



# TECHNIQUES FOR TRACKING, EVALUATING, AND REPORTING THE IMPLEMENTATION OF NONPOINT SOURCE CONTROL MEASURES

## AGRICULTURE

EPA/841-B-97-010  
September 1997

**TECHNIQUES FOR TRACKING, EVALUATING,  
AND REPORTING THE IMPLEMENTATION  
OF NONPOINT SOURCE CONTROL  
MEASURES**

**I. AGRICULTURE**

**Final  
September 1997**

Prepared for

Steve Dressing

Nonpoint Source Pollution Control Branch  
United States Environmental Protection Agency

Prepared by

Tetra Tech, Inc.  
EPA Contract No. 68-C3-0303  
Work Assignment No. 4-51

## TABLE OF CONTENTS

<b>Chapter 1</b>	<b>Introduction</b>	
1.1	Purpose of Guidance	1-1
1.2	Background	1-1
1.3	Types of Monitoring	1-3
1.4	Quality Assurance and Quality Control	1-4
1.5	Data Management	1-5
<b>Chapter 2</b>	<b>Sampling Design</b>	
2.1	Introduction	2-1
2.1.1	Study Objectives	2-1
2.1.2	Probabilistic Sampling	2-2
2.1.3	Measurement and Sampling Errors	2-8
2.1.4	Estimation and Hypothesis Testing	2-11
2.2	Sampling Considerations	2-13
2.2.1	Farm Ownership and Size	2-13
2.2.2	Location and Other Physical Characteristics	2-14
2.2.3	Farm Type and Agricultural Practices	2-15
2.2.4	Sources of Information	2-15
2.3	Sample Size Calculations	2-18
2.3.1	Simple Random Sampling	2-20
2.3.2	Stratified Random Sampling	2-24
2.3.3	Cluster Sampling	2-27
2.3.4	Systematic Sampling	2-27
<b>Chapter 3</b>	<b>Methods for Evaluating Data</b>	
3.1	Introduction	3-1
3.2	Comparing the Means from Two Independent Random Samples	3-2
3.3	Comparing the Proportions from Two Independent Samples	3-3
3.4	Comparing More Than Two Independent Random Samples	3-4
3.5	Comparing Categorical Data	3-4
<b>Chapter 4</b>	<b>Conducting the Evaluation</b>	
4.1	Introduction	4-1
4.2	Choice of Variables	4-2
4.3	Expert Evaluations	4-7
4.3.1	Site Evaluations	4-7
4.3.2	Rating Implementation of Management Measures and Best Management Practices	4-9
4.3.3	Rating Terms	4-10
4.3.4	Consistency Issues	4-12
4.3.5	Postevaluation Onsite Activities	4-13

4.4	Self-Evaluations .....	4-13
4.4.1	Methods .....	4-13
4.4.2	Cost .....	4-14
4.4.3	Questionnaire Design .....	4-17
4.5	Aerial Reconnaissance and Photography .....	4-19
<b>Chapter 5 Presentation of Evaluation Results</b>		
5.1	Introduction .....	5-1
5.2	Audience Identification .....	5-2
5.3	Presentation Format .....	5-2
5.3.1	Written Presentations .....	5-3
5.3.2	Oral Presentations .....	5-3
5.4	For Further Information .....	5-4
<b>References</b> .....		R-1
<b>Glossary</b> .....		G-1
<b>Index</b> .....		I-1
<b>Appendix A: Statistical Tables</b> .....		A-1

List of Tables

Table 2-1 Applications of four sampling designs for implementation monitoring ..... 2-3

Table 2-2 Errors in hypothesis testing ..... 2-12

Table 2-3 Acres of harvested cropland in Virginia from USDOC's 1992 Census of Agriculture ..... 2-14

Table 2-4 Definitions used in sample size calculation equations ..... 2-19

Table 2-5 Comparison of sample size as a function of various parameters ..... 2-21

Table 2-6 Common values of  $(Z_{\alpha} + Z_{2\beta})^2$  for estimating sample size ..... 2-23

Table 2-7 Allocation of Samples ..... 2-26

Table 2-8 Number of farms implementing recommended BMPs ..... 2-28

Table 3-1 Contingency table of observed operator type and implemented BMP ..... 3-5

Table 3-2 Contingency table of expected operator type and implemented BMP ..... 3-6

Table 3-3 Contingency table of implemented BMP and rating of installation and maintenance ..... 3-7

Table 3-4 Contingency table of implemented BMP and sample year ..... 3-8

Table 4-1 General types of information obtainable with self-evaluations and expert evaluations ..... 4-3

Table 4-2 Example variables for management measure implementation analysis ..... 4-6

List of Figures

Figure 2-1 Simple random sampling from a list and a map ..... 2-4

Figure 2-2 Stratified random sampling from a list and a map ..... 2-6

Figure 2-3 Cluster sampling from a list and a map ..... 2-7

Figure 2-4 Systematic sampling from a list and a map ..... 2-9

Figure 2-5 Graphical presentation of the relationship between bias, precision, and accuracy ..... 2-11

Figure 2-6 Example route for a county transect survey ..... 2-29

Figure 4-1 Potential variables and examples of implementation standards and specifications ..... 4-5

Figure 4-2 Sample draft survey for confined animal facility management evaluation ..... 4-15

Figure 5-1 Example of presentation of information in a written slide ..... 5-4

Figure 5-2 Example of representation of data using a combination of a pie chart and a horizontal bar chart ..... 5-5

Figure 5-3 Example representation of data in the form of a pie chart ..... 5-6

## CHAPTER 1. INTRODUCTION

### 1.1 PURPOSE OF GUIDANCE

This guidance is intended to assist state, regional, and local environmental professionals in tracking the implementation of best management practices (BMPs) used to control agricultural nonpoint source pollution. Information is provided on methods for selecting sites for evaluation, sample size estimation, sampling, and results evaluation and presentation. The focus of the guidance is on the statistical approaches needed to properly collect and analyze data that are accurate and defensible. A properly designed BMP implementation monitoring program can save both time and money. For example, there are over 37,000 farms in the state of Virginia. To determine the status of BMP implementation on each of those farms would easily exceed most budgets and thus statistical sampling of sites is needed. This document provides guidance for sampling representative farms to yield summary statistics at a fraction of the cost of a comprehensive inventory.

Some nonpoint source projects and programs combine BMP implementation monitoring with water quality monitoring to evaluate the effectiveness of BMPs at protecting water quality (Meals, 1988; Rashin et al., 1994; USEPA, 1993b). For this type of monitoring to be successful, the scale of the project must be small (e.g., a watershed of a few hundred to a few thousand acres). Accurate records of all the sources of pollutants of concern and a census of how all BMPs are operating are very important for this type of monitoring effort. Otherwise, it can be extremely difficult to

---

*The focus of this guide is on the design of monitoring programs to assess agricultural management measure and best management practice implementation, with particular emphasis on statistical considerations.*

---

correlate BMP implementation with changes in stream water quality. This guidance does not address monitoring the implementation and effectiveness of all BMPs in a watershed. This guidance does provide information to help program managers gather statistically valid information to assess implementation of BMPs on a more general (e.g., statewide) basis. The benefits of implementation monitoring are presented in Section 1.3.

### 1.2 BACKGROUND

Pollution from nonpoint sources—sediment deposition, erosion, contaminated runoff, hydrologic modifications that degrade water quality, and other diffuse sources of water pollution—is the largest cause of water quality impairment in the United States (USEPA, 1995). Congress passed the Coastal Zone Act Reauthorization Amendments of 1990 (CZARA) to help address nonpoint source pollution in coastal waters. CZARA provides that each state with an approved coastal zone management program develop and submit to the U.S. Environmental Protection Agency (EPA) and National Oceanic and Atmospheric Administration (NOAA) a Coastal Nonpoint Pollution Control Program (CNPCP). State programs must “provide for the

implementation” of management measures in conformity with the EPA *Guidance Specifying Management Measures For Sources Of Nonpoint Pollution In Coastal Waters*, developed pursuant to section 6217(g) of CZARA (USEPA, 1993a). Management measures (MMs), as defined in CZARA, are economically achievable measures to control the addition of pollutants to coastal waters, which reflect the greatest degree of pollutant reduction achievable through the application of the best available nonpoint pollution control practices, technologies, processes, siting criteria, operating methods, or other alternatives. Many of EPA’s MMs are combinations of BMPs. For example, depending on site characteristics, implementation of the Confined Animal Facility MM might involve use of the following BMPs: Construction of a waste storage pond, installation of grassed waterways, protection of heavily-used areas, management of roof runoff, and construction of a composting facility.

CZARA does not specifically require that states monitor the implementation of MMs and BMPs as part of their CNPCPs. State CNPCPs must however, provide for technical assistance to local governments and the public for implementing the MMs and BMPs. Section 6217(b) states:

Each State program . . . shall provide for the implementation, at a minimum, of management measures . . . and shall also contain . . . (4) The provision of technical and other assistance to local governments and the public for implementing the measures . . . which may include assistance . . . to predict and assess the effectiveness of such measures . . . .

EPA and NOAA also have some responsibility under section 6217 for providing technical assistance to implement state CNPCPs. Section 6217(d), Technical assistance, states:

[NOAA and EPA] shall provide technical assistance . . . in developing and implementing programs. Such assistance shall include: . . . (4) methods to predict and assess the effects of coastal land use management measures on coastal water quality and designated uses.

This guidance document was developed to provide the technical assistance described in CZARA sections 6217(b)(4) and 6217(d), but the techniques described can be used for other similar programs and projects. For instance, monitoring projects funded under Clean Water Act (CWA) section 319(h) grants, efforts to implement total maximum daily loads developed under CWA Section 303(d), stormwater permitting programs, and other programs could all benefit from knowledge of BMP implementation.

Methods to assess the implementation of MMs and BMPs, then, are a key focus of the technical assistance to be provided by EPA and NOAA. Implementation assessments can be done on several scales. Site-specific assessments can be used to assess individual BMPs or MMs, and watershed assessments can be used to look at the cumulative effects of implementing multiple MMs. With regard to “site-specific” assessments, individual BMPs must be assessed at the appropriate scale for the BMP of interest. For example, to assess the implementation of MMs and BMPs for animal waste handling and disposal on a farm, only the structures, areas, and practices implemented specifically for animal waste

management (e.g., dikes, diversions, storage ponds, composting facility, and manure application records) would need to be inspected. In this instance the animal waste storage facility would be the appropriate scale and “site.” To assess erosion control, the proper scale might be fields over 10 acres and the site could be 100-meter transect measurements of crop residue. For nutrient management, the scale and site might be an entire farm. Site-specific measurements can then be used to extrapolate to a watershed or statewide assessment. It is recognized that some studies might require a complete inventory of MM and BMP implementation across an entire watershed or other geographic area.

### 1.3 TYPES OF MONITORING

The term *monitor* is defined as “to check or evaluate something on a constant or regular basis” (Academic Press, 1992). It is possible to distinguish among various types of monitoring. Two types, implementation and trend (i.e., trends in implementation) monitoring, are the focus of this guidance. These types of monitoring can be used to address the following goals:

- Determine the extent to which MMs and BMPs are implemented in accordance with relevant standards and specifications.
- Determine whether there has been a change in the extent to which MMs and BMPs are being implemented.

In general, implementation monitoring is used to determine whether goals, objectives, standards, and management practices are being implemented as detailed in implementation

plans. In the context of BMPs within state CNPCPs, implementation monitoring is used to determine the degree to which MMs and BMPs required or recommended by the CNPCPs are being implemented. If CNPCPs call for voluntary implementation of MMs and BMPs, implementation monitoring can be used to determine the success of the voluntary program (1) within a given monitoring period (e.g., 1 or 2 years); (2) during several monitoring periods, to determine any temporal trends in BMP implementation; or (3) in various regions of the state.

Trend monitoring involves long-term monitoring of changes in one or more parameters. As discussed in this guidance, public attitudes, land use, or the use of different agricultural practices are examples of parameters that could be measured with trend monitoring. For example, the Conservation Technology Information Center tracks trends in the implementation of different tillage practices from year to year (CTIC, 1994). Isolating the impacts of MMs and BMPs on water quality requires tracking MM and BMP implementation over time, i.e., trend monitoring.

Because trend monitoring involves measuring a change (or lack thereof) in some parameter over time, it is necessarily of longer duration and requires that a baseline, or starting point, be established. Any changes in the measured parameter are then detected in reference to the baseline.

Implementation and the related trend monitoring can be used to determine (1) which MMs and BMPs are being implemented, (2) whether MMs and BMPs are being implemented as designed, and

(3) the need for increased efforts to promote or induce use of MMs and BMPs. Data from implementation monitoring, used in combination with other types of data, can be useful in meeting a variety of other objectives, including the following (Hook et al., 1991; IDHWH, 1993; Schultz, 1992):

- To evaluate BMP effectiveness for protecting soil and water resources.
- To identify areas in need of further investigation.
- To establish a reference point of overall compliance with BMPs.
- To determine whether farmers are aware of BMPs.
- To determine whether farmers are using the advice of agricultural BMP experts.
- To identify any BMP implementation problems specific to a category of farm.
- To evaluate whether any agricultural practices cause environmental damage.
- To compare the effectiveness of alternative BMPs.

MacDonald et al. (1991) describes additional types of monitoring, including effectiveness monitoring, baseline monitoring, project monitoring, validation monitoring, and compliance monitoring. As emphasized by McDonald and others, these monitoring types are not mutually exclusive and the distinctions among them are usually determined by the purpose of the monitoring.

Effectiveness monitoring is used to determine whether MMs or BMPs, as designed and implemented, are effective in meeting management goals and objectives.

Effectiveness monitoring is a logical follow-up to implementation monitoring, because it is essential that effectiveness monitoring include an assessment of the adequacy of the design and installation of MMs and BMPs. For example, the objective of effectiveness monitoring could be to evaluate the effectiveness of MMs and BMPs *as designed and installed*, or to evaluate the effectiveness of MMs and BMPs *that are designed and installed adequately or to standards and specifications*. Effectiveness monitoring is not addressed in this guide, but is the subject of another EPA guidance document, *Monitoring Guidance for Determining the Effectiveness of Nonpoint Source Controls* (USEPA, 1997).

#### 1.4 QUALITY ASSURANCE AND QUALITY CONTROL

An integral part of the design phase of any nonpoint source pollution monitoring project is quality assurance and quality control (QA/QC). Development of a quality assurance project plan (QAPP) is the first step of incorporating QA/QC into a monitoring project. The QAPP is a critical document for the data collection effort inasmuch as it integrates the technical and quality aspects of the planning, implementation, and assessment phases of the project. The QAPP documents how QA/QC elements will be implemented throughout a project's life. It contains statements about the expectations and requirements of those for whom the data is being collected (i.e., the decision maker) and provides details on project-specific data collection and data management procedures that are designed to

ensure that these requirements are met. Development and implementation of a QA/QC program, including preparation of a QAPP, can require up to 10 to 20 percent of project resources (Cross-Smieciniski and Stetzenback, 1994), but this cost is recaptured in lower overall costs due to the project being well planned and executed. A thorough discussion of QA/QC is provided in Chapter 5 of EPA's *Monitoring Guidance for Determining the Effectiveness of Nonpoint Source Controls* (USEPA, 1997).

### 1.5 DATA MANAGEMENT

Data management is a key component of a successful MM or BMP implementation monitoring effort. The data management system that is used—which includes the quality control and quality assurance aspects of data handling, how and where data are stored, and who manages the stored data—determines the reliability, longevity, and accessibility of the data. Provided that the data collection effort was planned and executed well, an organized and efficient data management system will ensure that the data can be used with confidence by those who must make decisions based upon it, the data will be useful as a baseline for similar data collection efforts in the future, the data will not become obsolete (or be misplaced!) quickly, and the data will be available to a variety of users for a variety of applications.

Serious consideration is often not given to a data management system prior to a data collection effort, which is precisely why it is so important to recognize the long-term value of a small investment of time and money in proper data management. Data management competes with other agency priorities for money, staff,

and time, and if the importance and long-term value of proper data management is recognized early in a project's development, the more likely it will be to receive sufficient funding. Overall, data management might account for only a small portion of a project's total budget, but the return on the investment is great when it is considered that the larger investment in data collection can be rendered virtually useless unless data is managed adequately.

Two important aspects of data that should be considered when planning the initial data collection effort and a data management system are data life cycle and data accessibility. The data life cycle can be characterized by the following stages:

(1) Data is collected; (2) data is checked for quality; (3) data is entered into a data base; (4) data is used, and (5) data eventually becomes obsolete. The expected usefulness and life span of the data should be considered during the initial stages of planning a data collection effort, when the money, staff, and time that are devoted to data collection must be weighed against its usefulness and longevity. Data with a limited use and that is likely to become obsolete soon after it is collected is a poorer investment decision than data with multiple applications and a long life span. If a data collection effort involves the collection of data of limited use and a short life span, it might be necessary to modify the data collection effort—either by changing its goals and objectives or by adding new ones—to increase the breadth and length of the data's applicability. A good data management system will ensure that any data that are collected will be useful for the greatest number of applications for the longest possible time.

Data accessibility is a critical factor in determining its usefulness. Data attains its highest value if it is as widely accessible as possible, if access to it requires the least amount of staff effort as possible, and if it can be used by others conveniently. If data are stored where those who might need it can obtain it with little assistance, it is more likely to be shared and used. The format for data storage determines how conveniently the data can be used. Electronic storage in a widely available and used data storage format makes it convenient to use. Storage as only a paper copy buried in a report, where any analysis requires entry into an electronic format or time-consuming manipulation, makes data extremely inconvenient to use and unlikely that it will be used.

The following should be considered for the development of a data management strategy:

- *What level of quality control should the data be subject to?* Data that will be used for a variety of purposes or that will be used for important decisions should receive a careful quality control check.
- *Where and how will the data be stored?* The options for data storage range from a printed final report on a bookshelf to an electronic data base accessible to government agencies and the public. Determining where and how data will be stored therefore also requires careful consideration of the question: *How accessible should the data be?*
- *Who will maintain the data base?* Data stored in a large data base might be managed by a professional data manager, while data kept in agency files might be

managed by people with various backgrounds over the course of time.

- *How much will data management cost?* As with all other aspects of a data collection effort, data management costs money and this cost must be balanced with all other costs involved in the project.

## CHAPTER 2. SAMPLING DESIGN

### 2.1 INTRODUCTION

This chapter discusses recommended methods for designing sampling programs to track and evaluate the implementation of nonpoint source control measures. This chapter does not address sampling to determine whether the management measures (MMs) or best management practices (BMPs) are effective since no water quality sampling is done. Because of the variation in agricultural practices and related nonpoint source control measures implemented throughout the United States, the approaches taken by various states to track and evaluate nonpoint source control measure implementation will differ. Nevertheless, all approaches can be based on sound statistical methods for selecting sampling strategies, computing sample sizes, and evaluating data. EPA recommends that states should consult with a trained statistician to be certain that the approach, design, and assumptions are appropriate to the task at hand.

As described in Chapter 1, implementation monitoring is the focus of this guidance. Effectiveness monitoring is the focus of another guidance prepared by EPA, *Monitoring Guidance for Determining the Effectiveness of Nonpoint Source Controls* (USEPA, 1997). The recommendations and examples in this chapter address two primary monitoring goals:

- Determine the extent to which MMs and BMPs are implemented in accordance with relevant standards and specifications.

- Determine whether there is a change in the extent to which MMs and BMPs are being implemented.

For example, state or county agriculture personnel might be interested in whether regulations for the exclusion of livestock from riparian areas are being adhered to in regions with particular water quality problems. State or county personnel might also be interested in whether, in response to an intensive state-wide effort to improve pesticide use practices and increase the use of integrated pest management practices, there is a detectable change in the pesticide practices being used by farmers.

#### 2.1.1 Study Objectives

To develop a study design, clear, quantitative monitoring objectives must be developed. For example, the objective might be to estimate the percent of farm owners or managers that use integrated pest management (IPM) to within  $\pm 5$  percent. Or perhaps a state is getting ready to perform an extensive 2-year outreach and cost-share effort to promote a fence-out or other program to reduce cattle wading through streams. In this case, detecting a 10 percent change in the farms that permit their cattle direct access to streams might be of interest. In the first example, summary statistics are developed to describe the current status, whereas in the second example, some sort of statistical analysis (hypothesis testing) is performed to determine whether a significant change has really occurred. This choice has an impact on how the data are collected. As an example, summary statistics might require unbalanced

sample allocations to account for variability such as farm size, type, and ownership, whereas balanced designs (e.g., two sets of data with the same number of observations in each set) are more typical for hypothesis testing.

### 2.1.2 Probabilistic Sampling

Most study designs that are appropriate for tracking and evaluating implementation are based on a probabilistic approach since tracking every farm is not cost-effective. In a probabilistic approach, individuals are randomly selected from the entire group. The selected individuals are evaluated, and the results from the individuals provide an unbiased assessment about the entire group. Applying the results from randomly selected individuals to the entire group is *statistical inference*. Statistical inference enables one to determine, for example, in terms of probability, the percentage of farms using IPM without visiting every farm. One could also determine whether the change in the number of farms with appropriate nutrient management is within the range of what could occur by chance or the change is large enough to indicate a real modification of farmer practices.

The group about which inferences are made is the population or *target population*, which consists of *population units*. The *sample population* is the set of population units that are directly available for measurement. For example, if the objective is to determine the degree to which adequate animal waste management has been established in agricultural operations, the population to be sampled would be agricultural operations for which animal waste management is an appropriate BMP (i.e., farms with livestock).

Statistical inferences can be made only about the target population available for sampling. For example, if implementation of grazing management is being assessed and only public grazing lands can be sampled, inferences cannot be made about the management of private grazing lands. Another example to consider is a mail survey. In most cases, only a percentage of survey forms is returned. The extent to which nonrespondents bias the survey findings should be examined: Do the nonrespondents represent those less likely to use IPM? Typically, a second mailing, phone calls, or visits to those who do not respond might be necessary to evaluate the impact of nonrespondents.

The most common types of sampling that should be used for implementation monitoring are summarized in Table 2-1. In general, probabilistic approaches are preferred. However, there might be circumstances under which targeted sampling should be used. Targeted sampling refers to using best professional judgement for selecting sample locations. For example, state or county agriculture personnel deciding to evaluate all farms in a given watershed would be targeted sampling. The choice of a sampling plan depends on study objectives, patterns of variability in the target population, cost-effectiveness of alternative plans, types of measurements to be made, and convenience (Gilbert, 1987).

Table 2-1. Applications of four sampling designs for implementation monitoring.

Sampling Design	Comment
Simple Random Sampling	Each population unit has an equal probability of being selected.
Stratified Random Sampling	Useful when a sample population can be broken down into groups, or strata, that are internally more homogeneous than the entire sample population. Random samples are taken from each stratum although the probability of being selected might vary from stratum to stratum depending on cost and variability.
Cluster Sampling	Useful when there are a number of methods for defining population units and when individual units are clumped together. In this case, clusters are randomly selected and every unit in the cluster is measured.
Systematic Sampling	This sampling has a random starting point with each subsequent observation a fixed interval (space or time) from the previous observation.

*Simple random sampling* is the most elementary type of sampling. Each unit of the target population has an equal chance of being selected. This type of sampling is appropriate when there are no major trends, cycles, or patterns in the target population (Cochran, 1977). Random sampling can be applied in a variety of ways including farm or field selection. Random samples can also be taken at different times at a single farm. Figure 2-1 provides an example of simple random sampling from a listing of farms and from a map.

If the pattern of MM and BMP implementation is expected to be uniform across the state, simple random sampling is appropriate to estimate the extent of implementation. If, however, implementation is homogeneous only within certain categories (e.g., federal, state, or private lands), stratified random sampling should be used.

