

Defining the Beginning: The Importance of Research Design

Justin D. Congdon

Savannah River Ecology Lab, P. O. Drawer E, Aiken, South Carolina 29802 USA; Tel: +1 (803) 725-5341; Fax: +1 (803) 725-3309; email: congdon@srel.edu

Arthur E. Dunham

Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104 USA; Tel: +1 (215) 898-4117; Fax: +1 (215) 898-8780; email: adunham@sas.upenn.edu

*A proverb: "Success is like a turtle climbing a mountain.
Failure is like water running down hill."*

Introduction

The probability of success of a research project is greatly enhanced when the "beginning" is correctly defined as a precise statement of goals and justification. Having accomplished this, the sequential steps necessary for writing a research plan and then successfully executing a research project are easier to identify and organize. Therefore, the message of this chapter is: by the time the laboratory is prepared or the first datum collected in the field, substantial effort should already have been invested in the conceptual and logistical framework of a project. This chapter discusses the steps that should result in a well-designed and integrated research plan.

A research plan consists of two general areas: research concepts and context (Table 1) and research logistics (Table 2). How well a research project is planned and how well the steps in the plan are integrated can make the difference between success or failure. The process of writing a research plan should start as soon as possible in the development of research ideas. For assistance with the process of writing research plans and research proposals, see Reis-Lehrer (1995).

Table 1. An outline of the concepts and context sections of a research plan.

-
1. Literature survey
 - a. Become familiar with the literature to identify a research problem and to explore the areas of biology that may impact the research plan.
 - b. Explore ways that enhance the integration of the study into broader biological disciplines and ways to collect, analyze, and present your data so that they are useful to others (e.g., for comparison among populations, or to use in models).
 2. Carefully and clearly state the problem in form of concise questions or as working or null hypotheses.
 3. Make a list of possible scenarios related to your research questions and then rank the most probable ones.
 4. Discuss all aspects of the research proposal with colleagues as you develop them. Investigators that work in the same or closely related fields are valuable sources of information.
 5. Write a detailed research proposal.
-

Table 2. An outline of the logistics section of a research plan.

1. Identify what information needs to be collected and how it will be collected.
 - a. Develop a logistical and quality control plan for the data collection, handling, and storage procedures, including who will be responsible for each procedure.
 - b. Design data sheets that prompt the person collecting data for each measurement. Data sheets should also be made with data recording, entry into computer files, editing and data analyses in mind.
 - c. Prior to actual data collection, carefully “walk through” as many of the techniques and procedures as feasible to detect problems with protocols and equipment.
 - d. Evaluate incoming data for recurring errors; frequent review may reveal unsuspected patterns that, if identified and responded to quickly, provide opportunities to improve research protocol or direction.
2. Talk to other investigators currently working with similar types of studies about logistics, research protocol, quality control plans, and types of data analyses.
3. As soon as possible following data collection, the data should be entered into computer files and then reviewed to detect problems associated with data sheets, data recording and entry, and with computer files.
4. Order supplies as far in advance as reasonable and funds are available.
5. Plan for the unexpected (equipment failure, accidents, illness, unusual weather).

Research Concepts and Context

The Literature

The ultimate quality and success of research are often a reflection of the time and effort invested in developing research ideas and concepts, a stage of planning that includes becoming familiar with the literature. Because the probability that a research idea will arise in a vacuum is vanishingly small, the probability of having a good research idea is higher for an investigator with experience and knowledge of the literature than it is for a novice.

The immediate goal of a literature survey is to determine if the research idea is worth pursuing (the research may have been done, or new publications may have revealed problems with the research idea). A second goal is to maximize the usefulness of re-

search results by integrating them into the specific research topics (*i.e.*, the results of others working in the same area) and into broader biological topics and disciplines (*e.g.*, life history, reproductive effort, parental investment, kin selection, conservation).

Placing a specific research project in a broader context requires a familiarity with: (1) literature outside of a specific research topic, (2) knowledge of current projects of other investigators, and (3) an historical perspective of the research problem and general topics. Scientific libraries at universities and individual reprint libraries are sources of review articles, book chapters on research topics and techniques, and articles on related topics of interest. To get started with sea turtle research, consult the Recent Papers section of the *Marine Turtle Newsletter*.

Research Questions

If the research idea is still viable after reviewing the literature, it is time to develop the research questions or hypotheses to be tested. The process of developing these details increases the probability of asking the right questions, and therefore, collecting the appropriate data. No amount of time spent, hard work, or elegant statistical methods will overcome the damage caused by a poorly framed question. Because research questions determine what, where, when and how data are collected, these questions represent an important link between the conceptual and logistic aspects of planning a research project.

