Review and Comment Concerning A Draft EPA Guidance Document:

"Guidance on Classification For Purposes of Disposal of Stored Natural Gas Pipe Which was not Part of a Pipe Removal Project Carried Out Under an EPA-Approved PCB Disposal Activity"

January 13, 1994

Prepared by:

William A. Huber, Ph.D.
Quantitative Decisions.
539 Valley View Road
Merion Station, PA 19066

TABLE OF CONTENTS

1.0 Introduction **1-1**

2.0 Review **2-1**

2.1 What the PCB Measurements Represent 2-1

2.1.1 Implications of different sample population definitions (pages 3-4) 2-1
2.1.2 Procedures for random sampling of pipes (pages 6-7) 2-2
2.1.3 What the wipe samples really measure (pages 5-6) 2-2

2.2 Statistical Elements of Rational Decision-Making 2-3

2.2.1 Basis for the "K-Factors" (pages 7-8) 2-5
2.2.2 Meaning and justification of the sample statistic "X BAR" (pages 4-5) 2-6
2.2.3 Rationale for using a tolerance interval (page 5) 2-7
2.2.4 What "Loss Function"? 2-7

2.3 Establishing Sample Sizes 2-8

2.3.1 Estimating sample size (page 6) 2-8
2.3.2 How the sample mean can be more or less "representative" (page 8) 2-11

3.0 Discussion and Comments **3-1**

3.1 The Guidance provides a good framework for regulating the disposal of stored pipeline 3-1

3.2 Some procedures specified by the Guidance could be harmful or costly to regulated companies 3-1

3.3 Alternative approaches to solving the pipe characterization problem 3-2

4.0 Recommendations **4-1**

5.0 References **5-1**

6.0 Appendices **6-1**

6.1 Appendix 1: The Draft Guidance 6-1

6.2 Appendix 2: The Exponential Distribution 6-12

6.2.1 Definition 6-12
6.2.2 Sampling Distribution of "X BAR:" The Gamma Distributions 6-13
Why X BAR is a Useful Statistic 6-14
6.2.3 Computing the K-Factors 6-14
6.2.4 Assessing the Power 6-16

1.0 INTRODUCTION

This draft review explains and interprets the statistical content of a draft EPA guidance document [EPA 1991], hereafter abbreviated as "Guidance." Its objective is to identify issues that may possibly be important to regulated facilities and to provide recommendations for addressing them.

The ten-page Guidance (attached as Appendix 1 to this report) consists of three sections:

I. Introduction

II. A Statistical Procedure For Sampling A Population of Stored Pipeline

III. Additional Information Requirements

The first section of the Guidance introduces 40 CFR 761, the Toxic Substances Control Act (TSCA), as its regulatory basis. TSCA governs the disposal of PCB-contaminated articles. The EPA allows different disposal procedures according to the concentrations of PCBs in pipe:

- If the concentration exceeds 500 parts per million (ppm), incinerate or landfill after draining liquids--40 CFR 761.60(b)(5)(i).
- If the concentration exceeds 50 ppm, but does not exceed 500 ppm, drain liquids and store according to 40 CFR 761.65 before disposal--40 CFR 761.60(b)(6).
- If the concentration exceeds 50 ppm, but does not exceed 500 ppm, dispose of the drained article in an unregulated fashion--40 CFR 761.60(b)(5)(ii).

The referenced regulation--40 CFR 761.65--also recognizes 50 ppm and 500 ppm as decision points without introducing any other reference concentrations.

The second section of the Guidance discusses how to classify stored pipeline according to these regulatory concentration thresholds. Its procedure contains the following important elements:

- Specification of a stored pipe population (pages 3-4)
- Random sampling of pipes (page 6-7)
- Measurement of PCB concentrations using wipe samples (pages 5-6)
- Assumption of an underlying exponential distribution of measured concentrations (pages 7-8)

- Use of the mean PCB concentrations in the wipe samples ("X BAR") to characterize the PCB concentration in the stored pipe population (pages 4-5)
- Classification of the stored pipe population using a one-sided upper tolerance limit (page 5)
- Specification of minimum sample sizes to achieve sufficient "power" and "representativeness" (pages 6, 8, and 9)

The review and comments in this report focus on the second section of the Guidance. Section 2 provides a detailed commentary. Section 3 is a discussion designed to place the details in perspective and to exhibit relationships among the concerns raised in Section 2 . Section 4 presents recommendations for responding to and complying with the Guidance. The two remaining sections supply references and technical appendices.

