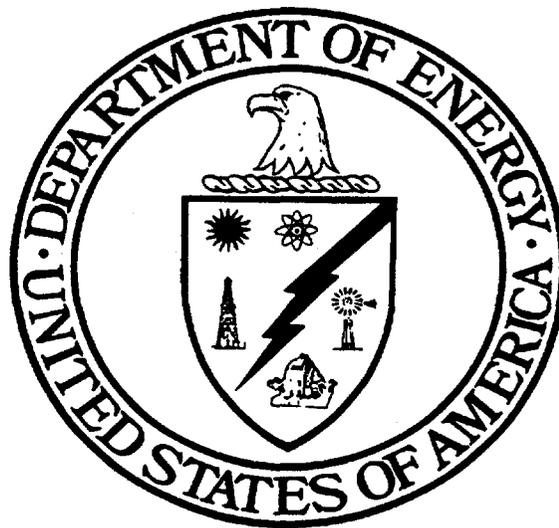


# Statistical and Cost-Benefit Enhancements to the DQO Process for Characterization Decisions



September 12, 1996

Prepared for  
Office of Transportation, Emergency Management & Analytical Services (EM 76)  
Office of Site Operations  
Environmental Management  
U.S. Department of Energy  
Washington, DC

**MASTER**

HH  
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

**DISCLAIMER**

**Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.**

### **DISCLAIMER**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

This paper was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or services by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

# STATISTICAL AND COST-BENEFIT ENHANCEMENTS TO THE DQO PROCESS FOR CHARACTERIZATION DECISIONS\*

Daniel Goodman  
Environmental Statistics Group  
Montana State University  
Bozeman, MT 59717

September 12, 1996

## Contents

1	INTRODUCTION . . . . .	2
1.1	Brief Overview of the Hypothesis Testing Approach . . . . .	3
1.2	Difficulties in Applying the Hypothesis Testing Approach . . . . .	4
1.3	Background to the Decision Theory Approach . . . . .	4
1.4	Perspectives on the Costs of Data and the Uses of Data . . . . .	5
1.5	Applications that would Benefit from a Decision Theory Approach . . . . .	5
1.5.1	Example 1. Patchy contamination of surface soil . . . . .	5
1.5.2	Example 2. Patchy contamination of deep soils causing groundwater contamination . . . . .	6
1.5.3	Example 3. Retrieving waste from buried tanks . . . . .	7
1.5.4	Example 4. Searches for "needle in haystack" contamina- tion sources . . . . .	7
1.5.5	Example 5. Routing D&D waste loads for disposal . . . . .	8
1.5.6	Common Features of the Five Examples . . . . .	9
2	THE LOGIC OF DECISION THEORY . . . . .	9

---

\* Prepared under cooperative agreement DE-FC07-94ID13317 to Montana State University. Technical project officer, Stan Blacker, MacTec; DOE project manager Dave Bottrell, DOE.

3	EXAMPLE ANALYSIS FOR A BINARY DECISION ON D&D WASTE ROUTING . . . . .	10
3.1	Preview of the notation . . . . .	12
3.2	Optimizing the decision about over- or under-the-action-level in a given component step . . . . .	13
3.2.1	Strategy . . . . .	13
3.2.2	Cost structure . . . . .	13
3.2.3	Probabilities of outcomes . . . . .	14
3.2.4	Probability-weighted costs of outcomes, and the "best" decision . . . . .	18
3.3	Optimizing the decision about collecting more data in a given component step . . . . .	19
3.3.1	Strategy . . . . .	19
3.3.2	Limiting the number of sample sizes to be evaluated: the cost of uncertainty . . . . .	21
3.3.3	Expected cost of the decision after more data are collected . . . . .	22
3.4	Information and calculation requirements . . . . .	25
3.5	Example solutions of some scenarios . . . . .	27
4	CONCLUSION . . . . .	30
5	REFERENCES . . . . .	30

## 1. INTRODUCTION

The costs of characterization can comprise a sizeable fraction of a remediation program budget. The DQO Process has been instituted at DOE to ensure that the investment in characterization adds net value to each remediation project. Thoughtful characterization can be very important to minimizing the total cost of a remediation. Strategic information gained by characterization can reduce the remediation costs by reducing the unproductive investment in unnecessary remediation of portions of a site that really don't need to be remediated, and strategic information can reduce remediation costs by reducing the frequency of expensive rework or emergency action that result when remediation has not been pursued to the extent that really is needed.

The DQO process attempts to create a rigorous, objective, stepwise framework for precisely defining the remediation problem at a site, identifying the type of data that might best resolve any questions about the remediation, and determining whether (and how much) collection of new data is warranted. Existing guidance

documents [USEPA, 1994] approach the formalization of the DQO process by use of *statistical hypothesis testing*. In the present document, we extend the tool kit for the DQO process to include *statistical decision theory*.

### 1.1. Brief Overview of the Hypothesis Testing Approach

The hypothesis testing approach to the DQO process for a binary decision (e.g., determining that the contamination concentration is *above* or *below* the action level) treats one of the possibilities in a remediation determination as the *null hypothesis*, and then statistically analyzes available data to *accept* or *reject* the null hypothesis based on satisfaction of a *significance level*, (1 minus the *confidence*), and to make decisions about the need for more data.

If the conclusion of the statistical test is to reject the null hypothesis, that conclusion becomes *the decision*, with the assurance (in some technically complicated sense) that the probability of this decision being wrong is no greater than the significance level of the test.

If the conclusion of the statistical test is to accept the null hypothesis, the question then arises: what is the probability of this alternative conclusion being wrong? This is addressed by carrying out a *statistical power analysis*, in which an *alternative hypothesis* (namely that the true concentration is a specified distance on the *other* side of the action level from the condition of the null hypothesis), and a maximum probability *beta*, (1 minus the *power*), is specified for tolerating mistakenly accepting the null hypothesis when the alternative hypothesis in fact is true, and when the *expected standard deviation* of the samples is a specified value.

If the power analysis determines that this requirement for statistical power is met with the existing data, and the existing data fail to reject the null hypothesis according to the significance test, then the conclusion is to accept the null hypothesis, and that conclusion becomes *the decision*. If, on the other hand, the requirement for statistical power is not met, then the conclusion is that more data are required, and the power analysis will indicate how many more samples are needed.

In this fashion, the hypothesis testing approach to the DQO process formulates a *decision rule* for drawing conclusions about remediation actions based on data, with known confidence and statistical power, and for determining how much, if any, new data is required to achieve this confidence and power.

## 1.2. Difficulties in Applying the Hypothesis Testing Approach

One important limitation of the hypothesis testing approach to the DQO process is that some of the key specifications are left to expert judgment. These specifications may prove problematic to DQO participants, since there is little technical guidance to indicate how they should be set.

There are four hypothesis testing specifications which can prove problematic. These are: (1) the critical significance level and (2) power requirement, that determine the false negative and false positive error tolerances; (3) the width of the "gray area" that defines the alternative hypothesis; and (4) the expected standard deviation that is used in the power test.

These four specifications for the DQO really matter. Changing the values used, even by a small amount, can change the conclusion drawn from data (*above* the action level versus *below* the action level) and can greatly change the calculated number of new samples required to meet the error tolerances. These changes in DQO results can greatly change the cost of a remediation and the cost of a characterization program, so it is desirable that the results be *right*. The measure of whether the specifications are right is whether the human health and environmental risk at the site is controlled to as low a level as required, while simultaneously minimizing cost.

## 1.3. Background to the Decision Theory Approach

One approach to making objective and technically correct specifications for the error tolerances in a DQO is to use *statistical decision theory* to set the error tolerances. Decision theory brings elements of cost-benefit analysis and risk assessment into the DQO process to adjust the error tolerances in such a way that costs on the average will be minimized while satisfying all risk constraints. This is done by explicitly using cost information and risk constraints as inputs that then determine the error tolerances according to a mathematical formula.

