Confidence, Cost, and Risk: Setting Performance Objectives for Environmental Decisions

Evan J. Englund*

Environmental Protection Agency, Las Vegas, Nevada, 89193

* Corresponding author phone: (702)798-2248; fax: (702)798-2107; e-mail: englund.evan@epa.gov

Abstract:

Need for decision performance is often the driving factor in the design of sampling plans that specify quantity and quality requirements for environmental data. Decision performance can be quantified by different measures that lead to different performance objectives, and in turn to different optimal sampling designs. Three performance measures relevant to environmental decision making are examined: decision confidence; cost of errors; and risks. The corresponding optimization objectives were applied to a hypothetical example involving a decision whether or not to treat waste before sending it to a landfill. The resulting optimal sample sizes ranged from 1 to 29. Optimal waste treatment costs varied by a factor of 5, and the optimal mean contaminant concentration in the landfill varied by more than a factor of 4.

Introduction

This paper deals with optimizing the process of environmental decision making at an operational level. A typical example might involve collecting samples to determine whether the concentration of a toxin in a mass of waste exceeds some threshold value, thus requiring treatment. It is assumed that the larger policy decisions have already been made: The waste must be managed; and an acceptable threshold level has been established.

Optimization in general involves obtaining the best possible result. Optimizing an environmental decision process typically weighs potential mitigation or prevention costs against corresponding environmental benefits in the form of lower human health risks and improved quality of lift. Optimization can be unconstrained, where both costs and benefits are adjusted to find the most cost-effective result. Alternately, constrained optimization involves either finding the lowest cost to obtain a required benefit or finding the maximum benefit for a fixed cost. Optimizing a decision making process with imperfect information is a matter of balancing the cost of additional information against the benefits of fewer errors.

111 recent years, there has been a growing consensus that the need for data quality in environmental decision-making is driven by a project*s objectives (1,2). A focus on performance rather than standardized methods and procedures provides the flexibility needed to adopt innovative, cost effective approaches. The number of samples, sampling times and locations, sampling methods, sub-sampling methods, analytical methods, statistical estimates, and final decision rule all combine to determine the ultimate decision quality. To use this approach it is necessary to set clear performance requirements that are appropriate to the problem at hand.

The simplest measure of decision quality is whether or not a decision is correct; the corresponding measure of a decision process is the probability that the decision will be correct. More complex measures value the costs and benefits of individual decisions and multiply by the decision probabilities to obtain an overall expected value for the decision process. Environmental decisions generally involve potential mitigation or prevention costs, with corresponding environmental benefits in the form of lower human health risks and improved quality of life.

A complicating factor in many environmental cases is that different stakeholders value costs and benefits differently (3,4). An industrial plant operator required to reduce emissions is likely to be more concerned with mitigation costs, while nearby residents worry more about health risks from continued exposure. Even when both sides agree that the other has valid concerns, it is not easy to arrive at the consensus needed to design an optimal decision process. This paper examines a simple hypothetical decision scenario and optimizes the decision process with respect to objectives based on three different measures of decision performance.

- Confidence: Minimize sampling cost with limits on decision error rates. (1,2,5)

- Cost: Minimize a loss function that includes sampling costs, expected treatment costs and expected environmental costs; or equivalently, maximize a utility function..(3,4,6)

- Risk: Minimize sampling and treatment costs with limits on posterior risk. (7-10)

For each optimal decision process all of the various performance measures are computed, This allows examination, for example, of the decision error rates and posterior risk level that result from an optimal cost-based design.

The example

A hypothetical example is used instead of an actual case study because true values are known. sampling variability can be controlled, and irrelevant details can be eliminated. The intent is to retain sufficient realism and complexity to make the results of the investigation instructive.

Assume that an industrial process generates 100-tonne batches of waste that will be deposited in an on-site landfill. A regulatory restriction has been imposed: If the concentration is less than 100 ppm, the waste can go directly to the landfill; otherwise, it must be treated to remove the arsenic, after which it can also be placed in the landfill.

The prior distribution of batch mean concentrations is shown in Figure 1. Within batches, the distribution of sample concentrations is known to be log-normal, with natural log standard deviation equal to 0.55. Sampling, preparation, and analysis cost $25 per sample. Landfill disposal of treated waste costs $10,000 more per 100-tonne batch than disposal of untreated waste. It will be assumed that the threshold value was set correctly and that the treatment always works, leaving decision error as the only uncertainty in the process.