In *stratified random sampling*, the target population is divided into groups called strata

for the purpose of obtaining a better estimate of the mean or total for the entire population. Simple random sampling is then used within each stratum. Stratification involves the use of categorical variables to group observations into more units, thereby reducing the variability of observations within each unit. For example, in a state with federal, state, and private rangelands that are used for grazing, there might be different patterns of BMP implementation. Lands in the state could be divided into federal, state, and private as separate strata from which samples would be taken. In general, a larger number of samples should be taken in a stratum if the stratum is more variable, larger, or less costly to sample than other strata. For example, if BMP implementation is more variable on private rangelands, a greater number of sampling sites might be needed in that stratum to increase the precision of the overall estimate. Cochran (1977) found that stratified random sampling provides a better estimate of the mean for a population with a trend, followed in order by systematic sampling (discussed later) and

Farm Catalog No.	Waterbody	Type	County Code
1	Stream	Crop	N3
2	Pond	Crop	S4
3	Pond	Livestock	S2
4	Stream	Crop/Livestock	E5
5	---	Livestock	S1
6	River	Crop	S7
7	Lake	Crop/Livestock	W18
8	---	Crop	E34
...	...	...	...
118	Stream	Crop/Livestock	S21
119	Stream	Crop/Livestock	W7
120	---	Crop/Livestock	W4
121	---	Livestock	N5
122	Bay	Crop/Livestock	N9
123	Bay	Crop	S3
124	Stream	Crop	W11
125	Pond	Crop/Livestock	E14
126	Stream	Crop/Livestock	S14
127	---	Livestock	S8
128	Pond	Crop	N13

Figure 2-1a. Simple random sampling from a listing of farms. In this listing, all farms are presented as a single list and farms are selected randomly from the entire list. Shaded farms represent those selected for sampling.

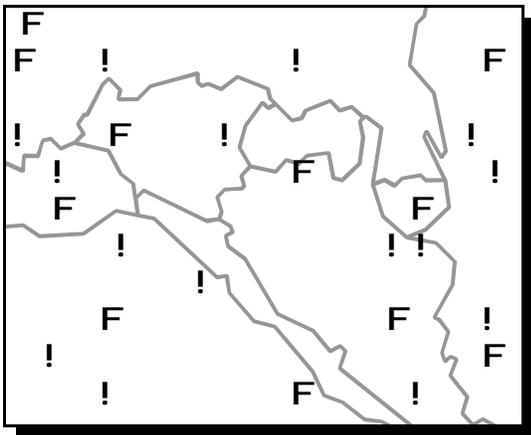


Figure 2-1b. Simple random sampling from a map. Dots represent farms. All farms of interest are represented on the map, and the farms to be sampled (open dots—F) were selected randomly from all of those on the map. The shaded lines on the map could represent county, watershed, hydrologic, or some other boundary, but they are ignored for the purposes of simple random sampling.

simple random sampling. He also noted that stratification typically results in a smaller variance for the estimated mean or total than that which results from comparable simple random sampling.

If the state believes that there will be a difference between two or more subsets of farms, such as between types of ownership or crop, the farms can first be stratified into these subsets and a random sample taken within each subset (McNew, 1990). The goal of stratification is to increase the accuracy of the estimated mean values over what could have been obtained using simple random sampling of the entire population. The method makes use of prior information to divide the target population into subgroups that are internally homogeneous. There are a number of ways to “select” farms (e.g., by farm ownership, farm size, farm type, hydrologic unit, soil type, or county), or sets of farms, to be certain that important information will not be lost, or that MM or BMP use will not be misrepresented as a result of treating all potential survey farms as equal. Figure 2-2 provides an example of stratified random sampling from a listing of farms and from a map.

It might also be of interest to compare the relative percentages of cropland classified as having high, medium, and low erosion potentials that are under conservation tillage. Highly erodible land might be responsible for a larger share of sediment losses, and it would usually be desirable to track the extent to which conservation tillage practices have been implemented on these land areas. A stratified random sampling procedure could be used to estimate the percentage of total cropland with different erosion potentials under conservation tillage.

*Cluster sampling* is applied in cases where it is more practical to measure randomly selected groups of individual units than to measure randomly selected individual units (Gilbert, 1987). In cluster sampling, the total population is divided into a number of relatively small subdivisions, or clusters, and then some of the subdivisions are randomly selected for sampling. For one-stage cluster sampling, the selected clusters are sampled totally. In two-stage cluster sampling, random sampling is performed within each cluster (Gaugush, 1987). For example, this approach might be useful if a state wants to estimate the proportion of farms less than 800 meters from a stream that are following state-approved nutrient management plans. All farms less than 800 meters from a particular stream (or portion of a stream) can be regarded as a single cluster. Once all clusters have been identified, specific clusters can be randomly chosen for sampling. Freund (1973) notes that estimates based on cluster sampling are generally not as good as those based on simple random samples, but they are more cost-effective. As a result, Gaugush (1987) believes that the difficulty associated with analyzing cluster samples is compensated for by the reduced sampling requirements and cost. Figure 2-3 provides an example of cluster sampling from a listing of farms and from a map.

Farm Catalog No.	Water Body	Type	County Code
1	Stream	Crop	N3
2	Pond	Crop	S4
6	River	Crop	S7
8	---	Crop	E34
...	...	...	...
123	Bay	Crop	S3
124	Stream	Crop	W11
128	Pond	Crop	N13
3	Pond	Livestock	S2
5	---	Livestock	S1
...	...	...	...
121	---	Livestock	N5
127	---	Livestock	S8
4	Stream	Crop/Livestock	E5
7	Lake	Crop/Livestock	W18
...	...	...	...
118	Stream	Crop/Livestock	S21
119	Stream	Crop/Livestock	W7
120	---	Crop/Livestock	W4
122	Bay	Crop/Livestock	N9
125	Pond	Crop/Livestock	E14
126	Stream	Crop/Livestock	S14

Figure 2-2a. Stratified random sampling from a listing of farms. Within this listing, farms are subdivided by type. Then, considering only one farm type (e.g., crop farms), some farms are selected randomly. The process of random sampling is then repeated for the other farm types (i.e., livestock, crop/livestock). Shaded farms represent those selected for sampling.

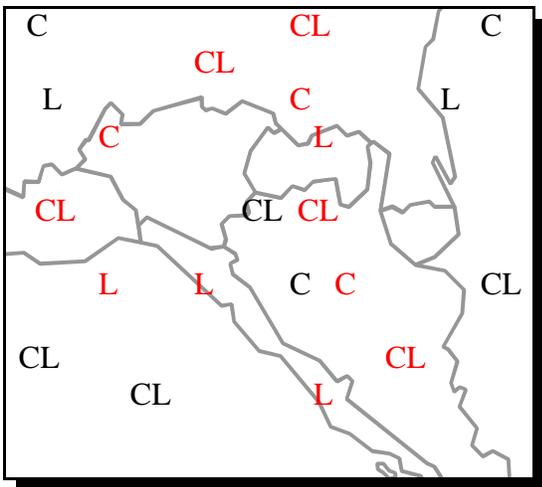


Figure 2-2b. Stratified random sampling from a map. Letters represent farms, subdivided by type (C = crop, CL = crop/livestock, L = livestock). All farms of interest are represented on the map. From all farms in one type category, some were randomly selected for sampling (highlighted farms). The process was repeated for each farm type category. The shaded lines on the map could represent county, soil type, or some other boundary, and could have been used as a means for separating the farms into categories for the sampling process.

Farm Catalog No.	Water Body	Type	County Code
1	Stream	Crop	N3
4	Stream	Crop/Livestock	E5
...	...	...	...
118	Stream	Crop/Livestock	S21
119	Stream	Crop/Livestock	W7
124	Stream	Crop	W11
126	Stream	Crop/Livestock	S14
2	Pond	Crop	S4
3	Pond	Livestock	S2
...	...	...	...
125	Pond	Crop/Livestock	E14
128	Pond	Crop	N13
6	River	Crop	S7
7	Lake	Crop/Livestock	W18
122	Bay	Crop/Livestock	N9
123	Bay	Crop	S3
5	---	Livestock	S1
8	---	Crop	E34
...	...	...	...
120	---	Crop/Livestock	W4
121	---	Livestock	N5
127	---	Livestock	S8

Figure 2-3a. One-stage cluster sampling from a listing of farms. Within this listing, farms are subdivided by the type of waterbody near them. Some of the waterbody types were then randomly selected (in this case streams and bays) and all farms with those waterbodies were selected for sampling. Shaded farms represent those selected for sampling.

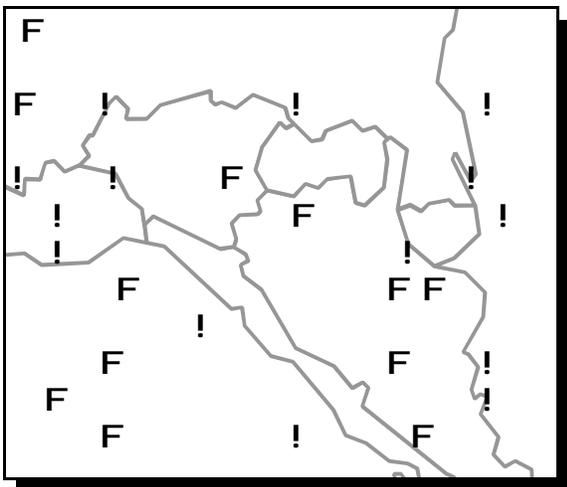


Figure 2-3b. Cluster sampling from a map. All farms in the area of interest are represented on the map (closed {!} and open {F} dots). Waterbody types were selected randomly, and farms with those waterbodies (closed dots {!}) were selected for sampling. Shaded lines could represent a type of boundary, such as soil type, county, or watershed, and could have been used as the basis for the sampling process as well.

*Systematic sampling* is used extensively in water quality monitoring programs because it is relatively easy to do from a management perspective. In systematic sampling the first sample has a random starting point and each subsequent sample has a constant distance from the previous sample. For example, if a sample size of 70 is desired from a mailing list of 700 farm owners, the first sample would be randomly selected from among the first 10 people, say the seventh person. Subsequent samples would then be based on the 17<sup>th</sup>, 27<sup>th</sup>, ..., 697<sup>th</sup> person. In comparison, a stratified random sampling approach might be to sort the mailing list by county and then to randomly select farm owners from each county. Figure 2-4 provides an example of systematic sampling from a listing of farms and from a map.

In general, systematic sampling is superior to stratified random sampling when only one or two samples per stratum are taken for estimating the mean (Cochran, 1977) or when there is a known pattern of management measure implementation. Gilbert (1987) reports that systematic sampling is equivalent to simple random sampling in estimating the mean if the target population has no trends, strata, or correlations among the population units. Cochran (1977) notes that on the average, simple random sampling and systematic sampling have equal variances. However, Cochran (1977) also states that for any single population for which the number of sampling units is small, the variance from systematic sampling is erratic and might be smaller or larger than the variance from simple random sampling.

Gilbert (1987) cautions that any periodic variation in the target population should be known before establishing a systematic

sampling program. Sampling intervals equal to or multiples of the target population's cycle of variation might result in biased estimates of the population mean. Systematic sampling can be designed to capitalize on a periodic structure if that structure can be characterized sufficiently (Cochran, 1977). A simple or stratified random sample is recommended, however, in cases where the periodic structure is not well known or if the randomly selected starting point is likely to have an impact on the results (Cochran, 1977).

Gilbert (1987) notes that assumptions about the population are required in estimating population variance from a single systematic sample of a given size. However, there are systematic sampling approaches that do support unbiased estimation of population variance, including multiple systematic sampling, systematic stratified sampling, and two-stage sampling (Gilbert, 1987). In multiple systematic sampling more than one systematic sample is taken from the target population. Systematic stratified sampling involves the collection of two or more systematic samples within each stratum.

### 2.1.3 Measurement and Sampling Errors

In addition to making sure that samples are representative of the sample population, it is also necessary to consider the types of bias or error that might be introduced into the study. *Measurement error* is the deviation of a measurement from the true value (e.g., the percent residue cover for a field was estimated as 23 percent and the true value was 26 percent). A consistent under- or overestimation of the true value is referred to as *measurement bias*. *Random sampling error* arises from the variability from one population unit to the next (Gilbert, 1987), explaining

Farm Catalog No.	Water Body	Type	County Code
1	Stream	Crop	N3
2	Pond	Crop	S4
3	Pond	Livestock	S2
4	Stream	Crop/Livestock	E5
5	---	Livestock	S1
6	River	Crop	S7
7	Lake	Crop/Livestock	W18
8	---	Crop	E34
...	...	...	...
118	Stream	Crop/Livestock	S21
119	Stream	Crop/Livestock	W7
120	---	Crop/Livestock	W4
121	---	Livestock	N5
122	Bay	Crop/Livestock	N9
123	Bay	Crop	S3
124	Stream	Crop	W11
125	Pond	Crop/Livestock	E14
126	Stream	Crop/Livestock	S14
127	---	Livestock	S8
128	Pond	Crop	N13

Figure 2-4a. Systematic sampling from a listing of farms. From a listing of all farms of interest, an initial site (Farm No. 3) was selected randomly from among the first ten on the list. Every fifth farm listed was subsequently selected for sampling. Shaded farms represent those selected for sampling.

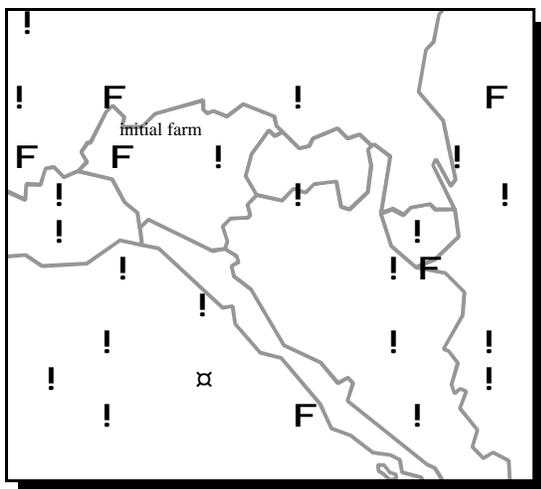


Figure 2-4b. Systematic sampling from a map. Dots (! and F) represent farms of interest. A single point on the map (⊠) and one of the farms were randomly selected. A line was stretched outward from the point to (and beyond) the selected farm. The line was then rotated about the map and every fifth dot that it touched was selected for sampling (open dots—F). The direction of rotation was determined prior to selection of the point of the line's origin and the initial farm. The shaded lines on the map could represent county boundaries, soil type, watershed, or some other boundary, but were not used for the sampling process.

why the proportion of farm owners using a certain BMP differs from one survey to another.

The goal of sampling is to obtain an accurate estimate by reducing the sampling and measurements errors to acceptable levels, while explaining as much of the variability as possible to improve the precision of the estimates (Gaugush, 1987). *Precision* is a measure of how close an agreement there is among individual measurements of the same population. The *accuracy* of a measurement refers to how close the measurement is to the true value. If a study has low bias and high precision, the results will have high accuracy. Figure 2-5 illustrates the relationship between bias, precision, and accuracy.

As suggested earlier, numerous sources of variability should be accounted for in developing a sampling design. Sampling errors are introduced by virtue of the natural variability within any given population of interest. As sampling errors relate to MM or BMP implementation, the most effective method for reducing such errors is to carefully determine the target population and to stratify the target population to minimize the nonuniformity in each stratum.

Measurement errors can be minimized by ensuring that interview questions or surveys are well designed. If a survey is used as a data collection tool, for example, the investigator should evaluate the nonrespondents to determine whether there is a bias in who returned the results (e.g., whether the nonrespondents were more or less likely to implement MMs or BMPs). If data are collected by sending staff out to inspect randomly selected fields, the approach for inspecting the fields should be consistent. For

example, how do survey personnel determine that at least 40 percent of the ground is covered by residuals, or what is the basis for determining whether a BMP has been properly implemented?

Reducing sampling errors below a certain point (relative to measurement errors) does not necessarily benefit the resulting analysis because total error is a function of the two types of error. For example, if measurement errors such as response or interviewing errors are large, there is no point in taking a huge sample to reduce the sampling error of the estimate since the total error will be primarily determined by the measurement error. Measurement error is of particular concern when farmer surveys are used for implementation monitoring. Likewise, reducing measurement errors would not be worthwhile if only a small sample size were available for analysis because there would be a large sampling error (and therefore a large total error) regardless of the size of the measurement error. A proper balance between sampling and measurement errors should be maintained because research accuracy limits effective sample size and vice versa (Blalock, 1979).

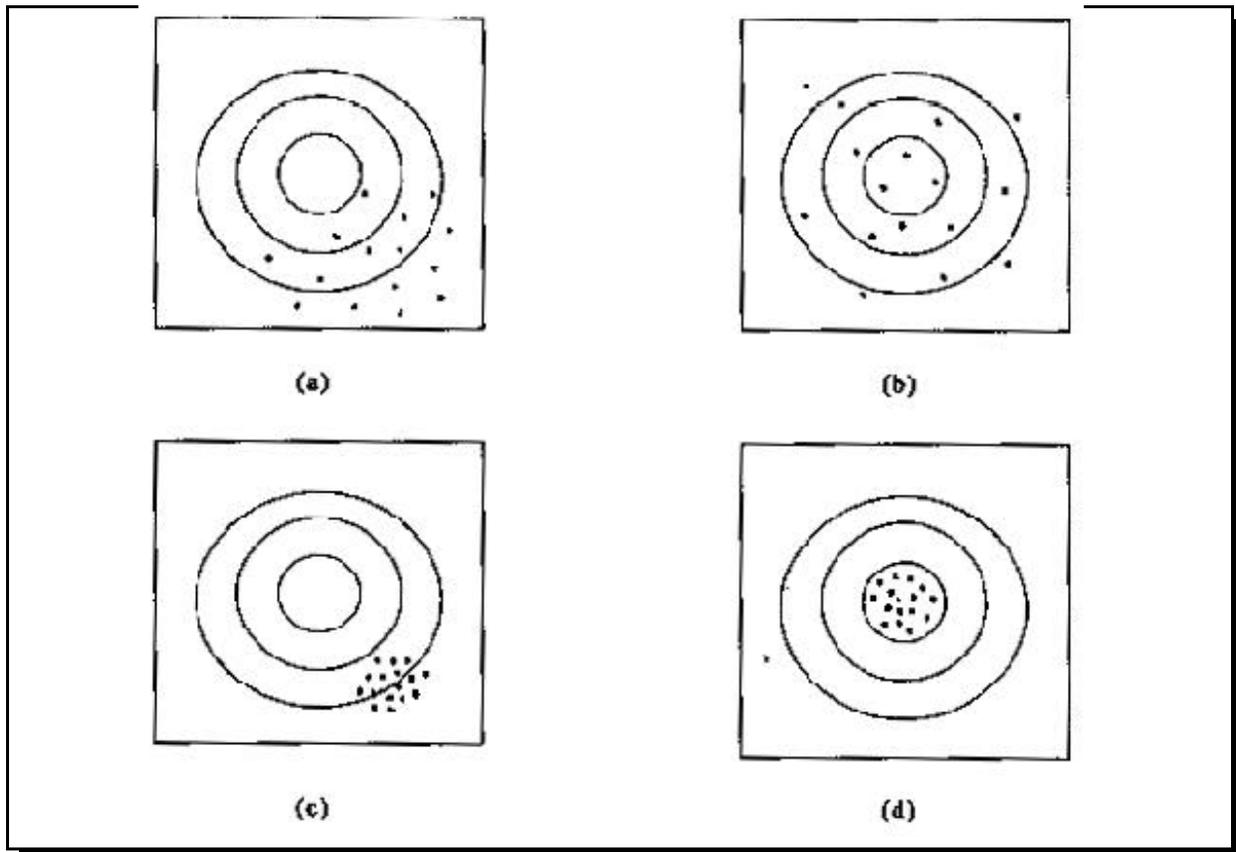


Figure 2-5. Graphical representation of the relationship between bias, precision, and accuracy (after Gilbert, 1987). (a): high bias + low precision = low accuracy; (b): low bias + low precision = low accuracy; (c): high bias + high precision = low accuracy; and (d): low bias + high precision = high accuracy.

#### 2.1.4 Estimation and Hypothesis Testing

Rather than presenting every observation collected, the data analyst usually summarizes major characteristics with a few descriptive statistics. Descriptive statistics include any characteristic designed to summarize an important feature of a data set. A point estimate is a single number that represents the descriptive statistic. Statistics common to implementation monitoring include proportions, means, medians, totals, and others. When estimating parameters of a population, such as the proportion or mean, it is useful to estimate the *confidence interval*.

The confidence interval indicates the range in which the true value lies for a stated confidence level. For example, if it is estimated that 65 percent of soybeans were planted using no-till and the 90 percent confidence limit is  $\pm 5$  percent, there is a 90 percent chance that between 60 and 70 percent of the soybeans were planted using no-till.

Hypothesis testing should be used to determine whether the level of MM and BMP implementation has changed over time. The *null hypothesis* ( $H_0$ ) is the root of hypothesis testing. Traditionally,  $H_0$  is a statement of no change, no effect, or no difference; for example, “the proportion of farm owners using IPM after the cost-share program is equal to the proportion of farm owners using IPM before the cost-share program.” The *alternative hypothesis* ( $H_a$ ) is counter to  $H_0$ , traditionally being a statement of change, effect, or difference. If  $H_0$  is rejected,  $H_a$  is accepted. Regardless of the statistical test selected for analyzing the data, the analyst must select the *significance level* ( $\alpha$ ) of the test. That is, the analyst must determine what error level is acceptable. There are two types of errors in hypothesis testing:

Type I:  $H_0$  is rejected when  $H_0$  is really true.

Type II:  $H_0$  is accepted when  $H_0$  is really false.

Table 2-2 depicts these errors, with the magnitude of Type I errors represented by  $\alpha$  and the magnitude of Type II errors represented by  $\beta$ . The probability of making a Type I error is equal to the  $\alpha$  of the test and is

selected by the data analyst. In most cases, managers or analysts will define  $1 - \alpha$  to be in the range of 0.90 to 0.99 (e.g., a confidence level of 90 to 99 percent), although there have been applications where  $1 - \alpha$  has been set to as low as 0.80. Selecting a 95 percent confidence level implies that the analyst will reject the  $H_0$  when  $H_0$  is true (i.e., a false positive) 5 percent of the time. The same notion applies to the confidence interval for point estimates described above:  $\alpha$  is set to 0.10, and there is a 10 percent chance that the true percentage of soybeans planted using no-till is outside the 60 to 70 percent range. This implies that if the decisions to be made based on the analysis are major (i.e., affect many people in adverse or costly ways) the confidence level needs to be greater. For less significant decisions (i.e., low-cost ramifications) the confidence level can be lower.

Type II error depends on the significance level, sample size, and variability, and which alternative hypothesis is true. *Power* ( $1 - \beta$ ) is defined as the probability of correctly rejecting  $H_0$  when  $H_0$  is false. In general, for a fixed sample size,  $\alpha$  and  $\beta$  vary inversely. For a fixed  $\alpha$ ,  $\beta$  can be reduced by increasing the sample size (Remington and Schork, 1970).

Table 2-2. Errors in hypothesis testing.

Decision	State of Affairs in the Population	
	$H_0$ is True	$H_0$ is False
Accept $H_0$	$1 - \alpha$ (Confidence level)	$\beta$ (Type II error)
Reject $H_0$	$\alpha$ (Significance level) (Type I error)	$1 - \beta$ (Power)

## 2.2 SAMPLING CONSIDERATIONS

In a document of this brevity, it is not possible to address all of the issues that face technical staff who are responsible for developing and implementing studies to track and evaluate the implementation of nonpoint source control measures. For example, when is the best time to implement a survey or do on-site visits? In reality, it is difficult to pinpoint a single time of the year. Some BMPs can be checked any time of the year, whereas others have a small window of opportunity. In northern areas, the time between fall harvest and winter snows might be the most effective time of year to assess implementation of a large number of erosion control practices.

If the goal of the study is to determine the effectiveness of a farmer education program, sampling should be timed to ensure that there was sufficient time for outreach activities and for the farmers to implement the desired practices. Also, farmers are more receptive to visits and participation in a survey during off-peak business times (i.e., not during planting, harvesting, livestock birthing, etc.). Furthermore, field personnel must have permission to perform site visits from each affected farm owner or manager prior to arriving at the farms. Where access is denied, a replacement farm is needed. This farm is selected in accordance with the type of farm selection being used, i.e., simple random, stratified random, cluster, or systematic.

From a study design perspective, all of these issues—study objectives, sampling strategy, allowable error, and formulation of hypotheses—must be considered together with determining the sampling strategy. This section describes common issues that the technical staff might consider in targeting their

sampling efforts or determining whether to stratify their sampling efforts. In general, if there is reason to believe that there are different rates of BMP or MM implementation in different groups, stratified random sampling should increase overall accuracy. Following the discussion, a list of resources that can be used to facilitate evaluating these issues is presented.

