



Data Quality Objectives Decision Error Feasibility Trials Software (DEFT) - USER'S GUIDE

EPA QA/G-4D

Quality

DISCLAIMER

The Data Quality Objectives Decision Error Feasibility Trials Software and documentation are provided “as is,” without guarantee or warranty of any kind, expressed, or implied. The Quality Staff, U.S. Environmental Protection Agency, or the United States Government will not be liable for any damages, losses, or claims consequent to use of the software or documentation.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not constitute or imply its endorsement, recommendation, or favoring by the U.S. Environmental Protection Agency or the United States Government.

FOREWORD

The U.S. Environmental Protection Agency (EPA) has developed the Data Quality Objectives Decision Error Feasibility Trials (DEFT) software (Windows Version 1.0) to support the application of the Data Quality Objectives (DQO) Process, a systematic planning process developed by EPA. The DQO Process is the Agency's preferred planning process when making decisions that involve selecting between opposing conditions. The DQO Process is an important tool for project managers and planners to define the type, quality, and quantity of data needed to make defensible decisions.

This document provides guidance to EPA program managers and planning teams. It does not impose legally binding requirements and may not apply to a particular situation based on the circumstances. EPA retains the discretion to adopt approaches on a case-by-case basis that differ from this guidance where appropriate. EPA may periodically revise this guidance without public notice.

This document is one of the *U.S. Environmental Protection Agency Quality System Series* documents. These documents describe the EPA policies and procedures for planning, implementing, and assessing the effectiveness of a quality system. Questions regarding this document or other *Quality System Series* documents should be directed to the Quality Staff at:

U.S. EPA
Quality Staff (2811R)
1200 Pennsylvania Avenue, NW
Washington, DC 20460
Phone: (202) 564-6830
Fax: (202) 565-2441
e-mail: quality@epa.gov

Copies of the EPA *Quality System Series* documents may be obtained from the Quality Staff or by downloading them from the Quality Staff Home Page:

<http://www.epa.gov/quality>

TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	i
CHAPTER 1. GETTING STARTED	1
1.1 INTRODUCTION	1
1.2 CONSIDERATIONS FOR DECIDING WHEN TO USE DEFT	3
1.3 INSTALLATION AND USE	6
1.4 RELATED SOFTWARE PRODUCTS	7
CHAPTER 2. USING THE SOFTWARE	9
2.1 ENTRY SCREENS	9
2.2 THE INPUT VERIFICATION SCREEN	16
2.3 THE DESIGN/DQO SUMMARY SCREEN	17
2.3.1 Modifying the DQOs	18
2.3.2 Selecting a New Sampling Design	19
2.3.3 Modifying Design-Specific Information	21
2.3.4 Specifying a Sample Size or Budget	21
2.3.5 Displaying the Decision Performance Goal Diagram	22
2.3.6 Saving the Current Information	22
2.3.7 Restoring the Original DQOs	22
2.4 THE DECISION PERFORMANCE GOAL DIAGRAM SCREEN	23
2.4.1 The Decision Performance Goal Diagram	24
2.4.2 Copying and Saving the Diagram	24
CHAPTER 3. EXAMPLES OF DEFT APPLICATIONS	25
3.1 TESTING A MEAN AGAINST A FIXED STANDARD	25
3.2 TESTING A PERCENTILE AGAINST A FIXED STANDARD	29
3.3 TESTING THE DIFFERENCE BETWEEN TWO MEANS	31
3.4 TESTING THE DIFFERENCE BETWEEN TWO PROPORTIONS	34
CHAPTER 4. EXTENDED APPLICATIONS OF DEFT	37
4.1 USING DEFT TO DETERMINE SAMPLE SIZES FOR ESTIMATION	37
4.2 USING DEFT TO RECONCILE SAMPLE DATA WITH PROJECT DQOS ...	38
4.2.1 Estimation Problems	38
4.2.2 Hypothesis Testing	38
4.3 USING DEFT FOR GRID SAMPLING DESIGNS	39
4.4 TESTING A PERCENTILE AGAINST A FIXED STANDARD	40

	<u>Page</u>
CHAPTER 5. ALGORITHMS USED IN DEFT	43
5.1 TESTING A MEAN AGAINST A FIXED STANDARD	43
5.1.1 Simple Random Sampling	43
5.1.2 Composite Sampling	44
5.1.3 Stratified Sampling	45
5.2 TESTING A PERCENTILE AGAINST A FIXED STANDARD	47
5.2.1 Simple Random Sampling	47
5.2.2 Stratified Sampling	48
5.3 TESTING THE DIFFERENCE BETWEEN TWO MEANS	50
5.4 TESTING THE DIFFERENCE BETWEEN TWO PROPORTIONS	51
5.5 ESTIMATING A POPULATION MEAN	52
5.6 ESTIMATING A POPULATION PROPORTION	52
REFERENCES	53

LIST OF FIGURES

	<u>Page</u>
Figure 1. DEFT and the DQO Process	2
Figure 2. Input Verification Screen	17
Figure 3. Example Design/DQO Summary Screen	18
Figure 4. Example Design Performance Goal Diagram Screen	23
Figure 5. Input Verification Screen for Example 1	27
Figure 6. Design/DQO Summary Screen for Example 1	27
Figure 7. Decision Performance Goal Diagram for Example 1	28
Figure 8. Input Verification Screen for Example 2	31
Figure 9. Design/DQO Summary Screen for Example 3	34
Figure 10. Decision Performance Goal Diagram for Example 4	36

LIST OF TABLES

	<u>Page</u>
Table 1. DQOs to Enter Into DEFT	10
Table 2. Sampling Designs Available in DEFT	19
Table 3. Summary of Design Information	21
Table 4. Using DEFT for Estimation	37
Table 5. Translating DQOs for Percentiles into DQOs for Proportions	41

CHAPTER 1

GETTING STARTED

1.1 INTRODUCTION

What is the DEFT software and this User's Guide?

The Decision Error Feasibility Trials (DEFT) software (Windows Version 1.0) was developed to assist in determining the feasibility of data quality objectives (DQOs) developed using the Data Quality Objectives Process. DEFT allows decision makers and members of a planning team¹ to quickly generate cost information about several simple sampling designs based on the DQOs. If necessary, the planning team can change the DQOs and evaluate the effect of these changes.

This user's guide contains detailed instructions on how to use DEFT. It is designed to supplement the *EPA Guidance on the Data Quality Objectives Process (QA/G-4)* (EPA, 2000c) which describes the DQO Process in detail. Therefore, this user's guide does not provide instructions on implementing the DQO Process, but instead contains information on how to use the DQOs generated through the DQO Process in DEFT.

How does DEFT determine feasibility?

DEFT uses the DQOs developed by a planning team to provide an estimate of sample number and cost. It determines feasibility based on economic considerations, not policy or other qualitative criteria.

How does this version of the software differ from the last [Version 4.0 (DOS)]?

This is the Windows Version 1.0 of DEFT. Major changes from the previous release of DEFT, which was DOS version 4.0, include:

- added capabilities for addressing hypotheses concerning population proportions and population percentiles;
- new routines for determining false acceptance error rates when a sample size is specified;
- the ability to save, print, and copy the decision performance goal diagram; and

¹ The DQO Process emphasizes using a multi-disciplinary team approach to offer different kinds of perspectives for reaching consensus about critical elements of the planning process, such as decision statements and decision error limits that are acceptable. An example DQO planning team might include a chemist, engineer, geologist, and toxicologist to support the project manager and QA officer.

- a Windows platform design (i.e., the software is now designed to run in a Windows environment instead of DOS).

Other minor changes have also been implemented, such as the ability to consider a fixed sampling cost or the use of a coefficient of variation instead of a standard deviation.

What is the DQO Process and how does DEFT assist in its implementation?

The DQO Process (Figure 1) is a 7-step systematic planning process developed by EPA (EPA 2000c). It provides a systematic procedure for defining the criteria that a data collection design should satisfy, including when to collect samples, where to collect samples, the tolerable level of decision errors for the study, and how many samples to collect. The DQO Process usually is conducted using a multi-disciplinary team approach.

Two difficult steps in the DQO Process are Step 6: Specify Tolerable Limits on Decision Errors, and Step 7: Optimize the Design. During Step 7, the DQOs are incorporated into a sampling design. If the DQOs are not feasible, it is necessary to iterate through one or more of the earlier steps of the DQO Process to revise or relax the criteria until the planning team is able to identify a sampling design that will meet the budget and generate data that are adequate for the decision. This iteration can be time-consuming and costly. DEFT reduces the need for this iteration by determining the feasibility of the DQOs before the final step of the DQO Process is implemented.

What are DQOs and how do they related to DEFT?

DQOs are qualitative and quantitative statements derived from the outputs of the first six steps of the DQO Process that:

- Clarify the study objective;
- Define the most appropriate type of data to collect;
- Determine the most appropriate conditions from which to collect the data; and

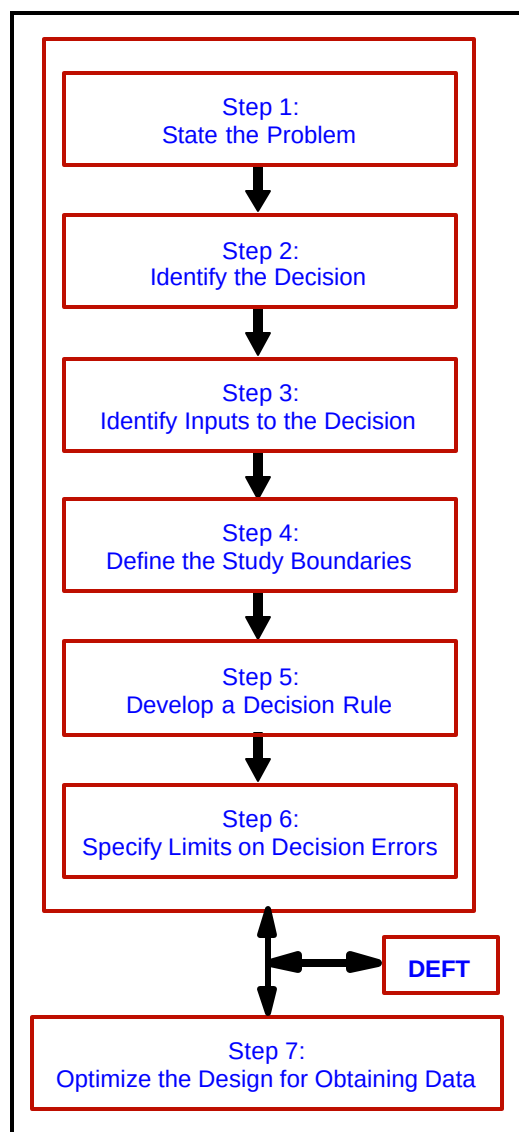


Figure 1. DEFT and the DQO Process

- Specify tolerable limits on decision errors which will be used as the basis for establishing the quantity and quality of data needed to support the decision.

The DQOs are then used to develop a scientific and resource-effective data collection design. DEFT helps determine the feasibility of the DQOs before a data collection design is developed.

Where can I find information on the DQO Process?

The DQO Process is described in the following two documents:

- *Guidance for the Data Quality Objectives Process (QA/G-4)* (EPA, 2000c)
- *The Data Quality Objectives Process for Hazardous Waste Site Investigations (QA/G-4HW)* (EPA, 2000a)

The first document provides general guidance; the second provides guidance for Superfund and Resource Conservation and Recovery Act applications.

This User's Guide does not describe the DQO Process or its outputs (DQOs) in detail because this information is contained in the documents listed above. It is strongly recommended that those who are unfamiliar with the DQO Process use the above documents and the help screens in DEFT to obtain more information on the DQO Process.

1.2 CONSIDERATIONS FOR DECIDING WHEN TO USE DEFT

When should I use DEFT during the DQO Process?

DEFT was developed primarily as a tool for the program manager and planning team to use before consulting with a statistician to develop a sampling design. The software is mostly used between Step 6: Specify Limits and Decision Errors and Step 7: Optimize the Design of the DQO Process. DEFT generates cost information about several simple sampling designs based on the outputs from the first six steps of the DQO Process. The planning team can use this information to evaluate whether the DQOs generate cost-feasible sample sizes before the sampling and analysis design team begins developing a final sampling design in the last step of the DQO process.

What are some additional applications of DEFT?

In addition to the standard application of the DQO process described above, DEFT can address alternative situations including:

- Estimation of population parameters;
- Reconciling project results with the DQOs;
- Testing hypothesis for percentiles; and
- Estimating sample sizes for grid sampling.

These topics are discussed in Chapter 4.

What planning should I do before using DEFT?

Before using DEFT, the planning team should complete Steps 1 through 6 of the DQO Process to define the DQOs required to achieve data of appropriate quality for its intended use. For example, the planning team should carefully define the decision rule to be tested (Step 5 of the DQO Process) in order to properly frame the use of the outputs from DEFT. The team should also carefully consider the consequences of decision errors and use this analysis to set the limits on decision errors (Step 6 of the DQO Process). Note that there are no rules for setting the limits on decision errors, and there is no easy way to select limits. EPA recommends setting the limits based on an analysis of the consequences. A lack of serious consideration about the consequences of making a false rejection or a false acceptance decision undermines the effectiveness of using DEFT to calculate sample size.

For what problems can DEFT generate sample sizes?