The true concentration of a batch is unknown to the operator and must be estimated by sampling and chemical analyses. At issue is how to deal with estimation uncertainty and its effect on the quality of decision-making. One approach is to reduce the magnitude of the uncertainty, at the expense of increased costs, by taking more and/or better samples and analyses. An alternate approach is to bias the decision to allow for error, as when an upper confidence limit is used instead of the mean. This bias reduces one type of decision error at the expense of increasing the other. Both uncertainty reduction and decision bias are used as design factors in the example. Sampling and analytical methods are fixed, so sample size is the only factor in estimation uncertainty. The decision rule is simple: If the sample mean exceeds the threshold concentration, treat the batch. Decision bias is introduced in the decision rule by simply changing the threshold from the original *target* threshold of 100 ppm to some other *operational* threshold or action level.

In the example, sample size and operational threshold are the only factors that determine a design. This suggests a convenient notation: A design with a sample size of 10 and an operational threshold of 90 ppm is referred Lo as [10, 90].

Emphasis here is not on how to optimize the design of a decision process, but on how the choice of performance measure affects the optimum. More realistic scenarios would add complexity without necessarily changing the fundamental issues. The ideal decision rule in this univariate example is represented by a point threshold separating the concentration axis into 'treat* and 'don*t treat* segments. In a multivariate case with k contaminants the ideal rule would be represented by a boundary surface threshold in k-dimensional space separating 'treat* from 'don*t treat* volumes. Similarly, the multivariate cost and risk models are more complex fuctions of k variables. The basic question, however is unchanged: Should the operational design be optimized with respect to confidence, cost, or risk?

More realistic scenarios might include the necessity of estimating batch and within batch variability. They might also consider alternative sampling and analytical methods, and alternative sampling schemes such as composite or adaptive sampling. Real data sets are likely to contain missing values and non-detects. Possible alternative decision rules could involve transformed data, statistical parameters other than the mean, or formal parametric or nonparametric hypothesis tests.

Design optimization

Sampling design optimization is done by trial and error using a Monte Carlo computer simulation of the decision process. An initial sample size n and operational action level are selected and the performance of the design is evaluated: For each batch in the distribution (Figure l), a sample of size n is drawn from the corresponding within-batch sample distribution. The sample mean is compared to the operational action level to make the decision, and the true batch mean determines whether the decision was correct. The various performance measures associated with all three optimization methods are tallied for later comparison. The process is repeated with different combinations of sample size and action level until the best combination is found.

Bayesian design optimization

Given the mean of a sample of size n, an operator could make the optimal batch treatment decision by using Bayes* rule (11). The prior probability distribution of batch means is given, as is the within batch distribution. The latter determines the likelihood function, that is, the probability of observing the particular sample mean given any batch mean. Bayes* rule gives the posterior* pr-probability distribution of the batch mean, which determines the expected value of the decision alternatives according to any particular performance measure. By solving Bayes* Rule for hypothetical future combinations of sample size and batch mean, an optimal design can be found.

The term “Bayesian” usually refers lo the sometimes controversial approach of using subjective knowledge to estimate prior distributions and likelihood functions (3,4). However, because those are given in the example, the posterior probability distribution from Bayes* rule is exact, as is the resulting optimal design solution. The Bayes optimal design was computed for two cases in the example, confirming the process simulation results. This is described in the supporting materials. Obtaining an optimal design is important; the method used to obtain it is not. The use of process simulation in this paper is the author*s preference.

Confidence: Minimize sampling cost with limits on decision error rates

In this approach, the performance objective is to make a decision with a high level of confidence that the decision is correct. The sampling design objective is to achieve the desired confidence level at minimal sampling cost.

The example follows the well documented (1, 12) “Data Quality Objectives” (DQO) process currently recommended by the EPA. The process is abbreviated here because the basic question. “What should be done with the waste?”, has already been answered: “Treat the waste if the true concentration exceeds 100 ppm.” The DQO process restates the decision to treat waste as equivalent to accepting or rejecting a null hypothesis. In the example, the two possible hypotheses are “the concentration is greater than or equal to 100 ppm” and “the concentration is less than 100 ppm.” One of these hypotheses is chosen as the null, and limits are then set on the allowable frequency of decision errors, based w an evaluation of the consequences of error.