### 2.2.1 Farm Ownership and Size

Farm ownership can be divided (i.e., stratified) into multiple categories for sampling purposes depending on the MM implementation being tracked. The *1992 Census of Agriculture* (USDOC, 1994) provides information by state on:

- Farms by type of ownership (individual or family, partnership, corporation, and other).
- Farms owned versus rented or leased.
- Farm owner characteristics.
- Farm gross income.
- Average farm size.

- Number of farms by size (1 to 9, 10 to 49, 50 to 179, 180 to 499, 500 to 999, 1,000 to 1,999, and 2,000 acres or more).

The Economic Research Section of the U.S. Department of Agriculture (USDA) also provides information on farm ownership, as do many state programs. For example, a sampling plan to determine the percentage of acres on which erosion control practices had been implemented could be designed based on the data shown in Table 2-3. (The units of interest are acres of harvested cropland.) However, it should be noted that if there is reason to believe that implementation of erosion control practices is not uniform among farm owners of farms of differing sizes, more intense sampling of one or more subpopulations (strata) might be warranted.

2.2.2 Location and Other Physical Characteristics

Selection of farms for sampling should ensure a representative sample of all appropriate areas of a state or coastal zone. Stratifying by county, watershed, hydrologic unit, or any other geographically or physically based area might increase overall accuracy. Other important considerations for selecting areas from which to sample include:

- Areas with different soil types.
- Areas with different erosion potentials (see USDA’s *National Resources Inventory*).
- Areas with different climates (i.e., differences in total rainfall or storm frequency).
- Areas with known degraded water quality conditions.

2.2.3 Farm Type and Agricultural Practices

To obtain a representative sample, data must first be collected on the types of agricultural

Table 2-3. Acres of harvested cropland in Virginia from USDOC’s 1992 Census of Agriculture.

Total Farm Size (acres)	Number of Farms	Harvested Cropland (acres)
1 to 49 acres	9,802	88,488
50 to 99 acres	7,690	158,089
100 to 500 acres	16,125	965,178
500 to 999 acres	2,515	551,639
1,000 to 2,999 acres	943	428,572
2,000 acres or more	257	215,010
<b>Total</b>	<b>37,332</b>	<b>2,406,976</b>

practices that occur in a designated sampling area. Once farms have been stratified by the types of MMs they should be implementing, farms can be selected for sampling. For example, if grazing management were the only practice being evaluated, farms with only cropland would be removed from the sample population. Alternatively, if the investigator is interested in agriculture practices that affect the delivery of nitrogen to surface waters, only farms where MMs or BMPs that affect nitrogen movement are being implemented would be selected. Numerous sources of information can be used to infer the sample population. These sources should be consulted before designing a monitoring plan. The U.S. Department of Commerce's (USDOC) *Census of Agriculture* provides information by state on:

- Acres of harvested cropland.
- Acres of irrigated cropland.
- Types of livestock (cattle, milk cows, hogs and pigs, chickens, etc.).
- Types of crops (corn, wheat, tobacco, soybeans, peanuts, hay, land in orchards, etc.).

USDA's *National Resources Inventory* provides statistical information by U.S. Geological Survey (USGS) cataloging unit on the acreage of different crop types and other land uses.

#### 2.2.4 Sources of Information

For a truly random selection of population units, it is necessary to access or develop a database that includes the entire target population. The *Census of Agriculture*

(USDOC, 1994) is a good source, but it is limited to some extent by confidentiality constraints. (Certain data are not included, except at the state level, for counties that have only a few operations or are dominated by a single operation.) Other currently available national data bases generally include only agricultural entities that participate in cost-share programs. A more inclusive source presently available is county land maps. These maps, however, generally lack data regarding the specific type of farm operation and therefore do not provide the information needed to perform simple random site selection.

The following are possible sources of information on farms, which can be used for identifying potential monitoring farms and obtaining other information for farm selection. Positive and negative attributes of each information source are included.

**1992 National Resource Inventory (USDA, 1994a):** The National Resource Inventory (NRI) is a data base composed of data on the natural resources on the nonfederal lands of the United States) 74 percent of the Nation's land area. Its focus is on the soil, water, and related resources of farms, nonfederal forests, and grazing lands. The data were collected from more than 800,000 sample sites nationwide and are statistically reliable for analysis at the national, regional, state, major land resource area, or multiple county level, though not at the county level. Data elements include land cover/use (cropland, pasture land, rangeland and its condition, forest land, barren land, rural land, urban, and built-up areas), land ownership, soil information, irrigation, water bodies, conservation practices, and cropping history. Data are available on CD-ROMs and can be integrated with other data

through spatial linkages in a geographic information system (GIS). To obtain the NRI data base, contact: NRCS National Cartography and Geospatial Center, Fort Worth Federal Center, Building 23, Room 60, P.O. Box 6567, Fort Worth, TX 76115-0567; 1-800-672-5559; <http://www.ncg.nrcs.usda.gov>.

**Census of Agriculture (USDOC, 1994):** The Census of Agriculture is the leading source of statistics about the Nation's agricultural production and the only source for consistent, comparable data at the county, state, and national levels. Data are collected on a 5-year cycle in years ending in "2" and "7" and are available on computer tapes and CD-ROMs. Data elements include farms (number and size), harvested cropland, irrigated land, market value of products, farm ownership, livestock and poultry, selected crops harvested, and more. The Census of Agriculture has been transferred to the National Agricultural Statistics Service (NASS), who funded the 1997 census. Information on obtaining the Census of Agriculture is available on the Internet at <http://www.census.gov>.

**USDA Farm Numbers:** USDA farm numbers are developed when a farmer receives any financial assistance from a USDA organization. Only farms participating in USDA programs are included in the data base.

**USGS Land Use and Land Cover (USGS, 1990):** Using these data, at a level 2 definition, provides information on four categories of agricultural land uses: (1) cropland and pasture; (2) orchards, groves, vineyards, nurseries, and ornamental horticulture areas; (3) confined feeding operations; and (4) other agricultural land.

Watershed, topography, soil types, and/or political boundary maps could be used in conjunction with this land use information. Information on obtaining land use and land cover maps is available on the Internet at <http://www.usgs.gov> or at <http://www.ncg.nrcs.usda.gov>.

**County Land Maps:** These maps can provide information on farm owners or managers and possibly land use. Selection of farms to determine the type of operations occurring would have to be made randomly.

**State Cooperative Extension Service:** Farms that received Extension Service grants or participated in Coop programs are included. These programs vary from state to state. As with the USDA farm numbers, nonparticipatory farms are not included, which could result in biased sampling.

**Complaint Records:** Complaint records could be used in combination with other sources. Such records represent farms that have had problems in the past, which will very likely skew the data set.

**National Agriculture Statistics Service (NASS):** This agency, a branch of the USDA, issues reports related to national forecasts and estimates of crops, livestock, poultry, dairy, prices, labor, and related agricultural items (USDA, undated). The agency has the most comprehensive national list of farms available. NASS could produce random lists of farmers through one of its two frames. The first frame is an area frame, which randomly selects land segments that average 1 square mile in size. In most states the area frame is stratified into four broad categories based on land use: (1) areas intensively cultivated for crops, (2) extensive areas used primarily for grazing and

producing livestock, (3) residential and business land in cities and towns, and (4) nonagricultural lands such as parks and military complexes. The second frame is the list frame, which consists of names and addresses of producers grouped by size and type of unit. In a list frame sample names are selected randomly (based on whatever stratification is desired) and mailed questionnaires. Phone calls or visits are made to those farmers who do not respond by mail. A disadvantage of NASS is that it does not release names to other agencies. If this method of selection were chosen, NASS would have to perform the sampling. Information on obtaining data from NASS is available on the Internet at <http://www.usda.gov/nass> or through the NASS hotline at 1-800-727-9540.

**Computer-aided Management Practices System (CAMPS):** This data base has records of all nutrient management plans developed by the USDA Natural Resource Conservation Service (formerly the Soil Conservation Service, or SCS).

**Field Office Computing System (FOCS):** The Field Office Computing System (FOCS) replaced CAMPS, and full conversion from CAMPS to FOCS was completed in all field offices of the Natural Resources Conservation Service by January 1996. The system contains information on client businesses, resource inventories, conservation plans, practice cost comparisons, and a variety of specialty applications. Some of these applications are SOILS, with county-level soils data; PLANTS, with state-level plant data; GLA (Grazing Land Applications), with forage, herd, grazing schedule, and feedstuff data; WEQ (Wind Erosion Equation), a tool to compute wind erosion; Crop Rotation Detail,

which includes planting, harvest, and tillage data; RUSLE (Revised Universal Soil Loss Equation), a tool to compute sheet/rill erosion; Nutrient Screening Tool, a tool for evaluating nitrogen and phosphorus leaching and surface runoff; Pesticide Screening Tool, a tool for evaluating potential for pesticide leaching and runoff; and Farm\*A\*Syst, software for evaluating the potential for surface and groundwater pollution. Information on FOCS is available through the Internet at <http://www.itc.nrcs.usda.gov/fchd/focs>.

**Farm Service Agency (FSA):** The Farm Service Agency (FSA), created when the Department of Agriculture reorganized in 1994, incorporates programs from the Agricultural Stabilization and Conservation Service (ASCS), the Federal Crop Insurance Corporation, and the Farmers Home Administration. FSA administers programs for commodity loans, commodity purchases, crop insurance, emergency and disaster relief, farm ownership and operation loans, and farmland conservation. The Conservation Reserve Program assists farmers in conserving and improving soil, water, and wildlife resources on farmland by converting highly erodible and other environmentally sensitive acreage from production to long-term cover. FSA also maintains a collection of aerial photographs of farmlands. Information on FSA can be obtained through the Internet at <http://www.fsa.usda.gov>, or at the following address: USDA FSA Public Affairs Staff, P.O. Box 2415, STOP 0506, Washington, DC, 20013, (202) 720-5237. For information on the collection of aerial photographs maintained by the agency, contact USDA FSA Aerial Photography Field Office, P.O. Box 30010, Salt Lake City, UT, 84130-0010, (801) 975-3503.

### 2.3 SAMPLE SIZE CALCULATIONS

This section describes methods for estimating sample sizes to compute point estimates such as proportions and means, as well as detecting changes with a given significance level.

Usually, several assumptions regarding data distribution, variability, and cost must be made to determine the sample size. Some assumptions might result in sample size estimates that are too high or too low.

Depending on the sampling cost and cost for not sampling enough data, it must be decided whether to make conservative or “best-value” assumptions. Because the cost of visiting any individual farm or group of farms is relatively constant, it is more economical to collect a few extra samples rather than realize you need to go back to collect additional data. In most cases, the analyst should probably consider evaluating a range of assumptions on the impact of sample size and overall program cost.

To maintain document brevity, some terms and definitions that will be used in the remainder of this chapter are summarized in Table 2-4. These terms are consistent with those in most introductory-level statistics texts, and more information can be found there. Those with some statistical training will note that some of these definitions include an additional term referred to as the *finite population correction term* ( $1-N$ ), where  $N$  is equal to  $n/N$ . In many applications, the number of population units in the sample population ( $N$ ) is large in comparison to the population units sampled ( $n$ ) and  $(1-N)$  can be ignored. However, depending on the number of units (farms for example) in a particular population,  $N$  can become quite small.  $N$  is determined by the definition of the sample population and the corresponding population

units. If  $N$  is greater than 0.1, the finite population correction factor should not be ignored (Cochran, 1977).

Applying any of the equations described in this section is difficult when no historical data set exists to quantify initial estimates of proportions, standard deviations, means, or coefficients of variation. To estimate these parameters, Cochran (1977) recommends four sources:

- Existing information on the same population or a similar population.
- A two-step sample. Use the first-step sampling results to estimate the needed factors, for best design, of the second step. Use data from both steps to

Table 2-4. Definitions used in sample size calculation equations.

N	=	total number of population units in sample population	$p = a/n$	$q = 1 - p$
n	=	number of samples		
$n_0$	=	preliminary estimate of sample size	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
a	=	number of successes	$s = \sqrt{s^2}$	$C_v = s/\bar{x}$
p	=	proportion of successes		
q	=	proportion of failures (1-p)		
$x_i$	=	$i^{\text{th}}$ observation of a sample		
$\bar{x}$	=	sample mean	$d =  \bar{x} - \mu $	$d_r = \frac{ \bar{x} - \mu }{\mu}$
$s^2$	=	sample variance		
s	=	sample standard deviation		
$N\bar{x}$	=	total amount	$s^2(\bar{x}) = \frac{s^2}{n}(1 - \phi)$	$s(\bar{x}) = \frac{s}{\sqrt{n}}(1 - \phi)^{0.5}$
$\mu$	=	population mean		
$\sigma^2$	=	population variance	$s(N\bar{x}) = \frac{Ns}{\sqrt{n}}(1 - \phi)^{0.5}$	$s(p) = \sqrt{\frac{pq}{n}}(1 - \phi)^{0.5}$
$\sigma$	=	population standard deviation		
$C_v$	=	coefficient of variation		
$s^2(\bar{x})$	=	variance of sample mean		
$\phi$	=	$n/N$ (unless otherwise stated in text)	$Z_\alpha$	= value corresponding to cumulative area of $1 - \alpha$ using the normal distribution (see Table A1).
$s(\bar{x})$	=	standard error (of sample mean)	$t_{\alpha,df}$	= value corresponding to cumulative area of $1 - \alpha$ using the student t distribution with df degrees of freedom (see Table A2).
$1 - \phi$	=	finite population correction factor		
d	=	allowable error		
$d_r$	=	relative error		

estimate the final precision of the characteristic(s) sampled.

- A “pilot study” on a “convenient” or “meaningful” subsample. Use the results to estimate the needed factors. Here the results of the pilot study generally cannot be used in the calculation of the final precision because often the pilot sample is not representative of the entire population to be sampled.

- Informed judgment, or an educated guess.

It is important to note that this document only addresses estimating sample sizes with traditional parametric procedures. The methods described in this document should be appropriate in most cases, considering the type of data expected. If the data to be sampled are skewed, as with much water quality data, the investigator should plan to transform the data to something symmetric, if not normal, before computing sample sizes (Helsel and Hirsch,

1995). Kupper and Hafner (1989) also note that some of these equations tend to underestimate the necessary sample because power is not taken into consideration. Again, EPA recommends that if you do not have a background in statistics, you should consult with a trained statistician to be certain that your approach, design, and assumptions are appropriate to the task at hand.

### 2.3.1 Simple Random Sampling

In simple random sampling, we presume that the sample population is relatively homogeneous and we would not expect a difference in sampling costs or variability. If the cost or variability of any group within the sample population were different, it might be

*What sample size is necessary to estimate the proportion of farms implementing IPM to within ±5 percent?*

*What sample size is necessary to estimate the proportion of farms implementing IPM so that the relative error is less than 5 percent?*

more appropriate to consider a stratified random sampling approach.

To estimate the proportion of farms implementing a certain BMP or MM, such that the allowable error,  $d$ , meets the study precision requirements (i.e., the true proportion lies between  $p-d$  and  $p+d$  with a  $1-\alpha$  confidence level), a preliminary estimate of sample size can be computed as (Snedecor and Cochran, 1980)

$$n_o = \frac{(Z_{1-\alpha/2})^2 p q}{d^2} \tag{2-1}$$

If the proportion is expected to be a low number, using a constant allowable error might not be appropriate. Ten percent plus/minus 5 percent has a 50 percent relative error. Alternatively, the relative error,  $d_r$ , can be specified (i.e., the true proportion lies between  $p-d_r p$  and  $p+d_r p$  with a  $1-\alpha$  confidence level) and a preliminary estimate of sample size can be computed as (Snedecor and Cochran, 1980)

$$n_o = \frac{(Z_{1-\alpha/2})^2 q}{d_r^2 p} \tag{2-2}$$

In both equations, the analyst must make an initial estimate of  $p$  before starting the study. In the first equation, a conservative sample size can be computed by assuming  $p$  equal to 0.5. In the second equation the sample size gets larger as  $p$  approaches 0 for constant  $d_r$ , thus an informed initial estimate of  $p$  is needed. Values of  $\alpha$  typically range from 0.01 to 0.10. The final sample size is then estimated as (Snedecor and Cochran, 1980)

$$n = \begin{cases} \frac{n_o}{1 + \phi} & \text{for } \phi > 0.1 \\ n_o & \text{otherwise} \end{cases} \tag{2-3}$$

where  $N$  is equal to  $n_o/N$ . Table 2-5 demonstrates the impact on  $n$  of selecting  $p$ ,  $\alpha$ ,  $d_r$ , and  $N$ . For example, 278 random samples are needed to estimate the proportion

Table 2-5. Comparison of sample size as a function of  $p$ ,  $\alpha$ ,  $d$ ,  $d_r$ , and  $N$  for estimating proportions using equations 2-1 through 2-3.

Probability of Success, $p$	Significance level, $\alpha$	Allowable error, $d$	Relative error, $d_r$	Preliminary sample size, $n_p$	Sample Size, $n$				
					Number of Population Units in Sample Population, $N$				
					500	750	1,000	2,000	Large $N$
0.1	0.05	0.050	0.500	138	108	117	121	138	138
0.1	0.05	0.075	0.750	61	55	61	61	61	61
0.5	0.05	0.050	0.100	384	217	254	278	322	384
0.5	0.05	0.075	0.150	171	127	139	146	171	171
0.1	0.10	0.050	0.500	97	82	86	97	97	97
0.1	0.10	0.075	0.750	43	43	43	43	43	43
0.5	0.10	0.050	0.100	271	176	199	213	238	271
0.5	0.10	0.075	0.150	120	97	104	107	120	120

of 1,000 farmers using IPM to within  $\pm 5$  percent ( $d=0.05$ ) with a 95 percent confidence level assuming roughly one-half of farmers are using IPM.

*What sample size is necessary to estimate the average number of acres per farm that are under conservation tillage to within  $\pm 25$  acres?*

*What sample size is necessary to estimate the average number of acres per farm that are under conservation tillage to within  $\pm 10$  percent?*

Suppose the goal is to estimate the average acreage per farm where conservation tillage is used. The number of random samples

required to achieve a desired margin of error when estimating the mean (i.e., the true mean lies between  $\bar{x}-d$  and  $\bar{x}+d$  with a  $1-\alpha$  confidence level) is (Gilbert, 1987)

$$n = \frac{(t_{1-\alpha/2, n-1} s/d)^2}{1 + (t_{1-\alpha/2, n-1} s/d)^2/N} \tag{2-4}$$

If  $N$  is large, the above equation can be simplified to

$$n = (t_{1-\alpha/2, n-1} s/d)^2 \tag{2-5}$$

Since the Student's  $t$  value is a function of  $n$ , Equations 2-4 and 2-5 are applied iteratively. That is, guess at what  $n$  will be, look up  $t_{1-\alpha/2, n-1}$  from Table A2, and compute a revised  $n$ . If the initial guess of  $n$  and the revised  $n$  are different, use the revised  $n$  as the new

guess, and repeat the process until the computed value of  $n$  converges with the guessed value. If the population standard deviation is known (not too likely), rather than estimated, the above equation can be further simplified to:

$$n = (Z_{1-\alpha/2}\sigma/d)^2 \quad (2-6)$$

To keep the relative error of the mean estimate below a certain level (i.e., the true mean lies between  $\bar{x}-d_r$  and  $\bar{x}+d_r$  with a  $1-\alpha$  confidence level), the sample size can be computed with (Gilbert, 1987)

$$n = \frac{(t_{1-\alpha/2, n-1} C_v/d_r)^2}{1 + (t_{1-\alpha/2, n-1} C_v/d_r)^2/N} \quad (2-7)$$

$C_v$  is usually less variable from study to study than are estimates of the standard deviation, which are used in Equations 2-4 through 2-6. Professional judgment and experience, typically based on previous studies, are required to estimate  $C_v$ . Had  $C_v$  been known,  $Z_{1-\alpha/2}$  would have been used in place of  $t_{1-\alpha/2, n-1}$  in Equation 2-7. If  $N$  is large, Equation 2-7 simplifies to:

$$n = (t_{1-\alpha/2, n-1} C_v/d_r)^2 \quad (2-8)$$

For County X, farms range in size from 20 to 4,325 acres although most are less than 500 acres in size. The goal of the sampling program is to estimate the average number of cropland acres using minimum tillage. However, the investigator is concerned about skewing the mean estimate with the few large farms. As a result, the sample population for this analysis is the 430 cropland farms with less than 500 total acres of cropland. The investigator also wants to keep the relative

error under 15 percent (i.e.,  $d_r < 0.15$ ) with a 90 percent confidence level.

Unfortunately, this is the first study that County X has done and there is no information about the coefficient of variation,  $C_v$ . The investigator, however, is familiar with a recent study done by another company. Based on that study, the investigator estimates the  $C_v$  as 0.6 and  $s$  equal to 30. As a first-cut approximation, Equation 2-6 is applied with  $Z_{1-\alpha/2}$  equal to 1.645 and assuming  $N$  is large:

$$\begin{aligned} n &= (1.645 * 0.6 / 0.15)^2 \\ &= 43.3 \approx 44 \text{ samples} \end{aligned}$$

Since  $n/N$  is greater than 0.1 and  $C_v$  is estimated (i.e., not known), it is best to reestimate  $n$  with Equation 2-7 using 44 samples as the initial guess of  $n$ . In this case,  $t_{1-\alpha/2, n-1}$  is obtained from Table A2 as 1.6811.

$$\begin{aligned} n &= \frac{(1.6811 \times 0.6 / 0.15)^2}{1 + (1.6811 \times 0.6 / 0.15)^2 / 430} \\ &= 40.9 \approx 41 \text{ samples} \end{aligned}$$

Notice that the revised sample is somewhat smaller than the initial guess of  $n$ . In this case it is recommended to reapply Equation 2-7 using 41 samples as the revised guess of  $n$ . In this case,  $t_{1-\alpha/2, n-1}$  is obtained from Table A2 as 1.6839.

$$\begin{aligned} n &= \frac{(1.6839 \times 0.6 / 0.15)^2}{1 + (1.6839 \times 0.6 / 0.15)^2 / 430} \\ &= 41.0 \approx 41 \text{ samples} \end{aligned}$$

Since the revised sample size matches the estimated sample size on which  $t_{1-\alpha/2, n-1}$  was based, no further iterations are necessary. The proposed study should include 41 farms randomly selected from the 430 cropland farms with less than 500 total acres of cropland in County X.

*What sample size is necessary to determine whether there is a 20 percent difference in BMP implementation before and after a cost-share program?*

*What sample size is necessary to detect a 30-acre increase in average conservation tillage acreage per farm between farm owners that own versus rent land?*

When interest is focused on whether the level of BMP implementation has changed, it is necessary to estimate the extent of implementation at two different time periods. Alternatively, the proportion from two different populations can be compared. In either case, two independent random samples are taken and a hypothesis test is used to

determine whether there has been a significant change in implementation. (See Snedecor and Cochran (1980) for sample size calculations for matched data.) Consider an example in which the proportion of highly erodible land under conservation tillage will be estimated at two time periods. What sample size is needed?

To compute sample sizes for comparing two proportions,  $p_1$  and  $p_2$ , it is necessary to provide a best estimate for  $p_1$  and  $p_2$ , as well as specifying the significance level and power (1- $\beta$ ). Recall that power is equal to the probability of rejecting  $H_0$  when  $H_0$  is false. Given this information, the analyst substitutes these values into (Snedecor and Cochran, 1980)

$$n_o = (Z_\alpha + Z_{2\beta})^2 \frac{(p_1q_1 + p_2q_2)}{(p_2 - p_1)^2} \quad (2-9)$$

where  $Z_\alpha$  and  $Z_{2\beta}$  correspond to the normal deviate. Although this equation assumes that  $N$  large, it is acceptable for practical use (Snedecor and Cochran, 1980). Common values of  $(Z_\alpha \text{ and } Z_{2\beta})^2$  are summarized in Table 2-6. To account for  $p_1$  and  $p_2$  being

Table 2-6. Common values of  $(Z_\alpha + Z_{2\beta})^2$  for estimating sample size for use with equations 2-9 and 2-10.

Power, 1- $\beta$	$\alpha$ for One-sided Test			$\alpha$ for Two-sided Test		
	0.01	0.05	0.10	0.01	0.05	0.10
0.80	10.04	6.18	4.51	11.68	7.85	6.18
0.85	11.31	7.19	5.37	13.05	8.98	7.19
0.90	13.02	8.56	6.57	14.88	10.51	8.56
0.95	15.77	10.82	8.56	17.81	12.99	10.82
0.99	21.65	15.77	13.02	24.03	18.37	15.77

estimated,  $Z$  could be substituted with  $t$ . In lieu of an iterative calculation, Snedecor and Cochran (1980) propose the following approach: (1) compute  $n_o$  using Equation 2-9; (2) round  $n_o$  up to the next highest integer,  $f$ ; and (3) multiply  $n_o$  by  $(f+3)/(f+1)$  to derive the final estimate of  $n$ .