The third section of the Guidance is brief and unrelated to the statistical issues.

Added 1 February 2001:

Federal Register: June 29, 1998 (Volume 63, Number 124), page 35397, supersedes the Guidance:

> One commenter asked EPA to clarify the relationship between the proposed regulations and EPA Technical Guidance Documents (TGDs). The commenter requested that EPA allow regulated entities the option of using TGDs and ADPs to meet the proposed requirements (e.g., Secs. 761.30(i)(5) and 761.60), particularly with respect to using existing PCB concentrations rather than presumed concentrations. The three TGDs for declassification, abandonment, and classification of stored pipe (Refs. 11, 12, and 13) were developed to implement EPA's presumption policy of PCB contamination at 500 ppm. As discussed above, today's rule eliminates the presumption policy and allows natural gas pipeline systems to be managed based on actual PCB concentration. Therefore, today's regulations supersede these guidance documents.

(Reference 12 is the Guidance.) The full text is available on the web at http://www.epa.gov/opptintr/pcb/pcbdisp.txt. This change renders some of the criticism in the present document of historical interest only, but the analysis of the proposed tolerance limit test and the discussion of the attendant statistical issues remain valid.

2.0 REVIEW

2.1 What the PCB Measurements Represent

2.1.1 Implications of different sample population definitions (pages 3-4)

The Guidance provides a good discussion of the definition of a stored pipe population. It distinguishes "general" from "specific" populations. A *general population* of stored pipe consists of "all pipe which is within the property boundaries of the storage site and for which their source and sequence information is not available." Reading the Guidance strictly, a *specific population* consists of pipe having a common diameter. However, the EPA's intent evidently is to cause pipe to be grouped according to all factors that could be related to potential PCB contamination.

Important factors could include:

- Material and construction
- Size
- Date of manufacture or installation into the distribution system
- Date of removal from the distribution system
- Visual status (staining or presence of liquids, for example)
- Location within the distribution system
- Proximity to sections of installed pipeline known to exhibit contamination

When too many factors are considered for defining specific populations, each population will have a small amount of pipe. This will increase the expense of sampling, because the ability of measurements to characterize a population depends only on the number of measurements, not on the size of the population (when the measurements do not exhaustively sample the population). Thus, where ten measurements might suffice to characterize a general population, one hundred measurements would be needed if it were divided into ten specific populations.

One purpose of dividing a general population into specific populations is to minimize the risk of mis-characterizing the PCB distribution. PCB distributions in a heterogeneous population could be multi-modal (for example) and difficult to describe statistically. The EPA's fundamental assumption for statistical analysis, that the distribution of PCB concentrations is exponential, would then not be true.

The Guidance allows companies to propose different criteria for defining the stored general pipe populations. Presumably this offer extends to defining specific populations, too: the Guidance specifies the factors to be observed--source, sequence, and diameter--but does not specify how they are to be used for defining specific populations.

2.1.2 Procedures for random sampling of pipes (pages 6-7)

The Guidance requires that the "samplers shall select individual pipe units from the population according to a random sampling procedure ... in a statistically valid way." Its advice is good, because any non-randomized procedure cannot be defended against a charge of (implicit) bias. We would add three more related recommendations.

1. The sample selection procedure should be documented in detail *before* any samples are obtained. For many reasons, some of them unconscious, samplers can circumvent sampling plans: for example, some pipe may be relatively inaccessible, or some may appear more "appropriate" for sampling. It is too easy after the fact to make such judgment sampling appear to be the result of random sampling. More subtly, it is possible to generate several (or many) truly random samples, and then select which set of random samples to observe based on judgment. Selecting from sets of random samples produces a sample that is no longer random, although the documentation looks excellent: random numbers were used for selection.
2. If pieces of pipe within a population have greatly varying lengths, it may be worthwhile to account for these differences. The reason for doing so would depend on the origin of the pipe. When there is no prior reason to suspect contamination in two identical uniform sections of pipe, one of which is 10 feet long and the other 40 feet, then sampling should be four times more intense in the longer section.
3. PCB concentrations within a pipe could vary greatly. The PCBs are carried in the liquids that flowed along the bottom of the pipe. To identify possible contamination, it is reasonable to sample the bottom of the inner portion of each pipe. However, such samples would not be representative of PCB concentrations in the pipe material as a whole. Hence, relative sample location within each pipe should be considered as another factor when defining sampling populations.