This version of DEFT will generate sample sizes using different sampling designs for the following questions:

- *Is the population mean greater/less than a fixed standard?* For example, does the mean concentration of hazardous waste in a drum exceed the regulatory threshold?
- *Is the population proportion/percentile greater/less than a fixed standard?* For example, in a storage yard where waste drums of many types have been placed, does the proportion of drums containing hazardous waste exceed 50%?
- *Is the difference between two population means significant?* For example, does the mean concentration of radioactive soil contaminants at the former fuel processing facility exceed the mean concentration of radioactive soil contaminants in the downtown city park?
- *Is the difference between two population proportions/percentiles significant?* For example, does the 98th percentile of daily PM₁₀ particulate concentration measurements taken during 1998 in St. Louis differ significantly from the same measurements taken during 1999?

Chapter 2 describes the sampling designs available in DEFT that can be used in designing a study to answer these questions and Chapter 3 provides examples of these applications using DEFT.

When shouldn't I use DEFT?

DEFT is not an expert system that considers the appropriateness of the DQOs or ensures an optimal (or even feasible) sampling design. Therefore, the software should not be used to validate the DQOs or to select a final sample size. DEFT should be used only to evaluate the feasibility of the DQOs generated through Step 6 of the DQO Process. In Step 7 of the DQO process, more sophisticated tools may be used to aid in design optimization, which may yield a lower-cost design.

There is no easy method for developing an optimal sampling design. Factors such as environmental medium, parameter of interest, contaminant of interest, and sampling boundaries as well as components of cost and variance all affect the choice of a sampling design. The application of DEFT for calculating a number of samples is straightforward for random sampling across space when the population remains relatively static over time. For example, DEFT is particularly applicable for calculating sample sizes when investigating slow-moving contaminants in surface soil because the samples can be collected randomly across space, and the concentrations do not change much over time. On the other hand, when investigating contaminants in ground water, sampling locations may need to be restricted to locations where wells currently exist, and the concentrations at any given location may vary greatly over relatively short periods of time, making the problem much more dynamic. DEFT is not designed to handle problems that involve streams of data over time, which require careful consideration of how correlations affect the analysis.

Volatile contaminants may present complex challenges because they may move quickly through an environmental medium, thereby creating a dynamic sampling problem in the field, while also posing difficulties in implementing analytical methods, thereby creating measurement problems in both the field and the laboratory. DEFT does not address these types of problems involving dynamic fate and transport for processes such as volatilization, retardation, or decay.

DEFT has capabilities that can be misused as well. A composite sampling design is applicable for testing hypotheses concerning the mean; however, it is not applicable for testing hypotheses concerning percentiles. An optimal sampling design accounts for all factors relevant to the problem at hand, and is practical, feasible, and satisfies the DQOs. DEFT cannot take all of these factors into account, hence it should not be used to determine the sampling design or final sample size.

What statistical assumptions does DEFT make?

For the one-population cases, it is assumed that the action level is fixed (i.e., the action level is a known quantity) and that there is only one infinite (or extremely large) population. For the two-

population cases, it is assumed that both sample sizes are large and that the variability of the two populations are approximately equal. For example, the one-population case tests whether the mean concentration of a contaminant at a site exceeds a health-based standard. A two-population case tests whether or not the mean concentration of a contaminant at an industrial site exceeds the concentration at a nearby residential site (each site is considered to have a separate population of interest, and samples from both sites are required for the calculation). DEFT also assumes that a design comparable to either a simple random sampling design or a stratified simple random sampling design is feasible. For example, DEFT is not designed to be used to determine the number of drinking water wells to be selected when the sample wells will be selected on the basis of hydrogeology instead of selected randomly.

What quality control procedures has DEFT been subjected to?

This software was peer reviewed and incorporates revisions recommended by the reviewers. It has been tested extensively including an analysis of the inputs, processes and expected outputs for each routine. This testing is documented in the *Test Plan for the Data Quality Objectives Decision Error Feasibility Trials (DQO/DEFT) Software* (Flanagan and Aanstoos, 2001).

1.3 INSTALLATION AND USE

What computers will run DEFT?

Any computer running Windows 95 or Windows NT or their successors should be able to run DEFT. Its memory and disk requirements are negligible compared with other Windows applications. The minimum graphic resolution required is 640x480, with 800x600 recommended.

How do I install DEFT?

To install DEFT, save the file g4d-final.exe to your computer. Then

1. Select "Run" from the Taskbar Start menu
2. Enter "x:\g4d-final.exe" substituting the location where you saved the g4d-final.exe file for x. If you received the DEFT software on a floppy disk, enter the drive letter for x.

The DEFT software is then installed in the in the default folder c:\deft.

How do I start DEFT?

After DEFT has been installed:

1. Click on the *Start* button found on the Windows task bar
2. Select “Run”
3. Enter “c:\deft\deft.exe” by the ‘Open’ prompt and then press “OK”.

Alternatively, you can run DEFT directly from a floppy disk by entering “x:\deft.exe” (where x is the letter of your disk drive) in Step 3. Also, you can manually create an icon from which to launch DEFT and place it on your desktop and/or Start menu. See your Windows documentation for directions.

When DEFT is launched it displays an opening screen with general information about the program’s purpose and proper use. After you click the *OK* button, DEFT then prompts you for your initial DQO inputs which are described in Chapter 2.

How do I skip the entry screens?

DEFT prompts the user to enter the information from the DQO Steps 1 through 6 based on a series of five entry screens. The first entry screen determines the parameter of interest (see Section 2.1). To skip the remaining entry screens, click on the *Summary* button in the bottom right corner of the second entry screen. This will take you directly to the Design/DQO Summary Screen (Section 2.3) using the default values contained in DEFT.

How do I start a new analysis?

To exit DEFT through the Design/DQO Summary Screen, press the *Exit* button on the bottom right-hand corner of the Design/DQO Summary Screen. DEFT will then ask if you want to “*Start a new DQO analysis?*” Pressing the *No* button will exit the program.

Where can I get help?

An electronic copy of this User’s Guide is accessible by clicking on the *Help* button contained on each DEFT window or dialog box.

How do I exit DEFT?

To exit the software at anytime, click on the close button (the X in the upper right hand button) of an DEFT window or dialog box.

1.4 RELATED SOFTWARE PRODUCTS

In addition to DEFT, there are several free computer-based programs available that are related to the DQO Process. Each of the programs listed below operate on an IBM PC with a VGA monitor.

- *Visual Sample Plan (VSP)* (dco.pnl.gov/VSP/Index.htm) — VSP is designed to select the number of samples and provide random or gridded sampling locations, based on various sampling schemes, overlaid on a site map. VSP includes SampTOOL, an Internet tool to guide the user in selecting an appropriate sampling design given the type of problem and environmental medium (surface soil, subsurface soil, sediment, surface water, groundwater, air, biota, contaminated material, and surface).
- *DQO-PRO* (www.acs-envchem.duq.edu/dqopro.htm) — DQO-PRO helps a user understand the significance of the DQOs by showing the relationship between the numbers of samples and DQO parameters such as confidence levels for false acceptance and false rejection decision errors; tolerable error versus analyte concentration, standard deviation, etc.; and confidence levels versus sampling area grid size. DQO-PRO can be used to calculate the number of samples required to meet the DQOs or satisfy the desired confidence interval widths.
- *GEOPACK* (www.epa.gov/ada/csmos/models/geopack.html) — GEOPACK is a comprehensive geostatistical software system for conducting analysis of the spatial variability of one or more random functions. GEOPACK is menu-driven, user-friendly, requires a minimum number of input data, and includes on-line help.

In addition to the programs listed above, there are numerous statistical packages that are useful in implementing the DQO process.

CHAPTER 2 USING THE SOFTWARE

DEFT uses the DQOs defined in Steps 1 - 6 of the DQO Process to determine their feasibility based on several simple sampling designs. This is done in three steps:

- 1) the information from DQO Steps 1-6 are entered into DEFT (Entry Screens),
- 2) this information is then verified and saved (the Input Verification Screen), and
- 3) finally DEFT uses this information with different sampling designs to estimate sample size and costs (the Design/DQO Summary Screen).

These steps are described in detail in this Chapter and examples are provided in Chapter 3.

Note: The information below describes information required for the software and constraints related to the software. It does not describe the DQO Process or its outputs in any detail. For this information, consult the EPA Guidance for the Data Quality Objectives Process (EPA, 2000c).

2.1 ENTRY SCREENS

DEFT prompts the user to enter the information from the DQO Steps 1 through 6 based on a series of five entry screens. This information is described below and summarized in Table 1. For each item, the relevant step of the DQO Process is provided.

Note: To skip all entry screens after the first, click on the Summary button in the bottom right corner of the second entry screen. This will take you directly to the Design/DQO Summary Screen (see Section 2.3) using the default values contained in DEFT and described below.

The software automatically starts with a simple random sampling design so that the user only enters the minimum amount necessary to generate a sample size. On each entry screen there is a *NEXT* button which must be clicked in order to accept the values shown on that screen and advance to the next screen. (Note: The *ENTER* key will NOT advance to the next screen; you must click *NEXT*.) On all but the first entry screen there also appears a *BACK* button which will allow you to back up to the previous screen. When each screen initially appears, the fields on it are filled in with default values which you can either accept or change. To move between fields for the purpose of entering or changing values, you may either click in the desired field with the mouse or use the *TAB* key to move to the next field.

Table 1. DQOs to Enter Into DEFT

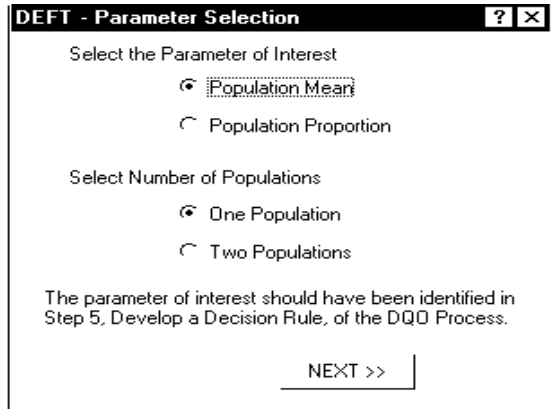
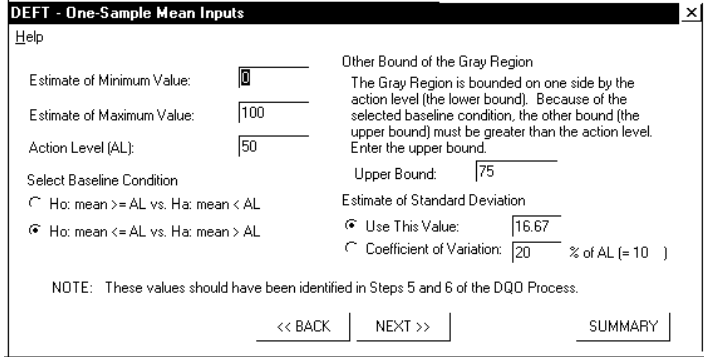
DQOs	Valid Entries	DQO Step	Entry Screen
Parameter of Interest	<u>Parameter</u> Mean or Proportion <u>Number of Populations</u> One or Two	5	
Minimum Value of the Parameter of Interest (MIN)	<u>For Means</u> MIN < MAX <u>For Proportions</u> 0 if 1 population -1 if 2 populations	6	
Maximum Value of the Parameter of Interest (MAX)	<u>For Means</u> MAX > MIN <u>For Proportions</u> 1 for both 1 and 2 populations	6	

Table 1. DQOs to Enter Into DEFT

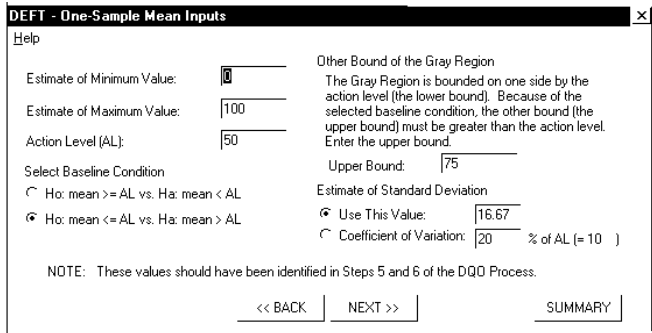
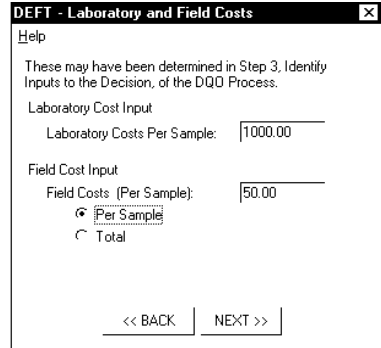
DQOs	Valid Entries	DQO Step	Entry Screen
Action Level (AL)	$\text{MIN} < \text{AL} < \text{MAX}$ For two populations, $\text{AL} = 0$.	5	
Baseline and Alternative Conditions	1. H_0 : parameter $\geq$ AL vs. H_a : parameter $<$ AL 2. H_0 : parameter $\leq$ AL vs. H_a : parameter $>$ AL	6	
Bounds of the Gray Region (GR)	$\text{MIN} < \text{GR} < \text{AL}$ or $\text{AL} < \text{GR} < \text{MAX}$	6	
Estimate of Standard Deviation (SD)	$0 < \text{SD} < 2 * (\text{MAX} - \text{MIN})$	3	
Sampling and Analysis Costs	Costs ≥ 0	3	