To detect a difference in proportions of 0.20 with a two-sided test,  $\alpha$  equal to 0.05,  $1-\beta$  equal to 0.90, and an estimate of  $p_1$  and  $p_2$  equal to 0.4 and 0.6,  $n_o$  is computed as

$$\begin{aligned} n_o &= 10.51 \frac{[(0.4)(0.6) + (0.6)(0.4)]}{(0.6 - 0.4)^2} \\ &= 126.1 \end{aligned}$$

Rounding 126.1 to the next highest integer,  $f$  is equal to 127, and  $n$  is computed as  $126.1 \times 130/128$  or 128.1. Therefore 129 samples in each random sample, or 258 total samples, are needed to detect a difference in proportions of 0.2. Beware of other sources of information that give significantly lower estimates of sample size. In some cases the other sources do not specify  $1-\beta$ ; otherwise, be sure that an “apples-to-apples” comparison is being made.

To compare the average from two random samples to detect a change of  $\delta$  (i.e.,  $\bar{x}_2 - \bar{x}_1$ ), the following equation is used:

$$n_o = (Z_\alpha + Z_{2\beta})^2 \frac{(s_1^2 + s_2^2)}{\delta^2} \quad (2-10)$$

Common values of  $(Z_\alpha \text{ and } Z_{2\beta})^2$  are summarized in Table 2-6. To account for  $s_1$  and  $s_2$  being estimated,  $Z$  should be replaced with  $t$ . In lieu of an iterative calculation, Snedecor and Cochran (1980) propose the following approach: (1) compute  $n_o$  using Equation 2-10; (2) round  $n_o$  up to the next highest integer,  $f$ ; and (3) multiply  $n_o$  by  $(f+3)/(f+1)$  to derive the final estimate of  $n$ .

Continuing the County X example above, where  $s$  was estimated as 75 acres, the investigator will also want to compare the average number of cropland acres using minimum tillage now to the average number of minimum tillage acres in a few years. To demonstrate success, the investigator believes that it will be necessary to detect a 50-acre increase. Although the standard deviation might change after the cost-share program, there is no particular reason to propose a different  $s$  after the cost-share program. To detect a difference of 50 acres with a two-sided test,  $\alpha$  equal to 0.05,  $1-\beta$  equal to 0.90, and an estimate of  $s_1$  and  $s_2$  equal to 75,  $n_o$  is computed as

$$\begin{aligned} n_o &= 10.51 \frac{(75^2 + 75^2)}{50^2} \\ &= 47.3 \end{aligned} \quad (2-11)$$

Rounding 47.3 to the next highest integer,  $f$  is equal to 48, and  $n$  is computed as  $47.3 \times 51/49$  or 49.2. Therefore 50 samples in each random sample, or 100 total samples, are needed to detect a difference of 50 acres.

### 2.3.2 Stratified Random Sampling

The key reason for selecting a stratified random sampling strategy over simple random sampling is to divide a heterogeneous population into more homogeneous groups. If populations are grouped based on size (e.g., farm size) when there is a large number of

*What sample size is necessary to estimate the average number of acres per farm that are under conservation tillage when there is a wide variety of farm sizes?*

small units and a few larger units, a large gain in precision can be expected (Snedecor and Cochran, 1980). Stratifying also allows the investigator to efficiently allocate sampling resources based on cost. The stratum mean,  $\bar{x}_h$ , is computed using the standard approach for estimating the mean. The overall mean,  $\bar{x}_{st}$ , is computed as

$$\bar{x}_{st} = \sum_{h=1}^L W_h \bar{x}_h \quad (2-12)$$

where  $L$  is the number of strata and  $W_h$  is the relative size of the  $h^{\text{th}}$  stratum.  $W_h$  can be computed as  $N_h/N$  where  $N_h$  and  $N$  are the number of population units in the  $h^{\text{th}}$  stratum and the total number of population units across all strata, respectively. Assuming that simple random sampling is used within each stratum, the variance of  $\bar{x}_{st}$  is estimated as (Gilbert, 1987)

$$s^2(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^L N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_h^2}{n_h} \quad (2-13)$$

where  $n_h$  is the number of samples in the  $h^{\text{th}}$  stratum and  $s_h^2$  is computed as (Gilbert, 1987)

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{h,i} - \bar{x}_h)^2 \quad (2-14)$$

There are several procedures for computing sample sizes. The method described below allocates samples based on stratum size, variability, and unit sampling cost. If  $s^2(\bar{x}_{st})$  is specified as  $V$  for a design goal,  $n$  can be obtained from (Gilbert, 1987)

$$n = \frac{\left( \sum_{h=1}^L W_h s_h \sqrt{c_h} \right)^2}{V + \frac{1}{N} \sum_{h=1}^L W_h s_h^2} \quad (2-15)$$

where  $c_h$  is the per unit sampling cost in the  $h^{\text{th}}$  stratum and  $n_h$  is estimated as (Gilbert, 1987)

$$n_h = n \frac{W_h s_h / \sqrt{c_h}}{\sum_{h=1}^L W_h s_h / \sqrt{c_h}} \quad (2-16)$$

In the discussion above, the goal is to estimate an overall mean. To apply a stratified random sampling approach to estimating proportions, substitute  $p_h$ ,  $p_{st}$ ,  $p_h q_h$ , and  $s^2(p_{st})$  for  $\bar{x}_h$ ,  $\bar{x}_{st}$ ,  $s_h^2$ , and  $s^2(\bar{x}_{st})$  in the above equations, respectively.

To demonstrate the above approach, consider the County X example again. In addition to the 430 farms that are less than 500 acres, there are 100 farms that range in size from 501 to 1,000 acres, 50 farms that range in size from 1,001 to 2,000 acres, and 20 farms that range in size from 2,001 to 4,500 acres. Table 2-7 presents three basic scenarios for estimating sample size. In the first scenario,  $s_h$  and  $c_h$  are assumed equal among all strata. Using a design goal of  $V$  equal to 100 and applying Equation 2-15 yields a total sample size of 51.4 or 52. Since  $s_h$  and  $c_h$  are uniform, these samples are allocated proportionally to  $W_h$ , which is referred to as *proportional allocation*. This allocation can be verified by comparing the percent sample allocation to  $W_h$ . Due to rounding up, a total of 53 samples are allocated.

Under the second scenario, referred to as the *Neyman allocation*, the variability between strata changes, but unit sample cost is constant. In this example,  $s_h$  increases by 50 between strata. Because of the increased variability in the last three strata, a total of

Table 2-7. Allocation of samples.

Farm Size (acres)	Number of Farms ( $N_h$ )	Relative Size ( $W_h$ )	Standard Deviation ( $s_h$ )	Unit Sample Cost ( $c_h$ )	Sample Allocation	
					Number	%
A) Proportional allocation ( $s_h$ and $c_h$ are constant)						
20-80	430	0.7167	30	1	31	70.5
81-200	100	0.1667	30	1	7	15.9
201-300	50	0.0833	30	1	4	9.1
301-400	20	0.0333	30	1	2	4.5
Using Equation 2-15, $n$ is equal to 41.9. Applying Equation 2-16 to each stratum yields a total of 44 samples after rounding up to the next integer.						
B) Neyman allocation ( $c_h$ is constant)						
20-80	430	0.7167	30	1	35	56.5
81-200	100	0.1667	45	1	13	21.0
201-300	50	0.0833	60	1	9	14.5
301-400	20	0.0333	75	1	5	8.1
Using Equation 2-15, $n$ is equal to 59.3. Applying Equation 2-16 to each stratum yields a total of 62 samples after rounding up to the next integer.						
C) Allocation where $s_h$ and $c_h$ are not constant						
20-80	430	0.7167	30	1.00	38	61.3
81-200	100	0.1667	45	1.25	12	19.4
201-300	50	0.0833	60	1.50	8	12.9
301-400	20	0.0333	75	2.00	4	6.5
Using Equation 2-15, $n$ is equal to 60.0. Applying Equation 2-16 to each stratum yields a total of 62 samples after rounding up to the next integer.						

79.1 or 80 samples are needed to meet the same design goal. So while more samples are taken in every strata, proportionally fewer samples are needed in the smaller farm size group. For example, using proportional allocation nearly 70 percent of the samples are

taken in the 20 to 500-acre farm size stratum, whereas approximately 54 percent of the samples are taken in the same stratum using the Neyman allocation.

Finally, introducing sample cost variation will also affect sample allocation. In the last scenario it was assumed that it is twice as expensive to evaluate a farm from the largest farm size stratum than to evaluate a farm from the smallest farm size stratum. In this example, roughly the same total number of samples are needed to meet the design goal, yet more samples are taken in the smaller size stratum.

### 2.3.3 Cluster Sampling

Cluster sampling is commonly used when there is a choice between the size of the sampling unit (e.g., fields versus farms). In general, it is cheaper to sample larger units than smaller units, but these results tend to be less accurate (Snedecor and Cochran, 1980). Thus, if there is not a unit sampling cost advantage to cluster sampling, it is probably better to use simple random sampling. To decide whether to perform a cluster sample, it will probably be necessary to perform a special investigation to quantify sampling errors and costs using the two approaches.

Perhaps the best approach to explaining the difference between simple random sampling and cluster sampling is to consider an example set of results. In this example, the investigator did a field evaluation of BMP implementation along a stream to evaluate whether recommended BMPs had been implemented and maintained. Since the watershed was quite large, the investigator elected to inspect 10 farms at each site. Table 2-8 presents the number of farms at each site that had implemented and maintained recommended BMPs. The overall mean is 5.6; a little more than one-half of the farms have implemented recommended BMPs. However, note that since the population unit corresponds to the 10

farms collectively, there are only 30 samples and the standard error for the proportion of farmers using recommended BMPs is 0.035. Had the investigator incorrectly calculated the standard error using the random sampling equations, he or she would have computed 0.0287, nearly a 20 percent error.

Since the standard error from the cluster sampling example is 0.035, it is possible to estimate the corresponding simple random sample size to get the same precision using

$$n = \frac{pq}{s(p)^2} = \frac{(0.56)(0.44)}{0.035^2} = 201 \quad (2-17)$$

Is collecting 300 samples using a cluster sampling approach cheaper than collecting about 200 simple random samples? If so, cluster sampling should be used; otherwise simple random sampling should be used.

### 2.3.4 Systematic Sampling

It might be necessary to obtain a baseline estimate of the proportion of farms where nutrient management practices have been implemented using a mailed questionnaire or phone survey. Assuming a record of farms in the state is available in a sequence unrelated to the manner in which nutrient management plans are implemented by individual farms (e.g., in alphabetical order by the farm owner's name), a systematic sample can be obtained in the following manner (Casley and

Table 2-8. Number of farms (out of 10) implementing recommended BMPs.

3	9	5	7	6	4	6	3	5	5
5	7	7	4	7	5	3	8	4	6
8	4	7	4	5	3	3	9	9	7
Grand Total = 168									
$\bar{x} = 5.6$		$p = 5.6/10=0.560$							
$s = 1.923$		$s = 1.923/10=0.1923$							
Standard error using cluster sampling: $s(p)=0.1923/(30)^{0.5}=0.035$									
Standard error if simple random sampling assumption had been incorrectly used: $s(p)=((0.56)(1-0.56)/300)^{0.5} =0.0287$									

Lury, 1982):

1. Select a random number  $r$  between 1 and  $n$ , where  $n$  is the number required in the sample.
2. The sampling units are then  $r, r + (N/n), r + (2N/n), \dots, r + (n-1)(N/n)$ , where  $N$  is total number of available records.

If the population units are in random order (e.g., no trends, no natural strata, uncorrelated), systematic sampling is, on average, equivalent to simple random sampling.

Once the sampling units (in this case, specific farms) have been selected, a questionnaire can be mailed to the farm owner or a telephone inquiry made about nutrient management practices being followed by the farm owner.

In another example, the Conservation Technology Information Center (CTIC), with the assistance of the Natural Resource Conservation Service (NRCS, formerly the

Soil Conservation Service), randomly selects approximately 3,100 sites for its annual *National Crop Residue Management Survey* (CTIC, 1994). A method for randomly selecting sites to fit local data needs was recently developed for assessing implementation of conservation tillage practices (CTIC, 1995). This method, the County Transect Survey, involves establishing a driving route that passes through all regions heavily used for crop production. Large urbanized areas and heavily traveled federal and state highways are avoided where possible. The direction of the route is not significant. In a recent application of the method in Illinois, the route was 110 miles long and included 456 cropland observation sites. Data were collected at set predetermined intervals. Data on rainfall, slope, soil erodibility, soil loss tolerance ( $T$ ), contouring, ephemeral erosion, and crop rotation/tillage system employed were also collected. Figure 2-6 presents the type of random route used in the survey. The county transect survey method has also been used successfully in Minnesota, Ohio, and Indiana

(CTIC, 1995), and is being considered for use in Pennsylvania.

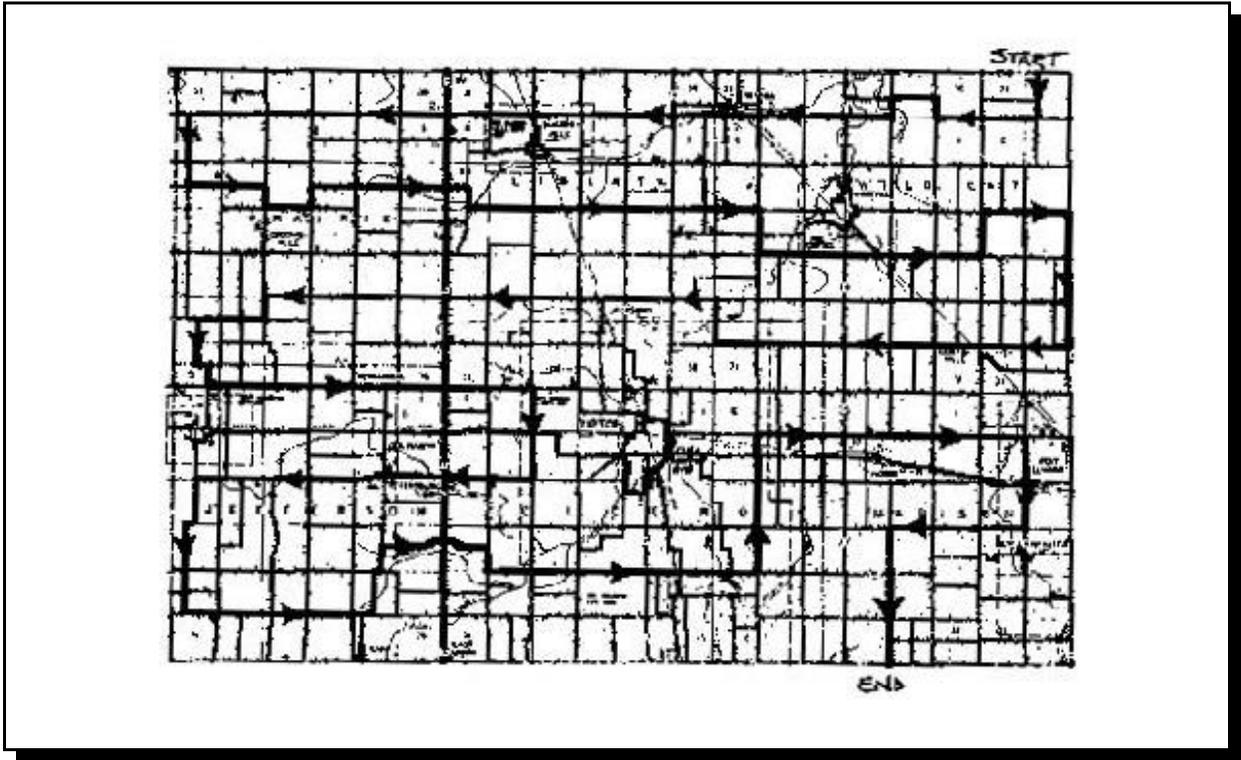


Figure 2-6. Example route for a county transect survey (CTIC, 1995).

## CHAPTER 3. METHODS FOR EVALUATING DATA

### 3.1 INTRODUCTION

Once data have been collected, it is necessary to statistically summarize and analyze the data. EPA recommends that the data analysis methods be selected before collecting the first sample. Many statistical methods have been computerized in easy-to-use software that is available for use on personal computers. Inclusion or exclusion in this section does not imply an endorsement or lack thereof by the U.S. Environmental Protection Agency. Commercial-off-the-shelf software that covers a wide range of statistical and graphical support includes SAS, Statistica, Statgraphics, Systat, Data Desk (Macintosh only), BMDP, and JMP. Numerous spreadsheets, database management packages, and other graphics software can also be used to perform many of the needed analyses. In addition, the following programs, written specifically for environmental analyses, are also available:

- SCOUT: A Data Analysis Program, EPA, NTIS Order Number PB93-505303.
- WQHYDRO (WATER QUALITY/HYDROLOGY GRAPHICS/ANALYSIS SYSTEM), Eric R. Aroner, Environmental Engineer, P.O. Box 18149, Portland, OR 97218.
- WQSTAT, Jim C. Loftis, Department of Chemical and Bioresource Engineering, Colorado State University, Fort Collins, CO 80524.

Computing the proportion of sites implementing a certain BMP or the average number of acres that are under a certain BMP

follows directly from the equations presented in Section 2.3 and is not repeated. The remainder of this section is focused on evaluating changes in BMP implementation. The methods provided in this section provide only a cursory overview of the type of analyses that might be of interest. For a more thorough discussion on these methods, the reader is referred to Gilbert (1987), Snedecor and Cochran (1980), and Helsel and Hirsch (1995). Typically the data collected for evaluating changes will typically come as two or more sets of random samples. In this case, the analyst will test for a shift or step change.

Depending on the objective, it is appropriate to select a one- or two-sided test. For example, if the analyst knows that BMP implementation will only go up as a result of a cost-share program, a one-sided test could be formulated. Alternatively, if the analyst does not know whether implementation will go up or down, a two-sided test is necessary. To simply compare two random samples to decide if they are significantly different, a two-sided test is used. Typical null hypotheses ( $H_0$ ) and alternative hypotheses ( $H_a$ ) for one- and two-sided tests are provided below:

#### One-sided test

$H_0$ : BMP Implementation(Post cost share)  $\leq$   
BMP Implementation(Pre cost share)

$H_a$ : BMP Implementation(Post cost share)  $>$   
BMP Implementation(Pre cost share)

#### Two-sided test

$H_0$ : BMP Implementation(Post cost share) =  
BMP Implementation(Pre cost share)

$H_a$ : BMP Implementation(Post cost share)  $\neq$   
BMP Implementation(Pre cost share)

Selecting a one-sided test instead of a two-sided test results in an increased power for the same significance level (Winer, 1971). That is, if the conditions are appropriate, a corresponding one-sided test is more desirable than a two-sided test given the same " and sample size. The manager and analyst should take great care in selecting one- or two-sided tests.

### 3.2 COMPARING THE MEANS FROM TWO INDEPENDENT RANDOM SAMPLES

The Student's  $t$  test for two samples and the Mann-Whitney test are the most appropriate tests for these types of data. Assuming the data meet the assumptions of the  $t$  test, the two-sample  $t$  statistic with  $n_1+n_2-2$  degrees of freedom is (Remington and Schork, 1970)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{3-1}$$

where  $n_1$  and  $n_2$  are the sample size of the first and second data set and  $\bar{x}_1$  and  $\bar{x}_2$  are the estimated means from the first and second data set, respectively. The pooled standard deviation,  $s_p$ , is defined by

$$s_p = \left[ \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2} \right]^{0.5} \tag{3-2}$$

where  $s_1^2$  and  $s_2^2$  correspond to the estimated variances of the first and second data set,

### Tests for Two Independent Random Samples

Test*	Key Assumptions
Two-sample $t$	<ul style="list-style-type: none"> <li>Both data sets must be normally distributed</li> <li>Data sets should have equal variances†</li> </ul>
Mann-Whitney	<ul style="list-style-type: none"> <li>None</li> </ul>

\* The standard forms of these tests require independent random samples.  
 † The variance homogeneity assumption can be relaxed.

respectively. The difference quantity ( $\Delta_0$ ) can be any value, but here it is set to zero.  $\Delta_0$  can be set to a non-zero value to test whether the difference between the two data sets is greater than a selected value. If the variances are not equal, refer to Snedecor and Cochran (1980) for methods for computing the  $t$  statistic. In a two-sided test, the value from Equation 3-1 is compared to the  $t$  value from Table A2 with  $n/2$  and  $n_1+n_2-2$  degrees of freedom.

The Mann-Whitney test can also be used to compare two independent random samples. This test is very flexible since there are no assumptions about the distribution of either sample or whether the distributions have to be the same (Helsel and Hirsch, 1995). Wilcoxon (1945) first introduced this test for equal-sized samples. Mann and Whitney (1947) modified the original Wilcoxon's test to apply it to different sample sizes. Here, it is determined whether one data set tends to have larger observations than the other.

If the distributions of the two samples are similar except for location (i.e., similar spread and skew),  $H_a$  can be refined to imply that the median concentration from one sample is

“greater than,” “less than,” or “not equal to” the median concentration from the second sample. To achieve this greater detail in  $H_a$ , transformations such as logs can be used.

Tables of Mann-Whitney test statistics (e.g., Conover, 1980) may be consulted to determine whether to reject  $H_0$  for small sample sizes. If  $n_1$  and  $n_2$  are greater than or equal to 10 observations, the test statistic can be computed from the following equation (Conover, 1980):

$$T_1 = \frac{T - n_1 \frac{n+1}{2}}{\sqrt{\frac{n_1 n_2}{n(n-1)} \sum_{i=1}^n R_i^2 - \frac{n_1 n_2 (n+1)^2}{4(n-1)}}} \quad (3-3)$$

where

- $n_1$  = number of observations in sample with fewer observations,
- $n_2$  = number of observations in sample with more observations,
- $n$  =  $n_1 + n_2$ ,
- $T$  = sum of ranks for sample with fewer observations, and
- $R_i$  = rank for the  $i$ th ordered observation used in both samples.

$T_1$  is normally distributed and Table A1 can be used to determine the appropriate quantile. Helsel and Hirsch (1995) and USEPA (1997) provide detailed examples for both of these tests.

### 3.3 COMPARING THE PROPORTIONS FROM TWO INDEPENDENT SAMPLES

Consider the example in which the proportion of highly erodible land under conservation tillage has been estimated during two time periods to be  $p_1$  and  $p_2$  using sample sizes of  $n_1$  and  $n_2$ , respectively. Assuming a normal

approximation is valid, the test statistic under a null hypothesis of equivalent proportions (no change) is

$$\frac{p_1 - p_2}{\sqrt{p(1-p) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (3-4)$$

where  $p$  is a pooled estimate of proportion and is equal to  $(x_1 + x_2) / (n_1 + n_2)$  and  $x_1$  and  $x_2$  are the number of successes during the two time periods. An estimator for the difference in proportions is simply  $p_1 - p_2$ .

In an earlier example, it was determined that 129 observations in each sample were needed to detect a difference in proportions of 0.20 with a two-sided test,  $\alpha$  equal to 0.05, and  $1 - \beta$  equal to 0.90. Assuming that 130 samples were taken and  $p_1$  and  $p_2$  were estimated from the data as 0.6 and 0.4, the test statistic would be estimated as

$$\frac{0.6 - 0.4}{\sqrt{0.5(0.5) \left( \frac{1}{130} + \frac{1}{130} \right)}} = 3.22 \quad (3-5)$$

Comparing this value to the  $t$  value from Table A2 ( $\alpha/2 = 0.025$ ,  $df=258$ ) of 1.96,  $H_0$  is rejected.

