the variation associated with the differences between groups is required. *For one group to be considered different from another group, the differences between the two groups should be larger than the differences between the experimental units within the groups. Otherwise the differences between the groups would be indistinguishable.* The determination of a 'statistically significant difference' is made by comparing the size of an estimated difference to the size of the associated variation. The difference is considered statistically significant if it is larger than the associated variation.

The size of the variation associated with a difference is somewhat under the investigators' control. They can increase the precision of an estimated difference by increasing the number of true replications associated with each level. Many replications are required when the response is highly variable in order to detect differences and, when the response is consistent with a small amount of natural variation, fewer replications are required. However, in a practical sense, the resources available for the study are what determine the number of replications to be used.

An Example

In order to illustrate representation, blocking, randomization and replication I will show how they are used in the design of the hypothetical comparison of boron treatments study. To review, the question of interest is:

> *"Does the foliar boron concentration in Douglas-fir seedlings differ between the nursery grown seedlings that receive standard fertilizers with 0 lb/ac of boron, 1 lb/ac of boron, 2 lb/ac of boron, and 4 lb/ac of boron?"*

Representation

The investigators want the results of their study to apply to any Douglas-fir seedling grown in a specific Olympia, Washington nursery. They have no interest in whether or not the effects will be different at other nurseries. Within the particular nursery however, they know that some beds are better environments for Douglas-fir seedlings than other beds. They want their study to address effects across all potential beds within the nursery. Hence their study needs to be carried out only in the Olympia nursery but it needs to encompass the range of variability in all the beds at the nursery.

Replication – Part

In this proposed study, suppose that the fertilizer treatments must be applied to sections of nursery beds. The number of seedlings per section has not been selected yet. The investigators know that even if the seedlings appear uniform, there can be variation in their response to uniform conditions. This is one important reason to have more than one seedling in each section of a bed that receives one fertilizer level. Thus, the smallest piece of material that will receive an independent application of the treatment will be a group of seedlings in a section of the bed. This defines the group of seedlings as the experimental unit and therefore there must be replication of groups of seedlings that receive a single level of fertilizer. Before they decide on the number of replications for each level of fertilizer, the investigators need to consider whether blocking is helpful.

Blocking

When considering the idea of representation, the investigators noted that seedlings in some beds grew better than seedlings in other beds. There are two ways of selecting beds. Investigators could select beds known to be good environments for the seedlings or they could select beds encompassing the range in variation of growth environment. To determine the best option, review the question of interest. Since the investigators' questions apply to all beds in the nursery, the first option is not a good one. Here's the first connection among the study's components that we've seen. If they use the first option to select beds, then the question of interest will be implicitly changed. The focus will change from all nursery grown seedlings to all seedlings grown in the good environment beds in the nursery. So the question of interest dictates part of the design structure, i.e., the beds that will be used.

Since the study needs to represent all beds in the nursery, the investigators will randomly select a group of beds to use for the study from the set of all the beds in the nursery. Each bed would be considered a block. Since beds are typically very long, it is also conceivable that beds could be divided into large sections. In one scenario, one section of a bed could be used as a block and only one section per bed is chosen. While not entirely practical, this method does provide good representation across the set of all beds in the

nursery. And if growing conditions vary among beds, this method provides an efficient blocking factor.

But in a more practical scenario, the investigators are likely to choose more than one large section of each bed to use. If blocking for different growing conditions in beds is to be effective then there must be representation from more than one bed. By the definition of a block, the sections need to differ among themselves more than the small plots within the sections of the bed differ. Then, within each block (either a section of a bed or an entire bed), each level of the fertilizer (0, 1, 2 and 4 lb/ac) will be applied to a section of that block. Notice that since every level of the fertilizer occurs in every block, the number of replications used will be equal to the number of blocks. The investigators may also want to use part of the bed as a buffer around the edges of the bed or between the application rates. Now all that remains is how to decide how many blocks, how many seedlings per block and which part of each block gets which level of fertilizer. Note that blocking will be helpful in accounting for the variation due to different growing conditions among beds only if the sections of the beds that are used come from different beds. If some blocks come from the same bed then the efficiency of blocking will be reduced to some extent. The fewer beds that are used the greater the reduction in efficiency. There is likely to be a trade-off in any study between this efficiency and the choice of the blocks.