Table 1. DQOs to Enter Into DEFT

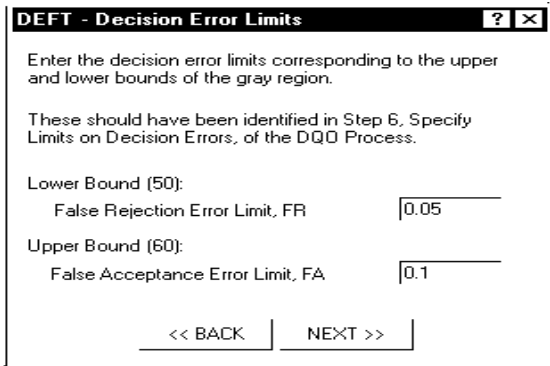
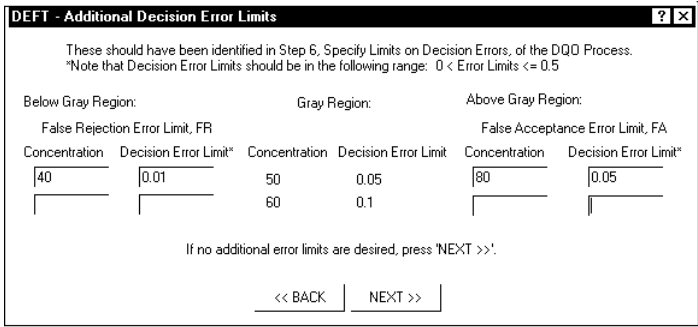
DQOs	Valid Entries	DQO Step	Entry Screen
False Rejection (FR) and False Acceptance (FA) error limits at the bounds of the gray region	$0 < FR \leq 0.5$ $0 < FA \leq 0.5$	6	

Table 1. DQOs to Enter Into DEFT

DQOs	Valid Entries	DQO Step	Entry Screen
Additional Error Limits Above and Below the Gray Region (x = Concentration/Proportion, p = Probability associated with x)	<u>Below the Gray Region</u> $\text{MIN} < x < \text{GR}$ or $\text{MIN} < x < \text{AL}$ <u>Above the Gray Region</u> $\text{GR} < x < \text{MAX}$ or $\text{AL} < x < \text{MAX}$ (Limit of two additional entries above and below.) $0 < p \leq 0.5$	6	

- **Parameter of Interest:** The parameter of interest is a descriptive measure of some characteristic or attribute of the statistical population. Defining the parameter of interest consists of two parts - selecting the parameter (valid entries are mean and proportion²) and identifying the number of populations (either a single population for comparing the parameter to a fixed standard, or two populations for determining the difference between the parameters from each population). Note: Once the parameter of interest has been selected, it may not be changed. (DQO Process Step 5)
- **Minimum and Maximum Values (Range) of the Parameter of Interest:** If the parameter of interest is a population mean or the difference between two population means, estimates of the minimum and maximum possible values are necessary for scaling, graphing, and computing default values. The range of the population mean must fall within the range of possible concentrations. If the parameter of interest is a proportion, the minimum value is automatically set to 0 and the maximum value is automatically set to 1. If the parameter of interest is the difference between two proportions, the minimum value is automatically set to -1 and the maximum value is automatically set to 1. These values are referred to throughout the rest of DEFT as the “minimum” and “maximum” concentrations. (DQO Process Step 6)
- **Action Level:** The action level is a value that provides the criterion for selecting among alternative actions. For the one sample case, this software assumes that the action level is fixed, such as a regulatory standard. For the two sample case, the default action level is zero to indicate “no difference between the two population parameters.” (DQO Process Step 5)
- **Baseline and Alternative Hypotheses:** The baseline (H_0) and alternative (H_a) conditions are used to identify which error is a false rejection error and which is a false acceptance error. There are two choices for the baseline and alternative conditions:
 - 1) H_0 : Parameter $\geq$ Action Level vs. H_a : Parameter $<$ Action Level and
 - 2) H_0 : Parameter $\leq$ Action Level vs. H_a : Parameter $>$ Action Level.

Once selected, these may not be changed. Because the alternative condition is the opposite of the baseline condition, DEFT will only state the baseline condition after this selection is made. (DQO Process Step 6)

²Note that determining sample sizes for testing hypotheses concerning percentiles is equivalent to determining sample sizes for hypotheses concerning proportions. Therefore, only proportions are displayed in the software. Chapter 3 describes the process of translating hypotheses concerning percentiles into hypotheses concerning proportions.

- **Gray Region:** The gray region is a range of possible parameter values where the consequences of a false acceptance decision error are relatively minor. The gray region is bounded on one side by the action level and on the other side by that parameter value where the consequences of making a false acceptance decision error begin to become significant. The program will automatically determine whether this bound should be less than or greater than the action level, based on your choice of baseline condition. (DQO Process Step 6)
- **Estimate of Standard Deviation:** When the parameter of interest is a population mean or the difference between two population means, an estimate of the standard deviation of the population of interest is necessary for computing sample sizes. (If the parameter of interest is a proportion or the difference between two proportions, an estimate of the standard deviation is not required.) The standard deviation is the square root of the variance. An estimate of this value may be available from a pilot study or the user can use the DEFT default value.³ If the difference between two means is the parameter of interest, DEFT assumes that the standard deviations of both populations are equal. The standard deviation must be greater than zero and less than or equal to two times the range of the population parameter (i.e., the standard deviation must be less than or equal to two times the maximum concentration minus the minimum concentration). Alternatively, the standard deviation may be specified as a fixed percentage of the action level. This percentage is sometimes referred to as the coefficient of variation, and this option may be chosen from the standard deviation entry screen. (DQO Process Step 3)
- **Sampling And Analysis Costs:** The average unit cost of analyzing a sample and the average unit cost of collecting a sample in the field are used to compute the total cost of a sampling design. The average cost of analyzing a sample is referred to as the “laboratory cost” and the average unit cost of collecting a sample is referred to as the “field cost” in DEFT. Both the laboratory and field costs must be greater than or equal to zero. For the field sampling cost, an alternative to specifying a per-sample cost is to specify the total cost for all samples regardless of their number. For the case where sample collection and measurement analysis are one process, you should enter the cost

³If there is no estimate available, then the (Maximum Concentration - Minimum Concentration) / 6 may be used as a rough approximation of the standard deviation. The default value assumes the population is normally distributed, i.e., that 99% of the values are represented by the mean $\pm 3\sigma$, and, therefore, the max-min is equivalent to 6σ . Note that this approximation is based on the range of the population, not the range of the population parameter and it should only be used if there is absolutely no other information available. The approximation is only valid for the purposes of DEFT, i.e., determining the feasibility of the DQOs. You should consult a statistician before developing an estimate for use in the actual sampling design.

of this process as the either the laboratory cost or the field cost and set the other cost equal to zero. (DQO Process Step 3)

- ***Probability Limits on Decision Errors for the Bounds of the Gray Region:*** Limits on the probability of false rejection error and a false acceptance error must be specified in order to compute sample sizes. DEFT will prompt you to enter these probabilities after it automatically determines which error is a false acceptance error and which is a false rejection error. Both probabilities must be greater than 0 and less than or equal to 0.5. (DQO Process Step 6)
- ***Additional Limits on Decision Errors:*** The DQO Process allows the planning team to set additional limits on decision errors besides those on the bounds of the gray region, although this is not required. In general, tolerable limits for making a decision error should decrease as the consequences of a decision error become more severe farther away from the Action Level. For example, the economics of making a false acceptance decision error may become more important as the true concentration is farther from the Action Level and the limits on decision error may be reduced at this point. DEFT will allow you to enter up to two additional limits below the lower bound of the gray region and up to two additional limits above the upper bound of the gray region. All probabilities must be greater than 0 and less than or equal to 0.5. (DQO Process Step 6)

2.2 THE INPUT VERIFICATION SCREEN

Once the DQOs are entered, DEFT displays the Input Verification Screen (Figure 2). This screen is used to verify the inputs from the entry screens. Any incorrect values can be corrected at this time by pressing the *Change* button underneath that value. For example, press the *Change Input Values* button to change the minimum possible value for the parameter of interest. Once the information has been verified and corrected if necessary, press *NEXT* to advance to the Design/DQO Summary Screen.

Note: *This is the last chance to adjust the minimum, maximum, and baseline condition as these can not be changed on the Design/DQO Summary Screen.*

The information on the Input Verification Screen is saved as the “Original DQOs,” as this information represents the DQOs of the planning team. This gives you the opportunity to select a sampling design, evaluate the performance of the design based on these original DQOs, then modify the DQOs to improve the performance of the sampling design. You may then select a different sampling

DEFT - Input Verification			Decision Error Limits		
Inputs			concentration	prob(error)	type
Parameter of Interest:	Mean				
Minimum Mean Concentration:	0				FR
Maximum Mean Concentration:	100				
Action Level:	50		40	.01	FR
Baseline condition:	mean <= 50.00		50*	0.05	FR
Gray Region:	50 - 60		60*	0.01	FA
Standard Deviation:	20		80	0.05	FA
					FA
Change Input Value(s)					
Cost of Sample					
Analysis in the Laboratory:	1000.00	Change Costs	* Use the Change Input Value(s) button to change these values.		
Collection in the Field:	50.00				
		NEXT >>			

Figure 2. Input Verification Screen

design and restore the original DQOs in order to evaluate the new sampling design's performance against the original DQOs (i.e., the DQOs of the planning team).

2.3 THE DESIGN/DQO SUMMARY SCREEN

After you verify the DQOs on the Input Verification Screen, DEFT estimates sample size, computes the total cost, and verifies that the decision error limits are satisfied using a Simple Random Sampling Design. This information is then displayed on the Design/DQO Summary Screen (Figure 3) along with the DQOs and information on the current sampling design. You can investigate the feasibility of the DQOs and save your analysis by:

- Modifying the DQOs (Section 2.3.1)
- Selecting a New Sampling Design (Section 2.3.2)
- Modifying Design-Specific Information (Section 2.3.3)
- Specifying a Sample Size or Budget (Section 2.3.4)
- Displaying the Decision Performance Goal Diagram (Section 2.3.5)
- Saving the Current Information (Section 2.3.6)
- Restoring the Original DQOs (Section 2.3.7)

DEFT has a sample size limitation of 30,000 total samples. If the sample size required to meet the DQOs exceeds this number, DEFT informs you of this in a pop-up error message. You will then need change the DQOs (such as reduce the false rejection error rate and/or the false acceptance error

DEFT (Beta) - Design/DQO Summary

Sampling Parameters
 Sampling Design: Simple Random Sampling Change Sampling Design
 Design-Specific Inputs:
 No design-specific inputs for Simple Random Sampling.

Number of Samples (per population): 36 Update

Total Cost: \$37800.00

DQO Input Summary

Costs
 Laboratory Costs Per Sample: 1000.00
 Field Costs Per Sample: 50.00 Change Costs

Input Values
 Parameter of Interest: Mean
 Action Level: 50
 Gray Region: 50 - 60
 Baseline condition: mean <= 50.00
 Standard Deviation: 20 Change Input Value(s)

Decision Error Limits

concentration	prob(error)	type
		FR
40	0.01	FR
50*	0.05	FR
60*	0.1	FA
	0.05	FA
		FA

* Use Change Input Value(s) button Update

Graph Save Original DQOs Exit

Figure 3. Example Design/DQO Summary Screen

rate or increase the width of the gray region) before continuing with the DQO constraint feasibility analysis.

The sample size formulas used in DEFT guarantee that the decision error limits set on the bounds of the gray region are satisfied. However, the sample size formulas do not account for any additional decision error limits. Therefore, DEFT verifies that these additional limits are satisfied. If a limit is not satisfied, the limit is marked “NS” in the Decision Error Limits Table.

2.3.1 Modifying the DQOs

- The **minimum**, **maximum**, and **baseline condition** can not be changed at this point (see Section 2.2).
- The **Action Level** can be modified by selecting the *Change Input Value(s)* button. This will display a screen where this item may be changed. Once the change is made, press the *NEXT* button to return to the Design/DQO Summary Screen. Sample sizes and costs are automatically updated.
- The **Other Bound of the Gray Region** can be modified by selecting the *Change Input Value(s)* button. This will display a screen where this item may be changed.

Once the change is made, press the *NEXT* button to return to the Design/DQO Summary Screen. Sample sizes and costs are automatically updated.

- The **Estimate of the Standard Deviation** can be modified by selecting the *Change Input Value(s)* button. This will display a screen where this item may be changed. Once the change is made, press the *NEXT* button to return to the Design/DQO Summary Screen. Sample sizes and costs are automatically updated.
- **Additional Decision Error Limits** can be adjusted by changing the values in the space provided and new limits may be added by entering them in the space available (in this case, both a concentration and probability must be entered). After changing or entering new limits, you must press the *Update* button to determine the new sample size and if the additional Decision Error Limits are satisfied. If a limit is not satisfied, the limit is marked “NS” in the Decision Error Limits Table.
- **Laboratory cost** and **field cost** estimates can be changed by selecting the *Change Costs* button to reflect the potential costs of a different sampling and/or analysis method. This will display a screen where these items may be changed. Once your changes are made, press the *NEXT* or *BACK* button to return to the Design/DQO Summary Screen. Sample sizes and costs are automatically updated.

2.3.2 Selecting a New Sampling Design

DEFT always starts with a simple random sampling design but allows you to consider other sampling designs which may perform more efficiently. To investigate other sampling designs, press the *Change Sampling Design* button. You will then be prompted to select from the relevant sampling designs shown in Table 2. For hypotheses about a single population mean, you may select either composite sampling or stratified sampling. For hypotheses about a single population proportion, you may select stratified sampling. These sampling designs, along with design-specific information, are discussed in Chapter 4.