### 3.4 COMPARING MORE THAN TWO INDEPENDENT RANDOM SAMPLES

The analysis of variance (ANOVA) and Kruskal-Wallis are extensions of the two-sample  $t$  and Mann-Whitney tests, respectively, and can be used for analyzing more than two independent random samples when the data are continuous (e.g., mean acreage). Unlike the  $t$  test described earlier,

the ANOVA can have more than one factor or explanatory variable. The Kruskal-Wallis test accommodates only one factor, whereas the Friedman test can be used for two factors. In addition to applying one of the above tests to determine if one of the samples is significantly different from the others, it is also necessary to do postevaluations to determine which of the samples is different. This section recommends Tukey's method to analyze the raw or rank-transformed data only if one of the previous tests (ANOVA, rank-transformed ANOVA, Kruskal-Wallis, Friedman) indicates a significant difference between groups. Tukey's method can be used for equal or unequal sample sizes (Helsel and Hirsch, 1995). The reader is cautioned, when performing an ANOVA using standard software, to be sure that the ANOVA test used matches the data. See USEPA (1997) for a more detailed discussion on comparing more than two independent random samples.

### 3.5 COMPARING CATEGORICAL DATA

In comparing categorical data, it is important to distinguish between nominal categories (e.g., land ownership, county location, type of BMP) and ordinal categories (e.g., BMP implementation rankings, low-medium-high scales).

The starting point for all evaluations is the development of a contingency table. In Table 3-1, the preference of three BMPs is compared to operator type in a contingency table. In this case both categorical variables are nominal. In this example, 45 of the 102 operators that own the land they till used BMP<sub>1</sub>. There were a total of 174 observations.

To test for independence, the sum of the squared differences between the expected ( $E_{ij}$ )

and the observed ( $O_{ij}$ ) count summed over all cells is computed as (Helsel and Hirsch, 1995)

$$\chi_{ct} = \sum_{i=1}^m \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (3-6)$$

where  $E_{ij}$  is equal to  $A_i C_j / N$ .  $P_{ct}$  is compared to the  $1 - \alpha$  quantile of the  $\chi^2$  distribution with  $(m-1)(k-1)$  degrees of freedom (see Table A3).

In the example presented in Table 3-1, the symbols listed in the parentheses correspond to the above equation. Note that  $k$  corresponds to the three types of BMPs and  $m$  corresponds to the three different types of

Table 3-1. Contingency table of observed operator type and implemented BMP.

Operator Type	BMP <sub>1</sub>	BMP <sub>2</sub>	BMP <sub>3</sub>	Row Total, A <sub>i</sub>
Rent	10 (O <sub>11</sub> )	30 (O <sub>12</sub> )	17 (O <sub>13</sub> )	57 (A <sub>1</sub> )
Own	45 (O <sub>21</sub> )	32 (O <sub>22</sub> )	25 (O <sub>23</sub> )	102 (A <sub>2</sub> )
Combination	8 (O <sub>31</sub> )	3 (O <sub>32</sub> )	4 (O <sub>33</sub> )	15 (A <sub>3</sub> )
Column Total, C <sub>j</sub>	63 (C <sub>1</sub> )	65 (C <sub>2</sub> )	46 (C <sub>3</sub> )	174 (N)

Key to Symbols:

O<sub>ij</sub> = number of observations for the *i*th operator and *j*th BMP type

A<sub>i</sub> = row total for the *i*th operator type (total number of observations for a given operator type)

C<sub>j</sub> = column total for the *j*th BMP type (total number of observations for a given BMP type)

N = total number of observations

operators. Table 3-2 shows computed values of  $E_{ij}$  and  $(O_{ij}-E_{ij})^2/E_{ij}$  in parentheses for the example data.  $P_{ct}$  is equal to 14.60. From Table A3, the 0.95 quantile of the  $P^2$  distribution with 4 degrees of freedom is 9.488.  $H_0$  is rejected; the selection of BMP is not random among the different operator types. The largest values in the parentheses in Table 3-2 give an idea as to which combinations of operator type and BMP are noteworthy. In this example, it appears that BMP<sub>2</sub> is preferred to BMP<sub>1</sub> for those operators that rent the land they till.

Now consider that in addition to evaluating information regarding the operator and BMP type, we also recorded a value from 1 to 5 indicating how well the BMP was installed and maintained, with 5 indicating the best results. In this case, the BMP implementation rating is ordinal. Using the same notation as before, the average rank of observations in row *x*,  $R_x$ , is equal to (Helsel and Hirsch, 1995)

$$R_x = \sum_{i=1}^{x-1} A_i + (A_x + 1)/2 \quad (3-7)$$

where  $A_i$  corresponds to the row total. The average rank of observations in column *j*,  $D_j$ , is equal to

$$D_j = \frac{\sum_{i=1}^m O_{ij} R_i}{C_j} \quad (3-8)$$

where  $C_j$  corresponds to the column total. The Kruskal-Wallis test statistic is then computed as

$$K = (N-1) \frac{\sum_{j=1}^k C_j D_j^2 - N \left[ \frac{N+1}{N} \right]^2}{\sum_{i=1}^m A_i R_i^2 - N \left[ \frac{N+1}{N} \right]^2} \quad (3-9)$$

where  $K$  is compared to the  $P^2$  distribution with  $k-1$  degrees of freedom. This is the most general form of the Kruskal-Wallis test since it is a comparison of distribution shifts

Table 3-2. Contingency table of expected operator type and implemented BMP. (Values in parentheses correspond to  $(O_{ij}-E_{ij})^2/E_{ij}$ .)

Operator Type	BMP <sub>1</sub>	BMP <sub>2</sub>	BMP <sub>3</sub>	Row Total
Rent	20.64 (5.48)	21.29 (3.56)	15.07 (0.25)	57
Own	36.93 (1.76)	38.10 (0.98)	26.97 (0.14)	102
Combination	5.43 (1.22)	5.60 (1.21)	3.97 (0.00)	15
Column Total	63	65	46	174

rather than shifts in the median (Helsel and Hirsch, 1995).

Table 3-3 is a continuation of the previous example indicating the BMP implementation rating for each BMP type. For example, 29 of the 70 observations that were given a rating of 4 are associated with BMP<sub>2</sub>. The terms inside the parentheses of Table 3-3 correspond to the terms used in Equations 3-7 to 3-9. Note that k corresponds to the three types of BMPs and m corresponds to the five different levels of BMP implementation. Using Equation 3-9 for the data in Table 3-3, K is equal to 14.86. Comparing this value to 5.991 obtained from Table A3, there is a significant difference in the quality of implementation between the three BMPs.

The last type of categorical data evaluation considered in this chapter is when both variables are ordinal. The Kendall  $J_b$  for tied data can be used for this analysis. The statistic  $J_b$  is calculated as (Helsel and Hirsch, 1995)

$$\tau_b = \frac{S}{\frac{1}{2}\sqrt{(N^2 - SS_a)(N^2 - SS_b)}} \quad (3-10)$$

where S, SS<sub>a</sub>, and SS<sub>c</sub> are computed as

$$S = \sum_{all\ xy} \left[ \sum_{i>x} \sum_{j>y} o_{xy} o_{ij} - \sum_{i<x} \sum_{j<y} o_{xy} o_{ij} \right] \quad (3-11)$$

$$SS_a = \sum_{i=1}^m A_i^2 \quad (3-12)$$

$$SS_c = \sum_{j=1}^k C_j^2 \quad (3-13)$$

To determine whether  $J_b$  is significant, S is modified to a normal statistic using

$$Z_S = \begin{cases} \frac{S-1}{\sigma_S} & \text{if } S > 0 \\ \frac{S+1}{\sigma_S} & \text{if } S < 0 \end{cases} \quad (3-14)$$

Table 3-3. Contingency table of implemented BMP and rating of installation and maintenance.

BMP Implementation Rating				Row Total, $A_i$
	BMP <sub>1</sub>	BMP <sub>2</sub>	BMP <sub>3</sub>	
1	1 ( $O_{11}$ )	2 ( $O_{12}$ )	2 ( $O_{13}$ )	5 ( $A_1$ )
2	7 ( $O_{21}$ )	3 ( $O_{22}$ )	5 ( $O_{23}$ )	15 ( $A_2$ )
3	15 ( $O_{31}$ )	16 ( $O_{32}$ )	26 ( $O_{33}$ )	57 ( $A_3$ )
4	32 ( $O_{41}$ )	29 ( $O_{42}$ )	9 ( $O_{43}$ )	70 ( $A_4$ )
5	8 ( $O_{51}$ )	15 ( $O_{52}$ )	4 ( $O_{53}$ )	27 ( $A_5$ )
Column Total, $C_j$	63 ( $C_1$ )	65 ( $C_2$ )	46 ( $C_3$ )	174 ( $N$ )

Key to Symbols:

$O_{ij}$  = number of observations for the  $i$ th BMP implementation rating and  $j$ th BMP type

$A_i$  = row total for the  $i$ th BMP implementation rating (total number of observations for a given BMP implementation rating)

$C_j$  = column total for the  $j$ th BMP type (total number of observations for a given BMP type)

$N$  = total number of observations

where

$$\sigma_s = \sqrt{\frac{N^3}{9} \left( 1 - \sum_{i=1}^m a_i^3 \right) \left( 1 - \sum_{j=1}^k c_j^3 \right)} \quad (3-15)$$

where  $Z_s$  is zero if  $S$  is zero. The values of  $a_i$  and  $c_j$  are computed as  $A_i/N$  and  $C_j/N$ , respectively.

Table 3-4 presents the BMP implementation ratings that were taken in three separate years. For example, 15 of the 57 observations that were given a rating of 3 are associated with Year 2. Using Equations 3-11 and 3-15,  $S$  and  $F_s$  are equal to 2,509 and 679.75, respectively. Therefore,  $Z_s$  is equal to  $(2509-1)/679.75$  or 3.69. Comparing this value to a value of 1.96 obtained from Table A1 ( $\alpha/2=0.025$ ) indicates that BMP implementation is improving with time.

Table 3-4. Contingency table of implemented BMP and sample year.

BMP Implementation Rating				Row Total, $A_i$	$a_i$
	Year 1	Year 2	Year 3		
1	2 ( $O_{11}$ )	1 ( $O_{12}$ )	2 ( $O_{13}$ )	5 ( $A_1$ )	0.029
2	5 ( $O_{21}$ )	7 ( $O_{22}$ )	3 ( $O_{23}$ )	15 ( $A_2$ )	0.086
3	26 ( $O_{31}$ )	15 ( $O_{32}$ )	16 ( $O_{33}$ )	57 ( $A_3$ )	0.328
4	9 ( $O_{41}$ )	32 ( $O_{42}$ )	29 ( $O_{43}$ )	70 ( $A_4$ )	0.402
5	4 ( $O_{51}$ )	8 ( $O_{52}$ )	15 ( $O_{53}$ )	27 ( $A_5$ )	0.155
Column Total, $C_j$	46 ( $C_1$ )	63 ( $C_2$ )	65 ( $C_3$ )	174 (N)	
$c_j$	0.264	0.362	0.374		

## Key to Symbols:

 $O_{ij}$  = number of observations for the  $i$ th BMP implementation rating and  $j$ th year $A_i$  = row total for the  $i$ th BMP implementation rating (total number of observations for a given BMP implementation rating) $C_j$  = column total for the  $j$ th BMP type (total number of observations for a given year)

N = total number of observations

 $a_i = A_i/N$  $c_j = C_j/N$

## CHAPTER 4. CONDUCTING THE EVALUATION

### 4.1 INTRODUCTION

This chapter addresses the process of determining whether agricultural MMs or BMPs are being implemented and whether they are being implemented according to approved standards or specifications. Guidance is provided on what should be measured to assess MM and BMP implementation, as well as methods for collecting the information, including physical farm or field evaluations, mail- and/or telephone-based surveys, personal interviews, and aerial reconnaissance and photography. Designing survey instruments to avoid error and rating MM and BMP implementation are also discussed.

Evaluation methods are separated into two types: Expert evaluations and self-evaluations. Expert evaluations are those in which actual field investigations are conducted by trained personnel to gather information on MM or BMP implementation. Self-evaluations are those in which answers to a predesigned questionnaire or survey are provided by the person being surveyed, usually a farm owner or manager. The answers provided are used as survey results. Self-evaluations might also include examination of materials related to a farm, such as applications for cost-share programs or crop histories. Extreme caution should be exercised when using data from self-evaluations as the basis for assessing MM or BMP compliance since they are not typically reliable for this purpose. Each of these evaluation methods has advantages and disadvantages that should be considered prior to deciding which one to use or in what

combination to use them. Aerial reconnaissance and photography can be used to support either evaluation method.

Self-evaluations are useful for collecting information on the level of awareness that farm owners or managers have of MMs or BMPs, dates of planting or harvest, field or crop conditions, which MMs or BMPs were implemented, and whether the assistance of a state or county agriculture professional was used. However, the type of or level of detail of information that can be obtained from self-evaluations might be inadequate to satisfy the objectives of a MM or BMP implementation survey. If this is the case, expert evaluations might be called for. Expert evaluations are necessary if the information on MM or BMP implementation that is required must be more detailed or more reliable than that that can be obtained with self-evaluations. Examples of information that would be obtained reliably only through an expert evaluation include an objective assessment of the adequacy of MM or BMP implementation, the degree to which site-specific factors (e.g., type of crop, soil type, or presence of a water body) influenced MM or BMP implementation, or the need for changes in standards and specifications for MM or BMP implementation. Sections 4.3 and 4.4 discuss expert evaluations and self-evaluations, respectively, in more detail.

Other important factors to consider when choosing variables include the time of year when the BMP compliance survey will be conducted and when BMPs were installed. Some agriculture BMPs, or aspects of their

implementation that can be analyzed vary with time of year and phase of farming operations. Variables that are appropriate to these factors should be chosen. The nutrient management and pesticide management MMs in particular might not lend themselves to direct on-site analysis except at specific times of year, such as during or soon after fertilizer and pesticide applications, respectively. Concerning BMPs that have been in place for some time, the adequacy of implementation might be of less interest than the adequacy of the operation and maintenance of the BMP. For example, it might be of interest to examine fences along streams for structural integrity (i.e., holes that would allow cattle to pass through) rather than to calculate the miles of stream along which the fences were installed. Similarly, waste storage structure might be inspected for the amount of freeboard when operating at capacity rather than analyzing their construction for adherence to construction specifications. If numerous BMPs are being analyzed during a single farm visit, variables that relate to different aspects of BMP installation, operation, and maintenance might be chosen separately for each BMP to be inspected.

Aerial reconnaissance and photography is another means available for collecting information on farming practices, though some of the MMs and BMPs employed for agriculture might be difficult if not impossible to identify on aerial photographs. Aerial reconnaissance and photography are discussed in detail in Section 4.5.

The general types of information obtainable with self-evaluations are listed in Table 4-1. Regardless of the approach(es) used, proper and thorough preparation for the evaluation is the key to success.

## 4.2 CHOICE OF VARIABLES

Once the objectives of a BMP implementation or compliance survey have been clearly defined, the most important factor in the assessment of MM or BMP implementation is the determination of which variable(s) to measure. A good variable provides a direct measure of how well a BMP was or is being implemented. Individual variables should provide measures of different factors related to BMP implementation. The best variables are those which are measures of the adequacy of MM or BMP implementation and are based on quantifiable expressions of conformance with state standards and specifications. As the variables that are used become less directly related to actual MM or BMP implementation, their accuracy as measures of BMP implementation decreases.

Examples of useful variables include the tons and percentage per day of animal manure captured and treated by wastewater facilities associated with confined animal facilities and the cattle-hours per day during which livestock are excluded from stream banks, both of which would be expressed in terms of conformance with applicable state standards and specifications. Less useful variables measure factors that are related to MM and BMP implementation but do not necessarily provide an accurate measure of their implementation. Examples of these types of variables are the number of manure storage facilities constructed in a year and the number of farms with approved pesticide management plans. Other poor variables would be the passage of legislation requiring MM or BMP application on farms, development of an information education program for nutrient management, or the number of requests for information on nutrient management. Although these

Table 4-1. General types of information obtainable with self-evaluations and expert evaluations.

### **Information Obtainable from Self Evaluations**

#### *Background Information*

- Type of facility installed (e.g., confined animal facility, wastewater storage and/or treatment facility)
- Capacity of facility
- Square feet of facilities
- Type and number of animals and/or crops on farm
- Cropping history
- Yield data and estimates
- Field limitations
- Pest problems on farm
- Soil test results
- Map

#### *Management Measures/Best Management Practices*

- Management measures used on farm
- BMPs installed
- Dates of MM/BMP installation
- Design specifications of BMPs
- Type of waterbody or area protected
- Previous management measures used

#### *Management Plans*

- Preparation of management plans (e.g., nutrient, grazing, pesticide, irrigation water)
- Dates of plan preparation and revisions
- Date of initial plan implementation
- Total acreage under management

#### *Equipment*

- Types of equipment used on farm
- Dates of equipment calibration
- Application rates
- Timing of applications
- Substances applied (e.g., pesticides, fertilizers)
- Ambient conditions during applications
- Location of mixing, loading, and storage areas

### **Information Requiring Expert Evaluations**

- Design sufficiency
- Installation sufficiency
- Adequacy of operation/management
- Confirmation of information from self evaluation

variables relate to MM or BMP implementation, they provide no real information on whether MMs or BMPs are actually being implemented or whether they are being implemented properly.

Variables generally will not directly relate to MM implementation, as most agriculture MMs are combinations of several BMPs. Measures of MM implementation, therefore, usually will be based on separate assessments of two or more BMPs, and the implementation of each BMP will be based on a unique set of variables. Some examples of BMPs related to the EPA's Grazing Management Measure, variables for assessing compliance with the BMPs, and related standards and specifications that might be required by state agriculture authorities are presented in Figure 4-1. Because farm owners and managers choose to implement or not implement MMs or BMPs based on site-specific conditions, it is also appropriate to apply varying weights to the variables chosen to assess MM and BMP implementation to correspond to site-specific conditions. For example, variables related to animal waste disposal practices might be de-emphasized—and other, more applicable variables emphasized more—on farms with relatively few animals. Similarly, on a farm with a water body, variables related to livestock access to the water body, sediment runoff, and chemical deposition (pesticide use, fertilizer use) might be emphasized over other variables to arrive at a site-specific rating of the adequacy of MM or BMP implementation.

The purpose for which the information collected during a MM or BMP implementation survey will be used is another important consideration when selecting variables. An implementation survey can serve many purposes beyond the primary

purpose of assessing MM and BMP implementation. For instance, variables might be selected to assess compliance with each category of BMP that is of interest and to assess overall compliance with BMP specification and standards. In addition, other variables might be selected to assess the effect that each has on the ability or willingness of farm owners or managers to comply with BMP implementation standards or specifications. The information obtained from evaluations using the latter type of variable could be useful for modifying MM or BMP implementation standards and specifications for application to particular farm types or conditions.

Table 4-2 provides examples of good and poor variables for the assessment of MM or BMP implementation of the agricultural MMs developed by EPA (USEPA, 1993a). The variables listed in the table are only examples, and local or regional conditions should ultimately dictate what variables should be used.

## GRAZING MANAGEMENT MEASURE

Protect range, pasture, and other grazing lands:

- (1) By implementing one or more of the following to protect sensitive areas (such as stream banks, wetlands, estuaries, ponds, lake shores, and riparian zones):
  - (a) Exclude livestock,
  - (b) Provide stream crossings or hardened watering access for drinking,
  - (c) Provide alternative drinking water locations,
  - (d) Locate salt and additional shade, if needed, away from sensitive areas, or
  - (e) Use improved grazing management (e.g., herding)
 to reduce the physical disturbance and reduce direct loading of animal waste and sediment caused by livestock; *and*
  
- (2) By achieving either of the following on all range, pasture, and other grazing lands not addressed under (1):
  - (a) Implement the range and pasture components of a Conservation Management System (CMS) as defined in the Field Office Technical Guide of the USDA-NRCS by applying the progressive planning approach of the USDA-NRCS to reduce erosion,
  - or*
  - (b) Maintain range, pasture, and other grazing lands in accordance with activity plans established by either the Bureau of Land Management of the U.S. Department of the Interior or the Forest Service of USDA.

*Related BMPs, measurement variables, and standards and specifications:*

Management Measure Practice	Potential Measurement Variables	Example Related Standards and Specifications
<ul style="list-style-type: none"> <li>• Postpone grazing or rest grazing land for a prescribed period</li> </ul>	<ul style="list-style-type: none"> <li>• Percent ground cover</li> <li>• Stubble height</li> </ul>	<ul style="list-style-type: none"> <li>• Recommended percent ground cover for grazing</li> <li>• Recommended stubble height for grazing</li> </ul>
<ul style="list-style-type: none"> <li>• Alternate water source installed to convey water away from riparian areas</li> </ul>	<ul style="list-style-type: none"> <li>• Presence of alternative water source</li> <li>• Distance from water body of water provided to livestock</li> </ul>	<ul style="list-style-type: none"> <li>• Guidelines for provision of alternative sources of water on farms with water bodies</li> </ul>
<ul style="list-style-type: none"> <li>• Livestock excluded from an area not intended for grazing</li> </ul>	<ul style="list-style-type: none"> <li>• Cattle-hours per day of exclusion of livestock from water bodies</li> </ul>	<ul style="list-style-type: none"> <li>• Guidelines for protection of water quality for specific types of water bodies.</li> </ul>
<ul style="list-style-type: none"> <li>• Range seeded to establish adapted plants on native grazing land</li> </ul>	<ul style="list-style-type: none"> <li>• Percent ground cover</li> <li>• Plant species</li> </ul>	<ul style="list-style-type: none"> <li>• Recommended amount of ground cover for grazing</li> <li>• Acceptable plant species for the region</li> </ul>

Figure 4-1. Potential variables and examples of implementation standards and specifications that might be useful for evaluating compliance with the Grazing Management Measure.

Table 4-2. Example variables for management measure implementation analysis.

Management Measure	Useful Variables	Less Useful Variables	Appropriate Sampling Unit
Erosion and Sediment Control	<ul style="list-style-type: none"> <li>• Area on which reduced tillage or terrace systems are installed</li> <li>• Area of runoff diversion systems or filter strips per acre of cropland</li> <li>• Area of highly erodible cropland converted to permanent cover</li> </ul>	<ul style="list-style-type: none"> <li>• Number of approved farm soil and erosion management plans</li> <li>• Number of grassed waterways, grade stabilization structures, filter strips installed</li> </ul>	<ul style="list-style-type: none"> <li>• Field</li> <li>• Acre</li> </ul>
Facility Wastewater and Runoff from Confined Animal Facilities	<ul style="list-style-type: none"> <li>• Quantity and percentage of total facility wastewater and runoff that is collected by a waste storage or treatment system</li> </ul>	<ul style="list-style-type: none"> <li>• Number of manure storage facilities</li> </ul>	<ul style="list-style-type: none"> <li>• Confined animal facility</li> <li>• Animal unit</li> </ul>
Nutrient Management	<ul style="list-style-type: none"> <li>• Number of farms following and acreage covered by approved nutrient management plans</li> <li>• Percent of farmers keeping records and applying nutrients at rates consistent with management recommendations</li> <li>• Quantity and percent reduction in fertilizer applied</li> <li>• Amount of fertilizer and manure spread between spreader calibrations</li> </ul>	<ul style="list-style-type: none"> <li>• Number of farms with approved nutrient management plans</li> </ul>	<ul style="list-style-type: none"> <li>• Farm</li> <li>• Field</li> <li>• Application</li> </ul>
Pesticide Management	<ul style="list-style-type: none"> <li>• Number of farms with complete records of field surveys and pesticide applications and following approved pest management plans</li> <li>• Number of pest field surveys performed on a weekly (or other time frame) basis</li> <li>• Quantity and percent reduction in pesticides use</li> </ul>	<ul style="list-style-type: none"> <li>• Number of farms with approved pesticide management plans</li> </ul>	<ul style="list-style-type: none"> <li>• Field</li> <li>• Farm</li> <li>• Application</li> </ul>
Grazing Management	<ul style="list-style-type: none"> <li>• Number of cattle-hours of access to riparian areas per day</li> <li>• Miles of stream from which grazing animals are excluded</li> </ul>	<ul style="list-style-type: none"> <li>• Miles of fence installed</li> </ul>	<ul style="list-style-type: none"> <li>• Stream mile</li> <li>• Animal unit</li> </ul>

### 4.3 Expert Evaluations

#### 4.3.1 Site Evaluations

Expert evaluations are the best way to collect reliable information on MM and BMP implementation. They involve a person or team of people visiting individual farms and speaking with farm owners and/or managers to obtain information on MM and BMP implementation. For many of the MMs, assessing and verifying compliance will require a farm visit and evaluation. The following should be considered before expert evaluations are conducted:

- *Obtaining permission of the farm owner or manager.* Without proper authorization to visit a farm from a farm owner or manager, the relationship between farmers and the agriculture agency, and any future regulatory or compliance action could be jeopardized.
- *The type(s) of expertise needed to assess proper implementation.* For some MMs, a team of trained personnel might be required to determine whether MMs have been implemented properly.
- *The activities that should occur during an expert evaluation.* This information is necessary for proper and complete preparation for the farm visit, so that it can be completed in a single visit and at the proper time.
- *The method of rating the MMs and BMPs.* MM and BMP rating systems are discussed below.
- *Consistency among evaluation teams and between farm evaluations.* Proper training

and preparation of expert evaluation team members are crucial to ensure accuracy and consistency.