2.1.3 What the wipe samples really measure (pages 5-6)

According to the Guidance, the wipe sample measurements may have no bearing on the PCB content of the pipe: "there is no empirically derived comparison between surface level concentrations of 100 μg/100cm^2 and liquid or non-liquid concentrations of 500 ppm." This is a serious and fundamental deficiency.

For physical reasons, one cannot expect any consistent relationship between surface level concentrations and concentrations within the pipe material itself. Old absorbent pipe material may hold much more PCBs than can be collected from a surface sample. Unabsorbent material, such as steel, may confine PCBs to the surface. The resulting wipe sample concentrations will grossly overestimate the total amount of PCBs.

The primary exposure pathway for PCB-contaminated pipe would be inhalation. The dermal contact route, which could be a concern for PCB contamination on other surfaces, is essentially nonexistent because of the pipe geometry. Surface wipe sample results generally have no correlation with airborne concentrations. Therefore the EPA's claims that "surface level concentrations represent the only reasonable alternative for assessing risk" and that its "legal authority to deregulate pipe based on surface level concentrations is well established" appear unfounded.

2.2 Statistical Elements of Rational Decision-Making

The Guidance provides a procedure for making a decision--classifying pipe for disposal--on the basis of measurements of randomly selected portions of pipe. This review shall later demonstrate that the Guidance adopts contradictory views of this procedure, resulting in over-regulation that will be unnecessarily burdensome to the regulated community. This subsection presents a scientific basis for evaluating decision procedures, a basis to be used later for carrying out the promised demonstration.

A *decision procedure*, or statistical test, associates a decision with every possible measurement result ("outcome"). In practice, this association is unique, and is specified using a statistical "recipe," such as:

> Multiply the sample mean by the factor found in the table. If the result is less than 100, decide that the pipe is unregulated for disposal; otherwise, decide that it is regulated.

Four ingredients are needed to completely specify a decision procedure [Kiefer, 1987]:

1. A rigorous description of the set of possible measurement results ("sample space")
2. A complete description of the possible decisions ("decision space")
3. A set of possible probability laws to model the random behavior of the results ("state space")

4. For every possible combination of probability law F and decision D, an explicit evaluation of the potential merit of making decision D when the true probability law is F ("loss function")

Given the four ingredients, a statistical "risk function" can be computed for any possible decision procedure. The risk is the loss expected in the long run based on consistently using the procedure. Because the actual probability law, or "state of nature," is unknown, and will vary from one application to another, the risk must be expressed as a function of the state space. It is possible that no decision procedure will be the best under every circumstance. The risk function is a valuable tool for evaluating decision procedures because it reveals this fundamental difficulty, enabling potential users of a statistical procedure to understand its strengths and limitations. The Guidance expresses risk in terms of *confidence* and *power,* which are defined and discussed below.

The first two ingredients have already been described in the section 2.1:

1. The measurement results are wipe sample concentrations. It will be simplest, for the sake of discussion, to assume that (1) measurement results are reported as numerical values by the analytical laboratory, so that no "nondetect" or "unquantified" results will be received; and (2) all reported numerical values will be positive. Upon obtaining PCB measurements of N randomly located samples, the sample space of all possible results will consist of the set of unordered collections of N values between 0 and 100%. (The assumptions about nondetects are not essential, but relaxing them would complicate the exposition of the ideas.)
2. The Guidance describes a very simple decision space having two elements: either the population of pipe should be "regulated for disposal" or not. Hereafter these decisions will be referred to as "regulated" and "unregulated," respectively. The TSCA regulations clearly indicate that the decision should be based on whether concentrations in the pipe material exceed 500 ppm. As previously discussed, a surrogate for that criterion will be whether wipe sample concentrations in the pipe population exceed 100 $\mu g/100cm^2$.