Table 2. Sampling Designs Available in DEFT

	One Population	Two Populations
Mean	- Simple Random Sampling - Composite Sampling - Stratified Sampling	- Simple Random Sampling
Proportion	- Simple Random Sampling - Stratified Sampling	- Simple Random Sampling

The first time a sampling design is selected, you are prompted to enter the design-specific information. For example, an estimate of the proportion of measurement variability to the total variability is required for the composite random sampling design. The design-specific information is described below by sampling design and summarized in Table 3. The information first specified is saved as part of the “Original DQOs” (see Section 2.2). You may then modify the design-specific information to evaluate the DQOs in relation to this sampling design.

Composite Sampling: DEFT uses composite samples with a simple random sampling design, which is referred to as “composite sampling.” The software computes the number of composite samples required to meet the DQOs based on a given number of individual samples per composite. To determine the number of composite samples, DEFT requires the following design-specific information:

- An estimate of the ratio of the relative standard deviation of measurement error to total standard deviation. This ratio must be less than one and greater than zero.
- The number of individual samples to be mixed to form each composite sample. This number should be greater than one.
- The cost of combining the individual samples to form a composite.

Stratified Sampling: DEFT uses stratification with a simple random sampling design within each strata, which is referred to as “stratified sampling.” The software computes the number of samples required per strata to meet the DQOs. To estimate the sample size required for a stratified design, DEFT requires the following design-specific information:

- The number of strata. This number must be greater than one and less than six. (There is a limit of six strata in DEFT because the software only demonstrates feasibility of the DQOs and five strata should be sufficient for this purpose.)
- A weight factor (weight) for each stratum. The stratum weight is the proportion of the volume or area of the environmental medium contained in the stratum in relation to the total volume or area of the study site. The sum of the strata weights must be 1, so the program automatically computes the weight of the final stratum. The default weight corresponds to equal weighing among the strata.
- If the population parameter is a single mean, an estimate of the standard deviation is needed for each stratum. The estimated standard deviation for each stratum must be greater than zero and less than two times the range of the population parameter, and the default value is the estimated total standard deviation.

- If the population parameter is a single proportion, an estimate of the stratum proportion is needed for each stratum. Each estimate must be greater than zero and less than one.

Table 3. Summary of Design Information

Sampling Design	Design Information	Limits	Default
Tests for a Single Mean			
Composite Sampling	Ratio (r) of measurement SD to total SD	$0 < r < 1$	0.10
	Number of individual samples (m) per composite	$1 < m$	4
	Cost (c) of composting	$c > \$0.00$	\$5
Stratified Sampling	Number of strata (L)	$2 \leq L \leq 5$	2
	Stratum weights (W_h)	$0 < W_h < 1$	Equal weights
	Stratum standard deviation ($\hat{\sigma}_h$)	$0 < \hat{\sigma}_h < 2(\text{max-min})$	$\hat{\sigma}_{\text{Total}}$
Tests for a Single Proportion			
Stratified Sampling	Number of strata (L)	$2 \leq L \leq 5$	2
	Stratum weights (W_h)	$0 < W_h < 1$	Equal weights
	Estimated stratum proportions (P_h)	$0 < P_h < 1$	Action Level

2.3.3 Modifying Design-Specific Information

To update design-specific information, select the *Change Design Inputs* button. DEFT will prompt you through changing the design specific information described in Section 2.3.2.

2.3.4 Specifying a Sample Size or Budget

Sometimes the total budget available for sampling and analysis may be set in advance. With this information, you can determine the total number of samples allowed and then determine what decision error limits are possible within this budget. For this situation, you specify the number of

samples and then DEFT adjusts the probability of a false acceptance decision error to meet your sample size.⁴ You can do this by changing either the “*Number of Samples*” field or the “*Total Cost*” field (in which case DEFT will compute the number of samples afforded by this cost using the cost input data). To change either of these fields, first click in its box and edit the value currently appearing. Then press the *Update* button to use the new value. Note that it is not valid to change both the Total Cost and the Number of Samples before clicking the *Update* button — if you do, then DEFT will ignore your cost entry and use your number of samples in computing the result. You may enter any sample size greater than 1 and less than or equal to 30,000.

2.3.5 Displaying the Decision Performance Goal Diagram

You can display the Decision Performance Goal Diagram by pressing the *Graph* button in the bottom, right-hand corner of the Design/DQO Summary Screen. This diagram is discussed in detail in Section 2.4.

2.3.6 Saving the Current Information

Once it has been determined that the DQOs are feasible for a sampling design, you may save the DQOs and design information to a plain text file by pressing the *Save* button in the lower right-hand corner of the Design/DQO Summary Screen. This text file can then be imported into any standard word processor.

The first time the Save option is clicked the user is prompted for a file name into which the text summary is saved. If the file name chosen is the same as an existing file, you will be asked if you want to overwrite the existing file. If you indicate you do not want to overwrite, you will be asked to select a new name. Once a new filename has been selected, all subsequent uses of the SAVE option (until the program is exited) cause a new summary to be appended to the same file. A new file is not created for each summary, and previous results of the current session are not overwritten.

2.3.7 Restoring the Original DQOs

Selecting the *Original DQOs* button on the bottom right-hand corner of the Design/DQO Summary Screen will restore the original DQOs (Section 2.2). This is useful for comparing variations of several sampling designs. For instance, if a sampling design is too expensive to satisfy the DQOs, so you may relax some constraints to obtain a feasible sample size. After this is complete, you may want to examine the performance of another sampling design using the Original DQOs. This option saves you from re-entering the original information manually.

⁴When specifying a sample size, DEFT may adjust the sample size to be slightly larger than the value provided by you, due to the way DEFT performs the calculations under these conditions.

2.4 THE DECISION PERFORMANCE GOAL DIAGRAM SCREEN

To display the Decision Performance Goal Diagram and associated options, press the *Graph* button in the bottom, right-hand corner of the Design/DQO Summary Screen. This will bring up a Performance Goal Diagram Screen like the one shown in Figure 4. This screen contains the following:

- **Decision Performance Goal Diagram** - See Section 2.4.1 for a discussion of this diagram.
- **DQO Summary button** - This button will return you to the Design/DQO Summary Screen described in Section 2.3.
- **Print Graph button** - This button lets you print the diagram using the standard Windows print dialog.
- **Copy Graph button** - This button allows you to copy the graph to other Windows applications (Section 2.4.2).
- **Save Graph button** - This button will let you save the diagram to a picture file (Section 2.4.2)
- **Help** - This option displays help and version information.
- **Exit** - This option will exit DEFT.

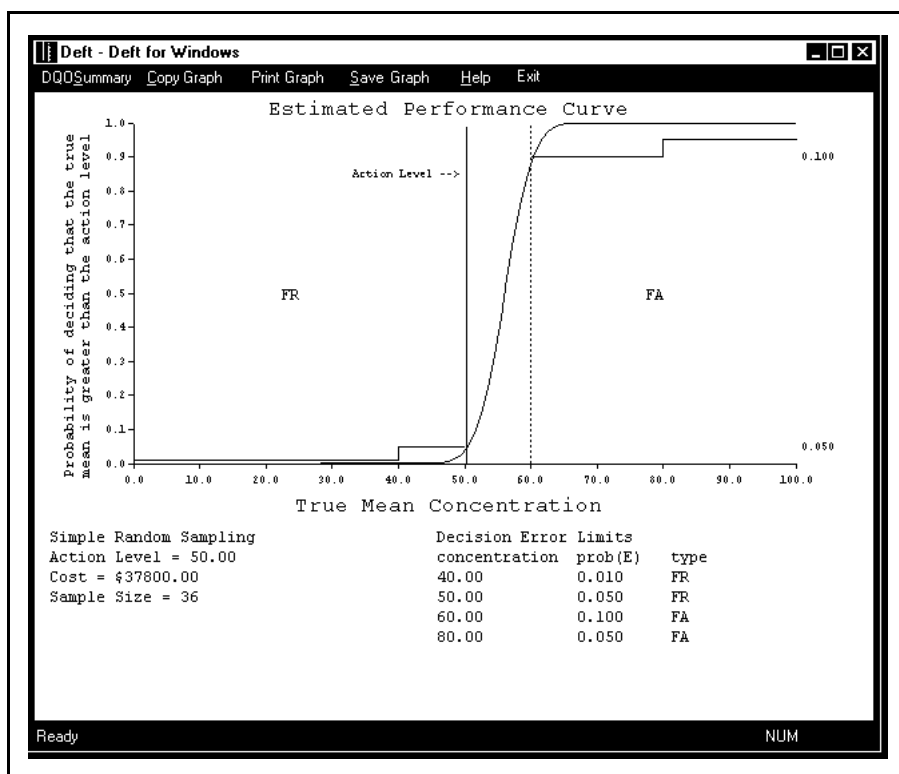


Figure 4. Example Design Performance Goal Diagram Screen

2.4.1 The Decision Performance Goal Diagram

DEFT has an option available to view the DQOs and design performance graphically on a separate screen. This is done using a decision performance goal diagram with the performance curve overlaid. The performance goal diagram summarizes the gray region, the limits on decision errors, and the action level. Information on the sample size and cost of the design are also summarized on this screen. The performance curve can be used to determine how well a design performs in relation to the limits on decision errors.

The sample size reported by DEFT is always greater than or equal to 2 so that an estimate of the standard deviation can be calculated from the data collected. In this case, the performance curve may satisfy a more stringent false acceptance decision error rate at the bound of the gray region than that displayed by the software. If so, use the option to specify the sample size (Section 2.3.4) to select a sample size of 2 to determine the exact decision error rate satisfied by the two samples.

Note that the performance curve displayed by DEFT is an estimate of the performance curve of the design⁵. Therefore, the curve may appear to show that a decision error limit is satisfied when it is not. The calculations performed in the software to determine if a particular error limit is satisfied are more accurate than those used to draw the performance curve. Therefore, you should use the text indication (“NS”) in the Decision Error Limits Table to determine whether or not a limit is satisfied.

The performance curve is always the probability of deciding that the true parameter value (such as a mean or proportion) is greater than the action level, irrespective of the directions of the baseline and alternative hypotheses. Thus the curve always starts at the lower left hand corner and rises to the upper right hand corner. This is in contrast to a statistical power curve. For more information regarding the performance curve, see the *Guidance on the Data Quality Objectives Process (EPA QA/G-4)* (EPA, 2000c).

2.4.2 Copying and Saving the Diagram

To save the Performance Goal Diagram, click on *Copy Graph* on the menu bar of the Performance Goal Diagram Screen. This will copy the current diagram to Windows Clipboard which allows it to be pasted into any Windows application that supports the pasting of bitmap pictures. To paste the diagram, open the Windows application and use the “*Paste*” command. This diagram can then be saved in any format allowed by the Windows application you are using. The diagram can also be printed or saved to a file as a Window bitmap (.bmp), using the appropriate options available on the menu bar.

⁵DEFT uses a normal distribution to approximate the power curve which is actually based on a non-central t-distribution.

CHAPTER 3

EXAMPLES OF DEFT APPLICATIONS

This chapter contains four examples where DEFT is used to determine the feasibility of the DQOs. Each example explains the planning team's choice of DQOs and shows, screen-by-screen, what inputs were entered into DEFT. Actual text from DEFT is shown in italics and quotations; actual buttons from DEFT are shown in italics. Note: The purpose of the examples is to show how DEFT may be used to generate data based on various scenarios and assumptions. Although the examples refer to various EPA requirements and standards, these are used for illustrative purposes only; they are not examples of EPA-approved decision error limits or other data quality objectives.

3.1 TESTING A MEAN AGAINST A FIXED STANDARD – CADMIUM IN FLY ASH

A waste incineration facility located in the Midwest routinely removes fly ash from its flue gas scrubber system and disposes of it in a municipal landfill. Previously, the waste fly ash was not hazardous according to Federal environmental regulations. Due to treatment of a new waste stream, representatives of the incineration company are concerned that the waste fly ash now contains hazardous levels of cadmium. If the fly ash meets the Federal standard and thus is considered non-hazardous, it can be disposed of in a municipal landfill. If not, then the ash would have to be sent to a higher-cost special hazardous waste disposal landfill.

Entry Screens

Parameter of Interest: The planning team considered the population mean to be the appropriate parameter of interest because there is a large mixing effect when collecting the ash. The planning team is interested in looking at potential scenarios in preparation for making a decision for each load of fly ash so that only hazardous loads are disposed of in a special landfill. Hence, each load of fly ash is a separate population for which a decision is needed.

Entry Screen 1: DEFT – Parameter Selection. Select 'Population Mean' under "*Select the Parameter of Interest*" and select 'One Population' under "*Select Number of Populations.*" Press the *NEXT* button.

Minimum and Maximum Values (Range) of the Parameter of Interest: The possible minimum value of cadmium is 0.0 mg/L and the team agreed to use a possible maximum value of 2.0 mg/L for planning purposes.

Action Level: The regulatory standard for cadmium concentration in the leachate resulting from Toxicity Characteristic Leaching Procedure (TCLP) extraction is 1.0 mg/L.

Baseline and Alternative Conditions: The baseline condition is specified under the regulations as the case where the fly ash is considered hazardous (Baseline: Mean $\geq$ 1.0 mg/L) and the alternative condition as the case where the waste is not considered hazardous (Alternative: Mean $<$ 1.0 mg/L).

Gray Region: The gray region is the area adjacent to the Action Level of 1.0 mg/L where the planning team considers the consequences of a false acceptance decision error to be minimal. A false acceptance error would result in unnecessary and costly disposal in a special landfill. The planning team specified a width of .25 mg/L for the gray region based on their preferences to guard against false acceptance decision errors at .75 mg/L.