Unlike the case for medical tests or scientific experiments, the choice of null hypothesis is not always obvious for environmental decisions. The “greater than” hypothesis applies if the health and ecological consequences of failure to treat contaminated waste are considered more severe than the costs of unnecessary treatment; otherwise, the “less than” hypothesis would be chosen. Figure 2 shows DQO performance limits for a typical “greater than” null hypothesis (1). Any decision process that meets these performance requirements must have a treatment probability within the acceptable area at all concentrations. In statistical terms, the upper and lower acceptable limits in Figure 2 correspond to false positive (alpha) and false negative (beta) error rates of 0.05, while the width of the gray region is called the “minimum detectable difference”.

In this example, the operator could meet the permitting requirements by doing no sampling and treating every batch at a cost of $10,000 per batch. However, with perfect knowledge, the operator could expect to treat approximately 50% of the batches, assuming the future batch distribution is similar to Figure 1. This would reduce the average batch cost to about $5,000.
The $5,000 difference is the maximum potential benefit from sampling, and thus an upper limit on sampling cost. EPA guidance recommends setting allowable error rates at 0.01 initially; only relaxing the requirement if adequate sampling designs are too expensive. Initially, gray region limits were set at 80 and 100 as in Figure 2, with error limits at 0.01. The number of samples needed to achieve these initial performance limits was estimated to be 144 (13, supporting material), which would cost $3600 per batch - a large fraction of the potential benefit. By increasing error rates to 0.1 and dropping the lower gray region bound to 75 ppm, the estimated sample size drops to a more reasonable 29, where sampling cost is about 15% of the potential benefit.

The optimal design (Figure 3) was [29, 87]. The "less than" hypothesis was evaluated in a similar manner. In this case the lower bound of the gr-ay region was set at 100 ppm and the upper bound at 133 ppm, with error rates of 0.1. The optimal design was [29, 116].

Cost: Minimize a cost function that includes sampling costs, treatment costs and environmental costs

The objective here is to design the most cost effective operation. In the example, sampling and treatment costs are known; "Environmental costs" of not treating a batch of waste must be determined. Environmental costs may include health and ecological effects and intangibles such as public dissatisfaction. Here, these costs are not estimated directly; rather, they are inferred from the ideal decision rule. Simply put, if we decide to treat a batch of waste, we must believe it is worth treating. The target threshold is taken to be the break-even point separating what is worth treating from what is not, that is, where the treatment cost equals the environmental cost. Thus the environmental cost of a batch of waste containing 100 ppm arsenic, if untreated, is $10,000.

If the environmental cost of no arsenic is zero and the environmental cost is proportional to the amount of arsenic, we obtain the linear cost model show in Figure 4. Various combinations of sample size and operational action level, and the resulting total cost response surface was contoured in Figure 5. The optimum design is [5,95]; the performance curve is shown in Figure 6. The environmental cost model need not be linear. The exponential model in Figure 4 imposes higher costs for larger treatment failure errors, perhaps reflecting loss of public confidence and good will. The optimal design for this case is [8,93]. Other cost factors not considered here, such as fines or penalties, might add discrete step increases to the underlying environmental cost model.

Risk; Minimize sampling and treatment cost with limits on posterior risk

In this approach the objective is to find the lowest cost decision process that adequately controls *posterior risk,* that is, the expected risk given the operation of the process. *Prior risk* is the risk that would occur without any decision process, and consequently, no treatment. Posterior risk is the sum, over all possible outcomes of the process, of the risk associated wilh an outcome times the probability of that outcome.

In the example, absolute risk limits are not used; instead, risk limits are set relative to the posterior risk from a perfect process. The maximum posterior risk if decisions were perfect equals the prior risk from an arsenic concentration of 100 ppm. This is taken as a threshold for acceptable posterior risk. The performance of the decision process will be designed so posterior risk will be equal to or less than the threshold. Risk is assumed to be proportional to concentration.