- *The collection of information while at a farm.* Information collection should be facilitated with preparation of data collection forms that include any necessary MM and BMP rating information needed by the evaluation team members.
- *The content and format of post-evaluation discussions.* Site evaluation team members should bear in mind the value of postevaluation discussion among team members. Notes can be taken during the evaluation concerning any items that would benefit from group discussion.

Evaluators might consist of a single person suitably trained in agricultural expert evaluation to a group of professionals with various expertise. The composition of evaluation teams will depend on the types of MMs or BMPs being evaluated. Potential team members could include:

- Agricultural engineer
- Agriculture extension agent
- Agronomist
- Hydrologist
- Pesticide specialist
- Soil scientist
- Water quality expert

The composition of evaluation teams can vary depending on the purpose of the evaluation, available staff and other resources, and the geographic area being covered. All team members should be familiar with the required MMs and BMPs, and each team should have a member who has previously participated in an expert evaluation. This will ensure familiarity

with the technical aspects of the MMs and BMPs that will be rated during the evaluation and the expert evaluation process.

Training might be necessary to bring all team members to the level of proficiency needed to conduct the expert evaluations. State or county agricultural personnel should be familiar with agriculture regulations, state BMP standards and specifications, and proper BMP implementation, and therefore are generally well qualified to teach these topics to evaluation team members who are less familiar with them. Agricultural agents or other specialists who have participated in BMP implementation surveys might be enlisted to train evaluation team members about the actual conduct of expert evaluations. This training should include identification of BMPs particularly critical to water quality protection, analysis of erosion potential, and other aspects of BMP implementation that require professional judgement, as well as any standard methods for measurements to judge BMP implementation against state standards and specifications.

Alternatively, if only one or two individuals will be conducting expert evaluations, their training in the various specialties, such as those listed above, necessary to evaluate the quality of MM and BMP implementation could be provided by a team of specialists who are familiar with agricultural practices and nonpoint source pollution.

In the interest of consistency among the evaluations and among team members, it is advisable that one or more mock evaluations take place prior to visiting selected sample farms. These “practice sessions” provide team members with an opportunity to become familiar with MMs and BMPs as they should

be implemented under different farm conditions, gain familiarity with the evaluation forms and meanings of the terms and questions on them, and learn from other team members with different expertise. Mock evaluations are valuable for ensuring that all evaluators have a similar understanding of the intent of the questions, especially for questions whose responses involve a degree of subjectivity on the part of the evaluator.

Where expert evaluation teams are composed of more than two or three people, it might be helpful to divide up the various responsibilities for conducting the expert evaluations among team members ahead of time to avoid confusion at the farm and to be certain that all tasks are completed but not duplicated. Having a spokesperson for the group who will be responsible for communicating with the farm owner or manager—prior to the expert evaluation, at the expert evaluation if they are present, and afterward—might also be helpful. A county agriculture representative is generally a good choice as spokesperson because he/she can represent the state and county agriculture authorities. Newly-formed evaluation teams might benefit most from a division of labor and selection of a team leader or team coordinator with experience in expert evaluations who will be responsible for the quality of the expert evaluations. Smaller teams might find that a division of responsibilities is not necessary, as might larger teams that have experience working together. If responsibilities are to be assigned, mock evaluations can be a good time to work out these details.

#### ***4.3.2 Rating Implementation of Management Measures and Best Management Practices***

Many factors influence the implementation of MMs and BMPs, so it is sometimes necessary to use best professional judgment (BPJ) to rate their implementation and BPJ will almost always be necessary when rating overall BMP compliance at a farm. Site-specific factors such as soil type, crop rotation history, topography, tillage, and harvesting methods affect the implementation of erosion and sediment control BMPs, for instance, and must be taken into account by evaluators when rating MM or BMP implementation. Implementation of MMs will often be based on implementation of more than one BMP, and this makes rating MM implementation similar to rating overall BMP implementation at a farm or ranch. Determining an overall rating involves grouping the ratings of implementation of individual BMPs into a single rating, which introduces more subjectivity than rating the implementation of individual BMPs based on standards and specifications. Choice of a rating system and rating terms, which are aspects of proper evaluation design, is therefore important in minimizing the level of subjectivity associated with overall BMP compliance and MM implementation ratings. When creating overall ratings, it is still important to record the detailed ratings of individual BMPs as supporting information.

Individual BMPs, overall BMP compliance, and MMs can be rated using a binary approach (e.g., pass/fail, compliant/ noncompliant, or yes/no) or on a scale with more than two choices, such as 1 to 5 or 1 to 10 (where 1 is the worst—see *Example*). The simplest method of rating MM and BMP implementation is the use of a binary approach. Using a binary approach, either an entire farm or individual MMs or BMPs are rated as being in compliance or not in

compliance with respect to specified criteria. Scale systems can take the form of ratings from poor to excellent, inadequate to adequate, low to high, 1 to 3, 1 to 5, and so forth.

Whatever form of scale is used, the factors that would individually or collectively qualify a farm, MM, or BMP for one of the rankings should be clearly stated. The more choices that are added to the scale, the smaller and smaller the difference between them becomes and each must therefore be defined more specifically and accurately. This is especially important if different teams or individuals rate farms separately. Consistency among the ratings then depends on each team or individual evaluator knowing precisely what the criteria for each rating option mean. Clear and precise explanations of the rating scale can also help avoid or reduce disagreements

**Example...**of a rating scale (adapted from Rossman and Phillips, 1992).

- A possible rating scale from 1 to 5 might be:*
- 5 = Implementation exceeds requirements*
  - 4 = Implementation meets requirements*
  - 3 = Implementation has a minor departure from requirements*
  - 2 = Implementation has a major departure from requirements*
  - 1 = Implementation is in gross neglect of requirements*

*where:*

*Minor departure is defined as “small in magnitude or localized,” major departure is defined as “significant magnitude or where the BMPs are consistently neglected” and gross neglect is defined as “potential risk to water resources is significant and there is no evidence that any attempt is made to implement the BMP.”*

among team members. This applies equally to a binary approach. The factors, individually or collectively, that would cause a farm, MM, or BMP to be rated as not being in compliance with design specifications should be clearly stated on the evaluation form or in support documentation.

Rating farms or MMs and BMPs on a scale requires a greater degree of analysis by the evaluation team than does using a binary approach. Each higher number represents a better level of MM or BMP implementation. In effect, a binary rating approach is a scale with two choices; a scale of low, medium, and high (compliance) is a scale with three choices. Use of a scale system with more than two rating choices can provide more information to program managers than a binary rating approach, and this factor must be weighted against the greater complexity involved in using one. For instance, a survey that uses a scale of 1 to 5 might result in one MM with a ranking of 1, five with a ranking of 2, six with a ranking of 3, eight with a ranking of 4, and five with a ranking of 5. Precise criteria would have to be developed to be able to ensure consistency within and between survey teams in rating the MMs, but the information that only one MM was implemented poorly, 11 were implemented below standards, 13 met or were above standards, and 5 were implemented very well might be more valuable than the information that 18 MMs were found to be in compliance with design specifications, which is the only information that would be obtained with a binary rating approach.

If a rating system with more than two ratings is used to collect data, the data can be analyzed either by using the original rating data or by first transforming the data into a

binomial (i.e., two-choice rating) system. For instance, ratings of 1 through 5 could be reduced to two ratings by grouping the 1s, 2s, and 3s together into one group (e.g., inadequate) and the 4s and 5s into a separate group (e.g., adequate). If this approach is used, it is best to retain the original rating data for the detailed information they contain and to reduce the data to a binomial system only for the purpose of statistical analysis. Chapter 3, Section 3.5, contains information on the analysis of categorical data.

### 4.3.3 Rating Terms

The choice of rating terms used on the evaluation forms is an important factor in ensuring consistency and reducing bias, and the terms used to describe and define the rating options should be as objective as possible. For a rating system with a large number of options, the meanings of each option should be clearly defined. It is suggested to avoid using terms such as “major” and “minor” when describing erosion or pollution effects or deviations from prescribed MM or BMP implementation criteria because they might have different connotations for different evaluation team members. It is easier for an evaluation team to agree upon meaning if options are described in terms of measurable criteria and examples are provided to clarify the intended meaning. It is also suggested not to use terms that carry negative connotations. Evaluators might be disinclined to rate a MM or BMP as having a “major deviation” from an implementation criterion, even if justified, because of the negative connotation carried by the term. Rather than using such a term, observable conditions or effects of the quality of implementation can be listed and specific ratings (e.g., 1-5 or compliant/noncompliant

for the criterion) can be associated with the conditions or effects. For example, instead of rating an animal waste management facility as having a “major deficiency,” a specific deficiency could be described and ascribed an associated rating (e.g., “Waste storage structure is designed for no more than 70% of the confined animals = noncompliant”).

Evaluation team members will often have to take specific notes on farms, MMs, or BMPs during the evaluation, either to justify the ratings they have ascribed to variables or for discussion with other team members after the survey. When recording notes about the farms, MMs, or BMPs, evaluation team members should be as specific as the criteria for the ratings. A rating recorded as “MM deviates highly from implementation criteria” is highly subjective and loses specific meaning when read by anyone other than the person who wrote the note. Notes should therefore be as objective and specific as possible.

An overall farm rating is useful for summarizing information in reports, identifying the level of implementation of MMs and BMPs, indicating the likelihood that environmental protection is being achieved, identifying additional training or education needs; and conveying information to program managers, who are often not familiar with MMs or BMPs. For the purposes of preserving the valuable information contained in the original ratings of farms, MMs, or BMPs, however, overall ratings should summarize, not replace, the original data. Analysis of year-to-year variations in MM or BMP implementation, the factors involved in MM or BMP program implementation, and factors that could improve MM or BMP implementation and MM or BMP program

success are only possible if the original, detailed farm, MM, or BMP data are used.

Approaches commonly used for determining final BMP implementation ratings include calculating a percentage based on individual BMP ratings, consensus, compilation of aggregate scores by an objective party, voting, and voting only where consensus on a farm or MM or BMP rating cannot be reached. Not all systems for arriving at final ratings are applicable to all circumstances.

#### **4.3.4 Consistency Issues**

Consistency among evaluators and between evaluations is important, and because of the potential for subjectivity to play a role in expert evaluations, consistency should be thoroughly addressed in the quality assurance and quality control (QA/QC) aspects of planning and conducting an implementation survey. Consistency arises as a QA/QC concern in the planning phase of an implementation survey in the choice of evaluators, the selection of the size of evaluation teams, and in evaluator training. It arises as a QA/QC concern while conducting an implementation survey in whether evaluations are conducted by individuals or teams, how MM and BMP implementation on individual fields or farms is documented, how evaluation team discussions of issues are conducted, how problems are resolved, and how individual MMs and BMPs or whole farms are rated.

Consistency is likely to be best if only one to two evaluators conduct the expert evaluations and the same individuals conduct all of the

evaluations. If, for statistical purposes, many farms (e.g., 100 or more) need to be evaluated, use of only one to two evaluators might also be the most efficient approach. In this case, having a team of evaluators revisit a subsample of the farms that were originally evaluated by one to two individuals might be useful for quality control purposes.

If teams of evaluators conduct the evaluations, consistency can be achieved by keeping the membership of the teams constant. Differences of opinion, which are likely to arise among team members, can be settled through discussions held during evaluations, and the experience of team members who have done past evaluations can help guide decisions. Pre-evaluation training sessions, such as the mock evaluations discussed above, will help ensure that the first few expert evaluations are not “learning” experiences to such an extent that those farms must be revisited to ensure that they receive the same level of scrutiny as farms evaluated later.

If different farms are visited by different teams of evaluators or if individual evaluators are assigned to different farms, it is especially important that consistency be established before the evaluations are conducted. For best results, discussions among evaluators should be held periodically during the evaluations to discuss any potential problems. For instance, evaluators could visit some farms together at the beginning of the evaluations to promote consistency in ratings, followed by expert evaluations conducted by individual evaluators. Then, after a few farm or MM evaluations, evaluators could gather again to discuss results and to share any knowledge gained to ensure continued consistency.

As mentioned above, consistency can be established during mock evaluations held before the actual evaluations begin. These mock evaluations are excellent opportunities for evaluators to discuss the meaning of terms on rating forms, differences between rating criteria, and differences of opinion about proper MM or BMP implementation. A member of the evaluation team should be able to represent the state’s position on the definition of terms and clarify areas of confusion.

Descriptions of MMs and BMPs should be detailed enough to support any ratings given to individual features and to the MM or BMP overall. Sketching a diagram of the MM or BMP helps identify design problems, promotes careful evaluation of all features, and provides a record of the MM or BMP for future reference. A diagram is also valuable when discussing the MM or BMP with the farm owner or identifying features in need of improvement or alteration. Farm owners or managers can also use a copy of the diagram and evaluation when discussing their operations with state or county agriculture personnel. Photographs of MM or BMP features are a valuable reference material and should be used whenever an evaluator feels that a written description or a diagram could be inadequate. Photographs of what constitutes both good and poor MM or BMP implementation are valuable for explanatory and educational purposes; for example, for presentations to managers and the public.

#### ***4.3.5 Postevaluation Onsite Activities***

It is important to complete all pertinent tasks as soon as possible after the completion of an expert evaluation to avoid extra work later and to reduce the chances of introducing error

attributable to inaccurate or incomplete memory or confusion. All evaluation forms for each farm should be filled out completely before leaving the farm. Information not filled in at the beginning of the evaluation can be obtained from the farm owner or manager if necessary. Any questions that evaluators had about the MMs and BMPs during the evaluation can be discussed, notes written during the evaluation can be shared and used to help clarify details of the evaluation process and ratings. The opportunity to revisit the farm will still exist if there are points that cannot be agreed upon among evaluation team members.

Also, while the evaluation team is still on the farm, the farm owner or manager should be informed about what will follow; for instance, whether he/she will receive a copy of the report, when to expect it, what the results mean, and his/her responsibility in light of the evaluation, if any. Immediately following the evaluation is also an excellent time to discuss the findings with the farm owner or manager if he/she was not present during the evaluation.

#### 4.4 SELF-EVALUATIONS

##### 4.4.1 *Methods*

Self-evaluations, while often not a reliable source of MM or BMP implementation data, can be used to augment data collected through expert evaluations or in place of expert evaluations where the latter cannot be conducted. In some cases, state agriculture authority staff might have been involved directly with BMP selection and implementation and will be a source of useful information even if an expert evaluation is not conducted. Self-evaluations are an appropriate survey method for obtaining background

information from farmers or persons associated with farming operations, such as county extension agents.

Mail, telephone, and mail with telephone follow-up are common self-evaluation methods. Mail and telephone surveys are useful for collecting general information, such as the management measures that specific agricultural operations should be implementing. County extension agents or other state or local agricultural agents can be interviewed or sent a questionnaire that requests very specific information. Recent advances in and increasing access to electronic means of communication (i.e., e-mail and the Internet) might make these viable survey instruments in the future.

Mail surveys with a telephone follow-up and/or farm visit are an efficient method of collecting information. The USDA National Agricultural Statistics Service has found that 10 to 20 percent of farm owners or managers will respond to crop production questionnaires that are mailed. Approximately two-thirds of the questionnaires that are not returned are completed by telephone and the remainder are completed by personal visits to farms (USDA, undated). The entire NASS survey effort, from designing the questionnaire to reporting the results, takes approximately 6 months.

The level of response obtained by NASS is probably higher than would be obtained for MM or BMP implementation monitoring because NASS has developed a high level of trust with farmers through years of cooperation. In addition, NASS is prohibited by law from releasing information on individual farm operations, a fact of which most farmers are aware.

To ensure comparability of results, information that is collected as part of a self-evaluation—whether collected through the mail, over the phone, or during farm visits—should be collected in a manner that does not favor one method over the others. Ideally, telephone follow-up and on-site interviews should consist of no more than reading the questions on the questionnaire, without providing any additional explanation or information that would not have been available to those who responded through the mail. This approach eliminates as much as possible any bias associated with the different means of collecting the information. Figure 4-2 presents an example of an animal waste management survey questionnaire modeled after a NASS crop production questionnaire. The questionnaire design is discussed in Section 4.4.3.

It is important that the accuracy of information received through mail and phone surveys be checked. Inaccurate or incomplete responses to questions on mail and/or telephone surveys commonly result from survey respondents misinterpreting questions and thus providing misleading information, not including all relevant information in their responses, not wanting to provide some types of information, or deliberately providing some inaccurate responses. Therefore, the accuracy of information received through mail and phone surveys should be checked by selecting a subsample of the farmers surveyed and conducting follow-up farm visits.

#### **4.4.2 Cost**

Cost can be an important consideration when selecting an evaluation method. Farm visits can cost several hundred dollars per farm visited, depending on the type of farming

involved, the information to be collected, and the number of evaluators used. Mail and/or telephone surveys can be an inexpensive means of collecting information, but their cost must be balanced with the type and accuracy of information that can be collected through them. Other costs also need to be figured into the overall cost of mail and/or telephone surveys, including follow-up phone calls and farm visits to make up for a poor response to mailings and for accuracy checks. NASS has found that a mail survey with a telephone follow-up costs \$6 to \$10 per farm. Farm visits can cost several hundred dollars per farm depending on the complexity of the operation and the desired information. Additionally, the cost of questionnaire design must be considered, as a well-designed questionnaire is extremely important to the success of self-evaluations. Questionnaire design is discussed in the next section.

The number of evaluators used for farm visits has an obvious impact on the cost of a MM or BMP implementation survey. Survey costs can be minimized by having one or two evaluators visit farms instead of having multiple-person teams visit each farm. If the expertise of many specialists is desired, it might be cost-effective to have multiple-person teams check the quality of evaluations conducted by one or two evaluators. This can usually be done at a subsample of farms after they have been surveyed.

An important factor to consider when determining the number of evaluators to include on farm visitation teams, and how to balance the use of one to two evaluators versus multiple-person teams, is the objectives of the survey. Cost notwithstanding, the teams conducting the expert evaluations must be sufficient to meet the objectives of the survey,

### Animal Waste Management Survey

**Purpose of Survey:** To determine conformity with the following criteria for the control of runoff from confined animal facilities:

*States would put their standards here.*

Limit the discharge from the confined animal facility to surface waters by:

- (1) Storing both the facility wastewater *and* the runoff from confined animal facilities that are caused by storms up to and including a 25-year, 24-hour frequency storm. Storage structures should:
  - (a) Have an earthen lining or plastic membrane lining, or
  - (b) Be constructed with concrete, or
  - (c) Be a storage tank.
- (2) Managing stored runoff and accumulated solids from the facility through an appropriate waste utilization system.

**Population of interest:** Farms in the coastal zone with new or existing confined animal facilities that contain the following number of animals or more:

Animal Type	Number	Animal Units
Beef Cattle	300	300
Stables (horses)	200	400
Dairy Cattle	70	98
Layers	15,000	150 <sup>1</sup> 495 <sup>2</sup>
Broilers	15,000	150 <sup>1</sup> 495 <sup>2</sup>
Turkeys	13,750	2,475
Swine	80	200

Facilities that have been required by federal regulation 40 CFR 122.23 to obtain an NPDES discharge permit are excluded.

**Level of analysis:** *States should determine the level of analysis necessary.*

<sup>1</sup>If facility has a liquid manure system.

<sup>2</sup>If facility has continuous overflow watering.

Figure 4-2. Sample draft survey for confined animal facility management evaluation.

*Items of interest: This may vary depending on the type of facilities found within a state and the state's program for addressing this issue.*

***Land Use and Ownership***

Total acres operated	nnn
Land owned	nnn
Land rented	nnn

***Demographic Characteristics of Farm Operators***

Years farming	nnn
Years farming this operation	nnn
Years of formal education	nnn
Age	nnn

***Peak Number of Livestock***

Beef cattle	nnn
Horses	nnn
Dairy cattle	nnn
Layers (in facility with liquid manure systems)	nnn
Layers (in facility with continuous overflow watering)	nnn
Broilers (in facility with liquid manure systems)	nnn
Broilers (in facility with continuous overflow watering)	nnn
Turkeys	nnn
Swine	nnn

***Animal waste management practices***

Do you have a facility for wastewater and runoff from your animal operation?	y/n
Did an engineer, extension agent, or other professional assist in the design of the facility?	y/n/na
Was the facility designed to accommodate the peak amount of waste entering it?	y/n/na
Does the facility store both the wastewater and runoff caused by a 25-year, 24-hour frequency storm?	y/n/na/unknown
Does the facility have a earthen lining or plastic membrane?	y/n/na
Is the facility constructed with concrete?	y/n/na
Is the facility a storage tank?	y/n/na
Are the stored runoff and accumulated solids used as fertilizer?	y/n/na
If yes, what type of system is used?	nnnnnnnnnnnnnnnn

Figure 4-2. Sample draft survey. (continued)

and if the required teams would be too costly, then the objectives of the survey might need to be modified.

Another factor that contributes to the cost of a MM or BMP implementation survey is the number of farms to be surveyed. Once again, a balance must be reached between cost, the objectives of the survey, and the number of farms to be evaluated. Generally, once the objectives of the study have been specified, the number of farms to be evaluated is determined statistically to meet required data quality objectives. If the number of farms that is determined in this way would be too costly, then it would be necessary to modify the study objectives or the data quality objectives. Statistical determination of the number of farms to evaluate is discussed in Section 2.3.

#### 4.4.3 *Questionnaire Design*

Many books have been written on the design of data collection forms and questionnaires (e.g., Churchill, 1983; Ferber et al., 1964; Tull and Hawkins, 1990), and these can provide good advice for the creation of simple questionnaires that will be used for a single survey. However, for complex questionnaires or ones that will be used for initial surveys as part of a series of surveys (i.e., trend analysis), it is strongly advised that a professional in questionnaire design be consulted. This is because while it might seem that designing a questionnaire is a simple task, small details such as the order of questions, the selection of one word or phrase over a similar one, and the tone of the questions can significantly affect survey results. A professionally-designed questionnaire can yield information beyond that contained in the responses to the questions themselves, while a poorly-designed questionnaire can invalidate the results.

The objective of a questionnaire, which should be closely related to the objectives of the survey, should be extremely well thought out prior to its being designed. Questionnaires should also be designed at the same time as the information to be collected is selected to ensure that the questions address the objectives as precisely as possible. Conducting these activities simultaneously also provides immediate feedback on the attainability of the objectives and the detail of information that can be collected. For example, an investigator might want information on the extent of grazing in riparian areas, but might discover while designing the questionnaire that the desired information could not be obtained through the use of a questionnaire, or that the information that could be collected would be insufficient to fully address the chosen objectives. In such a situation the investigator could revise the objectives and questions before going further with questionnaire design.

Tull and Hawkins (1990) identified seven major elements of questionnaire construction:

1. Preliminary decisions
2. Question content
3. Question wording
4. Response format
5. Question sequence
6. Physical characteristics of the questionnaire
7. Pretest and revision.

Preliminary decisions include determining exactly what type of information is required, determining the target audience, and selecting the method of communication (e.g., mail, telephone, farm visit). These subjects are addressed in other sections of this guidance.

The second step is to determine the content of the questions. Each question should generate one or more of the information requirements identified in the preliminary decisions. The ability of the question to elicit the necessary data needs to be assessed. “Double-barreled” questions, in which two or more questions are asked as one, should be avoided. Questions that require the respondent to aggregate several sources of information should be subdivided into several specific questions or parts. The ability of the respondent to answer accurately should also be considered when preparing questions. Some respondents might be unfamiliar with the type of information requested or the terminology used. Or a respondent might have forgotten some of the information of interest, or might be unable to verbalize an answer. Consideration should be given to the willingness of respondents to answer the questions accurately. If a respondent feels that a particular answer might be embarrassing or personally harmful, (e.g., might lead to fines or increased regulation), he or she might refuse to answer the question or might deliberately provide inaccurate information. For this reason, answers to questions that might lead to such responses should be checked for accuracy whenever possible.