Estimate of Standard Deviation: The planning team conducted a pilot study of the fly ash to determine the variability in the concentration of cadmium within loads of fly ash. This study showed that each load of fly ash is fairly homogenous and the standard deviation in the concentration of cadmium among samples within loads of ash is approximately 0.6 mg/L.

Entry Screen 2: DEFT – One-Sample Mean Inputs. Enter 0.0 for “*Estimate of Minimum Value,*” 2.0 for “*Estimate of Maximum Value,*” and 1.0 for “*Action Level.*” Under “*Select Hypotheses*” select ‘ H_0 : mean $\geq$ AL vs. H_a : mean $<$ AL.’ Enter 0.75 for the ‘*Lower Bound*’ and 0.6 for “*Estimate of Standard Deviation*” by “*Use this Value.*” Press the *NEXT* button.

Sampling And Analysis Costs: The cost of selecting a sample is \$10. The cost of TCLP analysis is \$150 a sample.

Entry Screen 3: DEFT – Laboratory and Field Costs. Enter 150.00 for “*Laboratory Costs per Sample,*” 10.00 for “*Field Costs per Sample,*” and check the “*Per Sample*” box. Press the *NEXT* button.

False Rejection Error Limit: Regulations specify a 5% false rejection decision error. Consequences of a false rejection error (deciding that the waste is not hazardous when it is truly hazardous) are that the incineration company disposes of the hazardous waste in a sanitary landfill, possibly endangering human health and the environment.

False Acceptance Error Limit: The planning team set the maximum tolerable probability of making a false acceptance error at 20% at the bound of the gray region (0.75 mg/L). Since the baseline condition and the false rejection error limit are fixed by regulation, this is the only error limit the planning team can adjust and its primary consequence is economic. The consequence of a false acceptance error is an increase in unnecessary expenses from using a special disposal facility when it is not needed.

Entry Screen 4: DEFT – Decision Error Limits. Enter 0.20 for the “*False Acceptance Error Limit*” (under “*Lower Bound*”) and 0.05 for “*False Rejection Error Limit*” (under “*Upper Bound*”) Press the *NEXT* button.

Additional Limits on Decision Errors: The planning team wanted to use additional decision error limits and set the maximum tolerable probability of making a false acceptance error at 10% when the true mean is below 0.25 mg/L.

Entry Screen 5: DEFT – Additional Decision Error Limits. In the “*Below Gray Region*” section, enter 0.25 under “*Concentration*” and 0.10 under “*Decision Error Limit.*” Press the *NEXT* button.

Input Verification Screen

The input verification screen (Figure 5) is used to verify the inputs from the previous entry screens.

Input Verification Screen. Use the appropriate *Change* button to make any changes. Press the *NEXT* button.

Inputs		Decision Error Limits		
Parameter of Interest:	Mean	concentration	prob(error)	type
Minimum Mean Concentration:	0			FA
Maximum Mean Concentration:	2			FA
Action Level:	1	0.25	0.1	FA
Null Hypothesis (Ho):	mean >= 1	0.75*	0.2	FA
Gray Region:	0.75 - 1	1*	0.05	FR
Standard Deviation:	0.4			FR
				FR

Cost of Sample Analysis in the Laboratory: 150.00
Collection in the Field: 10.00

Change Costs

NEXT >>

* Use the Change Input Value(s) button to change these values.

Figure 5. Input Verification Screen for Example 1

Design/DQO Summary Screen

The DEFT – Design/DQO Summary Screen (Figure 6) shows that under simple random sampling, the minimum number of observations needed to satisfy the decision error limits is 37 and the total cost is \$5,920. The incineration company would like to hold the study costs to around \$2,500 per load of fly ash, so the planning team decided to investigate composite sampling to see if it meets their DQOs.

DEFT - Design/DQO Summary

Help

Sampling Parameters

Sampling Design: Simple Random Sampling

Design-Specific Inputs:

No design-specific inputs for Simple Random Sampling.

Number of Samples: 37

Total Cost: \$5920.00

Update Screen Data

Update

DQO Input Summary

Costs

Laboratory Costs Per Sample: 150.00

Field Costs (Per Sample): 10.00

Change Costs

Input Values

Parameter of Interest:	Mean	concentration	prob(error)	type
Action Level:	1			FA
Gray Region:	0.75 - 1			FA
Baseline Condition:	mean >= 1.000	0.25	0.1	FA
Standard Deviation:	0.6	0.75*	0.2	FA
		1*	0.05	FR
				FR
				FR

* Use Change Input Value(s) button

Change Input Value(s)

Graph Save Original DQOs Exit

Figure 6. Design/DQO Summary Screen for Example 1

For composite sampling, the planning team needed to consider some additional parameters. Using the results of the pilot study, the variability among subsamples within a composite sample is expected to be negligible. Thus, the measurement standard deviation was estimated to be a very small proportion of the total standard deviation (.0001). Also, the planning team decided that the load of fly ash could be easily divided into eight strata of equal size. To form each composite sample, the containers will be divided into eight strata of equal size, a sample taken randomly from within each stratum, and then the eight samples would be composited. The planning team assumed the cost for the compositing would be minimal so \$0 was used as the compositing cost.

Design/DQO Summary Screen. Press the *Change Sampling Design* button. On the box that appears, select “*Composite Sampling*” and press the *OK* button. The “*Composite Design Inputs*” box will appear. Enter 0.0001 for “*Measurement SD/Total SD,*” 8 for “*Aliquots per Composite Sample,*” and 0.0 for “*Cost for Compositing the Aliquots.*” Press the *OK* button.

Press the *Graph* button to see the Decision Performance Goal Diagram shown in Figure 7. To return to the Design/DQO Summary Screen, press the *DQO Summary* button.

For this composite sampling design, the number of samples is 6 and the cost is \$1,380. Therefore, these data quality objectives are feasible and the planning team can continue with Step 7 of the DQO Process, Optimize the Design.

During Step 7 of the DQO Process, the planning team decided to take eight composite samples to improve the likelihood that their error limits would be satisfied for every load of fly ash. This design came to a total cost of \$1,840 and the false acceptance error rate has decreased from 0.2 (20%) to 0.082 (8.2%). Note: this is one potential scenario, the planning team could have specified a different width for the gray region or a different false acceptance error rate depending on their concern about costs. However, the risk to human health (controlled by the selection of the baseline condition and the false rejection error rate) can not change as this is specified by EPA through the regulations.

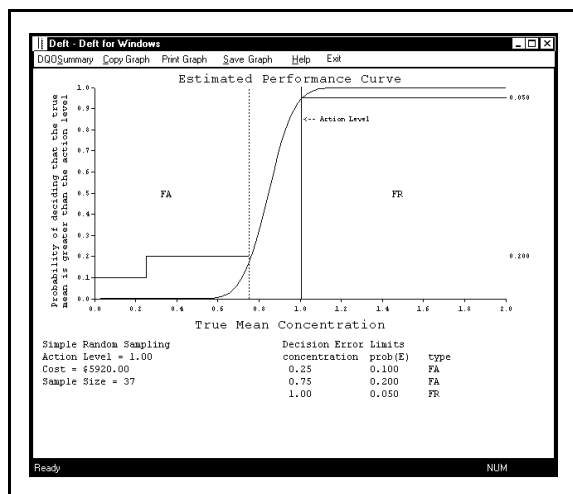


Figure 7. Decision Performance Goal Diagram for Example 1

Design/DQO Summary Screen. To changing the composite sample size from 6 to 8, enter 8 for the “*Number of Samples.*” Then press the *Update* button to see the effect on cost and the false acceptance error rate.

3.2 TESTING A PERCENTILE AGAINST A FIXED STANDARD – URBAN AIR QUALITY COMPLIANCE

Representatives of a metropolitan area in the Northeast want to determine if their area will meet the $PM_{2.5}$ ($PM_{2.5}$, particulate matter of aerodynamic diameter less than or equal to 2.5 micrometers) standard over the next year. Federal regulations specify the 24-hour standard $PM_{2.5}$ as a concentration of no more than $65 \mu g/m^3$, based on the 3-year average of the annual 98th percentiles. Their sampling network consists of three fixed-site multiple-filter gravimetric devices for measuring daily concentrations (24-hr averages) of $PM_{2.5}$. Each of the three monitors measures concentrations once every 3 days, for a total of 365 measurements per year.

Entry Screens

Parameter of Interest: The population parameter of interest to the planning team was the 98th percentile of $PM_{2.5}$ concentrations, as specified in the regulations. However, the sample size required to estimate a population percentile is usually determined by calculating the sample size needed to estimate the corresponding population proportion (see Section 4.4). Thus, the planning team formulated their study design requirements in terms of estimating a population proportion.

Entry Screen 1: DEFT – Parameter Selection. Select ‘Population Proportion’ under “Select the Parameter of Interest” and select ‘One Population’ under “Select Number of Populations.” Press the *NEXT* button.

Minimum and Maximum Values (Range) of the Parameter of Interest: For tests of a single proportion, the minimum value is 0 and the maximum is 1.

Action Level: Because the 24-hour standard for $PM_{2.5}$ is a concentration of no more than $65 \mu g/m^3$ based on the 3-year average of the annual 98th percentiles, the action level is the 98th percentile, or a proportion of 0.98.

Baseline and Alternative Conditions: The planning team designated the baseline condition as the case of non-attainment (98th percentile at or above $65 \mu g/m^3$) and the alternative condition as the case of attainment. Determining whether the 98th percentile of $PM_{2.5}$ is greater than or equal to $65 \mu g/m^3$ is equivalent to determining whether the proportion of daily concentrations less than or equal to $65 \mu g/m^3$ is less than or equal to 0.98. Thus, the planning team agreed on the equivalent conditions: $H_0: P \leq 0.98$ vs. $H_a: P > 0.98$, where P represents the proportion of daily concentrations less than or equal to $65 \mu g/m^3$.

Gray Region: The gray region is the area adjacent to the Action Level (0.98) where the planning team considers the consequences of a false acceptance decision error to be minimal. A false acceptance error would result in the implementation of unnecessary and costly control strategies. The planning team specified a width of .015 for the gray region based on their preferences to guard against false acceptance decision errors, thereby establishing a gray region of 0.98 to 0.995.

Entry Screen 2: DEFT – One-Sample Proportion Inputs. Enter 0.98 for the “Action Level”, select ‘ H_0 : proportion $\leq$ AL vs. H_a : proportion $>$ AL’ under “Select Hypotheses,” and enter 0.995 for the “Upper Bound.” Press the *NEXT* button.

Sampling And Analysis Costs: There are no costs for the air sampling and analysis because the air monitoring system is already operational.

Entry Screen 3: DEFT – Laboratory and Field Costs. Enter 0.0 for both the “Laboratory Costs per Sample” and the “Field Costs per Sample.” Check the “Per Sample” box. Press the *NEXT* button.

False Rejection Error Limit: The planning team agreed that the tolerable false rejection error rate should be no more than 10%. While lowering the tolerable bound on such error was desirable, the planning team believed that a significantly smaller error rate was unobtainable for all but the most extensive and costly network designs.

False Acceptance Error Limit: The team wanted to protect against unnecessary and costly control strategies (i.e., incorrectly failing to reject the baseline condition), but was willing to tolerate a greater probability of making this false acceptance decision error. They decided the limit should be no more than 30% at proportions above the upper bound of the gray region.

Entry Screen 4: DEFT – Decision Error Limits. Enter 0.1 for “False Rejection Error Limit” (under “Lower Bound”) and 0.3 for the “False Acceptance Error Limit” (under “Upper Bound”). Press the *NEXT* button.

Additional Limits on Decision Errors: No additional limits on decision errors were specified by the planning team.

Entry Screen 5: DEFT – Additional Decision Error Limits. Press the *NEXT* button.

Input Verification Screen

The input verification screen (Figure 8) is used to verify the inputs from the previous entry screens.

Input Verification Screen. Use the appropriate *Change* button to make changes. Once the information is correct, press the *NEXT* button.

Design/DQO Summary Screen

The Design/DQO Summary Screen shows that the minimum number of observations needed to satisfy the decision error limits with a simple random sampling design is 209. If each of the three monitors in the network continues to sample once every 3 days, the planning team will have a total of 365 samples for the year which will be more than sufficient.

The planning team then continues with Step 7 of the DQO Process, “Optimize the Design.” Once the sampling design is optimized, the planning team documents the design and quality objectives and submit this information to the appropriate regulatory body for approval.

Design/DQO Summary Screen. Press the *Save* button to save the Design/DQO Summary Screen to a file. Press the *Exit* button to exit DEFT.

Figure 8. Input Verification Screen for Example 2

3.3 TESTING THE DIFFERENCE BETWEEN TWO MEANS – CYANIDE CONTAMINATION IN GROUND WATER

EPA is concerned that storage of waste materials at an abandoned factory had resulted in environmental contamination. Test wells at the site showed high concentrations of cyanide were found in these wells, ranging from 1.5 parts per million (ppm) to over 300 ppm (at the plant site). Prior activities at the site included moving the waste to a containment area and removal of the surface and subsurface soils where the waste had been stored. EPA will now determine if cyanide in the ground water has decreased to levels comparable to an area that was never contaminated (a reference site). For purposes of this example, sampling costs will be less than \$12,500. Note: sampling costs may vary depending upon the particular scenario and applicable requirements.

Entry Screens

Parameter of Interest: The planning team considered the difference in population means to be the appropriate parameter of interest. The team is comparing the remediated site to a reference site to determine if the remediated site's levels of cyanide are the same as the reference levels; so, there are two populations of interest.