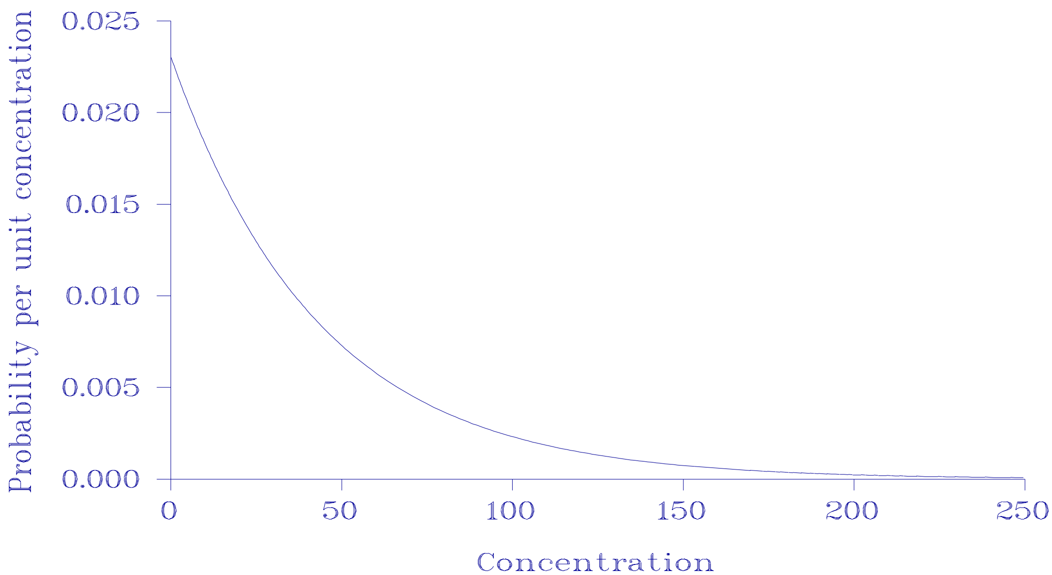
The remainder of this section discusses the remaining ingredients: describing the state space ("exponential distribution"), selecting a decision procedure ("tolerance interval" and "X BAR"), and the loss function (not discussed in the Guidance).

2.2.1 Basis for the "K-Factors" (pages 7-8)

The K-factors are the result of developing a test of hypothesis for an exponential distribution. This subsection discusses the exponential distribution. A later subsection, 2.3.1, discusses the test of hypothesis.

The Guidance makes the "assumption that there is an exponential distribution of PCB levels throughout the pipeline." It supplies no justification, nor does it explain the properties of the exponential distribution. Assuming an exponential distribution is unusual, but not necessarily invalid. Figure 1 illustrates the exponential distribution. Appendix 2 provides a more detailed discussion of the properties of the exponential distribution and derives the K-factor values.

Figure 1
Exponential Distribution: Probability Density Function



This figure accurately shows the probability density function for an exponential distribution. Its formula is

$$y = e^{-x/\theta} / \theta,$$

where θ is the *parameter* of the distribution. Changing θ changes the scale (and location) of the distribution, but not its shape. In the figure, θ was chosen to make exactly 10% of the probability lie above 100 concentration units: $\theta = 100/\ln(10)$, which is approximately 43.43. The true mean and true standard deviation of an exponential distribution with parameter θ are both equal to θ.

The merits of assuming an exponential distribution include:

- It is mathematically tractable, enabling direct calculation of important statistics such as tolerance limits, confidence limits, and power curves.
- It has a skewed shape, similar to distributions often exhibited by environmental concentration measurements.
- All statistical information about a random sample from a population is contained in the sample mean (X BAR) alone.

- It is simply described and depends on just one parameter.

Some possible disadvantages and objections are:

1. The exponential distribution might not accurately model the true behavior of the PCB levels.
2. Having only one parameter that simultaneously specifies both scale and location may be too restrictive to be realistic.
3. The theory of the exponential distribution is not well known, nor are tables of statistics (such as the "K-factors") widely published.
4. The range of the exponential distribution is infinite, whereas the largest possible concentration is 100%.