Coincidentally, the decision theory approach eliminates the need for defining an alternative hypothesis with a gray area, and it automates the estimation of the expected standard deviation, removing the need to specify the standard deviation as a separate user-defined input.

Taken together, these features of the decision theory approach reduce the role of expert judgment that otherwise would introduce an element of subjectivity in the DQO process, and they lead to a DQO that reduces the cost of the remediation within the constraints set by risk limits.

The decision theory approach enhances the ability of the DQO process to achieve greater efficiency and consistency in remediation by defining correctly how much data is needed and how the remediation decision rule should respond to uncertainty.

#### **1.4. Perspectives on the Costs of Data and the Uses of Data**

The costs of data collection can be very high; but the data can be more than worth their cost if the additional data result in a better remediation by eliminating wasteful unnecessary remediation or by avoiding costly corrective action that might be necessary if an initial remediation were inadequate. The key to the success of the decision theory approach lies in balancing the costs of data against the cost reductions that data can confer. In calculating the required certainty for making a remediation decision, statistical decision theory considers the costs and probabilities of all the possible outcomes of the remediation decision (including the possible mistakes). In calculating the merits of more sampling, statistical decision theory considers the costs of the data, the probability that the new data will lead to a better decision, and the cost saving that would be attributable to a better decision.

#### **1.5. Applications that would Benefit from a Decision Theory Approach**

It may be helpful to illustrate the opportunities for application of the decision theory approach to the DQO process by presenting some example situations where the roles of the cost of data and the value of data to the decision are very important.

##### **1.5.1. Example 1. Patchy contamination of surface soil**

Imagine a site where spills have contaminated the surface soil with  $Cs^{137}$ . The Cs binds to soil, so there has not been much migration, but there have been many spills over the years and the locations of the spills are now uncertain. In order to lift access restrictions in the area, the surface soil contamination must be below a mean of "W" pCi/gm for each 5000 sq ft unit of the site. Because of patchiness of the contamination, it is not certain whether any given 5000 sq ft unit is genuinely above or below the action level, even though it is a simple matter to compute whether the average from the samples is above or below the action level. How wide a margin of safety should be allowed in deciding about remediation of a unit

where the samples are near the action level? How many additional samples should be taken to increase the certainty? The correct answers to these questions depend on the costs of remediation, the costs of mistaken decisions (e.g., the cost of an unscheduled cleanup if later monitoring reveals that an area that was not initially scheduled for planned remediation really is above the action level), and the costs of samples. Statistical decision theory finds the correct margin of safety and the correct investment in sampling to minimize the total cost.

#### **1.5.2. Example 2. Patchy contamination of deep soils causing groundwater contamination**

Imagine a site where  $\text{Sr}^{90}$  leaks from underground tanks and pipes have caused contamination hot spots at a variety of locations and depths. The locations and depths are not known for all of the hot spots. Core samples indicate the curie loading at the sampled locations and depths. There is a commitment to remediate the site to a total curie inventory below "X" Ci, since any more than that results in an unacceptable rate of radioisotope contamination to a nearby stream into which groundwater drains from the site. Calculations indicate how large and hot a hot spot can be tolerated and still meet this total inventory criterion. Any hot spot that is larger or hotter must be removed. If, because of uncertainties in the sampling data, too much contamination is left behind, the mistake will be detected later by monitoring of radiation in the nearby stream, and this mistake will trigger an emergency remediation that is much more expensive than a planned surgical excavation and removal of known hot spots. How much core sampling (which is quite expensive) is warranted to locate all the hot spots that are above threshold? Where should samples be located to obtain the most useful information from each sample? More sampling than necessary wastes money because the core drilling operation is very expensive. Remediating every "suspect" location wastes money because many of these locations will turn out not to be that hot, but the excavation of relatively clean dirt is still very expensive. Failing to find enough of the hot spots wastes money because later monitoring will reveal that too much contamination was left behind, triggering a very expensive emergency remediation. Statistical decision theory finds the right balance between the various costs and uncertainties to minimize the total cost of the remediation.

### 1.5.3. Example 3. Retrieving waste from buried tanks

Imagine a site where a large number of buried waste tanks are scheduled for retrieval of their contents. It is not practical or necessary to remove absolutely all the waste contents, because some amount of material may be left in the tank for grout disposal in place, provided the remaining material is within the specifications for grout, and provided the remaining inventory in each tank does not exceed "Y" Ci. Trying to remove all the waste material is outlandishly expensive, because the removal technology of choice (simple sluicing) functions well only when the depth of remaining waste exceeds a few inches. The decision about how much material to leave in each tank is based on sampling information from that tank, which is used to calculate how much material can be left while still meeting the specifications for in-place grout disposal. Because of extreme heterogeneity in the waste material, the sampling information is uncertain. Mistakenly leaving too much material in the tank will have extremely costly consequences when it is discovered that the grout mixture is out of specifications, necessitating expensive rework. Mistakenly attempting to retrieve more of the tank contents than was really necessary can be very expensive if this forces use of exotic retrieval technologies. It is not practical to overwhelm the uncertainty and heterogeneity with "saturation" sampling, because samples are quite expensive. Statistical decision theory takes the various costs and uncertainties into account in making the right decisions about how many samples are worthwhile and how much waste material can safely be left behind in each tank.

### 1.5.4. Example 4. Searches for "needle in haystack" contamination sources

Imagine a site where known Sr<sup>90</sup> contamination of groundwater indicates a source, but where limited soil sampling has failed to locate the source. Continued soil sampling does not seem promising to find the "needle in a haystack," but the source must be found, because the level of groundwater contamination is unacceptable. A commitment is made to do more groundwater sampling and combine this with modeling to attempt to "back trail" the water contamination plume to the unknown source. Wells are quite expensive. How many new wells should be sunk, and in what locations, in order to locate the source as inexpensively as possible in a reasonable amount of time? Would there be a significant cost savings in a "sequential" sampling pattern, where a coarse grid of a small number of initial wells was used to roughly delimit the plume, and this information was then

used to target the locations for the next set of new wells to zero in on the source. Sequentially drilling new wells, analyzing their samples, and then deciding about where to place the next wells, is bound to be more efficient, since the stepwise process can identify areas where there is no longer any need to sample; but the sequential process will of course take more time. Statistical decision theory can determine the best number of "iterations" of sampling and analysis steps, to get the most efficient search for the source completed within a set amount of time.

#### **1.5.5. Example 5. Routing D&D waste loads for disposal**

Imagine a D&D project generating a large number of truck loads of material that must be disposed of. Material with a radiation level below "Z" Ci/kg may be disposed of relatively inexpensively in a landfill. Material that is above this contamination threshold must be disposed of at a more secure site where there is more control of access and a high degree of containment and monitoring, all of which naturally increases costs. It is desirable to send as much material as possible to the less expensive landfill disposal, but mistakenly sending waste that is of too high a level to qualify for this less expensive disposal will result in very high costs of later remediation at the landfill site when unacceptable radiation levels are detected there. The D&D waste is somewhat heterogeneous in composition. This presents an opportunity to minimize the number of loads sent to high level disposal by distributing the various waste fractions to keep each load just below the threshold allowing for landfill disposal (essentially diluting a small volume of high level fractions with low level material). But because of the heterogeneity of composition, a considerable amount of sampling is required to achieve a high degree of certainty about the total radioactivity inventory of a load. For the same reason, a large amount of sampling of the D&D waste stream during the demolition process would be necessary to mix the various components for minimizing the number of high level loads. Given the costs of sampling, the costs of the two disposal routes, and the costs of mistakenly contaminating a landfill disposal site to the point that it requires subsequent remediation, what is the optimal unit waste volume to be sampled for deciding how it should be routed for mixing with other material into a load that then gets routed for the correct type of disposal, and how much sampling is optimal for each unit? Can these decisions about sampling, mixing, and routing, be stratified, so that less sampling is invested in units whose status is obvious, reserving more of the sampling effort to decide the status of borderline volumes? Statistical decision theory can determine the best

size volume for making mixing decisions, can define a routing rule that takes into account the costs of a mistake, and can adjust the sampling effort based on initial cheap screening scans so that the detailed sampling is only done where needed.