Replication – Part

Practical considerations limit the number of replications (blocks). If the nursery can free up only 5 sections of beds (blocks) for use in the study and enough seedlings to populate the 5 sections, no further replication is possible. The number of blocks might also be limited by the number of seedlings available for use, or the amount of time required to collect the information from the seedlings. In each of these cases, the implementation step of the study is affecting the design step.

For example, if 5000 seedlings are available for use, and 1000 seedlings are required to fill each 12' long block to the required density then there can't be any more than 5 blocks used. Or, suppose that 6 blocks will be used and blocks are 21' long and 4' wide; and seedling density is 125 seedlings per bed foot. Then 21*125*6 = 15750 seedlings are required for the entire study (2625 per block). The time and resources required to record data from each application rate may be limited so it is not required that all seedlings in each application rate be measured. Suppose that 1' buffers are required between each set of treated seedlings. Then 4 bed feet of seedlings received one fertilizer application (Figure 1). Each 4' x 4' teatment section contains approximately 500 seedlings, but due to edge effects and time constraints, only 50 seedlings from the interior of the bed were used in the data collection step.

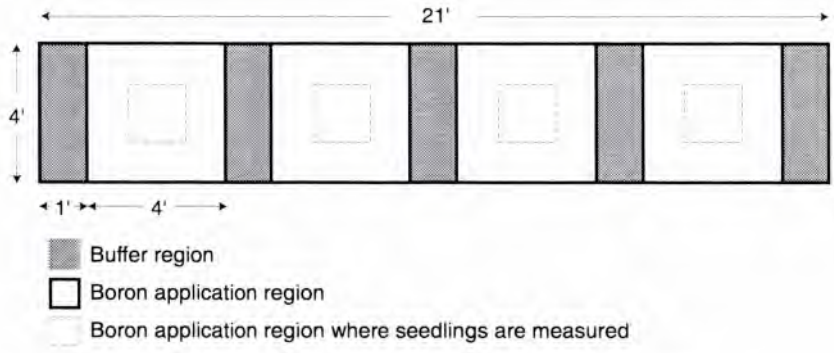

Figure 1. Example of experimental design to compare fertilizer applications.

Randomization

How are levels of fertilizer assigned to the sections of the bed?

Since each section corresponds to an experimental unit, levels of fertilizer should be randomly assigned to sections of the bed within each block. Since no effects within blocks have been identified, the random assignment of the fertilizer levels to sections of the block is justified and it will keep any systematic effects across blocks from being assigned to the same fertilizer level. Note that if investigators had suspected that within block effects existed they would have used a different kind of design to account for those effects.

At this point in the planning process, the investigators know that the study will compare 4 fertilizer application rates using about 15750 seedlings in 6 sections of the nursery beds. The bed sections will be randomly selected from the set of all beds in the nursery. The bed sections will be divided into 4 sections of equal size and planted at a density of 125 seedlings per bed foot. Within each block or bed section, the 4 sections will be randomly assigned to one of the 4 levels of fertilizer application rate and the fertilizer will be applied to the 500 seedlings within the buffer strips. The investigators will collect foliar samples from 50 seedlings from each group of 500 seedlings. Since it is the variation between the replicate groups of 500 seedlings that is required to assess differences between the foliar boron levels, it is not necessary to generate a foliar boron concentration for each of the 50 foliar samples per replicate. Each group of 500 seedlings will only supply one average value of foliar boron for the comparison of the application rates. To save costs, foliar samples from the 50 trees would be composited and the composite sample sent to the laboratory for analysis.

Notice the interconnections between the five components of the study.

The question of interest affects the study design with respect to the use of the beds, the restriction of the study to the one nursery and the definition of the data to be collected from the seedlings. The potential implementation of the study affects the study design since the number of seedlings to measure was restricted due to constraints in resources. In addition, beds were used as the blocking factor since bed-to-bed effects were known to occur. If other

types of effects were identified, the blocking factor could have been different. It is also possible that the question of interest could have been revised due to limitation in resources. For example, if the original intention was for the study to apply to any nursery in western Washington but resources weren't available for such a large study, the question would have to be reworded to reflect the fact that the study would only apply to the single nursery.

3. Carrying out the Study

The implementation of every study will be different. Situations that occur during implementation can result in changes to the other components of the study. Sometimes, the data that are collected do not conform to the assumptions of the proposed data analysis so a different analysis must be used. Unforeseen field effects can eliminate one of the groups in the comparisons of interest. This changes to the question of interest that can be answered and the conclusions that will be drawn. Or an anticipated outcome never happens so that the planned question of interest is not answered. For example, in a study to compare growth of bigleaf maple trees, elk browsing eliminated the ability to compare heights. The question of interest was modified to use survival and not height as the response variable (Alison Luckett, personal communication). The better the investigators can anticipate what will happen, the less likely that catastrophic changes will have large impacts on the information that is obtained from a study. But they should anticipate unexpected occurrences and be ready with alternative plans.