The most serious objection is the first, because if true, all results would be meaningless. Fortunately, it is possible to test the assumption that PCB concentrations are exponentially distributed. How to do so is discussed in Chapter 3. The second objection can be tested with actual data. Expanding the exponential distribution to include other distributions, such as Gamma distributions, might overcome the single-parameter limitation. The other objections are minor. Appendix 2 fills the theoretical gap mentioned in the third objection, showing how to derive the K-factors. The last objection applies to conventional probability models, including the normal and lognormal distributions. Because all of these models assign infinitesimally small probabilities to very large concentrations, they are effective in practice.

2.2.2 Meaning and justification of the sample statistic "X BAR" (pages 4-5)

The theory of normally distributed random variables is popular in part because all the statistical information contained in a set of normally distributed measurements is described by only two statistics: the mean and standard deviation. In this regard, the theory of exponentially distributed random variables is even simpler: all the information is described by the sample mean alone. It is a "minimal sufficient statistic." Appendix 2 describes the basis for this claim. What it means in practice (technically, for "convex loss functions") is that any decision procedure based on a set of measurements can always be replaced by one at least as good (for every possible state of nature) that involves only the sample mean. Therefore, the use of "X BAR"--the sample mean--is thoroughly justified by theoretical considerations, *once the assumption of an exponential distribution is adopted*. (If this assumption is modified, then the sample mean may no longer be a sufficient statistic. For example, both the sample arithmetic mean and sample geometric mean are needed for making inferences about a gamma distribution.)

2.2.3 Rationale for using a tolerance interval (page 5)

A tolerance interval estimates a range of the underlying distribution of measurements. As used in the Guidance, all tolerance intervals are one-sided, beginning with zero and ending at some fixed upper percentile of the distribution. Therefore this review refers to them as tolerance *limits*. The limit proposed by the Guidance is based on the 90th percentile (its *coverage*), but in theory any percentile greater than 0 and strictly less than 100 could be applied. Using the 90th percentile means that the EPA wishes not to regulate the entire population of pipe, nor even an average aspect of it; instead, *it will regulate a population of pipe for disposal based on the highest 10% of concentrations*. This philosophy is consistent with other parts of the TSCA regulations, which are concerned with identifying potential "hot spots" of PCBs in spill areas [EPA 1985].

Use of a percentile higher than the 90th would be more stringent, because if a higher percentile of a distribution meets a fixed standard, then so will all percentiles lower than it. Conversely, use of a percentile lower than the 90th would be less stringent.

One advantage of using a high percentile is that some tolerance limit procedures can accommodate many nondetect results. Provided that all detection limits are less than the smallest quantifiable result reported, and provided that about 10 percent of the results are quantified (the exact amount depends on the procedure used to estimate the tolerance limit), a tolerance limit can be computed. However, when so many measurements are nondetect, it becomes difficult to test an underlying distributional assumption without obtaining many (more than 100) samples. Note that the tolerance limit procedure of the Guidance would require modification or replacement to handle nondetects.

2.2.4 What "Loss Function"?

A loss function, in the setting established by the Guidance, describes how one assesses the "badness" of the two possible decisions--treat the pipe as regulated or not--for every possible underlying exponential distribution. It is rare for a loss function to be explicitly given in this kind of setting. Doing so would clarify exactly how the EPA balances the cost of unnecessarily disposing uncontaminated pipe against the potential harm in failing to dispose of contaminated pipe. (For a clear discussion of this issue, see [EPA 1993].) Instead, the EPA seeks to reduce the frequency with which contaminated pipe is mistakenly classified as unregulated. As one would expect, this frequency is very low for highly contaminated pipe because of the high probability that most samples will yield high measurements. The frequency increases when pipe exhibits only slightly more than 10% contamination.

Conversely, the interests of the regulated party lie in reducing the frequency with which pipe meeting regulatory requirements is incorrectly identified as subject to disposal. This frequency will be lowest for pipe completely free of PCBs and highest for pipe marginally below the Guidance's threshold of 10% contamination at the 100 μg/100cm^2 level. The only means of changing these frequencies are:

1. Obtain a different number of samples: having more measurements always reduces the risk of error.
2. Change the decision procedure.