An enlightening *a priori* exercise entails listing all possible answers to specific research questions (this list should not be influenced by whatever biases you may bring to the research). From the list of possible answers, rank those that appear the most probable answers or outcomes. By definition, the list of ranked answers should be shorter than the list of possible answers. The list and rankings should be kept with the research plan and then examined for possible surprises when the project is completed.

Hypotheses

Specific hypotheses or research questions depend on the context of a study. Testing a scientific hypothesis (a statement that attempts to predict how a particular feature of nature works), or answering a scientific question, almost always involves testing at least one, and often several, statistical hypotheses. A statistical hypothesis is a statement that attempts to predict the parameters of one (or more than one) probability distribution; for example, that the means of

two distributions are not different. Both experimental and observational studies should be designed with the ultimate statistical tests in mind.

Clear statements of statistical questions or hypotheses should be made early in the process of planning research and should be formulated so that they are concerned with the parameters of well-defined statistical populations. The statistical populations must be the population sampled, or the methods of statistical inference will not apply. Consultations with statisticians about experimental design, data collection, and statistical analyses should be made early in the process of developing the research design. Consultations with a statistician will be more profitable if the research questions or hypotheses are clearly stated and if some prior effort has been made to understand experimental design.

Sample Size

One area that is often overlooked in the design of ecological studies concerns the amount of data to collect. Data collection is often difficult, expensive, and may involve unavoidable destructive sampling of animals. Obtaining sufficient data to provide robust statistical tests of hypotheses may often conflict with logistical and ethical considerations concerning data acquisition. In such cases, sample sizes need to be large enough to provide adequate tests of important experimental effects, but they should not be unnecessarily large. Excellent sources for details of experimental design and sampling protocol can be found in Manly (1992), Sokal and Rohlf (1995), Winer (1971).

Statistical Power

The probability of rejecting a null hypothesis when it is false is termed the power of a statistical test, and calculation of sample size necessary to detect effects of a particular magnitude is called power analysis. If lack of consideration of statistical power results in inadequate sample sizes, confidence intervals about parameter estimates will be too wide to provide support for a conclusion that a null hypothesis was not rejected.

All research programs should include calculations of the power of statistical tests or width of confidence intervals (or both) that will result from planned sample sizes. Data necessary for these calculations (the variation expected for a given parameter) may not be available for the system that the researcher wants to study. Pilot studies or estimates based on previous studies

of similar systems may provide the data for power analysis. Details of how to conduct power analyses for various statistical tests and sampling designs are complex and beyond the scope of this chapter; however, detailed treatments may be found in most texts on experimental design (Winer, 1971; Montgomery, 1984; Manly, 1992; Sokal and Rohlf, 1995).

Development of the Research Plan

Two steps should be ongoing in the development of both the conceptual and logistic areas of a research plan. First, informal conversation with colleagues should be undertaken. Each colleague will bring a different viewpoint to proposed research that can improve research questions and generate new ones. In addition, many logistical problems that may hamper a new research program have already been experienced and solved by others. Because formal reviews are time consuming and are a courtesy, requests for these reviews should not be made until all conceptual and logistic steps in the research design are completed.

Second, the early development of a research plan should include the beginning of a detailed written research proposal. The process of writing a detailed proposal will help identify problems with research concepts, questions, and logistics and will enhance integration of various aspects of the proposed research. Project design and management programs are available for personal computers that assist project organization and time budgeting; these programs can be of value organizing research and writing the research plan.

Research Logistics

Research Quality

Once the questions or hypotheses have identified the data that are necessary, a plan should be developed for data collection (Table 2). Even though the precise statement of the research goals has identified the correct data needed to answer research questions, the quality of data collected depends on consistency of collection procedures, completeness, and accuracy of measurements. Therefore, each step of the logistical portion of the research plan should be based on how to assure the quality of the data collected (Table 3). The goal should always be to obtain the best data possible; however, each step in a research plan should consider the safety of the investigators and the welfare of the study organisms.

Table 3. Topics for the data acquisition and quality control section of a research plan.

1. Data collection.
2. Data recording.
3. The number of people making measurements and recording data.
4. The number and kinds of quality control measures necessary to verify the accuracy and consistency of data collected.
5. Data entry into computer files.
6. Storage of original and copies of data sheets and computer files.
7. Assignment of tasks and responsibilities.
8. Data analyses.