#### **1.5.6. Common Features of the Five Examples**

These five scenarios have a number of important features in common:

- they all involve situations where additional data would represent a significant cost
- they all involve situations where proceeding with a remediation decision in the absence of additional data could lead to significant mistakes
- they all involve situations where the costs of the possible mistakes in a remediation decision (basically, too much or too little remediation) are readily quantified (as the cost of the unnecessary remediation or the cost of rework and damage control)
- they all accomplish the desired level of risk control by a commitment to later verification or continued monitoring, coupled to a commitment to correct any mistakes that the verification or monitoring reveal
- each remediation constitutes a "big picture" project that is made up of a series of repeated decision (about portions of the site, or loads of waste, or individual tanks) where each decision step involves a remediation decision based on existing information as well as a decision about possibly obtaining more data.

These features define a problem that lends itself to solution by the decision theory approach that minimizes cost while satisfying the requirement for risk control.

## **2. THE LOGIC OF DECISION THEORY**

The logic of decision theory is very simple in the abstract. At each step where a decision must be made, the analysis considers each possible outcome of each possible course of action that may be decided upon. Then for each course of action, the expected cost is computed as the sum of the cost of each outcome

weighted by its probability, where the probability is determined by statistical analysis of available data. The best course of action, if the decision must be made based only on the data already existing, is the course of action that the analysis shows to have the lowest expected cost. The merits of additional data are evaluated by computing (based on the data already in hand) the probabilities that the additional data will lead to a different (and probably better) decision and considering the expected cost of following that course of action. The correct number of additional samples is that for which the net benefit (expected cost reduction owing to the probabilities of different decisions minus the cost of the data themselves) is greatest.

In practice, of course, the calculation of these probabilities may be complicated, but computer tools are available to reduce the inconvenience [Goodman, 1996]. There is also the slight obstacle that these techniques of statistical decision theory are less familiar to most practitioners than are the techniques of conventional hypothesis testing. Nevertheless, there is a good literature available. The classic text on statistical decision theory is Raiffa and Schlaifer (1961). The standard modern text is Berger (1985). A good introduction on the theory of decision rules is DeGroot (1970). The new standard text on Bayesian inference (which is essential to the method we will illustrate) is O'Hagen (1994). And a survey of Empirical Bayes methods, which is our choice of method for obtaining prior distributions, is given by Maritz and Lwin (1989). An introduction to modern computational strategies for obtaining numerical solutions to these decision theory problems is given by Smith and Gelfand (1992).

### **3. EXAMPLE ANALYSIS FOR A BINARY DECISION ON D&D WASTE ROUTING**

We consider the DQO for a series of waste loads from a D&D project, where the remediation decision is to determine, load by load, whether the curie inventory for that load is above or below an action level that would preclude landfill disposal. We will assume that each load is subjected to an initial screening sampling that provides a rough estimate for the curie inventory. The results of that screening may be accurate enough for making some decisions about some loads that are either way above or way below the action level, but the accuracy may not be sufficient for other loads whose rough estimates are too close to the crucial threshold concentration for a comfortable conclusion as to which side of the action level they actually lie on. We will consider one property, total radioactivity, whose value for

the load (estimated by the sample mean) is the crucial quantity for making the decision about disposal for that load. We designate the action level as  $T$ , and the true (but unknown) average in the load as  $A$ .

Previous experience with similar D&D projects provides historical data on what sorts of values can be expected for the mean total radioactivity level for a waste load, and what sorts of values can be expected for the heterogeneity (measured as a standard deviation) of the measurements within the same load. Generally the pooled sample size of historical data on similar D&D projects, coupled with measured values from previous loads on the current D&D project, is large enough to allow reasonably convincing calculation of the average radiation level and within-load variability in radiation level for the entire collection of loads. This is because the pooled sample size is large enough to average out much of the uncertainty about the respective individual loads. But the existing information will not give as reliable an estimate of the mean (or the variance) for any *particular* load, since there are far fewer data from any one load than for the entire pooled archive of loads that have been examined. Put another way, the combined existing information from all similar D&D projects is often adequate to give a pretty good estimate of *how many* of the loads are above the action level (because that is an aggregate property, for which there are many data points), but it is not adequate to pin down exactly *which* of the loads are above the action (because that is a property of the individual load, for which there are few data).

Given the large number of loads about which decisions need to be made, there are likely to be loads for which the estimates from initial screening data are sufficient to conclude that the load is well above the action level; there are likely to be loads for which the estimates from initial screening data are sufficient to conclude that the tank is well below the action level; and there are likely to be some loads where the estimates from initial screening data leave enough uncertainty that the collection of some more samples from that load is warranted. Our problem is to find an objectively justifiable decision rule to define quantitatively how far an estimate needs to be above or below the action level for us to accept the conclusion from the screening data alone, and to define quantitatively how many more samples should be taken if the screening data are too close to borderline.

To simplify the illustration, we will assume that there are only two choices, in addition to collection of more data, as to the next action: choice  $O$  is appropriate if the load truly is *over* the action level  $T$ , and choice  $U$  is appropriate if the load truly is *under* the action level  $T$ . To further simplify, we will assume that the logistics of additional sample collection and analysis after the screening analysis for

a particular load make it impractical to conduct the sampling *sequentially* (taking one sample, analyzing that sample, considering the decision between choice  $O$  and  $U$ , and again deciding whether one more sample is warranted before the decision, and so forth). We will assume that essentially random sampling is feasible and appropriate. Finally, we will assume that the consequences of each choice,  $O$  or  $U$ , are of only two kinds, corresponding to whether the decision turns out to be "right" or "wrong" for that step.

In other situations there may be more options to choose among at a given step, and so there will be more categories than simply over- or under-the-action-level for making a decision; and the costs of decision consequences (especially the consequences of a mistake, where the mistake may result from a large or small discrepancy between the true value  $A$  and the estimate may also be more graded. Also, there may be situations when sequential sampling is a practical option. And for some decisions there may be a serious question about *where* to sample, in addition to the question about *how many* samples. All these complications can be accommodated by extending the same logic as presented in this simplified illustration.

### 3.1. Preview of the notation

The analysis that follows will be concerned with a *decision* about whether a load is over or under the action level. Since the decision is made under conditions of uncertainty, the load could *actually* be over or under the action level, regardless of what the decision was. Once the decision is made, subsequent events and ongoing monitoring will reveal the relationship between the decision made and the actual load status. But, at the time of the decision, the actual load status is unknown, so we can only refer to the *probabilities* of the consequences of any decision, based on the probabilities of the actual load status as computed from the data in hand at that time.

When we make a decision about collecting additional data, those additional data themselves are not yet in hand, so we can only quantify the *probabilities* of what those future data will show, the *probabilities* of how those data, when we get them, will cause us to revise our calculations about the *probabilities* that the load is over or under the action level, the *probabilities* of the various *decisions* we will then make once the additional data are in hand, and the *probabilities* that the future decisions will later turn out to have been right or wrong. Obviously, the hall-of-mirrors effect in this last sentence will require some very careful math-

emathical notation to keep track of actual but unknown load status, decisions (that may or may not prove correct) about the load, decision consequences that result from each possible combination of decision made and actual load status, data in hand, data not yet in hand, and probabilities of all of the above.