4. Compiling the Data

In the planning phase, the investigators must consider how to compile the data collected on the 50 individual trees from each replicate into a form that will address whether or not there are differences in the levels of foliar boron. First, the individual seedling measurements from the 50 seedlings that received the same fertilizer rate in each bed are compiled to obtain a single value representing the experimental unit. In this example, it seems reasonable to composite the leaf material from 50 seedlings to represent the foliar concentration for each group of 500 seedlings in the treated plot. In other studies, data could be collected on each individual tree and statistics such as the average, the median value, or the maximum value could be used instead of the average. The investigators decide which statistic to use. They require a statistic that will represent the group well in the context of the particular problem. After this initial summary step is taken, the 1200 values (50 trees * 6 beds * 4 fertilizers) will be reduced to 24 values (6 beds * 4 levels) for the final comparison.

Why reduce the number of data points from 1200 to 24? Why not use 1200 values? If the 1200 values were used, it would appear as if there are 50

trees * 6 beds = 300 replicate values for each treatment. But it has already been discussed that there are really only 6 replications for each treatment, not 300. Using the 1200 values for the study, instead of 24, misrepresents the amount of replication that was used and misrepresents the strength of the study. Using 300 values as replications, rather than 6 replications for each treatment, is to use pseudoreplication, rather than true replication (Hurlbert, 1983).

The next step is to compare the sets of 6 replicate values from each treatment to determine if differences exist between the treatment groups. One approach is to calculate the average for each treatment group and an estimate of the variation for each average and compare the averages. But the numerous tools for determining differences among groups are beyond the scope of this discussion and readers are referred to statistical textbooks for details (Ramsey and Schafer 1996, Peterson 1985, Steel and Torrie 1997).

Sometimes, the data that are collected aren't what was anticipated and proposed analyses are no longer appropriate. Even in the most conscientiously planned studies, unanticipated results can require investigators to reconsider planned analyses and make appropriate changes.

5. Drawing Conclusions

Throughout the planning stage, as the proposed design and analysis crystallizes, the investigators need to ask whether what is being proposed will result in appropriate results that allow them to draw usable conclusions. For example, if the original study was to make inference to all Douglas-fir seedlings grown in western Washington, but decisions were made to limit inference to one nursery, then the conclusions would not apply to all of western Washington. Or if conclusions were to show how foliar boron concentrations changed as the boron levels in the fertilizer changed, then data analyses must demonstrate whether or not the trend exists. The study question provides the springboard from which conclusions are generated. If the investigators want to make conclusions about trends, the question of interest should reflect that. Some investigators may suggest that thinking about conclusions before collecting data is premature. Thinking about specific conclusions might be. But determining whether the study plan will result in applicable information is not!

Planning a study is an iterative process. Having worked through the 5 steps that comprise a study, the investigators need to revisit each step again to be certain that all issues have been addressed and that resources are available to do what is planned. At the end of the planning process investigators will have a clear picture of how to proceed. In addition, if unforeseen events cause the study to deviate from the plan, the investigators will be able to understand the impact of the changes on all aspects of the study.

Literature Cited

Cox, D. R. 1958. Planning of Experiments. John Wiley and Sons, Inc., New York, New York.

Heffner, R.A., M.J. Butler W and C.K. Reilly. 1996. Pseudoreplication revisited, Ecology 77(8) : 2558-2562.

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

Manly, B.F. 1992. The design and analysis of research studies. Cambridge University Press, New York, New York.

Peterson, R. G. 1985. Design and Analysis of Experiments. Marcel Dekker, Inc., New York.

Ramsey, F. L. and D.W. Schafer. 1997. The statistical sleuth, a course in methods of data analysis. Duxbury Press, Wadsworth Publishing Co, Belmont, CA.

Skalski, J.R. and D. S. Robson. 1992. Techniques for Wildlife Investigations, Design and Analysis of Capture Data. Academic Press Inc, San Diego, California.

Steel, R.G., J.H. Torrie, and D.A. Dickey. 1997. Principles and procedures of statistics, a biometrical approach. McGraw-Hill, New York.