Entry Screen 1: DEFT – Parameter Selection. Select 'Population Mean' under "*Select the Parameter of Interest*" and select 'Two Population' under "*Select Number of Populations.*" Press the *NEXT* button.

Minimum and Maximum Values (Range) of the Parameter of Interest: Because the soil was remediated, the most that the reference levels could exceed the site levels is by 20 ppm and the most that site levels could exceed reference levels is by 100 ppm. Therefore, the range of the difference is -20 to 100 ppm.

Action Level: If the remediation methods worked, then the two sites should be approximately equal in average contamination levels, so the action level is "no difference" or 0.

Baseline and Alternative Conditions: The planning team designated the baseline condition as the case where the remediated site remains contaminated $H_0: \text{mean}_1 - \text{mean}_2 \geq 0$ where mean_1 is the mean of the remediated site and mean_2 is the mean of the reference area).

Gray Region: The gray region is the area adjacent to the Action Level where the planning team considers the consequences of a false acceptance decision error to be minimal. A false acceptance error would result in the implementation of unnecessary and costly higher-stage remediation efforts. The planning team specified a width of 5 ppm for the gray region based on their preferences to guard against false acceptance decision errors, thereby establishing a gray region of -5.0 to 0.0.

Estimate of Standard Deviation: The planning team conducted a pilot study of the cyanide in the remediated wells and determined the standard deviation be 3.5 ppm. They also assumed the standard deviation in the reference wells would be the same.

Entry Screen 2: DEFT – Two-Sample Mean Inputs. Enter -20.0 for "*Estimate of Minimum Value,*" 100.0 for "*Estimate of Maximum Value,*" and 0.0 for "*Action Level.*" Under "*Select Hypotheses*" select ' $H_0: \text{mean}_1 - \text{mean}_2 \geq \text{AL}$ vs. $H_a: \text{mean}_1 - \text{mean}_2 < \text{AL}$.' Enter -5.00 for the "*Lower Bound*" and 3.5 for "*Estimate of Standard Deviation*" by "*Use this Value.*" Press the *NEXT* button.

Sampling And Analysis Costs: The cost of selecting a sample is \$150 and the cost of analyzing a sample is \$500.

Entry Screen 3: DEFT – Laboratory and Field Costs. Enter 150.00 for “*Laboratory Costs per Sample*,” 500.00 for “*Field Costs per Sample*,” and check the “*Per Sample*” box. Press the *NEXT* button.

False Rejection Error Limit: The planning team determined that the tolerable false rejection decision error rate should be no more than 1% when the baseline condition is true. The planning team firmly believed that to protect human health and the environment it was necessary to have such a small false rejection error rate.

False Acceptance Error Limit: The team wants to protect against unnecessary and costly higher-stage remediation efforts (i.e. incorrectly failing to reject the baseline condition), and decided the error limit should be no more than 5%.

Entry Screen 4: DEFT – Decision Error Limits. Enter 0.05 for the “*False Acceptance Error Limit*” (under “*Lower Bound*”) and 0.01 for “*False Rejection Error Limit*” (under “*Upper Bound*”). Press the *NEXT* button.

Additional Limits on Decision Errors: No additional limits on decision errors were specified by the planning team.

Entry Screen 5: DEFT – Additional Decision Error Limits. Press the *NEXT* button.

Input Verification Screen

The input verification screen is used to verify the inputs from the previous entry screens.

Input Verification Screen. Use the appropriate *Change* button to make any changes. Once the information is correct, press the *NEXT* button.

Design/DQO Summary Screen

The DEFT – Design/DQO Summary Screen (Figure 9) shows that the minimum number of observations needed to satisfy the decision error limits with a simple random sampling design is 40 (20 per site) and the total cost is \$26,000. Since EPA would like to hold the study costs under \$12,500,

the DQOs are not feasible. To reduce the cost, the planning team decided to change the false acceptance error limit to 15% and the false rejection error limit to 5%.

Design/DQO Summary Screen. In the *Decision Error Limits* section under *prob(error)*, change 0.05 to 0.15 (next to -5.0) and change 0.01 to 0.05 (next to 0.0). Press the *Update* button.

The number of samples (per population) is now 36 (18 per site) and the total cost has decreased to \$11,700. The planning team now has DQOs that are feasible. These DQOs will be used in Step 7 of the DQO Process. Then the final sampling design and DQOs will be documented in a Quality Assurance Project Plan.

Figure 9. Design/DQO Summary Screen for Example 3

3.4 TESTING THE DIFFERENCE BETWEEN TWO PROPORTIONS – DIOXIN CONTAMINATION

At a hazardous waste site, EPA investigators must determine whether an area suspected to be contaminated with dioxin needs to be remediated. The potentially contaminated area (area 1) will be compared to a reference area (area 2) to see if dioxin levels in area 1 are greater than those in area 2. An inexpensive surrogate probe will be used to test each individual sample.

Entry Screens

Parameter of Interest: Because the presence of dioxin in the area will be represented as a proportion of all samples that are contaminated, the population proportion is the appropriate parameter of interest. The planning team is comparing the suspect site to a reference site to determine if the suspect site is contaminated with dioxin so there are two populations of interest.

Entry Screen 1: DEFT – Parameter Selection. Select ‘Population Proportion’ under “*Select the Parameter of Interest*” and select ‘Two Population’ under “*Select Number of Populations.*” Press the *NEXT* button.

Minimum and Maximum Values (Range) of the Parameter of Interest: When comparing two population proportions, the minimum difference is -1 and the maximum difference is 1.

Action Level: The health standard for dioxin is 1 ppb.

Baseline and Alternative Conditions: The planning team designated the baseline condition as the case where the suspect site is clean and the alternative condition as the case where the suspect site is contaminated (H_0 : $\text{mean}_1 - \text{mean}_2 \leq 0$ where mean_1 is the mean of the remediated site and mean_2 is the mean of the background area).

Gray Region: The gray region is the area adjacent to the Action Level of 0 where the planning team considers the consequences of a false acceptance decision error to be minimal. A false acceptance error would result in the failure to clean up the contaminated site, thereby posing risks to human health and the environment. The planning team specified a width of 0.10 for the gray region based on their preferences to guard against false acceptance decision errors, thereby establishing a gray region of 0.00 to 0.10.

Entry Screen 2: DEFT – Two-Sample Proportion Inputs. Under “*Select Hypotheses*” select ‘ H_0 : $\text{prop1} - \text{prop2} \leq \text{AL}$ vs. H_a : $\text{prop1} - \text{prop2} > \text{AL}$.’ Enter 0.10 for “*Upper Bound*.” Press the *NEXT* button.

Sampling And Analysis Costs: The cost of selecting a sample is \$0 and the cost of analyzing a sample is \$17.

Entry Screen 3: DEFT – Laboratory and Field Costs. Enter 17.00 for “*Laboratory Costs per Sample*,” 0.00 for “*Field Costs per Sample*,” and check the “*Per Sample*” box. Press the *NEXT* button.

False Rejection Error Limit: The team wants to protect against unnecessary and costly remediation efforts so it has specified that the false rejection decision error rate should be no more than 10%.

False Acceptance Error Limit: The planning team firmly believed that to protect human health and the environment it was necessary to have a small false acceptance error rate. The team decided that the probability of making this false acceptance decision error should be no more than 5%.

Entry Screen 4: DEFT – Decision Error Limits. Enter 0.05 for “*False Rejection Error Limit*” (under “*Lower Bound*”) and 0.10 for the “*False Acceptance Error Limit*” (under “*Upper Bound*”). Press the *NEXT* button.

Additional Limits on Decision Errors: No additional limits on decision errors were specified by the planning team.

Entry Screen 5: DEFT – Additional Decision Error Limits. Press the *NEXT* button.

Input Verification Screen

The input verification screen is used to verify the inputs from the previous entry screens.

Input Verification Screen. Use the appropriate *Change* button to make any changes. Once the information is correct, press the *NEXT* button.

Design/DQO Summary Screen

The “DEFT – Design/DQO Summary” Screen shows that the minimum number of observations needed to satisfy the decision error limits with a simple random sampling design is 644 (322 per site) and the total cost is \$10,948. Since this is close to the actual budget allocated of \$10,000, the planning team will proceed to Step 7 of the DQO process. (Note: there are other decision error limits that would meet project constraints such as cost. This is just one example of feasible DQOs for this problem.) During Step 7, the planning team expects that the actual budget will be met using these DQOs when the sampling design is optimized.

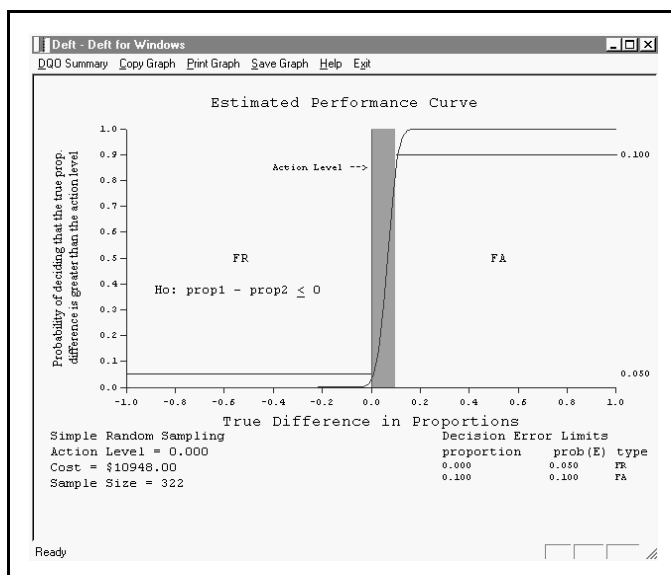


Figure 10. Decision Performance Goal Diagram for Example 4

Design/DQO Summary Screen. Press the *Graph* button to see the Decision Performance Goal Diagram shown in Figure 10. To return to the Design/DQO Summary Screen, press the *DQO Summary* button.

CHAPTER 4

EXTENDED APPLICATIONS OF DEFT

4.1 USING DEFT TO DETERMINE SAMPLE SIZES FOR ESTIMATION

Although DEFT has been designed to determine the minimum sample size needed for hypothesis testing problems, it also can be used to determine the minimum sample size needed to obtain a sufficiently precise estimate of a population mean or a population proportion. In either case, the estimation problem is to determine the minimum sample size needed to produce a 100 (1 - α)% confidence interval estimate (e.g., a 95% confidence interval where $\alpha = 0.05$) of a population mean or proportion such that the maximum width of the confidence interval is less than or equal to a pre-specified value. That is, the confidence interval estimate should be no wider than “the point estimate $\pm$ 1/2 of pre-specified width,” where the point estimate is the sample mean or proportion.

To use DEFT to estimate the sample size needed to obtain a confidence interval of a specified width, make the following adjustments to the DQOs entered into DEFT. First select an action level — for a mean, this can be any value you like; for a proportion, this should be a preliminary estimate of the proportion. Then set the maximum width for the confidence interval, determine what accuracy you would like the confidence interval (for example, a 90% confidence interval or a 95% confidence interval), and develop estimates of the standard deviation and sampling and analysis costs. The remaining DQOs are specified in Table 4.

Table 4. Using DEFT for Estimation

DQO	Use
Minimum Value	Action Level minus twice the maximum width (W), e.g., AL - 2W
Maximum Value	Action Level plus twice the maximum width (W), e.g., AL + 2W
Baseline Condition	mean $\leq$ AL
Gray Region	Action Level + 1/2 the maximum width
False Acceptance Error Rate	0.50
False Rejection Error Rate	α /2, so if you want a 90% confidence interval, $\alpha = .1$, and the false rejection error rate is 0.05
Additional Limits on Decision Errors	None

4.2 USING DEFT TO RECONCILE SAMPLE DATA WITH PROJECT DQOs

During data quality assessment, it is necessary to determine if the project DQOs have been achieved in order to properly interpret the results of the study. Guidance regarding this process is provided in *EPA Guidance on Data Quality Assessment (EPA/QA G-9)* (EPA, 2000b). The process of data quality assessment has two distinct phases:

1. Determining if the assumptions underlying the estimation and/or hypothesis testing procedures are satisfied; and
2. Determining if the sample size (number of observations) is sufficient to make a decision based on the data obtained.

DEFT can be used for the second phase, i.e., determining if the sample size is sufficiently large, assuming that the assumptions underlying the estimation or hypothesis testing procedure are satisfied (see Chapter 4 of EPA, 2000b). Use of DEFT for this determination is discussed in the below first for estimation and then for hypothesis testing.

4.2.1 Estimation Problems

If the study objective was to estimate a population parameter, the required sample size should have been determined by specifying the maximum allowable width for a confidence interval estimate of the parameter. In this case, one can determine if the sample size was sufficiently large by simply observing if the width of the actual confidence interval is less than or equal to the pre-specified maximum width. If so, the DQOs have been satisfied; if not, the sample size should be increased.

4.2.2 Hypothesis Testing

If the study objective was to test a hypothesis, the required sample size should have been determined by specifying the maximum probabilities for making decision errors — false rejection and false acceptance of the baseline condition. If the baseline condition is rejected, the probability of false rejection has been controlled by the critical value of the test statistic used to specify the threshold at which the decision was made to reject the baseline condition, and there is no need to determine if the sample size was adequate. However, if the baseline condition is not rejected, one needs to determine if the sample size is sufficiently large to provide adequate protection against false acceptance error. Guidance for determining if the sample size is sufficient is provided in Chapter 5 of the *EPA Guidance on Data Quality Assessment (EPA/QA G-9)* (EPA, 2000b) and DEFT can be used to implement the calculations that are required. The following discusses how to use DEFT to perform these computations for each type of hypothesis test that the software supports.