We will use a notation that does keep track of all these distinctions, but for just this reason, it will seem very tedious. This preview is intended to alert the reader as to why such a tedious notation is needed. Bear with us.

### 3.2. Optimizing the decision about over- or under-the-action-level in a given component step

#### 3.2.1. Strategy

The decision strategy is to minimize the summed cost of all the possible decision outcomes, each weighted by its respective probability. In the present example there are two decision categories (over or under the action level) and each has two categories of outcome (right or wrong), so there are a total of four outcomes to consider.

#### 3.2.2. Cost structure

Let the costs of each outcome be designated

- $C_{to}$  cost of implementing a decision that the load is over the action level, where this decision turns out to be correct (the mnemonic subscript *to* for this situation derives from the label "true over" paralleling the common terminology of "true positive" )
- $C_{tu}$  cost of implementing a decision that the load is under the action level, where this decision turns out to be correct (we subscript this as "true under" parallel to the common terminology of "true negative")
- $C_{fo}$  cost of implementing a decision that the load is over the action level, and then implementing corrective action when it turns out that this decision was mistaken (we subscript this as "false over" parallel to the common terminology of "false positive")

$C_{fu}$  cost of implementing a decision that the load is under the action level, and then implementing corrective action when it turns out that this decision was mistaken (we subscript this as "false under" parallel to the common terminology of "false negative")

### 3.2.3. Probabilities of outcomes

There are only two relevant categories of actual status of the load contents: actually over the action level ( $ao$ ) or actually under the action level ( $au$ ). Depending on the decision category chosen, nominally over the action level ( $O$ ) or nominally under the action level ( $U$ ), these determine the outcome (true over, true under, false over, or false under). Let the probabilities of the two relevant categories of actual status be designated

$P_{ao}$  probability, given the data, that the load contents are over the action level

$P_{au}$  probability, given the data, that the load contents are under the action level

Note that  $P_{au} = (1 - P_{ao})$ , since these exhaust the possibilities.

Probabilities, given the data, are technically different from the "significance" or "confidence" levels computed by conventional statistics intended for tests of hypotheses. Probabilities, given the data are obtained by Bayesian inference. In order for the probabilities to be correct long run frequencies of occurrence, as we require, Empirical Bayes methods often must be used, to avoid the difficulties posed by subjective theories of probability. In the Empirical Bayes approach, the *prior* probability is based on objective data; and provided there are enough such data, the calculated probabilities that result will have the desired properties. In the present example, the existing data (historical process information as well as sampling data) for the entire collection of loads forms the empirical basis for estimating prior probabilities.

In the present case, we will assume that the within-load heterogeneity gives rise to a normal distribution in the samples from a given load, where both the true mean,  $A$ , and the true variance,  $\sigma^2$ , of that distribution are unknown, but where the mean and variance of the observed samples from that load give some information on the true mean and variance. The decision about the load could be made without error if we knew the true mean concentration  $A$  for that load. As it

is, our statistical operations with the data will yield a joint probability distribution of the unknown load mean,  $A$ , and the unknown within-load variance,  $\sigma^2$ .

In the Bayes formulation, the distribution for the uncertain quantities,  $A$  and  $\sigma^2$ , given the data specific to the instance in question,  $x$ , is

$$p(A, \sigma^2|x) = k p(A, \sigma^2)p(x|A, \sigma^2) \quad (1)$$

where the result  $p(A, \sigma^2|x)$  is called the *posterior distribution*, and  $p(A, \sigma^2)$  is called the *prior distribution* based on all available information for related instances. The term  $p(x|A, \sigma^2)$  is a standard statistical function, called the *likelihood function*, relating the probability of the observed data to hypothetical values of the parameters of the distributions from which the data were drawn. In the present case, the likelihood function is simply the textbook distribution of samples from a normal distribution with mean  $A$  and variance  $\sigma^2$ . (If some other distribution, for example a lognormal, better describes the distribution of sample values within a load, then that would be used for the likelihood function.) The value  $k$  is a normalizing constant obtained from

$$k = \frac{1}{\int_0^{\infty} \int_0^{\infty} p(A, \sigma^2)p(x|A, \sigma^2)dAd\sigma^2} \quad (2)$$

We assume, for the calculations in this illustration, that the joint prior distribution  $p(A, \sigma^2)$ , describing the distributions of within-load means, and within-load variances, among similar D&D waste loads for which there are historical data, is a kind of 2-dimensional distribution, where one dimension describing the distribution of means is a family of normal distributions, and the second dimension describing the distribution of variances is a transformed gamma distribution (the gamma distribution is a natural distribution for variances). We specify the joint prior distribution in terms of four parameters: the grand mean (the average of the by-load means over all the loads); the variation in means (the standard deviation of the collection of by-load means); the mode of the distribution, over all the loads, of the within-load standard deviations (as the measure of within-load inhomogeneity); and the variation (standard deviation) of the distribution of within-load standard deviations.

In reality, the joint prior distribution will almost certainly have a more complicated shape that does not correspond nicely to a familiar (familiar to a statistician, at any rate) mathematical function. There are methods, that we will not go into here, for dealing with this complication. For purposes of this discussion, the

“nice” distribution has the advantage that it can be summarized in terms of four basic parameters that have some intuitive meaning, so we can show how changes in each of these understandable parameters leads to changes in the results of the analysis.

The quantity of interest for the decision, the mean  $A$  for the load in question, is itself one of the elements of the joint posterior distribution of load parameters computed from applying Bayes’ formula to the data. The distribution of the decision quantity  $A$  is obtained as a marginal of the joint distribution of the two uncertain parameters.

$$p(A|x) = \int_0^{\infty} p(A, \sigma^2|x) d\sigma^2 \quad (3)$$

The probability, given the data, that the true mean in the load in question exceeds the action level  $T$  is then computed simply by determining what fraction of the distribution of  $A$ , for this load, exceeds  $T$ ,

$$P_{ao} = \int_T^{\infty} p(A|x) dA \quad (4)$$

as shown graphically in Figure 1.

The calculation of  $P_{au}$  follows immediately.