The EPA is not solely concerned with error frequencies. When a decision procedure becomes onerous, such as by forcing regulated parties to gather too much expensive data, noncompliance and economic burden become important issues. Therefore it is proper that this Guidance consider sample sizes as well as error frequencies. This is the topic of the next subsection.

2.3 Establishing Sample Sizes

2.3.1 Estimating sample size (page 6)

The proper basis for estimating sample size should be consideration of statistical risk. The Guidance alludes to this by claiming to "ensure that the characterization test has adequate power." Yet the risk, as measured in the context of the decision-theoretic assumptions discussed previously, depends solely on the number of samples and not on the size of the pipeline population. Therefore, the procedure discussed in the Guidance, which bases sample sizes on pipeline populations, is not justifiable.

The use in the Guidance of terms such as "power" and "confidence" reveals a test-of-hypothesis approach to evaluating decision procedures. This approach views the types of decision errors unsymmetrically, depending on the "null hypothesis" that is adopted. It appears that the Guidance is inconsistent in its approach. This claim will now be demonstrated.

The two possible null hypotheses are that the (true) 90th percentile of PCB concentrations in the pipeline population is (1) less than or (2) greater than the target. Each hypothesis describes a set of possible states of nature. The test-of-hypothesis approach looks at the frequencies with which the procedure performs correctly:

- The minimum frequency with which the null hypothesis will be accepted when it is true is the test *confidence*.
- The probability of correctly deciding against the null hypothesis is the test *power*.

A "type I error," or "false positive," occurs when the test incorrectly decides against the null hypothesis. The confidence is 100% minus the maximum false positive frequency. A "type

II error," or "false negative," occurs when the test incorrectly decides for the null hypothesis. The power is 100% minus the false negative frequency. Hence confidence and power re-express a portion of the risk function.

The Guidance does not specify the null hypothesis. Two cases are therefore possible.

1. The null hypothesis is that the 90th percentile of the pipeline population is **less than** the target. Then (a) the confidence is the minimum frequency with which the test procedure will support this assertion when it is true, and (b) the power can be expressed as a graph showing the frequency with which the test will correctly decide that the pipe **should** be regulated for disposal. The horizontal axis of the graph will designate those exponential distributions whose 90th percentiles exceed the target.
2. The null hypothesis is that the 90th percentile of the pipeline population is **greater than (or equal to)** the target. Then (a) the confidence is the minimum frequency with which the test will correctly decide that the pipe should be regulated for disposal, and (b) the power can be expressed as a graph showing the frequency with which the test correctly decides that the pipe **should not** be regulated for disposal. The horizontal axis of the graph will designate those exponential distributions whose 90th percentiles fall below the target.

In the first case, the EPA will be primarily concerned with power, and in the second case, with confidence. However, it is not necessary--and maybe counterproductive--to be simultaneously concerned with both. When the null hypothesis is that the pipeline does not require regulation (Case 1), the null hypothesis will be accepted with the least frequency when the pipeline is just below the regulatory limit; that is, the 90th percentile of PCB measurements will be just below the target. In this circumstance, an *unbiased* estimate of the 90th percentile based on measurements will be low about half the time and high about half of the time. The null hypothesis will be correctly accepted only about 50% of the time. This confidence level is too low. To increase it, the estimated 90th percentile must be adjusted (or "biased") to fall below the target more often. The resulting estimator is known as a *lower confidence limit* for the 90th percentile.

To compensate for this intended bias, a sufficiently large number of samples should be obtained to reduce the risk of failing to identify contaminated pipeline. Therefore, in Case 1, the EPA would be justified in specifying a minimum sample size requirement.

In Case 2, the null hypothesis is that the pipeline does require regulation. Using similar reasoning, the estimate of the 90th percentile must be biased high: it will be an *upper confidence limit* for the 90th percentile. Appendix 2 shows that this is the basis for the table of "K-factors" given in the Guidance. Using these K-factors assures the EPA that the risk of failing to identify contaminated pipeline has been made acceptably small. Therefore, the EPA should not regulate sample sizes when these K-factors are used.

Table 1 summarizes the distinctions between the two test-of-hypothesis approaches.