Testing a mean against a fixed standard with simple random sampling. Use the actual standard deviation from the sample in place of the pre-specified population standard deviation (Section 2.1) to calculate the required sample size. The DQO requirements have been satisfied if the actual sample size is greater than or equal to the required sample size.

Testing a mean against a fixed standard with stratified simple random sampling. Use the actual standard deviation of the sample for each stratum instead of the pre-specified population standard deviation for each stratum to calculate the sample size required for each stratum (see Section 2.3.3). The DQOs have been satisfied if the actual sample size is greater than or equal to the required sample size for each stratum.

Testing a proportion (or percentile) against a fixed standard with simple random sampling. In this case, the DQOs depend on the population proportions that define the action level (specified in the baseline condition) and the boundary of the gray region for which the maximum false acceptance error rate was specified. Therefore, the power of the test does not depend on the actual sample proportion (or percentile), and the adequacy of the DQOs does not need to be verified.

Testing a proportion (or percentile) against a fixed standard with stratified simple random sampling. For each stratum, use the actual proportion from the sample in place of the pre-specified population proportion to calculate the sample size required for each stratum (see Section 2.3.3). The DQOs have been satisfied if the actual sample size is greater than or equal to the required sample size for each stratum.

Testing the difference between two means with simple random sampling. Use the pooled standard deviation [Box 3.3-1 of the *EPA Guidance on Data Quality Assessment (EPA/QA G-9)* (EPA, 2000b)] of the sample in place of the pre-specified population standard deviation to calculate the required sample size (see Section 2.3.3). The DQOs have been satisfied if the actual sample size for each population is greater than or equal to the required sample size.

Testing the difference between two proportions with simple random sampling. Use the actual proportions from the sample in place of the pre-specified population proportions to calculate the required sample size (see Section 2.3.3). The DQOs have been satisfied if the actual sample size for each population is greater than or equal to the required sample size.

4.3 USING DEFT FOR GRID SAMPLING DESIGNS

The simple random sampling option may be used to estimate the sample size for a randomized systematic sampling design (grid sampling with a random starting point). To do so, use DEFT to

develop a sample size and cost estimate for a Simple Random Sample (Chapter 2) and adjust the sampling protocols accordingly.

4.4 TESTING A PERCENTILE AGAINST A FIXED STANDARD

A population parameter commonly of interest in environmental studies is an upper percentile (upper proportion) because this parameter conservatively protects against extreme health affects. The median, a measure of central tendency, is the 50th percentile. A percentile provides information regarding extreme values and is useful when the population contains a large number of values less than the analytical method detection limit.

A population percentile represents the percentage of elements of a population having values less than or equal to some threshold C . Thus, if C is the 95th percentile of a population, the values of 95% of the elements of the population are less than or equal to C and 5% of the population have values greater than C . For example, if the 95th percentile of a chemical distribution is 40 ppm, then 95% of the concentration levels are less than or equal to 40 ppm.

Determining sample sizes for hypotheses concerning population percentiles is equivalent to determining sample sizes for hypotheses concerning the corresponding population proportions. As a result, DEFT only considers proportions. Therefore, to use DEFT for percentiles, the DQO inputs to DEFT must be transformed. Consider the decision to determine whether the 95th percentile of the cadmium concentration in a load of fly ash waste is less than 1 mg/L. The baseline condition in this case is that the 95th percentile of cadmium is less than or equal to 1 mg/L. Now, instead of considering the population (the load of fly ash) to consist of differing levels of cadmium, consider the population to consist of a binary variable that is '1' if the cadmium level at a particular point in the load of fly ash is 1 mg/L or less and is '0' if the level is above 1 mg/L. In this case, the hypothesis may be changed to a hypothesis for a proportion so that the baseline condition becomes "the proportion of cadmium levels 1 mg/L or less in the load of fly ash is greater than 0.95."

Once the hypotheses have been transformed, the other DQOs must also be transformed. This includes the other bound of the gray region and additional limits on decision errors. For example, the other bound of the gray region should have been specified as another percentile. This percentile will also need to be converted into a proportion. Table 5 describes the conversion necessary for all the DQO Process information required for DEFT; column 3 contains an example of the conversion from percentiles to proportions.

Table 5. Translating DQOs for Percentiles into DQOs for Proportions

DEFT Input	Translation from Percentiles to Proportions	Example Translation
Parameter of Interest	Use Population Proportion	
Minimum	0	
Maximum	1	
Action Level	Convert the action level to a proportion by dividing by 100.	The 95 th percentile becomes a proportion of 0.95.
Baseline and Alternative Conditions	Baseline condition, $H_0: Q^{\text{th}} \text{ Percentile} \geq x$, becomes $H_0: P \leq Q/100$ Baseline condition, $H_0: Q^{\text{th}} \text{ Percentile} \leq x$, becomes $H_0: P \geq Q/100$ where P is the proportion with observations coded as 1 if they are less than x and 0 otherwise.	The baseline condition $H_0: 95^{\text{th}} \text{ percentile} \geq 5 \text{ ppm}$ translates into $H_0: P \leq 0.95$ where P is the population proportion with observations coded as being 1 if they are less than or equal to 5 ppm and 0 otherwise.
Gray Region (GR)	Convert the percentile that describes the other bound of the gray region to a proportion by dividing by 100.	The 97.5 th percentile translates into a proportion of 0.975.
Sampling Cost	No change necessary.	No change necessary.
Analysis Cost	No change necessary.	No change necessary.
False Rejection Error Limit (FR)	No change necessary.	No change necessary. The false rejection error rate is still 5%.
False Acceptance Error Limit (FA)	No change necessary.	No change necessary. The false acceptance error rate is still 20%.

Table 5. Translating DQOs for Percentiles into DQOs for Proportions

DEFT Input	Translation from Percentiles to Proportions	Example Translation
Additional Error Limits Above or Below the Gray Region.	Convert the percentile to a proportion by dividing by 100. The probabilities remain the same.	An error limit of 10% at the 99 th percentile translates into a false acceptance error rate of 10% at a proportion of 0.99.

CHAPTER 5

ALGORITHMS USED IN DEFT

This chapter briefly describes the algorithms used in DEFT to calculate the minimum required sample sizes under the various sampling design options. For more information regarding these algorithms, see Gilbert (1989), Thompson (1992), or EPA (2000b).

5.1 TESTING A MEAN AGAINST A FIXED STANDARD

A population mean represents the center of a population. This parameter is useful when the action level is based on long-term average health effects (e.g., chronic conditions and carcinogenicity). The mean is most useful when the population is homogeneous and has a relatively small variance. Estimating the mean generally requires a smaller number of samples than estimating other population parameters. However, the mean is not a very representative measure of the center of the population if the underlying distribution of the population is highly skewed, or if the population contains a large proportion of values that are less than the analytical method detection limit.

5.1.1 Simple Random Sampling

The simplest probability sample is a simple random sample where every possible sampling point has an equal probability of being selected and each sample point is selected independently from all other sample points. Simple random sampling is appropriate when little or no information about a population is available. If some information is available, simple random sampling may not be the most cost-effective sampling design.

DEFT assumes that a t-test will be used to analyze the data. Therefore, the corresponding sample size formula is used in the computations:

$$n = \frac{\hat{\sigma}^2 (z_{1-\alpha} + z_{1-\beta})^2}{\Delta^2} + \frac{z_{1-\alpha}^2}{2} \quad (1)$$

where: $\hat{\sigma}^2$ = estimated variance;
 α = false rejection error rate;
 β = false acceptance error rate;
 z_p = the p^{th} percentile of the standard normal distribution (from standard statistical tables);
 Δ = the difference between the action level and the other bound of the gray region;
 n = the number of samples.

A derivation of this formula is contained in Appendix C of the *Guidance for the Data Quality Objectives Process* (EPA 2000c). The sample size reported by DEFT is always greater than or equal to 2 so that an estimate of the standard deviation can be calculated from the data collected. Therefore, if the formula above yields a value less than 2, DEFT will automatically report a sample size of 2. If the sample size calculated is greater than 30,000 DEFT will warn the user and make adjustments to the false rejection and false acceptance error rates (as discussed in Section 2.3.4).

The formula for computing the total cost of the simple random sampling design is:

$$\text{Total Cost} = n (\$ \text{ per field sample} + \$ \text{ per laboratory analysis})$$

The performance curve calculations are also based on the t-test. The software only approximates this performance curve instead of computing the exact curve. As a result of this approximation, the performance curve may appear to show that a decision error limit is satisfied when it is not, especially on the false rejection side of the gray region. Therefore, DEFT labels any decision error limit that is not satisfied as “NS.” This label should be used to determine whether or not a limit is satisfied, rather than the graph of the performance curve.

5.1.2 Composite Sampling

If analysis costs are high compared to sampling costs and the parameter of interest is a mean, then it may be appropriate to use composite samples to reduce analysis costs. A composite sample is a sample obtained by physically mixing (physically averaging) two or more samples before analysis. The use of composite samples in association with a sampling design can be a cost-effective way to select a large number of sampling units and provide better coverage of a population without analyzing each individual sample.

DEFT uses composite samples with a simple random sampling design, which is referred to as “composite sampling.” The software computes the number of composite samples, k , required to meet the DQOs based on a given number of individual samples, m , per composite sample. To determine the number of composite samples, an estimate of the ratio, r , which is the relative standard deviation of measurement error to total standard deviation, is required, along with the number of individual samples, m , to be mixed to form each composite sample. Note that $m \geq 1$, and $0 < r < 1$.

DEFT assumes that a t-test will be used to analyze the data. The software then uses the corresponding sample size formula to determine the required number of composite samples, k , of size m to satisfy the current DQOs. DEFT assumes that the total variability between composite samples can be represented as:

$$\sigma_T^2 = \sigma_X^2 + \sigma_e^2 \quad (2)$$

where σ_T^2 is the total variance, σ_x^2 is the true variance between composite samples (i.e., the “natural” variability with no measurement error), and σ_e^2 is the measurement error variance.

If one forms composite samples of size m , then the variance between the composite samples can be approximated as:

$$\hat{v}(m) = \hat{\sigma}_T^2 [(1-r^2)/m + r^2] \quad (3)$$

where $r = \hat{\sigma}_e / \hat{\sigma}_T$. DEFT uses this estimate of $\hat{v}(m)$ in place of $\hat{\sigma}^2$ in Equation 1 for simple random sampling. The resulting sample size is then the number of composite samples, k , of size m that should be selected in order to satisfy the DQOs.

The sample size reported by DEFT is always greater than or equal to 2 so that an estimate of the standard deviation may be calculated from the data collected. Therefore, if the formula above yields a value less than 2, DEFT will automatically report a sample size of 2. In addition, if the sample size calculated is greater than 30,000, DEFT will adjust the false acceptance error rate (see Section 2.3.4).

The formula for computing the total cost of the composite sampling design is:

$$\text{Total cost} = k [m(\$ \text{ per field sample}) + (\$ \text{ per lab analysis}) + (\$ \text{ per composite})]$$

The performance curve calculations are also based on the t-test. The software only approximates this performance curve instead of computing the exact curve. As a result of this approximation, the performance curve may appear to show that a decision error limit is satisfied when it is not, especially on the false rejection side of the gray region. Therefore, DEFT labels any decision error limit that is not satisfied as “NS.” This label should be used to determine whether or not a limit is satisfied, rather than the graph of the performance curve.

5.1.3 Stratified Sampling

Stratified random sampling can be used to improve the precision of a sampling design. To create a stratified sample, the study population is divided into two or more non-overlapping subsets, called strata, that cover the entire population. Strata should be defined so that physical samples within a stratum are more similar to each other than to samples from other strata. Previous data, information about concentration levels, previous studies, and knowledge about contamination sources can be used as the basis for creating strata. Once the strata are defined, DEFT assumes each stratum will be sampled separately using a simple random sampling design.

There is a limit of six strata in DEFT. (This limit was set because the software was designed only to demonstrate feasibility of the DQOs and six strata should be sufficient for this purpose.) To determine sample sizes for each stratum, a weighing factor (weight) and an estimate of the standard deviation is needed. The stratum weight is the proportion of the volume

or area of the environmental medium contained in the stratum in relation to the total volume or area of the study site. The sum of the strata weights must be 1, so the program automatically computes the weight of the final stratum. The default weight corresponds to equal weighing among the strata. The estimated standard deviation for each stratum must be less than two times the range of the population parameter, and the default value is the estimated total standard deviation.

DEFT assumes that a t-test will be used to analyze the data. Therefore, the corresponding sample size formula⁶ (repeated for each stratum) is used in the computations:

$$n_h = \left[\sum_{h=1}^L W_h \hat{\sigma}_h \right] \left[\frac{(z_{1-\alpha} + z_{1-\beta})^2}{\Delta^2} \right] W_h \hat{\sigma}_h \quad (4)$$

where n_h = the number of samples for stratum h;
 L = total number of strata;
 W_h = weight for stratum h;
 α = false rejection error rate;
 β = false acceptance error rate;
 $\hat{\sigma}_h$ = estimated standard deviation for stratum h;
 Δ = the difference between the action level and the other bound of the gray region;
 z_p = the p^{th} percentile of the standard normal distribution (from standard statistical tables).