Stating the objective in terms of risk has a major impact on design. Risk is a probability, not a tangible substance like arsenic. Posterior risk is a conditional probability dependent on the prior risk and the expected outcome of the process. Thus, while prior risk is proportional to true arsenic concentration, posterior risk is proportional to ***expected*** arsenic concentration. That is equivalent to the mean contribution to the landfill - the mean of the treated and untreated batches - over a large number of batches with the same initial concentration. The overall expectation of the process determines posterior risk, not any individual outcome.

The expected concentration of the batch is computed from the probabilities of all possible outcomes. Expected concentration (Ze) is the probability that the batch will go untreated (pu), times the original batch concentration (Z), plus the probability that the batch will be treated, Limes the concentration after treatment (Zt):

$$Ze = pu * Z + (1 - pu) * Zt,$$ or

$$pu = (Ze \text{ B } Zt)/(Z-Zt)$$

Substituting the maximum allowable expected concentration for Ze gives the maximum allowable value for pu, which is thus the bound on decision performance. Assuming perfect treatment,

$Zt = 0,$ giving

$pu = Ze/Z$, *and*

$pu[max] = 100/Z$

Plotting treatment probability (1 - pu[max]) versus concentration gives the curve shown in Figure 7 that establishes the risk based performance limit. As long as actual treatment probability is above this curve, the posterior risk will be below the threshold. The optimal design, *[1, 160]*, has the lowest sum of sampling and treatment costs of all designs that meet the requirement.

In a more realistic example, *Zt* might be non-zero or even a complex function of concentration and failure probabilities. It does not matter whether posterior risk is the result of decision error, or of treatment failure, or of incomplete treatment. Effective risk management requires that all contributions to posterior risk be managed. increasing *Zt* because of treatment failures and imperfections reduces the allowable error rate. Suppose the treatment process only removed 90% of the arsenic and completely failed 10%, of the time. The concentration after successful treatment is 0.1 Z, and after unsuccessful treatment is Z. The expected concentration after treatment is:

$Zt = 0.9*0.1Z + 0.1Z = 0.l9Z$, and

$pu[max] = (100\ B\ 0.19Z)/8.81Z$

In this case, the required treatment rate exceeds 100% for concentrations greater than 527 ppm, When this occurs, the treatment performance itself fails to meet the risk objective, even if decisions are perfect. Controlling risk at the target level requires improving the treatment.

Risk and Decision Scales

All of the analyses above were based on the application of the 100 ppm threshold to batch-scale decisions. This section will discuss the effect of changing the scale of concern on risk management strategy.

In a landfill scenario, risk is modeled as the integral of numerous release and failure possibilities over long periods of time. Consequently risk is determined by the total amount of contaminant in the landfill, 1
not by how it is distributed among batches of waste. Assuming the 100 ppm threshold came from risk modeling, it can be argued that the threshold properly applies to the mean concentration of the landfill. Decisions may be made on batches, but errors are significant only in their cumulative effect on the landfill mean. The decision process is then multi-scale, where decisions on batches are made to manage landfill risk.

The multi-scale risk management strategy was developed and implemented by the EPA for soil clean-up at the Piazza Road Superfund site (7,9,10) This was a pilot study led by EPA*s Quality Staff (then Quality Assurance Management Staff) to test the DQO process which was then being developed. Piazza Road demonstrated the importance of applying risk-based performance criteria at the appropriate risk management scale. The risk of concern was residential exposure, so the risk management unit was a yard-sized "exposure unit". Decisions on much smaller "rcmediation units" were made in order to treat the minimum required to briny the exposure unit below the threshold. This approach was shown to significantly reduce treatment costs while strictly controlling risk.

In the example the optimal design is the lowest cost combination of sample size and operational action level that adequately controls the expected landfill mean. As before, the optimum was determined by the simulation procedure. The decision performance curve for this optimal design [2, 325] is also shown in Figure 7. One problem with this approach is that it strongly depends on the assumption that the distribution of future batches is known. In practice, it would require monitoring the distribution of batch means and periodically reevaluating the action level.