Table 1
Distinctions Between The Two Test-of-Hypothesis Approaches

CONCEPT	**CASE 1**	**CASE 2**
Null Hypothesis:	The 90th percentile lies BELOW the target of 100 g/100cm^2	The 90th percentile lies ABOVE the target of 100 g/100cm^2
Meaning of 90% Confidence	When the pipe truly should be UNregulated, there is at least a 90% chance that the decision will be to leave the pipe UNregulated.	When the pipe truly should be regulated, there is at least a 90% chance that the decision will be to regulate it.
Meaning of Power	The ability to identify cases where MORE than 10% of the pipe is contaminated.	The ability to identify cases where LESS than 10% of the pipe is contaminated.
Type I Error ("False Positive")	The decision to regulate pipe that truly is not contaminated.	The decision NOT to regulate pipe that is contaminated.
Type II Error ("False Negative")	The decision NOT to regulate pipe that is contaminated.	The decision to regulate pipe that truly is NOT contaminated.

To summarize, the EPA is over-regulating when it simultaneously specifies K-factors for an upper confidence limit and requires a minimum sample size. A simple solution would be to retain the K-factors while allowing companies to select sample sizes that most economically balance sampling costs against the costs of possible unnecessary regulated disposal.

2.3.2 How the sample mean can be more or less "representative" (page 8)

The Guidance confuses "representativeness" with "power." According to the EPA Guidance on Data Quality Objectives [EPA 1987],

> "Representativeness expresses the degree to which sample data accurately and precisely represent a characteristic of a population ... ***[It] is a qualitative parameter which is most concerned with the proper design of the sampling program.*** The representativeness criterion is best satisfied by making certain

> that sampling locations are selected properly and a sufficient number of samples are collected" [page 4-18; emphasis added].

The mean of two randomly selected measurements is just as representative of the population as the mean of twenty; it will be much less precise, however. Moreover, one tends to be uncomfortable basing decisions on a very small number of measurements because they afford no opportunity to check the assumptions (such as an exponential distribution) upon which the decision procedure is based. We suspect that on pages 8 and 9 of the Guidance where the term "representative" is used, the word "power" should be understood.

3.0 Discussion and Comments

3.1 The Guidance provides a good framework for regulating the disposal of stored pipeline

The salient elements of this framework include:

- A precise definition of the stored pipeline population that is characterized by a sample
- Use of a statistically valid randomized sampling procedure
- Detailed documentation of the sample selection
- Specification of statistical assumptions
- Adoption of a simple statistical decision procedure
- Regulation through performance, expressed as coverage, confidence, and power, rather than through procedure, expressed as required sample sizes

This framework is beneficial both to the EPA and to the parties it regulates. The requirements for precision and documentation are not onerous and will provide scientific (and possibly legal) defense of decisions made from the samples. Specifying the statistical assumptions allows a sampler to monitor the performance of the decision procedure. Regulating the decision process through performance criteria provides regulated parties flexibility in how they meet the EPA's requirements.

3.2 Some procedures specified by the Guidance could be harmful or costly to regulated companies

Some of the procedures that could be harmful or costly are these:

1. <u>Defining "specific populations" that are very small:</u> Defining many small specific populations according to factors that are unrelated to possible PCB contamination would unnecessarily inflate the sampling cost.
2. <u>Using wipe samples as a surrogate for measuring PCB content in pipe:</u> The wipe sample results may not truly indicate PCB content in pipe or residual liquids.
3. <u>Assuming that PCB concentrations within a population are exponentially distributed:</u> The failure of the exponential assumption could result in over-regulation or under-regulation of pipe.

4. Adopting a null hypothesis that all pipeline populations are contaminated: Assuming a null hypothesis that all pipeline populations are contaminated results in large values for the "K-factors." Assuming the alternate hypothesis, which is conventional in all other EPA statistical regulations, produces K-factors that can be less than half of those published in the Guidance, resulting in a much reduced risk of falsely identifying contamination.

5. Regulating all pipeline based on the highest 10% of PCB concentrations, rather than on an average concentration: Regulating the pipeline based on the highest 10% of PCB concentrations, while consistent with the result of TSCA, has no basis in considerations of risk or disposability. It could greatly increase the amount of pipe requiring regulated disposal.

6. Unnecessarily specifying minimum sample sizes despite applying statistical performance requirements: The Guidance's specification of minimum sample sizes has no basis in statistical theory and is implicitly contradictory. The sample sizes specified tend to be too large.