The next step is the specific phrasing of the questions. Simple, easily understood language is preferred. The wording should not bias the answer or be too subjective. For instance, a question should not ask if grazing in riparian areas is a problem on the farm. Instead, a series of questions could ask if cattle are kept on the farm, if the farm has any riparian areas (which should be defined), if any means are provided along the riparian areas to exclude grazing animals, and what those means are. These questions all request factual information

of which a farmer should be knowledgeable and they progress from simple to more complex. All alternatives and assumptions should be clearly stated on the questionnaire, and the respondent’s frame of reference should be considered.

Fourth, the type of response format should be selected. Various types of information can best be obtained using open-ended, multiple-choice, or dichotomous questions. An open-ended question allows respondents to answer in any way they feel is appropriate. Multiple-choice questions tend to reduce some types of bias and are easier to tabulate and analyze; however, good multiple-choice questions can be more difficult to formulate. Dichotomous questions allow only two responses, such as “yes-no” or “agree-disagree.” Dichotomous questions are suitable for determining points of fact, but must be very precisely stated and unequivocally solicit only a single piece of information.

The fifth step in questionnaire design is the ordering of the questions. The first questions should be simple to answer, objective, and interesting in order to relax the respondent. The questionnaire should move from topic to topic in a logical manner without confusing the respondent. Early questions that could bias the respondent should be avoided. There is evidence that response quality declines near the end of a long questionnaire (Tull and Hawkins, 1990). Therefore, more important information should be solicited early. Before presenting the questions, the questionnaire should explain how long (on average) it will take to complete and the types of information that will be solicited. The questionnaire should not present the respondent with any surprises.

The layout of the questionnaire should make it easy to use and should minimize recording mistakes. The layout should clearly show the respondent all possible answers. For mail surveys, a pleasant appearance is important for securing cooperation.

The final step in the design of a questionnaire is the pretest and possible revision. A questionnaire should always be pretested with members of the target audience. This will preclude expending large amounts of effort and then discovering that the questionnaire produces biased or incomplete information.

#### 4.5 AERIAL RECONNAISSANCE AND PHOTOGRAPHY

Aerial reconnaissance and photography can be useful tools for gathering physical farm information quickly and comparatively inexpensively, and they are used in conservation for a variety of purposes. Aerial photography has been proven to be helpful for agricultural conservation practice identification (Pelletier and Griffin, 1988); rangeland monitoring (BLM, 1991); terrain stratification, inventory site identification, planning, and monitoring in mountainous regions (Hetzl, 1988; Born and Van Hooser, 1988); as well as for forest regeneration assessment (Hall and Alred, 1992) and forest inventory and analysis (Hackett, 1988). Factors such as the characteristics of what is being monitored, scale, and camera format determine how useful aerial photography can be for a particular purpose.

Pelletier and Griffin (1988) investigated the use of aerial photography for the identification of agriculture conservation practices. They found that practices that occupy a large area and have an identifiable pattern, such as

contour cropping, strip cropping, terraces, and windbreaks, were readily identified even at a small scale (1:80,000) but that smaller, single-unit practices, such as sediment basins and sediment diversions, were difficult to identify at a small scale. They estimated that 29 percent of practices could be identified at a scale of 1:80,000, 45 percent could be identified at 1:30,000, 70 percent could be identified at 1:15,000, and over 90 percent could be identified at a scale of 1:10,000.

Photographic scale and resolution must be taken into consideration when deciding whether to use aerial photography, and a photographic scale that produces good resolution of the items of importance to the monitoring effort must be chosen. The Bureau of Land Management (BLM) uses low-level, large-scale (1:1,000 to 1:1,500) aerial photography to monitor rangeland vegetation (BLM, 1991). The agency reports that scales smaller than 1:1,500 (e.g., 1:10,000, 1:30,000) are too small to monitor the classes of land cover (shrubs, grasses and forbs, bare soil, rock) on rangeland. Born and Van Hooser (1988) found that a scale of 1:58,000 was marginal for use in forestry resource inventorying and monitoring.

Camera format is a factor that also must be considered. Large-format cameras are generally preferred over small-format cameras (e.g., 35 mm), but are more costly to purchase and operate. The large negative size (9 cm x 9 cm) produced using a large-format camera provides the resolution and detail necessary for accurate photo interpretation. Large-format cameras can be used from higher altitudes than small-format cameras, and the image area covered by a large-format image at a given scale (e.g., 1:1,500) is much larger than the image area captured by a small-

format camera at the same scale. Small-scale cameras (i.e., 35 mm) can be used for identifications that involve large-scale features, such as mining areas, the extent of burning, and large animals in censuses, and they are less costly to purchase and use than large-format cameras, but they are limited in the altitude that the photographs can be taken from and the resolution that they provide when enlarged (Owens, 1988).

BLM recommends the use of a large-format camera because the images provide the photo interpreter with more geographical reference points, it provides flexibility to increase sample plot size, and it permits modest navigational errors during overflight (BLM, 1991). Also, if hiring someone to take the photographs, most photo contractors will have large-format equipment for the purpose.

A drawback to the use of aerial photography is that conservation practices that do not meet implementation or operational standards but that are similar to practices that do are indistinguishable from ones that do in an aerial photograph (Pelletier and Griffin, 1988). Also, practices that are defined by managerial concepts rather than physical criteria, such as irrigation water management or nutrient management, cannot be detected with aerial photographs.

Regardless of scale, format, or item being monitored, it is useful for photo interpreters to receive 2-3 days of training on the basic fundamentals of photo interpretation and that they be thoroughly familiar with the vegetation and land forms in the areas where the photographs that they will be interpreting were taken (BLM, 1991). A visit to the farms in the photographs is recommended to improve correlation between the interpretation

and actual farm characteristics. Generally, after a few visits and interpretations of photographs of those farms, photo interpreters will be familiar with the photographic characteristics of the vegetation in the area and the farm visits can be reserved for verification of items in doubt. A change in type of vegetation or physiography in photographs normally requires new visits until photo interpreters are familiar with the characteristics of the new vegetation in the photographs.

Information on obtaining aerial photographs is available from the Farm Service Agency and the Natural Resources Conservation Service. Contact the Farm Service Agency at: USDA FSA Aerial Photography Field Office, P.O. Box 30010, Salt Lake City, UT, 84130-0010, (801) 975-3503. The Farm Service Agency's Internet address is <http://www.fsa.usda.gov>. Contact the Natural Resources Conservation Service at: NRCS National Cartography and Geospatial Center, Fort Worth Federal Center, Bldg 23, Room 60, P.O. Box 6567, Fort Worth, TX 76115-0567; 1-800-672-5559. NRCS's Internet address is <http://www.ncg.nrcs.usda.gov>.

## CHAPTER 5. PRESENTATION OF EVALUATION RESULTS

### 5.1 INTRODUCTION

The first three chapters of this guidance presented techniques for the collection of information. Data analysis and interpretation are addressed in detail in Chapter 4 of EPA's *Monitoring Guidance for Determining the Effectiveness of Nonpoint Source Controls* (USEPA, 1997). This chapter provides ideas for the presentation of results.

The presentation of MM or BMP compliance survey results, whether written or oral, is an integral part of a successful monitoring study. The quality of the presentation of results is an indication of the quality of the compliance survey, and if the presentation fails to convey important information from the compliance survey to those who need the information, the compliance survey itself might be considered a failure.

The quality of the presentation of results is dependent on at least four criteria—it must be complete, accurate, clear, and concise (Churchill, 1983). Completeness means that the presentation provides all necessary information to the audience in the language that it understands; accuracy is determined by how well an investigator handles the data, phrases findings, and reasons; clarity is the result of clear and logical thinking and a precision of expression; and conciseness is the result of selecting for inclusion only that which is necessary.

Throughout the process of preparing the results of a MM or BMP compliance survey for presentation, it must be kept in mind that the study was initially undertaken to provide information for management

purposes—specifically, to help make a decision (Tull and Hawkins, 1990). The presentation of results should be built around the decision that the compliance survey was undertaken to support. The message of the presentation must also be tailored to that decision. It must be realized that there will be a time lag between the compliance survey and the presentation of the results, and the results should be presented in light of their applicability to the management decision to be made based on them. The length of the time lag is a key factor in determining this applicability. If the time lag is significant, it should be made clear during the presentation that the situation might have changed since the survey was conducted. If reliable trend data are available, the person making the presentation might be able to provide a sense of the likely magnitude of any change in the situation. If the change in status is thought to be insignificant, evidence should be presented to support this claim. For example, state that “At the time that the compliance survey was conducted, farmers were using BMPs with increasing frequency, and the lack of any changes in program implementation coupled with continued interaction with farmers provides no reason to believe that this trend has changed since that time.” It would be misleading to state “The monitoring study indicates that farmers *are* using BMPs with increasing frequency.” The validity and force of the message will be enhanced further through use of the active voice (*we believe*) rather than the passive voice (*it is believed*).

Three major factors must be considered when presenting the results of MM and BMP

implementation studies: Identifying the target audience, selecting the appropriate medium (printed word, speech, pictures, etc.), and selecting the most appropriate format to meet the needs of the audience.

## 5.2 AUDIENCE IDENTIFICATION

Identification of the audience(s) to which the results of the MM and BMP compliance survey will be presented determines the content and format of the presentation. For results of compliance survey studies, there are typically six potential audiences:

- Interested/concerned citizens
- Farm owners and managers
- Media/general public
- Policy makers
- Resource managers
- Scientists

These audiences have different information needs, interests, and abilities to understand complex data. It is the job of the person(s) preparing the presentation to analyze these factors prior to preparing a presentation. The four criteria for presentation quality apply regardless of the audience. Other elements of a comprehensive presentation, such as discussion of the objectives and limitations of the study and necessary details of the method, must be part of the presentation and must be tailored to the audience. For instance, details of the sampling plan, why the plan was chosen over others, and the statistical methods used for analysis might be of interest to other investigators planning a similar study, and such details should be recorded even if they are not part of any presentation of results because of their value for future reference when the monitoring is repeated or similar

studies are undertaken, but they are best not included in a presentation to management.

## 5.3 PRESENTATION FORMAT

Regardless of whether the results of a compliance survey are presented written or orally, or both, the information being presented must be understandable to the audience. Consideration of who the audience is will help ensure that the presentation is particularly suited to its needs, and choice of the correct format for the presentation will ensure that the information is conveyed in a manner that is easy to comprehend.

Most reports will have to be presented both written and orally. Written reports are valuable for peer review, public information dissemination, and for future reference. Oral presentations are often necessary for managers, who usually do not have time to read an entire report, only have need for the results of the study, and are usually not interested in the finer details of the study. Different versions of a report might well have to be written—for the public, scientists, and managers (i.e., an executive summary)—and separate oral presentations for different audiences—the public, farmers, managers, and scientists at a conference—might have to be prepared.

Most information can most effectively be presented in the form of tables, charts, and diagrams (Tull and Hawkins, 1990). These graphic forms of data and information presentation can help simplify the presentation, making it easier for an audience to comprehend than if explained exhaustively with words. Words are important for pointing out significant ideas or findings, and for interpreting the results where appropriate.

Words should not be used to repeat what is already adequately explained in graphics, and slides or transparencies that are composed largely of words should contain only a few essential ideas each. Presentation of too much written information on a single slide or transparency only confuses the audience. Written slides or transparencies should also be free of graphics, such as clever logos or background highlights—unless the pictures are essential to understanding the information presented—since they only make the slides or transparencies more difficult to read. Examples of graphics and written slides are presented in Figures 5-1 through 5-3.

Different types of graphics have different uses as well. Information presented in a tabular format can be difficult to interpret because the reader has to spend some time with the information to extract the essential points from it. The same information presented in a pie chart or bar graph can convey essential information immediately and avoid the inclusion of background data that are not essential to the point. When preparing information for a report, an investigator should organize the information in various ways and choose that which conveys only the information essential for the audience in the least complicated manner.

### 5.3.1 Written Presentations

The following criteria should be considered when preparing written material:

- *Reading level or level of education* of the target audience.
- *Level of detail necessary* to make the results understandable to the target audience) Different audiences require

various levels of background information to fully understand the study's results.

- *Layout.* The integration of text, graphics, color, white space, columns, sidebars, and other design elements is important in the production of material that the target audience will find readable and visually appealing.
- *Graphics.* Photos, drawings, charts, tables, maps, and other graphic elements can be used to effectively present information that the reader might otherwise not understand.

### 5.3.2 Oral Presentations

An effective oral presentation requires special preparation. Tull and Hawkins (1990) recommend three steps:

1. Analyze the audience, as explained above;
2. Prepare an outline of the presentation, and preferably a written script;
3. Rehearse it. Several dry runs of the presentation should be made, and if

## 5 Leading Sources of Water Quality Impairment in various types of water bodies

<u>RANK</u>	<u>RIVERS</u>	<u>LAKES</u>	<u>ESTUARIES</u>
1	<b>Agriculture</b>	<b>Agriculture</b>	Urban Runoff
2	STPs	STPs	STPs
3	Habitat Modification	Urban Runoff	<b>Agriculture</b>
4	Urban Runoff	Other NPS	Industry Point Sources
5	Resource Extraction	Habitat Modification	Petroleum Activities

Figure 5-1. Example of presentation of information in a written slide. (Source: USEPA, 1995)

possible it should be taped on a VCR and the presentation analyzed.

These steps are extremely important if an oral presentation is to be effective. Remember that oral presentations of ½ to 1 hour are often all that is available for the presentation of the results of months of research to managers who are poised to make decisions based on the presentation. Adequate preparation is essential if the oral presentation is to accomplish its purpose.

#### 5.4 FOR FURTHER INFORMATION

The provision of specific examples of effective and ineffective presentation graphics, writing styles, and organizations is beyond the

scope of this document. A number of resources that contain suggestions for how study results should be presented are available, however, and should be consulted. A listing of some references is provided below.

- *The New York Public Library Writer's Guide to Style and Usage* (NYPL, 1994) has information on design, layout, and presentation in addition to guidance on grammar and style.
- *Good Style: Writing for Science and Technology* (Kirkman, 1992) provides

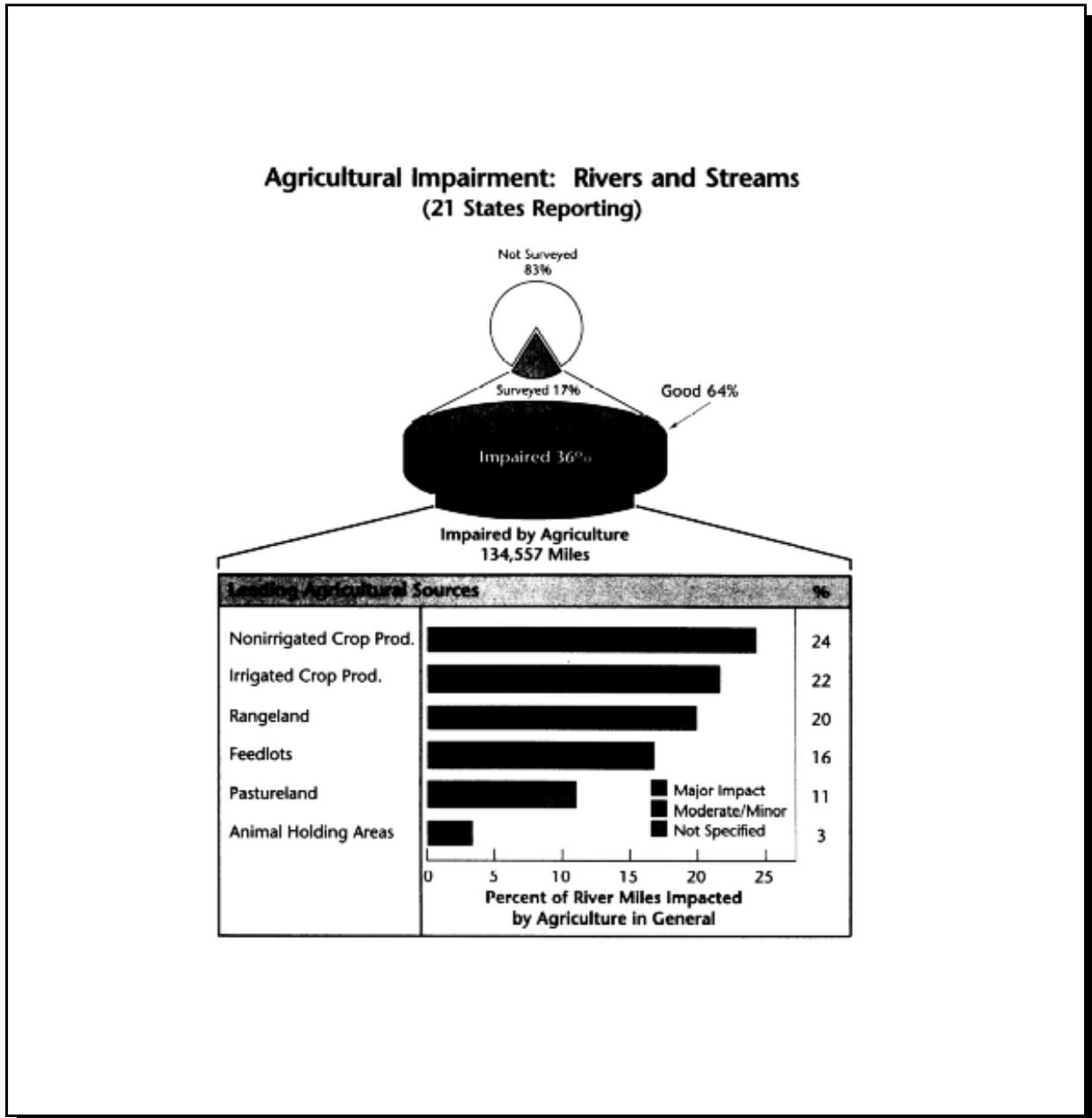


Figure 5-2. Example of representation of data using a combination of a pie chart and a horizontal bar chart. (Source: USEPA, 1995)

techniques for presenting technical material in a coherent, readable style.

into readable, well organized writing.

- *The Modern Researcher* (Barzun and Graff, 1992) explains how to turn research

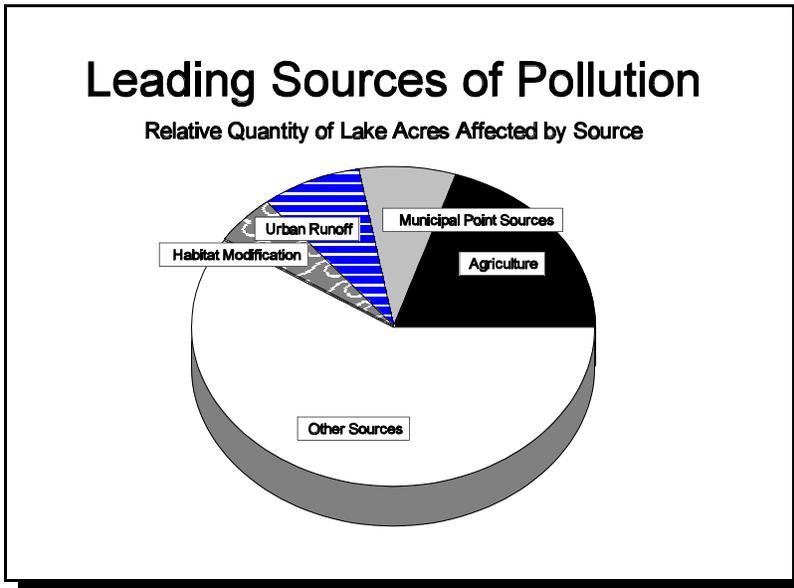


Figure 5-3. Example of representation of data in the form of a pie chart.

- *Writing with Precision: How to Write So That You Cannot Possibly Be Misunderstood*, 6th ed. (Bates, 1993) addresses communication problems of the 1990s.
- *Designer's Guide to Creating Charts & Diagrams* (Holmes, 1991) gives tips for combining graphics with statistical information.
- *The Elements of Graph Design* (Kosslyn, 1993) shows how to create effective displays of quantitative data.

## REFERENCES

- Academic Press. 1992. *Dictionary of Science and Technology*. Academic Press, Inc., San Diego, California.
- Barzun, J., and H.F. Graff. 1992. *The Modern Researcher*. 5th ed. Houghton Mifflin.
- Bates, J. 1993. *Writing with Precision: How to Write So That You Cannot Possibly Be Misunderstood*. 6th ed. Acropolis.
- Blalock, H.M., Jr. 1979. *Social Statistics*. Rev. 2nd ed. McGraw-Hill Book Company, New York, NY.
- BLM. 1991. *Inventory and Monitoring Coordination: Guidelines for the Use of Aerial Photography in Monitoring*. Technical Report TR 1734-1. Department of the Interior, Bureau of Land Management.
- Born, J.D., and D.D. Van Hooser. 1988. Intermountain Research Station remote sensing use for resource inventory, planning, and monitoring. In *Remote Sensing for Resource Inventory, Planning, and Monitoring*. Proceedings of the Second Forest Service Remote Sensing Applications Conference, Sidell, Louisiana, and NSTL, Mississippi, April 11-15, 1988.
- Casley, D.J., and D.A. Lury. 1982. *Monitoring and Evaluation of Agriculture and Rural Development Projects*. The Johns Hopkins University Press, Baltimore, MD.
- Churchill, G.A., Jr. 1983. *Marketing Research: Methodological Foundations*, 3rd ed. The Dryden Press, New York, New York.
- Cochran, W.G. 1977. *Sampling techniques*. 3rd ed. John Wiley and Sons, New York, New York.
- Cross-Smiecinski, A., and L.D. Stetzenback. 1994. *Quality planning for the life science researcher: Meeting quality assurance requirements*. CRC Press, Boca Raton, Florida.
- CTIC. 1994. *1994 National Crop Residue Management Survey*. Conservation Technology Information Center, West Lafayette, IN.
- CTIC. 1995. *Conservation IMPACT*, vol. 13, no. 4, April 1995. Conservation Technology Information Center, West Lafayette, IN.
- Ferber, R., D.F. Blankertz, and S. Hollander. 1964. *Marketing Research*. The Ronald Press Company, New York, NY.

- Freund, J.E. 1973. *Modern elementary statistics*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Gaugush, R.F. 1987. *Sampling Design for Reservoir Water Quality Investigations*. Instruction Report E-87-1. Department of the Army, US Army Corps of Engineers, Washington, DC.
- Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold, New York, NY.
- Hackett, R.L. 1988. Remote sensing at the North Central Forest Experiment Station. In *Remote Sensing for Resource Inventory, Planning, and Monitoring*. Proceedings of the Second Forest Service Remote Sensing Applications Conference, Sidell, Louisiana, and NSTL, Mississippi, April 11-15, 1988.
- Hall, R.J., and A.H. Alred. 1992. Forest regeneration appraisal with large-scale aerial photographs. *The Forestry Chronicle* 68(1):142-150.
- Helsel, D.R., and R.M. Hirsch. 1995. *Statistical Methods in Water Resources*. Elsevier. Amsterdam.
- Hetzel, G.E. 1988. Remote sensing applications and monitoring in the Rocky Mountain region. In *Remote Sensing for Resource Inventory, Planning, and Monitoring*. Proceedings of the Second Forest Service Remote Sensing Applications Conference, Sidell, Louisiana, and NSTL, Mississippi, April 11-15, 1988.
- Holmes, N. 1991. *Designer's Guide to Creating Charts & Diagrams*. Watson-Guptill.
- Hook, D., W. McKee, T. Williams, B. Baker, L. Lundquist, R. Martin, and J. Mills. 1991. *A Survey of Voluntary Compliance of Forestry BMPs*. South Carolina Forestry Commission, Columbia, SC.
- IDDHW. 1993. *Forest Practices Water Quality Audit 1992*. Idaho Department of Health and Welfare, Division of Environmental Quality, Boise, ID.
- Kirkman, J. 1992. *Good Style: Writing for Science and Technology*. Chapman and Hall.
- Kosslyn, S.M. 1993. *The Elements of Graph Design*. W.H. Freeman.
- Kupper, L.L., and K.B. Hafner. 1989. How appropriate are popular sample size formulas? *Am. Stat.* 43:101-105.
- MacDonald, L.H., A.W. Smart, and R.C. Wissmar. 1991. *Monitoring Guidelines to Evaluate the Effects of Forestry Activities on Streams in the Pacific Northwest and Alaska*. EPA/910/9-91-001. U.S. Environmental Protection Agency Region 10, Seattle, WA.