An alternative approach makes decisions that keep the expected cumulative landfill mean below some threshold. Decision errors produce a conditional bias (true landfill mean greater than the estimated mean), so the decision threshold must be below the risk threshold to compensate. Fewer samples mean more errors, a lower decision threshold, and thus higher treatment costs. Conversely, taking more samples reduces treatment costs, so design again involves finding the combination of decision threshold and sample size that keeps the true landfill mean below 100 ppm at minimum cost. The optimal design for this approach is [2, 92(cum)]. As shown in Figure 8, the landfill mean fluctuates initially when individual batch errors can have a significant effect. but ultimately converges to the expected value.

Results and discussion

Table I compares various performance measures for the optimal designs from the example. Three reference cases are also shown: perfect decision making; treating everything; and treating nothing.

For the example, the choice of performance measure and corresponding design objective was critical. Optimum designs varied by a factor of 29 in sampling cost, by 5 in treatment cost. and by more than 4 in mean concentration of the landfill. To the extent that these results can be generalized to real world cases, they suggest that considerable attention should be given to choosing the most appropriate objective for the case at hand.

Objective 1 controls decision error rates to ensure the desired level of confidence that a decision will be made correctly. This would obviously be appropriate when trying to establish a legal case "beyond reasonable doubt". It might also be indicated in highly publicized and controversial situations where reassuring the public is just as critical to a successful project as is managing the risk itself.

Objective 2 balances sampling, treatment, and environmental costs. This seems most appropriate when decisions are part of a routine operation, particularly if the threshold is a generic standard. The simple linear cost model produces an optimal design that appears quite good by most measures, compared to perfect performance. When compromise is sought between an operator*s desire for cost management and the regulators* mandate for environmental protection, this approach may provide a reasonable solution.

Objective 3 manages: risk. This is particularly appropriate when target thresholds have been determined by case specific risk analysis. It is the most scientifically defensible objective in such cases, because the risk management process is directly tied to the risk assessment. Design optimization becomes, in effect, an extension of the risk modeling process.

There is no reason why one of the three objectives must be selected to the exclusion of the others. Real world optimization often involves compromising among a number of desirable objectives, Optimizing designs for different performance criteria, as in this paper, can be advantageous. Comparing the various performance measures for alternate optimal designs will provide a clearer

Supporting materials

Supporting materials include additional discussions of hypothetical versus real world examples; process simulation versus Bayesian optimization, and estimating the sample size for objective 1. Optimizations and graphics were done using R software (supporting materials) - an open source freeware system for statistical computation. Commented examples of R command sequences arc provided for the process simulation optimization procedure as well as the alternative Bayesian optimization approach.

Acknowledgment

This work was conducted by EPA*s Office of Research and Development. It has not been reviewed by the EPA and does not necessarily reflect EPA policy. Dean Neptune of Neptune and Co, and John Warren of EPA*s Quality Staff contributed stimulating discussions and correspondence on the subject.

1. U.S. Environmental Protection Agency, *Guidance for the Data Quality Objectives Process (G-4)*, EPA/600/R-961005, August **2000**; http:www.epa.gov/quality/qa_doc.html.

2. U.S. Environmental Protection Agency, *OSWER PBMS Implementation Plan,* October **1998**; http://www.epa.gov/epaoswer/hazwaste/test/pdfs/pbms.pdf.

3. Freeze, R.A.; Massmann, J.; Smith, L.; Sperling, T.; James, B. *Ground Walter* **1990,** 38. 738-766.

4. Wolfson, L. J.; Kadane, J. B.; Small, M. J. . *Ecol. Applications* **1996**, 6, 1056-1066.

5. Fox, D. *Environtnetrics* 2001, 12, 437-439.

6. Englund, E. J.; Heravi, N. *Environ. Ecol. Stat.* 1994, 1, 247-263.

7. Blacker, S.; Goodman, D. *Environ. Sci. Technol.* 1994, 28, 466A-477A

8. Ginevan, M.E.; Splitstone, D. E. *Environ. Sci. Technol.* **1997**, 31, 92A- 96A.

9. Ryti, R. T.; Neptune, D.; Groskinsky, B. *Environmental Testing & Analysis* 1992, January/February.

10. Ryti, R. T., J. *Air Waste Manage. Assoc.* 1993,43, 197-202.

11. Winkler, R. L. *Introduction to Bayesian Inference and Decision,* Holt Rinehart and Winston, New York, **1972.**

12. Myers, J. *Geostatistical Error Management,* Van Norstrand Reinhold, New York, **1997.**

13. U.S. Environmental Protection Agency, *Decision Error Feasability Trials (DEFT) Software (G-4D),* EPA/240/B-011007, September **2001**; http://www.epa.gov/quality/qa_docs.html.