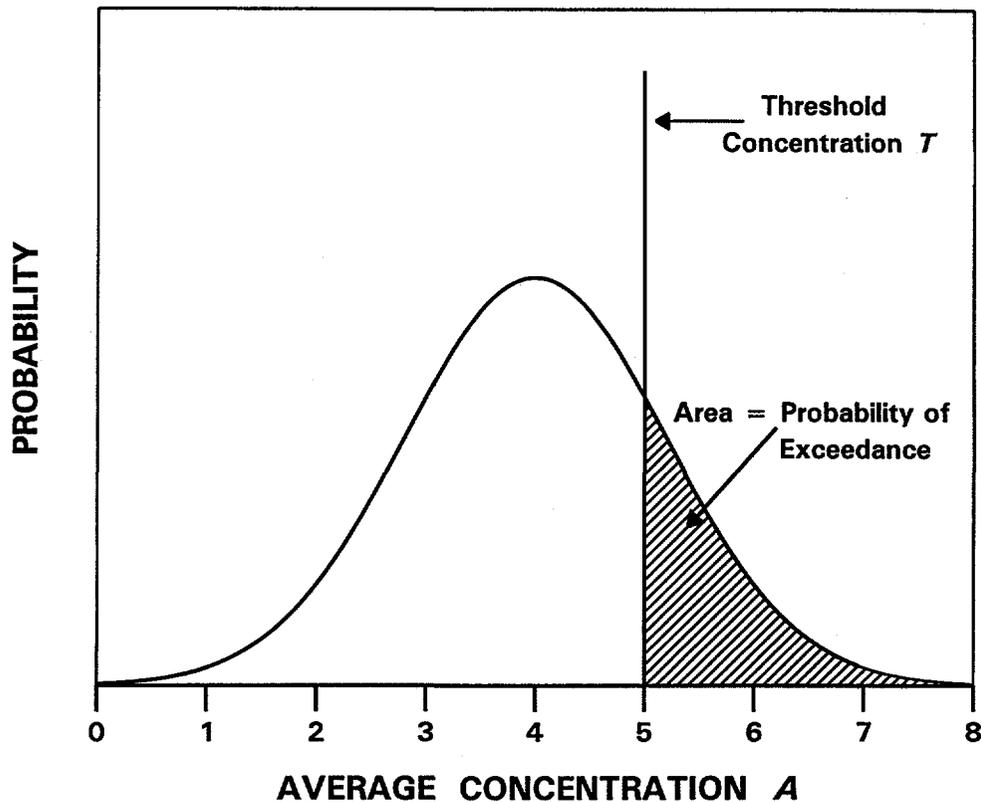


Figure 1. Relation of the distribution of the unknown true average concentration,  $A$ , to the probability that the load is in exceedance of the action level concentration  $T$ , for a case where the action level concentration is 5. The bell-shaped curve (it is actually a  $t$ -distribution, in this case) is the marginal posterior distribution for the unknown true average concentration,  $A$ , in the load, inferred from sample data from this load and comparative information from other loads, as per equations [1] through [3]. The probability of exceedance, as per equation [4], is the area under the curve of the tail of the distribution to the right of the action level  $T$ .

### 3.2.4. Probability-weighted costs of outcomes, and the "best" decision

We wish to compare the expected cost of the consequences of deciding that the load is nominally above the action level versus the cost of deciding that the load is nominally below the action level. In either case, the costs will be the sum of the probability-weighted costs of the consequences when it turns out that the decision was right or wrong.

Let

$D_o$  designate the expected cost of a decision to treat the case as nominally over action level

$D_u$  designate the expected cost of a decision to treat the case as nominally under action level

Forming the probability-weighted sums of the costs of the possible outcomes, we get

$$D_o = P_{ao}C_{to} + P_{au}C_{fo} = P_{ao}C_{to} + (1 - P_{ao})C_{fo} \quad (5)$$

$$D_u = P_{au}C_{tu} + P_{ao}C_{fu} = (1 - P_{ao})C_{tu} + P_{ao}C_{fu} \quad (6)$$

The break-even point for deciding one way or the other occurs when the costs of either decision are the same, which, equating the right hand sides of equations [5] and [6], and rearranging algebraically, gives:

$$P_{ao}^* = \frac{C_{tu} - C_{fo}}{C_{to} - C_{fo} + C_{tu} - C_{fu}} \quad (7)$$

where  $P_{ao}^*$  is the break-even point, the critical level for the probability of exceedance.

In other words, we optimize the decision about whether the load is over ( $O$ ) or under ( $U$ ) action level ( $T$ ) in a given component step, based on existing data, by applying the rule:

*Select the course of action associated with deciding that the load is over the action level whenever the calculated probability that the load is over the action level exceeds the critical probability solved in equation [7], and otherwise select the course of action associated with deciding that the load is under the action level.*

Formally, then, the condition for justifiably deciding that the load is over the action level is

$$P_{ao} > \frac{C_{tu} - C_{fo}}{C_{to} - C_{fo} + C_{tu} - C_{fu}} \quad (8)$$

We note that the degree of certainty required for deciding that the load is over the action level varies as a function of the respective costs of all four possible outcomes. Figure 2 shows how graphs of the cost of deciding ( $O$ ) or ( $U$ ), as functions of the current calculation of the probability of exceedance,  $P_{ao}$ , define a cross-over point,  $P_{ao}^*$ , such that when the probability of exceedance exceeds this value (as per equation [8]), the cost of deciding ( $O$ ) will be less than the cost of deciding ( $U$ ).

### 3.3. Optimizing the decision about collecting more data in a given component step

#### 3.3.1. Strategy

The decision strategy is to minimize the total cost of all the possible decision outcomes, each weighted by its respective probability. In the present example there are two general decision categories: (1) to proceed without additional data, using the rule developed above for optimizing the decision about over- or under-the-action-level in a given component step, or (2) collect more data, and then decide about over- or under-the-action-level, using the same rule developed above but with the larger data set. Under category (1) the expected cost has already been calculated in the preceding section: it is the lower of  $D_o$  or  $D_u$  (as given by equations [5] and [6]). Under category (2) the expected cost is more complicated, since it includes the cost of new data, plus the sum of the probability-weighted costs of the outcomes that will follow upon making the future decision using the new data.

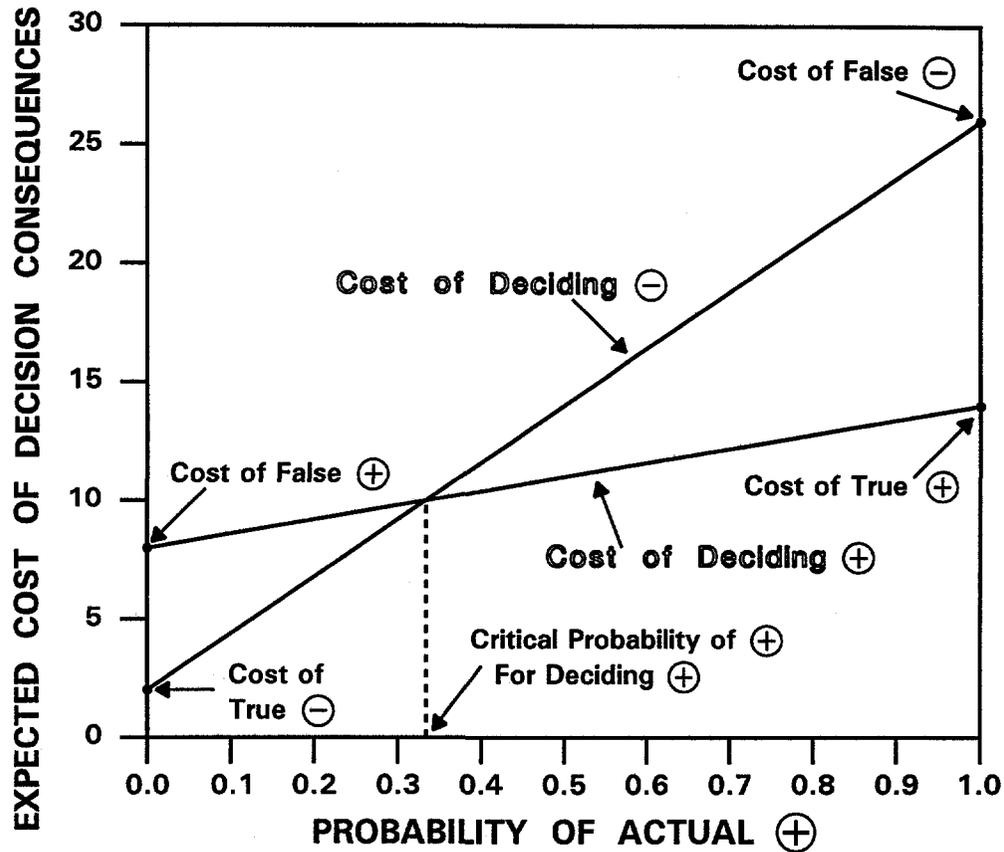


Figure 2. Relation between the expected cost of deciding *positive* or *negative* and the *probability of actual positive* (that is,  $P_{ao}$ , the probability of exceedance). Equation [5] shows why the expected cost of deciding *positive* describes a straight line connecting the unit costs of *false positive* and *true positive* outcomes. Equation [6] justifies the straight line relationship for the expected cost of deciding *negative*. The point where the two decision cost lines cross defines the *critical probability of exceedance* (that is  $P_{ao}^*$ ). If the probability of exceedance for a given load is larger than the critical level, the expected cost of deciding *positive* is lower than the expected cost of deciding *negative*, and therefore, in the absence of further information, the correct choice would be to decide *positive*.