- 
- Mann, H.B., and D.R. Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18:50-60.
- McNew, R.W. 1990. Sampling and Estimating Compliance with BMPs. In *Workshop on Implementation Monitoring of Forestry Best Management Practices*, Southern Group of State Foresters, USDA Forest Service, Southern Region, Atlanta, GA, January 23-25, 1990, pp. 86-105.
- Meals, D.W. 1988. *Laplatte River Watershed Water Quality Monitoring & Analysis Program*. Program Report No. 10. Vermont Water Resources Research Center, School of Natural Resources, University of Vermont, Burlington, VT.
- NYPL. 1994. *The New York Public Library Writer's Guide to Style and Usage*. A Stonesong Press book. HarperCollins Publishers, New York, NY.
- Owens, T. 1988. Using 35mm photographs in resource inventories. In *Remote Sensing for Resource Inventory, Planning, and Monitoring*. Proceedings of the Second Forest Service Remote Sensing Applications Conference, Sidell, Louisiana, and NSTL, Mississippi, April 11-15, 1988.
- Pelletier, R.E., and R.H. Griffin. 1988. An evaluation of photographic scale in aerial photography for identification of conservation practices. *J. Soil Water Conserv.* 43(4):333-337.
- Rashin, E., C. Clishe, and A. Loch. 1994. *Effectiveness of forest road and timber harvest best management practices with respect to sediment-related water quality impacts*. Interim Report No. 2. Washington State Department of Ecology, Environmental Investigations and Laboratory Services program, Watershed Assessments Section. Ecology Publication No. 94-67. Olympia, Washington.
- Remington, R.D., and M.A. Schork. 1970. *Statistics with applications to the biological and health sciences*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Rossmann, R., and M.J. Phillips. 1991. *Minnesota forestry best management practices implementation monitoring. 1991 forestry field audit*. Minnesota Department of Natural Resources, Division of Forestry.
- Schultz, B. 1992. *Montana Forestry Best Management Practices Implementation Monitoring*. The 1992 Forestry BMP Audits Final Report. Montana Department of State Lands, Forestry Division, Missoula, MT.
- Snedecor, G.W. and W.G. Cochran. 1980. *Statistical methods*. 7th ed. The Iowa State University Press, Ames, Iowa.
-

- Tull, D.S., and D.I. Hawkins. 1990. *Marketing Research. Measurement and Method*. Fifth edition. Macmillan Publishing Company, New York, New York.
- USDA. 1994a. *1992 National Resources Inventory*. U.S. Department of Agriculture, Natural Resource Conservation Service, Resources Inventory and Geographical Information Systems Division, Washington, DC.
- USDA. 1994b. *Agricultural Resources and Environmental Indicators*. Agricultural Handbook No. 705. U.S. Department of Agriculture, Economic Research Service, Natural Resources and Environmental Division, Herndon, VA.
- USDA. Undated. *Preparing Statistics for Agriculture*. U.S. Department of Agriculture, National Agricultural Statistics Service, Washington, DC.
- USDOC. 1994. *1992 Census of Agriculture*. U.S. Department of Commerce, Bureau of the Census. U.S. Government Printing Office, Washington, DC.
- USEPA. 1993a. *Guidance Specifying Management Measures For Sources Of Nonpoint Pollution In Coastal Waters*. EPA 840-B-92-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.
- USEPA. 1993b. *Evaluation of the Experimental Rural Clean Water Program*. EPA 841-R-93-005. U.S. Environmental Protection Agency, Office of Water, Washington, DC.
- USEPA. 1995. *National water quality inventory 1994 Report to Congress*. EPA 841-R-95-005. U.S. Environmental Protection Agency, Office of Water, Washington, DC.
- USEPA. 1997. *Monitoring Guidance for Determining the Effectiveness of Nonpoint Source Controls*. EPA 841-B-96-004. U.S. Environmental Protection Agency, Office of Water, Washington, DC. August.
- USGS. 1990. *Land Use and Land Cover Digital Data from 1:250,000- and 1:100,000-Scale Maps: Data Users Guide*. National Mapping Program Technical Instructions Data Users Guide 4. U.S. Department of the Interior, U.S. Geological Survey, Reston, VA.
- Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics* 1:80-83.
- Winer, B.J. 1971. *Statistical principles in experimental design*. McGraw-Hill Book Company, New York.

## GLOSSARY

*accuracy*: the extent to which a measurement approaches the true value of the measured quantity

*aerial photography*: the practice of taking photographs from an airplane, helicopter, or other aviation device while it is airborne

*allocation, Neyman*: stratified sampling in which the cost of sampling each stratum is in proportion to the size of the stratum but variability between strata changes

*allocation, proportional*: stratified sampling in which the variability and cost of sampling for each stratum are in proportion to the size of the stratum

*allowable error*: the level of error acceptable for the purposes of a study

*ANOVA*: see analysis of variance

*analysis of variance*: a statistical test used to determine whether two or more sample means could have been obtained from populations with the same parametric mean

*assumptions*: characteristics of a population of a sampling method taken to be true without proof

*bar graph*: a representation of data wherein data is grouped and represented as vertical or horizontal bars over an axis

*best professional judgement*: an informed opinion made by a professional in the appropriate field of study or expertise

*best management practice*: a practice or combination of practices that are determined to be the most effective and practicable means of controlling point and nonpoint pollutants at levels compatible with environmental quality goals

*bias*: a characteristic of samples such that when taken from a population with a known parameter, their average does not give the parametric value

*binomial*: an algebraic expression that is the sum or difference of two terms

*camera format*: refers to the size of the negative taken by a camera. 35mm is a small camera format

*chi-square distribution*: a scaled quantity whose distribution provides the distribution of the sample variance

*coefficient of variation*: a statistical measure used to compare the relative amounts of variation in populations having different means

*confidence interval*: a range of values about a measured value in which the true value is presumed to lie

*conservation tillage*: a method of conservation in which plant material is left on the ground after harvest to control erosion

*consistency*: conforming to a regular method or style; an approach that keeps all factors of measurement similar from one measurement to the next to the extent possible

*contour farming*: a farming method in which fields are tilled along the topographic contours of the land

*cumulative effects*: the total influences attributable to numerous individual influences

*degrees of freedom*: the number of residuals (the difference between a measured value and the sample average) required to completely determine the others

*design, balanced*: a sampling design wherein separate sets of data to be used are similar in quantity and type

*distribution*: the allocation or spread of values of a given parameter among its possible values

*e-mail*: an electronic system for correspondence

*erosion potential*: a measure of the ease with which soil can be carried away in storm water runoff or irrigation runoff

*error*: the fluctuation that occurs from one repetition to another; also *experimental error*

*estimate, baseline*: an estimate of baseline, or actual conditions

*estimate, pooled*: a single estimate obtained from grouping individual estimates and using the latter to obtain a single value

*finite population correction term*: a correction term used when population size is small relative to sample size

*Friedman test*: a nonparametric test that can be used for analysis when two variables are involved

*hydrologic modification*: the alteration of the natural circulation or distribution of water by the placement of structures or other activities

*hypothesis, alternative*: the hypothesis which is contrary to the null hypothesis

*hypothesis, null*: the hypothesis or conclusion assumed to be true prior to any analysis

*Internet*: an electronic data transmission system

*Kruskal-Wallis test*: a nonparametric test recommended for the general case with  $a$  samples and  $n_i$  variates per sample

*management measure*: an economically achievable measure for the control of the addition of pollutants from existing and new categories and classes of nonpoint sources of pollution, which reflect the greatest degree of pollutant reduction achievable through the application of the best available nonpoint pollution control practices, technologies, processes, siting criteria, operating methods, or other alternatives

*Mann-Whitney test*: a nonparametric test for use when a test is only between two samples

*mean, estimated*: a value of population mean arrived at through sampling

*mean, overall*: the measured average of a population

*mean, stratum*: the measured average within a sample subgroup or stratum

*measurement bias*: a consistent under- or overestimation of the true value of something being measured, often due to the method of measurement

*measurement error*: the deviation of a measurement from the true value of that which is being measured

*median*: the value of the middle term when data are arranged in order of size; a measure of central tendency

*monitoring, baseline*: monitoring conducted to establish initial knowledge about the actual state of a population

*monitoring, compliance*: monitoring conducted to determine if those who must implement programs, best management practices, or management measures, or who must conduct operations according to standards or specifications are doing so

*monitoring, project*: monitoring conducted to determine the impact of a project, activity, or program

*monitoring, validation*: monitoring conducted to determine how well a model accurately reflects reality

*navigational error*: errors in determining the actual location (altitude or latitude/longitude) of an airplane or other aviation device due to instrumentation or the operator

*nominal*: referred to by name; variables that cannot be measured but must be expressed qualitatively

*nonparametric method*: distribution-free method; any of various inferential procedures whose conclusions do not rely on assumptions about the distribution of the population of interest

*normal approximation*: an assumption that a population has the characteristics of a normally-distributed population

*normal deviate*: deviation from the mean expressed in units of  $\sigma$

*nutrient management plan*: a plan for managing the quantity of nutrients applied to crops to achieve maximum plant nutrition and minimum nutrient waste

*ordinal*: ordered such that the position of an element in a series is specified

*parametric method*: any statistical method whose conclusions rely on assumptions about the distribution of the population of interest

*physiography*: a description of the surface features of the Earth; a description of landforms

*pie chart*: a representation of data wherein data is grouped and represented as more or less triangular sections of a circle and the total is the entire circle

*population, sample*: the members of a population that are actually sampled or measured

*population, target*: the population about which inferences are made; the group of interest, from which samples are taken

*population unit*: an individual member of a target population that can be measured independently of other members

*power*: the probability of correctly rejecting the null hypothesis when the alternative hypothesis is false.

*precision*: a measure of the similarity of individual measurements of the same population

*question, dichotomous*: a question that allows for only two responses, such as "yes" and "no"

*question, double-barreled*: two questions asked as a single question

*question, multiple-choice*: a question with two or more predetermined responses

*question, open-ended*: a question format that requires a response beyond "yes" or "no"

*remote sensing*: methods of obtaining data from a location distant from the object being measured, such as from an airplane or satellite

*resolution*: the sharpness of a photograph

*sample size*: the number of population units measured

*sampling, cluster*: sampling in which small groups of population units are selected for sampling and each unit in each selected group is measured

*sampling, simple random*: sampling in which each unit of the target population has an equal chance of being selected

*sampling, stratified random*: sampling in which the target population is divided into separate subgroups, each of which is more internally similar than the overall population is, prior to sample selection

*sampling, systematic*: sampling in which population units are chosen in accordance with a predetermined sample selection system

*sampling error*: error attributable to actual variability in population units not accounted for by the sampling method

*scale (aerial photography)*: the proportion of the image size of an object (such as a land area) to its actual size, e.g., 1:3000. The smaller the second number, the larger the scale

*scale system*: a system for ranking measurements or members of a population on a scale, such as 1 to 5

*significance level*: Type I error expressed as a percentage, a probability, that measured values

*standard deviation*: a measure of spread; the positive square root of the variance

*standard error*: an estimate of the standard deviation of means that would be expected if a collection of means based on equal-sized samples of  $n$  items from the same population were obtained

*statistical inference*: conclusions drawn about a population using statistics

*statistics, descriptive*: measurements of population characteristics designed to summarize important features of a data set

*stratification*: the process of dividing a population into internally similar subgroups

*stratum*: one of the subgroups created prior to sampling in stratified random sampling

*streamside management area*: a designated area that consists of a waterbody (e.g., stream) and an adjacent area of varying width where management practices that might affect water quality, fish, or other aquatic resources are modified to protect the waterbody and its adjacent resources and to reduce the pollution effect of an activity on the waterbody

*Student's t test*: a statistical test used to test for significant differences between means when only two samples are involved

*subjectivity*: a characteristic of analysis that requires personal judgement on the part of the person doing the analysis

*target audience*: the population that a monitoring effort is intended to measure

*tillage*: the operation of implements through the soil to prepare seedbeds and rootbeds, control weeds and brush, aerate the soil, and cause faster breakdown of organic matter and minerals to release plant foods

*total maximum daily load*: a total allowable addition of pollutants from all affecting sources to an individual waterbody over a 24-hour period

*transformation, data*: manipulation of data such that it will meet the assumptions required for analysis

*Tukey's test*: a test to ascertain whether the interaction found in a given set of data can be explained in terms of multiplicative main effects

*unit sampling cost*: the cost of attributable to sampling a single population unit

*variance*: a measure of the spread of data around the mean

*watershed assessment*: an investigation of numerous characteristics of a watershed in order to describe its actual condition

*Wilcoxon's test*: a nonparametric test for use when a test is only between two samples

## INDEX

- accuracy* 2-10, 4-14
- allocation*
  - Neyman* 2-25, 2-27
  - proportional* 2-25
- analysis of variance* 3-4
  - rank-transformed* 3-4
- best professional judgement* 2-2
- bias, see error*
- BMP*
  - pass/fail rating system* 4-9
  - scale rating system* 4-9
- BMP implementation assessments*
  - site-specific* 1-2
  - watershed* 1-2
- camera format* 4-20
- Census of Agriculture* 2-13, 2-15, 2-16
- Clean Water Act*
  - Section 303(d)* 1-2
  - Section 319(h)* 1-2
- Coastal Nonpoint Pollution Control Program* 1-1
- Coastal Zone Act Reauthorization Amendments of 1990* 1-1
  - Section 6217(b)* 1-2
  - Section 6217(d)* 1-2
  - Section 6217(g)* 1-2
- complaint records* 2-16
- Computer-aided Management Practices System* 2-17
- Conservation Technology Information Center* 2-28
- consistency* 4-8, 4-12
- Cooperative Extension Service* 2-16
- cost of evaluations* 4-17
- County Transect Survey* 2-28
- County X example* 2-22, 2-24, 2-25
- data*
  - accessibility* 1-5, 1-6
  - electronic storage* 1-6
  - historical* 2-18
  - life cycle* 1-5
  - longevity* 1-5
  - management* 1-5
  - reliability* 1-5
  - transformation* 4-10
- Economic Research Section, USDA* 2-14
- error* 2-8
  - due to nonrespondents* 2-10
  - measurement* 2-8
  - reducing* 2-10
  - sampling* 2-10
  - Type I* 2-12
  - Type II* 2-12
- estimate*
  - point* 2-11
  - pooled* 3-3
- estimation* 2-11
- evaluations*
  - expert* 4-1, 4-7
  - information obtainable from* 4-1
  - mock* 4-8
  - self* 4-1, 4-13
  - site* 4-7
  - teams* 4-7
  - training for* 4-8
  - variable selection* 4-4
  - variables* 4-2
- farm numbers, USDA* 2-16
- Farm Service Agency* 2-17, 4-21
  - Aerial Photography Field Office* 4-21
- Field Office Computing System* 2-17
- finite population correction term* 2-18
- Friedman test* 3-4
- hypothesis*
  - alternative* 2-12
  - null* 2-12
- hypothesis testing* 2-12
- implementation rating* 4-9
- interviews, personal* 4-1
- Kruskal-Wallis test* 3-4
- Land Maps, county* 2-16
- Land Use and Land Cover, USGS* 2-16
- Management measures* 1-2
- Mann-Whitney test* 3-2
- monitor* 1-3
  - and CNPCPs* 1-3

*baseline 1-4*  
*compliance 1-4*  
*effectiveness 1-4*  
*implementation 1-3, 2-1*  
*objectives 1-4, 2-1*  
*project 1-4*  
*trend 1-3*  
*uses 1-4*  
*validation 1-4*  
*Monitoring Guidance for Determining the Effectiveness of Nonpoint Source Control Measures 1-4, 1-5, 2-1, 5-1*  
*National Agriculture Statistics Service 2-16, 4-14*  
*National Crop Residue Management Survey 2-28*  
*National Oceanic and Atmospheric Administration 1-1*  
*National Resources Inventory 2-15*  
*Natural Resources Conservation Service 4-21*  
*nonpoint source pollution, sources 1-1*  
*photographs 4-13*  
*aerial 2-18*  
*photography*  
*aerial 4-1, 4-20*  
*resolution 4-20*  
*scale 4-20*  
*population*  
*assumptions about 2-8*  
*sample, definition 2-2*  
*target, definition 2-2*  
*units, definition 2-2*  
*variation 2-8*  
*precision 2-10, 2-19*  
*presentations 5-1*  
*and time lag 5-1*  
*audience 5-2*  
*criteria 5-1*  
*format 5-2*  
*graphics 5-3*  
*major factors 5-2*  
*oral 5-2, 5-3*  
*resources 5-4*  
*written 5-2, 5-3*  
*quality assurance and quality control 1-4,*

*4-12*  
*quality assurance project plan 1-4*  
*questionnaires*  
*content 4-18*  
*design 4-17*  
*dichotomous 4-19*  
*elements 4-18*  
*layout 4-19*  
*multiple-choice 4-19*  
*objective 4-18*  
*open-ended 4-19*  
*ordering of questions 4-19*  
*phrasing 4-19*  
*pretest 4-19*  
*response format 4-19*  
*rating systems*  
*binary 4-9*  
*consistency 4-10*  
*overall rating 4-11*  
*scale 4-9*  
*terms 4-10*  
*sample size, estimation 2-18*  
*sampling*  
*cluster 2-5, 2-27*  
*per unit cost 2-25*  
*probabilistic 2-2*  
*simple random 2-3, 2-20*  
*strategy 2-13*  
*stratified random 2-3, 2-24*  
*systematic 2-8, 2-27*  
*timing 2-13*  
*scale, appropriate 1-3*  
*standard deviation, pooled 3-2*  
*statistical inference 2-2*  
*statistics*  
*confidence interval 2-11*  
*descriptive 2-11*  
*difference quantity 3-2*  
*overall mean 2-25*  
*parametric 2-19*  
*relative error 2-20*  
*significance level 2-12*  
*software 3-1*  
*stratum mean 2-25*  
*Student's t test 2-21, 3-2*  
*two-sample 3-2*

*Surveys*

*accuracy of information 4-14*

*mail 4-1, 4-14*

*telephone 4-1, 4-14*

*tests*

*one-sided, hypotheses 3-1*

*two-sided, hypotheses 3-1*

*Tukey's test 3-4*

*U.S. Environmental Protection Agency 1-1*

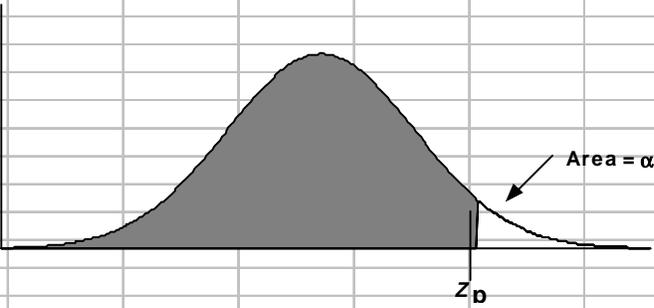
*Wilcoxon's test 3-3*

*[Note: Italicized page numbers indicate  
location of definitions of terms.]*

***APPENDIX A***

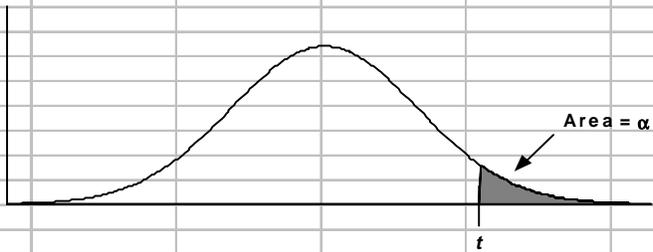
***Statistical Tables***

**Table A1. Cumulative areas under the Normal distribution (values of  $p$  corresponding to  $Z_p$ )**



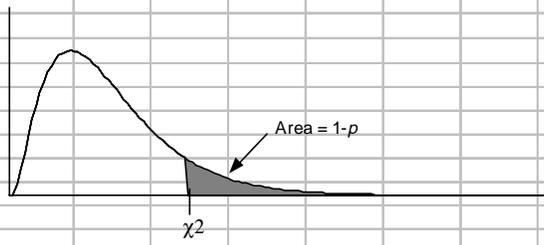
$Z_p$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

**Table A2. Percentiles of the  $t_{\alpha, df}$  distribution (values of  $t$  such that  $100(1-\alpha)\%$  of the distribution is less than  $t$ )**



<i>df</i>	$\alpha = 0.40$	$\alpha = 0.30$	$\alpha = 0.20$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.010$	$\alpha = 0.005$
1	0.3249	0.7265	1.3764	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.2887	0.6172	1.0607	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.2767	0.5844	0.9785	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.2707	0.5686	0.9410	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.2672	0.5594	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.2648	0.5534	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.2632	0.5491	0.8960	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.2619	0.5459	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.2610	0.5435	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.2602	0.5415	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.2596	0.5399	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.2590	0.5386	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.2586	0.5375	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.2582	0.5366	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.2579	0.5357	0.8662	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.2576	0.5350	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.2573	0.5344	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.2571	0.5338	0.8620	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.2569	0.5333	0.8610	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.2567	0.5329	0.8600	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.2566	0.5325	0.8591	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.2564	0.5321	0.8583	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.2563	0.5317	0.8575	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.2562	0.5314	0.8569	1.3178	1.7109	2.0639	2.4922	2.7970
25	0.2561	0.5312	0.8562	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.2560	0.5309	0.8557	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.2559	0.5306	0.8551	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.2558	0.5304	0.8546	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.2557	0.5302	0.8542	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.2556	0.5300	0.8538	1.3104	1.6973	2.0423	2.4573	2.7500
35	0.2553	0.5292	0.8520	1.3062	1.6896	2.0301	2.4377	2.7238
40	0.2550	0.5286	0.8507	1.3031	1.6839	2.0211	2.4233	2.7045
50	0.2547	0.5278	0.8489	1.2987	1.6759	2.0086	2.4033	2.6778
60	0.2545	0.5272	0.8477	1.2958	1.6706	2.0003	2.3901	2.6603
80	0.2542	0.5265	0.8461	1.2922	1.6641	1.9901	2.3739	2.6387
100	0.2540	0.5261	0.8452	1.2901	1.6602	1.9840	2.3642	2.6259
150	0.2538	0.5255	0.8440	1.2872	1.6551	1.9759	2.3515	2.6090
200	0.2537	0.5252	0.8434	1.2858	1.6525	1.9719	2.3451	2.6006
inf.	0.2533	0.5244	0.8416	1.2816	1.6449	1.9600	2.3264	2.5758

**Table A3. Upper and lower percentiles of the Chi-square distribution**



df	p											
	0.001	0.005	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995	0.999
1	....	....	....	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879	10.827
2	0.002	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597	13.815
3	0.024	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838	16.266
4	0.091	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860	18.466
5	0.210	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750	20.515
6	0.381	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548	22.457
7	0.599	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278	24.321
8	0.857	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955	26.124
9	1.152	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589	27.877
10	1.479	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188	29.588
11	1.834	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757	31.264
12	2.214	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300	32.909
13	2.617	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.736	27.688	29.819	34.527
14	3.041	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319	36.124
15	3.483	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801	37.698
16	3.942	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267	39.252
17	4.416	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718	40.791
18	4.905	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156	42.312
19	5.407	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582	43.819
20	5.921	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997	45.314
21	6.447	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401	46.796
22	6.983	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796	48.268
23	7.529	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181	49.728
24	8.085	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558	51.179
25	8.649	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928	52.619
26	9.222	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290	54.051
27	9.803	11.808	12.878	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645	55.475
28	10.391	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.994	56.892
29	10.986	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.335	58.301
30	11.588	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672	59.702
35	14.688	17.192	18.509	20.569	22.465	24.797	46.059	49.802	53.203	57.342	60.275	66.619
40	17.917	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766	73.403
50	24.674	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490	86.660
60	31.738	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952	99.608
70	39.036	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.43	104.21	112.32
80	46.520	51.172	53.540	57.153	60.391	64.278	96.578	101.88	106.63	112.33	116.32	124.84
90	54.156	59.196	61.754	65.647	69.126	73.291	107.57	113.15	118.14	124.12	128.30	137.21
100	61.918	67.328	70.065	74.222	77.929	82.358	118.50	124.34	129.56	135.81	140.17	149.45
200	143.84	152.24	156.43	162.73	168.28	174.84	226.02	233.99	241.06	249.45	255.26	267.54