Figure 1

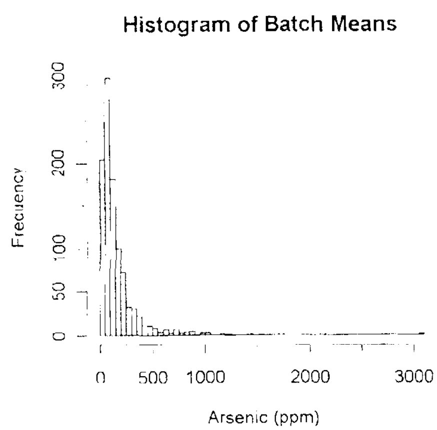

Figure 2

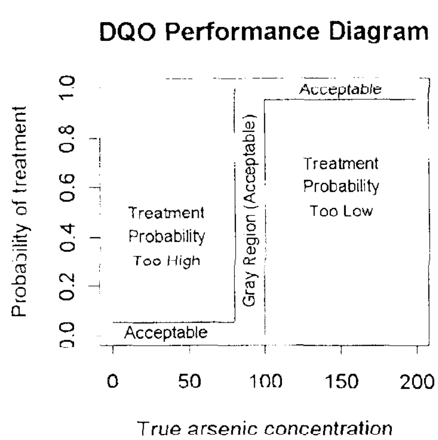

Figure 3

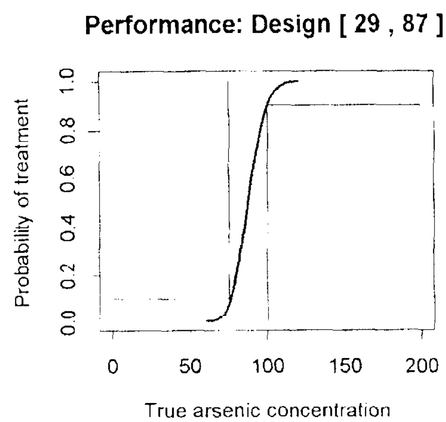

Figure 4

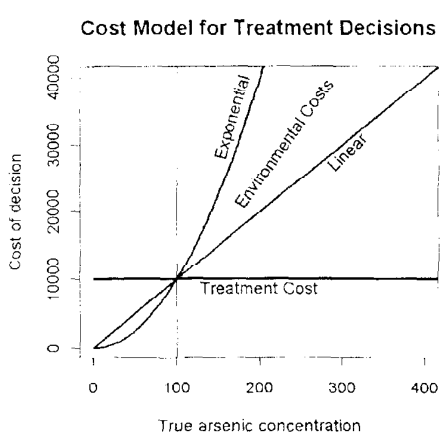

Figure 5

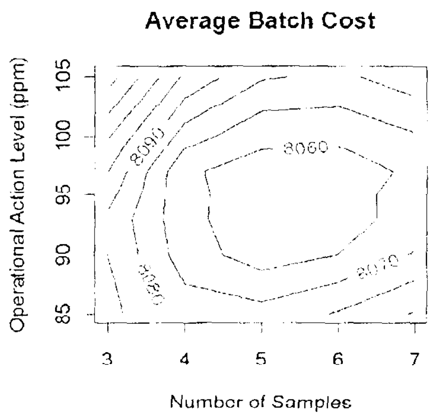

Figure 6

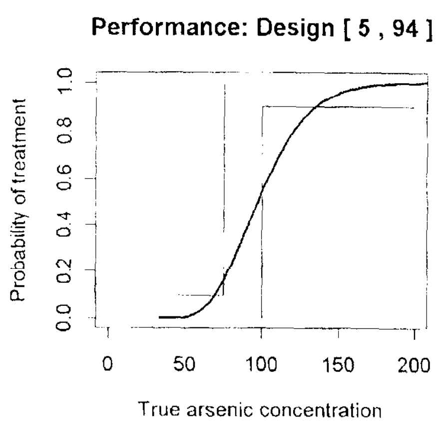

Figure 7

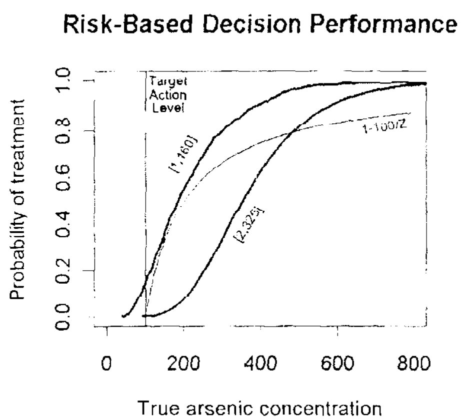

Figure 8

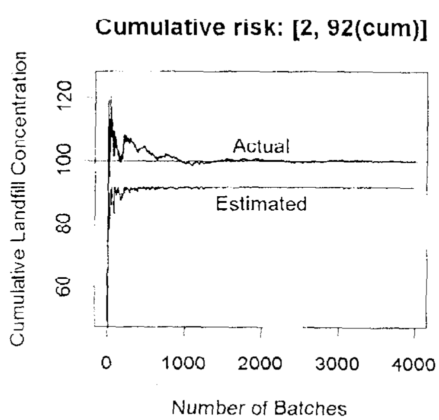

Figure Captions

Figure 1. Histogram of mean arsenic concentrations of 1000 hypothetical waste batches.

Figure 2. A schematic DQO performance goals diagram. The null hypothesis is that true arsenic concentration exceeds 100 ppm. Consequently the upper bound of the gl-ay region is set at 100 ppm, The performance curve for any successful design must be within the acceptable area.

Figure 3. Performance curve for the optimal confidence based design where the null hypothesis is that the waste batch mean is greater than 100 ppm.

Figure 4. The cost models used for cost based design optimization. Treatment cost was assumed to be independent of arsenic concentration. The environmental cost models were inferred by assuming that they are equal to the treatment costs at the ideal action level; the linear model is proportional to concentration and the exponential model proportional to concentration squared.

Figure 5. Total cost contoured as a function of the number of samples and the operational action level, assuming the linear environmental cost model. The minimum cost design is approximately [5, 95]. Total cost is the sum of sampling cost, treatment cost and environmental cost averaged over all batches.