### 3.3.2. Limiting the number of sample sizes to be evaluated: the cost of uncertainty

In determining the number of additional samples to collect, we balance the cost of information with the value of information. The cost of information is simply the cost of obtaining and analyzing the specified number of new samples: this is the unit cost of an additional sample times the number of new samples. The value of information is the cost reduction expected from making the decision with more (and therefore, on the average, better) information.

The decision to collect more samples really consists of a large family of decisions: 1 more sample, 2 more samples, 3 more samples, etc. We first restrict the search by asking whether there is an easy way to identify an upper limit to the number of samples that could be justified. If we could assign a cost to the uncertainty associated with the existing information, this will set an upper bound to the costs for additional data that could possibly be offset by the cost reductions that the information in those data confers.

Based on existing data, the respective probabilities that the mean concentration in the load is above or below the action level are  $P_{ao}$  and  $P_{au}$  (computed from equation [4]). This means that with *unlimited* additional data there is a probability  $P_{ao}$  that the data would show with no uncertainty that the load is above the action level, and there is a probability  $P_{au}$  that the data would show with no uncertainty that the load is below the action level, which exhausts the possibilities. So the expected cost of the decision,  $D_{\infty}$ , not including the cost of the data themselves, if there were no uncertainty, is the probability-weighted sum of the costs of the two possible correct decisions that would result:

$$D_{\infty} = P_{ao}C_{to} + P_{au}C_{tu}. \quad (9)$$

The difference between this cost (with no uncertainty) and the cost with the present uncertainty (the lower of  $D_o$  or  $D_u$ ) is, therefore, the cost of the present uncertainty. So the cost of the present uncertainty,  $I$ , is the lower of

$$I_o = D_o - (P_{ao}C_{to} + P_{au}C_{tu}) \quad (10)$$

and

$$I_u = D_u - (P_{ao}C_{to} + P_{au}C_{tu}) \quad (11)$$

If the unit cost of an additional sample is  $C_s$ , then the maximum number of additional samples worth considering is the largest whole number in the ratio

formed from the cost of the present uncertainty divided by  $C_s$ . Thus we only need search the expected total cost using additional sample sizes from 1 up to this easily calculated maximum,  $n_{max}$ , the smaller of

$$n_{max_o} = \text{integer}(I_o/C_s) \quad (12)$$

and

$$n_{max_u} = \text{integer}(I_u/C_s). \quad (13)$$

Figure 3 shows graphically how the cost of uncertainty relates to the other unit costs ( $C_{to}$ ,  $C_{fo}$ ,  $C_{tu}$ ,  $C_{fu}$ ) as a function of the current calculation of the probability of exceedance,  $P_{ao}$ .

### 3.3.3. Expected cost of the decision after more data are collected

The strategy for this calculation is readily understood, but the calculation itself is very involved. The strategy is to use the distribution that we already have for the uncertain parameters,  $A$  and  $\sigma^2$ , based on the existing information, to compute the probabilities of all possible future data that we could obtain if we decided to collect a particular number of additional samples. Then, the inferences that would be drawn from those future data, and the probabilities and costs of making right or wrong decisions based on those inferences, are systematically calculated, and averaged for the expected cost of all possible outcomes of collecting the additional data, including the cost of the new samples themselves. Whichever number of new samples is associated, on this basis, with the lowest expected cost, is the best number of new samples; and if that expected cost is lower than the expected cost with existing information, that number of new samples is the correct decision; if that expected cost is higher than the expected cost with existing information, then the correct decision is to proceed with selection of an action without collecting additional data at this step.

A more formal description of the calculations is as follows. We wish to calculate:

- $D_n$  expected total cost, including the cost of the new samples, of using the decision criterion from inequality [8] after collecting  $n$  new samples.

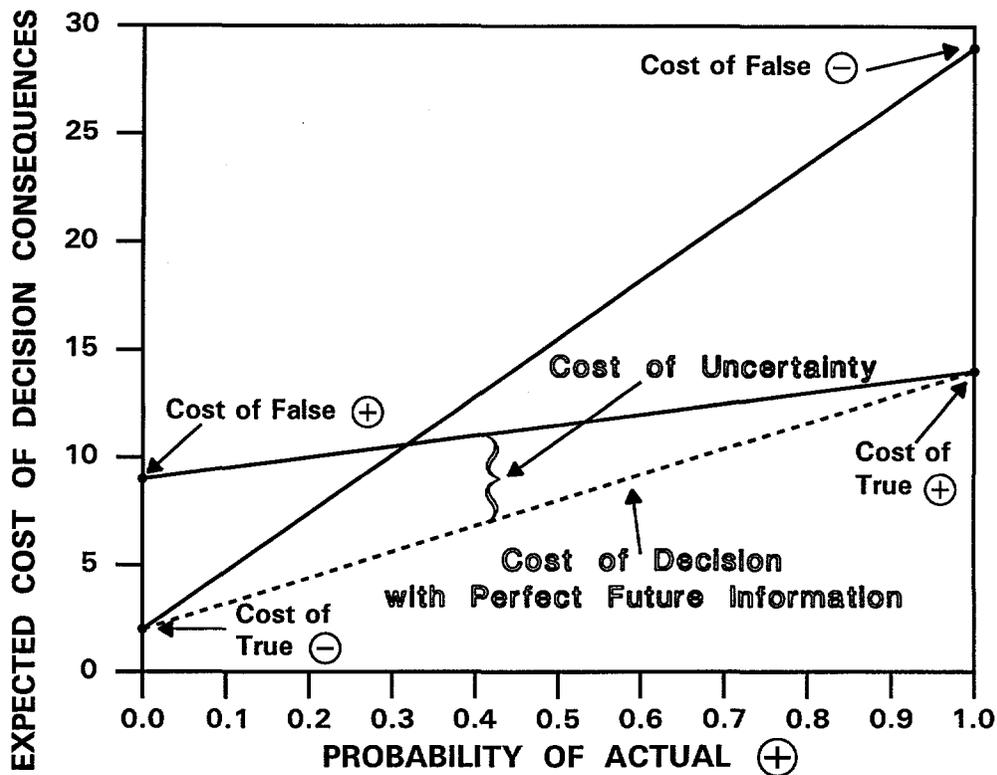


Figure 3. Relation between the *probability of actual positive* (that is,  $P_{ao}$ , the probability of exceedance) and the cost of uncertainty. Depending on whether the probability of exceedance is above or below the critical level,  $P_{ao}^*$  (the point where the two cost of decision lines cross), the choice with the lower expected cost will be *positive* or *negative* respectively, with an expected cost given by the location on the respective cost line corresponding to the probability of exceedance for that load with the current information, as illustrated in Figure 3. The expected cost of the decision that would result from having sufficient information to resolve all uncertainty about the load is given by the dashed straight line connecting the unit costs of *true positive* and *true negative* as per equation [9]. The *cost of uncertainty*, at a given value of probability of exceedance, is the difference between the cost of a decision with perfect information (dashed line) and the lower of the lines describing the decision cost for a *positive* or a *negative* decision at this location on the probability axis. The cost of uncertainty is zero where there is no present uncertainty (if the probability of exceedance is zero or one), and is greatest at the critical probability of exceedance defining the crossover point for deciding *positive* or *negative* with that cost structure.