Data Sheets

Data sheets should be designed to complement the data collection process and minimize mistakes and omissions. If data are collected in a sequence (date, time, location, animal identification, sex, body length, body mass), data columns should be organized in the same order to minimize recording errors. In addition, errors made during transfer of recorded data to computer files will be minimized if the structures of the data files are in the same order as the data sheets.

Quality Control

Many research efforts are beyond the scope of one investigator and some require many field and laboratory assistants. If more than one person is involved in making measurements or observations, and recording data, it is important that the results are consistent and repeatable among personnel. The accuracy (how close measurements are to the actual dimensions of the object) and repeatability (how close measurements are to each other if taken by more than one person) required for a particular parameter measurement will determine how much training of personnel is necessary and how frequently instrument calibration will be necessary.

Consistency of measurements among personnel is critical for reliable data collection because statistical detection of differences depends to a major degree on the variability of the parameter being compared among treatments, years, or sites. For example, the validity of comparisons of population numbers among years or between two areas depends on: 1) whether the level of effort and consistency of data collection were similar for each sample period or area, 2) whether techniques used to obtain the data and data analyses were similar, and 3) whether the same in-

vestigators or investigators with similar training collected the data each year. In addition, the reliability of statistical detection of differences depends on the degree of variability of parameters that are compared among years or between sites.

Data Management

If at all possible, a personal computer should be used to store, edit and manage data. A spread sheet program or a relational data base management program (there are many versions of both on the market) should be used. Some programs allow the user to develop a computer screen that looks just like a data sheet, a feature that can help reduce data entry errors. Data management programs also offer error detection procedures and data manipulation features such as data sorting or indexing (arranging data in specific ways), data queries (counts or displays of categories of data), and the ability to build in custom programs that summarize data automatically (means, minimums, maximums, and standard deviations). Many of these programs also contain some data graphing procedures that allow visual inspection of data. The internal features of these programs greatly assist management of data, and provide options for transferring data into other formats for use with statistical packages and sophisticated scientific graphing programs.

A master data file should be identified and should contain the latest data and represent the most recent editing changes. Entry of data should never be made directly into the master file, and data analyses should never be performed on the master file. Data should be entered into a separate file, edited, and then appended to the master file. Backups of the master file should be made and stored in separate locations. In addition, the person responsible for data editing, management, and storage of backups should be clearly identified.

Personnel

Personnel represent a major expense for many research projects. Hiring self-motivated and goal-oriented people and giving them adequate training and guidance has obvious benefits to a research effort. However, recognizing those characteristics in potential field assistants can be difficult. One exercise that conscientious people practice, or can be trained to conduct, is to ask themselves the following questions before beginning a task: 1) Do I clearly understand both the immediate and long-term goals of the task? 2) Do I know how to accomplish the goals? 3) Do I have the necessary supplies and equipment to

complete the task?

If each person (including the principal investigator) involved in a research project asks themselves these questions each day, the time wasted will be greatly reduced. After a task is completed, another question should always be asked: 4) Am I communicating appropriately and adequately with others involved with the project about problems, results, and decisions I made while completing the task? Lack of communication among investigators is one of the most widespread problems in research conducted by a team.

Summary

The value of a research project is determined not just by the new data obtained, but how the research complements previous investigations and contributes to our understanding of broad biological topics or to tests of broad ecological theories, concepts, or general problems in conservation and management of biodiversity. Just as individual research questions influence the quality of data collected, suites of related questions within a research project influence the quality of extended research goals such as synthesis of general topics (*e.g.*, causes of population regulation and dynamics; sources of variation in growth rates among individuals or populations) and development of new questions and hypotheses that will guide future research.

Acknowledgments

Previous versions of this chapter were improved by comments from Kurt Buhlmann, Nat Frazer, J. Whitfield Gibbons, Nancy Dickson, Mark Komoroski, and Roy Nagle. Writing was supported by Financial Assistance Award Number DE-FC09-96SR18546 from the U.S. Department of Energy to the University of Georgia Research Foundation.

Literature Cited

- Manly, B. F. 1992. *The Design and Analysis of Research Studies*. Cambridge University Press. Cambridge, England. 353 pp.
- Montgomery, D. C. 1984. *Design and Analysis of Experiments*. Second Edition. Wiley, New York.
- Reis-Lehrer, L. 1995. *Grant Application Writers Handbook*. Jones and Barlett, Boston. 472 pp.
- Sokal, R. R. and J. F. Rohlf. 1995. *Biometry: the Principles and Practice of Statistics in Biological Research*. Third Edition. W. H. Freeman, New York. 867 pp.
- Winer, B. J. 1971. *Statistical principles in experimental design*. Second Edition. New York, McGraw-Hill. xv + 907 p.