Figure 6. Performance curve for the minimum cost design from Figure 5. For reference, the DQO performance goals from Figure 3 are shown.

Figure 7. The light curved line is the upper risk based performance bound (treatment probability = 1-100/Z). There is no lower performance bound for risk. The heavy lines are performance

curves for two optimal designs. The [1,160] design is optimal at the batch scale - the expected contribution to the landfill of any batch is at or below 100 ppm. The [2.325] design is optimal at the landfill scale - the expected contribution to the landfill of *all batches combined* is at or below 100 ppm.

Figure: 8. Example of the performance of the cumulative mean decision rule with the optimal design of [2, 92(cum)]. Batches are treated if not treating them would raise the estimated cumulative landfill mean above 92 ppm. The estimated mean converges to 92 ppm, while the actual mean converges to an expected value of 100 ppm.

Table 1: Comparison of Decision

	Optimal Design	Treatment Cost	Sampling Cost	Environmental Cost	Total Cost	Probability of Treatment at: 75 ppm	100 ppm	133 ppm
Confidence								
Ho: Z > 100	[29, 86.6]	5631	725	2216	8572	0.08	0.9	1
Ho: Z < 100	[29, 115.5]	4305	725	3558	8588	0	0.08	0.9
Cost								
Linear	[5, 95]	5058	125	2864	8047	0.14	0.52	0.89
Risk								
Batch Mean	[1, 160]	2761	25	6645	9321	0.04	0.139	0.278
Landfill Mean	[2, 325]	1036	50	9980	11066	0.001	0.004	0.016
Cum. Landfill Mean	[2, 92(cum)]	1173	50	9991	11214	n.a.	n.a.	n.a.
Reference Cases								
Perfect	[0,100]	4970	0	2811	7781	0	1	1
Treat All	[0,0]	10000	0	0	10000	1	1	1
Treat None	[0,1000000]	0	0	15606	15606	0	0	0

Costs are given as average per batch

(Environmental Cost)/100 = mean landfill concentration