This requires calculation of the probabilities of the conclusions that will be drawn (correctly, and incorrectly) from the new data. Let  $p_n(P_{ao})$  designate the probability that when  $n$  new samples are collected the new probability assigned to the conclusion that the case is above the action level is  $P_{ao}$ . Note that for any particular value of  $P_{ao}$ , the anticipatory  $p_n(P_{ao})$  also gives the probability that the new data will lead to the inference that the probability of no exceedance is  $P_{au} = 1 - P_{ao}$ .

For a given pair of values of the uncertain parameters,  $A$  and  $\sigma^2$ , (whose joint probability is already defined, given the existing data, from equation [1]), the likelihood function (as used in equation [1]) can be used to give the probability of future data. The probability of each possible configuration of future data in turn allows recursive application of equation [1] and equation [3] (where the posterior distribution from the inference with the existing data serves as the prior distribution applied in equation [1] for inference from the future data). In the unintentionally droll terminology of statistical decision theory, this calculation of probabilities of inferences from data not yet in hand is called a *preposterior analysis*.

The distribution of probabilities of the decision quantity resulting from the preposterior analysis is then submitted to equation [4], to obtain a distribution, valid before the collection of the future data, for the probability  $p_n(P_{ao})$  that the future data will lead to a conclusion that the probability of the mean,  $A$ , being above the action level is the value  $P_{ao}$ .

The probability function  $p_n(P_{ao})$  completes the information needed for evaluating the expected worth of future data. Integrating over all possible values of  $P_{ao}$ , weighting each value by its probability, using inequality [8] to define the decision that will be made from inference on the new data, using the value of  $P_{ao}$  itself to calculate the probabilities that this future decision will later turn out to be correct or to require damage control, and using the unit cost information, then the expected total cost for taking  $n$  new samples and making the remediation decision for this step on the basis of the new data is

$$D_n = nC_s + \int_0^{P_{ao}^*} p_n(P_{ao}) [(1 - P_{ao})C_{tu} + P_{ao}C_{fu}] dP_{ao} + \int_{P_{ao}^*}^1 p_n(P_{ao}) [P_{ao}C_{to} + (1 - P_{ao})C_{fo}] dP_{ao} \quad (14)$$

where the critical probability,  $P_{ao}^*$ , is the break-even point defined in equation [7].

The value of  $D_n$  computed from equation [14] gives the expected total cost

of collecting  $n$  new samples and then using the information they yield to make the decision to select the action, in this step, consistent with the new calculation of the probability that the load average concentration,  $A$ , is over the action level  $T$ . This cost is then compared with the expected cost for other numbers of new samples,  $D_1, D_2, \dots, D_{n_{max}}$ , and with the expected cost for selecting an action without additional data. The design with the lowest cost is the correct choice for this component step of the remediation.

The expected cost  $D_n$  is a net cost. Equation [14] is readily partitioned to reveal the components of this net cost: the value of future information and the cost of future data, for a given number of additional samples. The term  $nC_s$  is the gross cost of the additional data, and the two integrals give the expected cost of the consequences of the future decision using the additional data. Subtracting the expected cost of future decision consequences from the smaller of  $D_o$  or  $D_u$  (the expected cost with no additional samples) gives the expected gross cost reduction attributable to the  $n$  additional samples—this is the *value of information*, shown graphically in Figure 4. Not surprisingly, we see that the value of information exhibits a pattern of diminishing returns (which makes sense, since the value of future information cannot exceed the current cost of uncertainty, no matter how many additional samples are proposed). The cost of the additional data is proportional to the number of new samples proposed, so the maximum expected net benefit (value minus cost) is achieved at an intermediate number of additional samples.

### 3.4. Information and calculation requirements

The amount of mathematics in the preceding sections may have looked so forbidding as to create the impression that this is only a theoretical approach that is too complicated to apply in the real world. In fact, the input information to use this method is not at all esoteric: it is the same information on costs and feasibility and sampling data that would be used for any common sense engineering analysis. And the calculations are natural candidates for standard computer methods.

The input quantities necessary for the analysis are nothing more than the various unit costs of defined activities or consequences ( $C_s, C_{to}, C_{tu}, C_{fo}, C_{fu}$ ), the sampling data ( $x$ ) bearing on the load in question, and the statistics (in this case the four parameters that define the prior distribution of comparative information) that describe the distributions of by-load mean concentrations and within-load heterogeneity over the entire collection of loads.

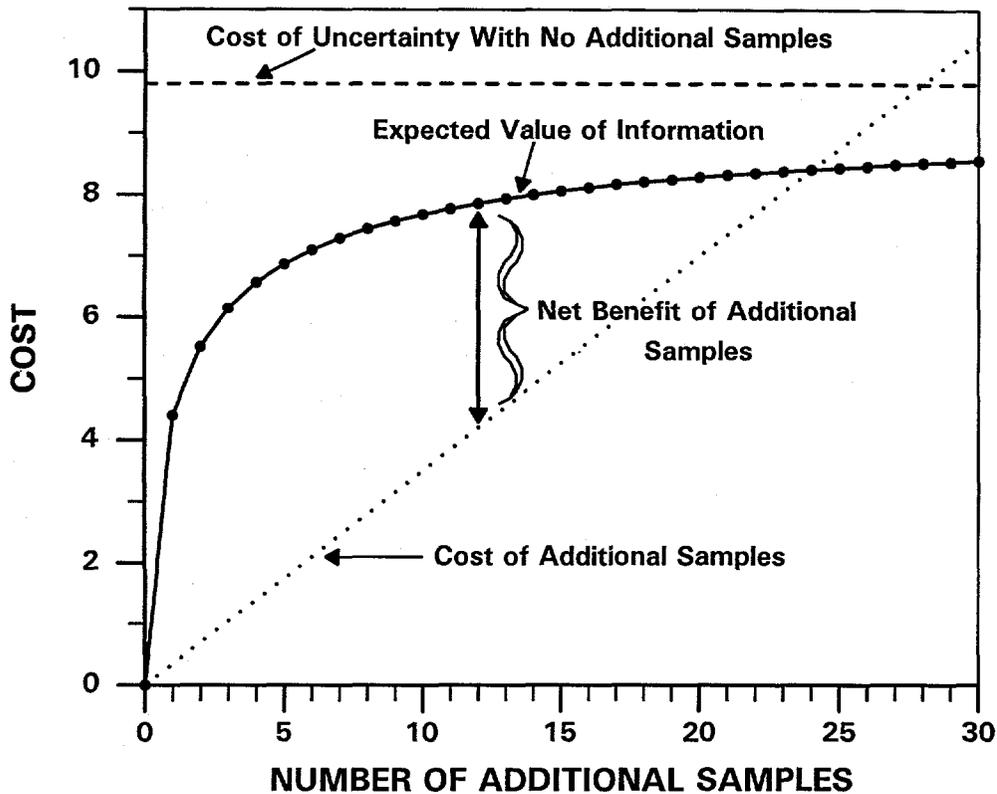


Figure 4. Value of future information and cost of future information as functions of the number of additional samples. The solid curve shows the expected value of information (the expected gross reduction of future cost of the decision) from additional samples, calculated from scenario (a) of Table 1, using the partitioning of equation [14] described in section 5.6. The straight dotted line shows the gross cost of the additional samples when the cost per new sample is 0.35. The net benefit of the additional samples is the difference between the value of information and the cost of the samples. In the present case, the net benefit is maximum at 4 additional samples, and that is the optimal number of new samples for this scenario. The straight dashed line is the cost of uncertainty with no additional samples, so the expected value of information will asymptotically approach this ceiling with increasing numbers of additional samples.

Quantifying the cost of sampling ( $C_s$ ) and the costs of responding to a correct selection of action ( $C_{to}$  or  $C_{tu}$ ) should present no special difficulty. Quantifying the costs of what turned out to be "false over" decisions ( $C_{fo}$ ) or "false under" decisions ( $C_{fu}$ ) is a little more complicated.