The sample size reported by DEFT is always greater than or equal to 2 so that an estimate of the standard deviation may be calculated from the data collected in each stratum. Therefore, if the formula above yields a value less than 2, DEFT will automatically report a sample size of 2. This means that the minimum sample size for a stratified design is equal to two times the number of strata. If the total sample size calculated is greater than 30,000 times the number of strata, DEFT will adjust the false acceptance error rate (see Section 2.3.4). In addition, this sample size formula assumes that the costs of sampling each stratum are the same. If not, see Chapter 6 of *Methods for Evaluating the Attainment of Cleanup Standards* (EPA, 1989) for a sample size formula that accounts for unequal stratum costs.

The formula for computing the total cost of the stratified sampling design is:

$$\text{Total Cost} = \sum_{h=1}^L n_h (\$ \text{ per field sample} + \$ \text{ per laboratory analysis})$$

⁶This sample size formula assumes that the standard deviation is known. Therefore, when the standard deviation is estimated and the calculated sample size is small, consider increasing the sample size by 2 or 3 samples per stratum.

The performance curve calculations are based on the t-test. The software only approximates this performance curve instead of computing the exact curve. As a result of this approximation, the performance curve may appear to show that a decision error limit is satisfied when it is not, especially on the false rejection side of the gray region. Therefore, DEFT labels any decision error limit that is not satisfied as “NS.” This label should be used to determine whether or not a limit is satisfied, rather than the graphical display of the performance curve.

5.2 TESTING A PROPORTION AGAINST A FIXED STANDARD

A proportion represents the number of objects in a population having (or not having) some characteristic divided by the total number of objects in the population. This characteristic may be qualitative, such as leaking drums versus non-leaking drums, or quantitative, such as the drums with concentration levels of a contaminant greater than some fixed level. A proportion is useful if the population consists of discreet objects such as drums or a population of fish. The following discussion assumes that the population is either infinite or extremely large.

5.2.1 Simple Random Sampling

The simplest probability sample is a simple random sample where every possible sampling point has an equal probability of being selected and each sample point is selected independently from all other sample points. Simple random sampling is appropriate when little or no information about a problem is available. If some information is available, it may not be the most cost-effective design.

DEFT assumes that a large-sample, normal approximation method will be used to analyze the data. Therefore, the corresponding sample size formula is used in the computations:

$$n = \left[\frac{z_{1-\beta} \sqrt{GR(1-GR)} + z_{1-\alpha} \sqrt{AL(1-AL)}}{\Delta} \right]^2 \quad (5)$$

where: α = false rejection error rate;
 β = false acceptance error rate;
 z_p = the p^{th} percentile of the standard normal distribution (from standard statistical tables);
 AL = action level;
 GR = other bound of the gray region; and
 n = the number of samples.

This formula is based on Box 7.2 on page 7-6 of *Methods for Evaluating the Attainment of Cleanup Standards: Volume I: Soils and Solid Media* (EPA, 1989). The sample size reported by DEFT is always greater than or equal to 2 so that an estimate of the standard deviation may be calculated from the data collected. Therefore, if the formula above yields a value less than 2,

DEFT will automatically report a sample size of 2. In addition, if the sample size calculated is greater than 30,000, DEFT will make adjustments to the false rejection and false acceptance error rates as explained in Section 2.3.4.

The formula for computing the total cost of the simple random sampling design is:

$$\text{Total Cost} = n (\$ \text{ per field sample} + \$ \text{ per laboratory analysis})$$

The performance curve calculations are also based on the large sample approximation. As a result of this approximation, the performance curve may appear to show that a decision error limit is satisfied when it is not, especially on the false rejection side of the gray region. Therefore, DEFT labels any decision error limit that is not satisfied as “NS.” This label should be used to determine whether or not a limit is satisfied, rather than the graphical display of the performance curve. Note that due to the approximation process, the performance curve may not intersect the false acceptance error limit. This is a result of a correction factor in the sample size formula that does not cancel out in the performance curve calculations. However, the sample size given by DEFT should satisfy the decision error limits regardless of the appearance of the performance curve unless these limits are otherwise marked.

5.2.2 Stratified Sampling

Stratified random sampling is used to improve the precision of a sampling design. To create a stratified sample, the study population is divided into two or more non-overlapping subsets, called strata, that cover the entire population. Strata should be defined so that physical samples within a stratum are more similar to each other than to samples from other strata. Previous data, information about concentration levels, previous studies, and knowledge about contamination sources or activities can be used as the basis for creating strata. Once the strata have been defined, DEFT assumes each stratum will be sampled separately using a simple random sampling design.

To estimate the sample size required for a stratified design, DEFT requires information regarding each individual stratum including a weighing factor (weight) and a preliminary estimate of the stratum proportion. The stratum weight is the proportion of the volume or area of the environmental medium contained in the stratum in relation to the total volume or area of the study population. The sum of the strata weights must be 1, so the program automatically computes the weight of the final stratum. The default weight corresponds to equal weighing among the strata. The estimated stratum proportions may be based on historical information. If there is no information available for estimating these proportions, use the action level or else the average of the action level and the other bound of the gray region. There is a limit of six strata in DEFT. (This limit was set because the software was designed only to demonstrate feasibility of the DQOs and six strata should be sufficient for this purpose.)

DEFT assumes that a large sample approximation will be used to analyze the data. Therefore, the corresponding sample size formula (repeated for each stratum) is used in the computations:

$$n_h = W_h P_h (1 - P_h) \left[\frac{z_{1-\alpha} + z_{1-\beta}}{\Delta} \right]^2 \quad (6)$$

where n_h = the number of samples for stratum h ;
 α = false rejection error rate;
 β = false acceptance error rate;
 Δ = the difference between the action level and the other bound of the gray region;
 W_h = weight for stratum h ;
 P_h = estimated proportion for stratum h ; and
 z_p = the p^{th} percentile of the standard normal distribution (from standard statistical tables).

This formula is based on Box 7.7 of *Methods for Evaluating the Attainment of Cleanup Standards* (EPA, 1989).

The sample size reported by DEFT is always greater than or equal to 2 so that an estimate of the standard deviation may be calculated from the data collected in each stratum. If the formula above yields a value less than 2, DEFT will automatically report a sample size of 2. This means that the minimum sample size for a stratified design is equal to two times the number of stratum. If the sample size calculated is greater than 30,000 times the number of strata, DEFT will adjust the false acceptance error rate (see Section 2.3.4). In addition, this sample size formula assumes that the costs of sampling each stratum are the same. If not, see Chapter 6 of *Methods for Evaluating the Attainment of Cleanup Standards* (EPA 1989) for a sample size formula that accounts for unequal stratum costs.

The formula for computing the total cost of the stratified sampling design is:

$$\text{Total Cost} = \sum_{h=1}^L n_h (\$ \text{ per field sample} + \$ \text{ per laboratory analysis})$$

where L represents the total number of strata.

The performance curve calculations are also based on the large sample approximation. As a result of this approximation, the performance curve may appear to show that a decision error limit is satisfied when it is not, especially on the false rejection side of the gray region. Therefore, DEFT labels any decision error limit that is not satisfied as “NS.” This label should be used to determine if a limit is satisfied, rather than the graph of the performance curve.

5.3 TESTING THE DIFFERENCE BETWEEN TWO MEANS

The simplest probability sample is a simple random sample where every possible sampling point has an equal probability of being selected and each sample point is selected independently from all other sample points. Simple random sampling is appropriate when little or no information about a population is available. If some information is available, simple random sampling may not be the most cost-effective sampling design.

DEFT assumes that a t-test will be used to analyze the data. Therefore, the corresponding sample size formula is used in the computations:

$$m = n = \frac{2\hat{\sigma}^2(z_{1-\alpha} + z_{1-\beta})^2}{\Delta^2} + \frac{z_{1-\alpha}^2}{4} \quad (7)$$

where: $\hat{\sigma}^2$ = estimated variance for both populations;
 α = false rejection error rate;
 β = false acceptance error rate;
 z_p = the p^{th} percentile of the standard normal distribution (from standard statistical tables);
 Δ = the difference between the action level and the other bound of the gray region;
 n = the number of samples for population 1; and
 m = the number of samples for population 2.

This formula is based on Section 3.3.1.1 of the *EPA Guidance for Data Quality Assessment: Practical Methods for Data Analysis (QA/G-9)* (EPA, 2000b).

The sample size reported by DEFT is always greater than or equal to 2 so that an estimate of the standard deviation may be calculated from the data collected. Therefore, if the formula above yields a value less than 2, DEFT will automatically report a sample size of 2. In addition, if the sample size calculated is greater than 30,000, DEFT will make adjustments to the false rejection and false acceptance error rates as described in Section 2.3.4.

The formula for computing the total cost of the simple random sampling design is:

$$\text{Total Cost} = (m + n) (\$ \text{ per field sample} + \$ \text{ per laboratory analysis})$$

The performance curve calculations are also based on the t-test. The software only approximates this performance curve instead of computing the exact curve. As a result of this approximation, the performance curve may appear to show that a decision error limit is satisfied when it is not, especially on the false rejection side of the gray region. Therefore, DEFT labels any decision error limit that is not satisfied as “NS.” This label should be used to determine whether or not a limit is satisfied, rather than the graphical display of the performance curve.

5.4 TESTING THE DIFFERENCE BETWEEN TWO PROPORTIONS

The simplest probability sample is a simple random sample where every possible sampling point has an equal probability of being selected and each sample point is selected independently from all other sample points. Simple random sampling is appropriate when little or no information about a population is available. If some information is available, simple random sampling may not be the most cost-effective sampling design.

DEFT assumes that a large sample normal approximation method will be used to analyze the data. Therefore, the corresponding sample size formula is used in the computations:

$$n = \frac{2(z_{1-\alpha} + z_{1-\beta})^2 \bar{P} (1 - \bar{P})}{(P_2 - P_1)^2} \quad (8)$$

where:

- $\bar{P}$ = $(P_1 + P_2) / 2$;
- α = false rejection error rate;
- β = false acceptance error rate;
- z_p = the p^{th} percentile of the standard normal distribution (from standard statistical tables);
- P_1 = the action level;
- P_2 = the other bound of the gray region;
- n = the number of samples for population 1; and
- m = the number of samples for population 2.

This formula is based on Box 3.3-5 of the *EPA Guidance for Data Quality Assessment: Practical Methods for Data Analysis (QA/G-9)* (EPA, 2000b).

The sample size reported by DEFT is always greater than or equal to 2 so that an estimate of the standard deviation may be calculated from the data collected. Therefore, if the formula above yields a value less than 2, DEFT will automatically report a sample size of 2. In addition, if the sample size calculated is greater than 30,000, DEFT will make adjustments to the false rejection and false acceptance error rates as discussed in Section 2.3.4.

The formula for computing the total cost of the simple random sampling design is:

$$\text{Total Cost} = 2n(\$ \text{ per field sample} + \$ \text{ per laboratory analysis})$$

The performance curve calculations are also based on the large sample approximation. As a result of this approximation, the performance curve may appear to show that a decision error limit is satisfied when it is not, especially on the false rejection side of the gray region. Therefore, DEFT labels any decision error limit that is not satisfied as “NS.” This label should be used to determine whether or not a limit is satisfied, rather than the graphical display of the

performance curve. Note that due to the approximation process, the performance curve may not intersect the false acceptance error limit. This is a result of a correction factor in the sample size formula that does not cancel out in the performance curve calculations. However, the sample size given by DEFT should satisfy the decision error limits regardless of the appearance of the performance curve unless these limits are otherwise marked.

5.5 ESTIMATING A POPULATION MEAN

The formula used by DEFT to calculate the sample size required for testing a mean against a fixed standard (Equation 1) is used to calculate a minimum sample size needed to generate a 100 (1- α)% confidence interval estimate (e.g., a 95% confidence interval where $\alpha = 0.05$) of a population mean with a specified maximum width. In Equation 1, Δ = Action Level + 1/2 the maximum width.

5.6 ESTIMATING A POPULATION PROPORTION

The formula used by DEFT to calculate the sample size required for testing a proportion against a fixed standard (Equation 5) is used to calculate the minimum sample size needed to generate a 100 (1 - α)% confidence interval estimate (e.g., a 95% confidence interval where $\alpha = 0.05$) of a population proportion with a specified maximum width. In Equation 1, Δ = Action Level + 1/2 the maximum width.

REFERENCES

- Flanagan, James B., and James V. Aanstoos, 2001. *Test Plan for the Data Quality Objectives Decision Error Feasibility Trials (DQO/DEFT) Software*, Research Triangle Institute Report RTI/07660/009/2.2F prepared under U.S. EPA Contract 68-C-99-246, Research Triangle Park, NC.
- Gilbert, R. O., 1987. *Statistical Methods for Environmental Pollution Monitoring*. John Wiley, New York, NY.
- Thompson, S. K., 1992. *Sampling*. John Wiley, New York, NY.
- U.S. Environmental Protection Agency, 2000a. *The Data Quality Objectives Process for Hazardous Waste Sites (QA/G-4HW)*, EPA/600/R-00/055, Office of Environmental Information.
- U.S. Environmental Protection Agency, 2000b. *Guidance for Data Quality Assessment: Practical Methods for Data Analysis - QA00 Update (QA/G-9)*, EPA/600/R-96/084, Office of Research and Development.
- U.S. Environmental Protection Agency, 2000c. *Guidance for the Data Quality Objectives Process (QA/G-4)*, EPA/600/R-96/055, Office of Research and Development.
- U.S. Environmental Protection Agency. 1989. *Methods for Evaluating the Attainment of Cleanup Standards: Volume I: Soils and Solid Media*. EPA/230/02-89-042, Office of Policy Planning and Evaluation.