The most favorable setting for weighing the costs of false positive ("false over") and false negative ("false under") decisions occurs where the choices are essentially among engineering alternatives. Then the costs of "errors" are quantified from an understanding of how an incorrect decision at a particular stage will be detected at some later stage, and understanding what recovery and replacement actions will be undertaken in consequence of the detection of that error, as described in section 1.5 of this report.

Needless to say, the mathematics of obtaining closed form solutions for the probabilities in this analysis can be very difficult. The equations in sections 3.2 and 3.3 already look rather daunting, owing to the integrations, but these forms of the equations don't even show the details of the expressions for terms such as the likelihood functions, or the joint distribution of the mean and the variance in the prior and posterior of Bayes' formula, or the recursive application of Bayes' formula to solve for  $p_n(P_{ao})$ . Nevertheless, all the steps are readily implemented in computer programs that utilize Monte Carlo methods for solving Bayes' formula.

The Monte Carlo numerical methods are computer intensive. For example, a pure Monte Carlo strategy for solving the example presented in this chapter required, for each scenario, two weeks of execution time on a work station (DEC 5000-260). This may seem outlandish, but it is well to bear in mind that whatever cost or delay this represents is very minor compared to the many millions that can be saved as the result of the two week calculation. For cases where there is a need to explore multiple scenarios in a shorter amount of time, it is possible to use a variety of explicit (but complicated and problem-specific) mathematical solutions to replace some of the Monte Carlo calculations for a much faster solution. The customized computer program used for the example in this chapter solves each scenario in matter of seconds on a standard PC.

### 3.5. Example solutions of some scenarios

Table 1 shows solutions for three example scenarios that could be plausible cost structures and data for some of the waste load decisions.

Scenario (a) has a comparatively high cost of false positives and false nega-

tives. It demands a higher probability of exceedance before deciding *positive* with existing data. The data on this load are very sparse, but they indicate that this load is just a hair over the critical probability for exceedance, triggering a positive decision, but under circumstances where the expected costs of a positive or negative decision are almost identical. Because the probability of exceedance is so close to the breakeven point, and given the high costs of false positive and false negative outcomes, the cost of the uncertainty with the present data is high. Since the cost of additional samples is not too high relative to the cost of uncertainty, additional samples are worthwhile. The optimal number of additional samples reduces the expected cost by about a third, which can be a very substantial amount.

Scenario (b) has a less threatening cost structure—in particular the cost of false positives is comparatively low, and so it demands less certainty of exceedance to trigger a positive decision. The data are still sparse, but they make a more convincing case that the load is below the action level—so much so that the probability of exceedance is below the critical probability, and so the decision with existing data is that the load is negative. Because of the less threatening relative costs of mistakes, and because the current probability of exceedance is not so near to the breakeven point, the cost of uncertainty is lower. Since the cost of additional samples is high, and the cost of uncertainty is modest, no additional samples are justified.

Table 1. Decision theory solutions for three example scenarios.

INPUT SPECIFICATIONS	Scenario		
	(a)	(b)	(c)
<b>Costs of Decision Consequences</b>			
COST OF TRUE POSITIVE $C_{to}$	20.0	10.0	10.0
COST OF FALSE POSITIVE $C_{fo}$	15.0	4.0	4.0
COST OF TRUE NEGATIVE $C_{tu}$	2.0	2.0	2.0
COST OF FALSE NEGATIVE $C_{fu}$	60.0	25.0	25.0
<b>Other Unit Costs</b>			
COST OF ADDITIONAL SAMPLE $C_s$	0.35	1.0	0.05
<b>Comparative Information</b>			
GRAND MEAN OF CONCENTRATIONS	4.50	4.50	5.50
STANDARD DEVIATION OF MEANS	1.50	1.50	1.50
MODAL WITHIN-LOAD STD DEV	1.00	1.50	1.50
STD DEV OF DISTRIBUTION OF STD DEVS	0.25	1.00	1.00
<b>Engineering Specifications</b>			
CONCENTRATION ACTION LEVEL $T$	5.0	5.0	5.0
<b>Sampling Data for This Load</b>			
NUMBER OF EXISTING OBSERVATIONS	1	2	3
OBSERVATION No. 1 $x_1$	4.40	3.00	3.00
OBSERVATION No. 2 $x_2$		4.00	4.00
OBSERVATION No. 3 $x_3$			3.50
<b>DECISION THEORY RESULTS</b>			
<b>Results of Cost Structure</b>			
CRITICAL PROBABILITY OF EXCEEDANCE $P_{ao}^*$	.245	.118	.118
<b>Decision Consequences with Existing Data</b>			
CURRENT PROBABILITY OF EXCEEDANCE $P_{ao}$	.246	.109	.191
EXPECTED COST OF DECIDING POSITIVE $D_o$	16.2	4.5	5.1
EXPECTED COST OF DECIDING NEGATIVE $D_u$	16.3	4.7	6.4
BEST ACTION DECISION ( $O$ or $U$ )	$O$	$U$	$O$
COST OF CURRENT UNCERTAINTY $I$	9.8	1.6	1.6
<b>Evaluation of Worth of Additional Data</b>			
OPTIMAL NUMBER OF ADDITIONAL SAMPLES $n$	4	0	7
EXPECTED COST WITH ADDITIONAL SAMPLES $D_n$	11.1	4.5	4.7

Scenario (c) has the same cost structure as scenario (b) for the costs of decision outcomes (and therefore the same critical probability of exceedance for triggering a positive decision) but it has a vastly cheaper unit cost for additional samples. Scenario (c) is also distinguished by a distribution of comparative information for the entire set of loads that creates a default presumption that the load is over the action level (rather than the reverse presumption in scenarios (a) and (b), above). Despite the slightly stronger data for this load (3 samples rather than the two samples from scenario (b), but with the same average), the presumption of exceedance owing to the comparative information leads to a decision with the existing data that the load is a positive. If the decision were made, instead, that the load is a negative, the expected costs would be about 25% higher. Because of the very low cost of additional samples, the optimal number of additional samples is substantial, and these additional samples would reduce the expected cost by about 10%.

#### 4. CONCLUSION

In this report we have shown how the technically rigorous methods of statistical decision theory can be used to bring a cost-benefit perspective to the DQO decisions about additional sampling needs for remediation decisions. These methods allow evaluation of the needs for specific kinds of information, the actual number of additional samples that would be optimal, and the cost advantage of this approach versus other less formal ways of planning.

#### 5. REFERENCES

- Berger, J.O. 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer Verlag. New York.
- DeGroot, M.H. 1970. *Optimal Statistical Decisions*. McGraw-Hill, New York.
- Goodman, D. 1996. *Statistical and Cost-Benefit Enhancements to the DQO Process for Characterization Decisions in TWRS. Report to the DOE Office of Hanford Tanks, EM 38*. Montana State University, Bozeman, Montana.
- Maritz, J.S., and T. Lwin. 1989. *Empirical Bayes Methods, 2nd ed.* Chapman and Hall, London.
- O'Hagen, A. 1994. *Kendall's Advanced Theory of Statistics. Vol. 2B. Bayesian Inference*. Edward Arnold. London.

- Raifa, H., and R. Schlaifer. 1961. *Applied Statistical Decision Theory*. Division of Research, Harvard Business School. Boston.
- Smith, A.F.M., and A.E. Gelfand. 1992. Bayesian statistics without tears: a sampling-resampling perspective. *The American Statistician* 46:84-88.
- USEPA. 1994. *Guidance for the Data Quality Objectives Process*. EPA QA/G-4. Quality Assurance Management Staff, Washington, DC.