3.3 Alternative approaches to solving the pipe characterization problem

The following alternative approaches can ameliorate or eliminate the difficulties expressed in the preceding section.

1. Specific populations could be too small. A better understanding of the factors that truly influence potential PCB concentration would help resolve the issue of how finely to characterize specific populations. A regulated company could develop this understanding during a long-term sampling program by collecting and analyzing information about the factors and the PCB concentrations.

2. Wipe samples do not measure true PCB concentrations. The preferred alternative would be any procedure that obtains a small, representative portion of the pipe material and measures its PCB content. *This issue could be crucial*, because the errors introduced by using wipe samples as surrogates for true PCB content could be enormous.

3. PCB concentrations might not be exponentially distributed. The statistical underpinnings of the Guidance could easily be extended to include the family of Gamma distributions (Appendix 2), which would allow for better characterization of the important "upper tail" of concentrations. Other statistical distributions, such as the Normal, Lognormal, and Weibull, could be considered without changing the intent or function of the Guidance.

4. <u>The K-factor procedure cannot handle nondetects.</u> Laboratories have difficulty measuring PCB concentrations to infinitesimally small levels. Nondetects should be expected. The procedure required by the Guidance cannot handle any nondetects. Simple modifications, such as replacing nondetects by a fixed multiple of the detection limit, could appreciably change the operating characteristics of the procedure--its coverage, confidence, and power. The Guidance should allow alternative statistical procedures.

5. <u>The null hypothesis that all pipeline populations are contaminated leads to unnecessarily stringent procedures.</u> An alternative is to adopt the hypothesis that pipeline populations are not contaminated unless the measurements demonstrate that they are. The EPA would then have to regulate the minimum sample size or, equivalently, the power of the test. Regulating power is more complicated, but the EPA has demonstrated its capacity to do so within recently developed guidance documents concerning other statistical applications [EPA 1992]. (A better alternative, however, may be to retain the hypothesis of contamination but to allow companies to determine the number of samples to use.)

6. <u>The minimum sample sizes are too large</u>. As an alternative to adopting a different null hypothesis, the Guidance could instead relax its unnecessary minimum sample size requirements. Each regulated party could select the sample size it needs to reduce the frequency with which uncontaminated stored pipe is regulated for disposal, because using the published K-factors already provides the protection the EPA needs against failing to regulate contaminated pipe.

7. <u>Regulating on the basis of 10% of the PCB concentrations is unnecessarily stringent</u>. One alternative would be simultaneously to regulate the mean and the maximum concentrations. Richard Gilbert has proposed establishing a "hot measurement" (HM) to serve as a threshold [Gilbert & Simpson 1992]. Any valid individual measurement exceeding the HM value would automatically be considered evidence of contamination, subject to subsequent verification sampling. The HM value would be higher than the target mean value (of 500 ppm), but low enough to detect individual "hot spots" that would concern the EPA.

4.0 RECOMMENDATIONS

QUANTITATIVE DECISIONS recommends that regulated facilities consider the following actions.

1. Comment, in writing, to the developers of the draft Guidance prior to its finalization. Address the points discussed in Chapter 3 of this review.

2. If the Guidance is finalized without change,

    (a) Implement a program to observe the factors that could determine specific populations and to analyze the factor and PCB measurement data statistically, assessing which factors are important for defining specific populations.

    (b) Characterize the statistical distributions of all PCB measurements to verify or refute the assumption of exponential distribution.

    (c) Consider simultaneous pipe-material measurements of PCB concentrations, to assess the degree of correspondence with the wipe sample results.

3. If the Guidance is changed to provide the opportunity for regulated parties to determine the proper sample sizes, then perform a cost-benefit analysis that accounts for sampling costs and disposal costs (for both regulated and unregulated disposal methods) to determine the optimal sample size.

4. Perform a preliminary statistical analysis of existing stored-pipe measurement data, if any exists, to evaluate the potential disposal outcome according to the draft Guidance. The results could indicate the degree to which regulated facilities should be concerned about the details of the Guidance.

5. Follow the recommendations of section 2.1.2 regarding procedures for random sampling of pipes.

5.0 REFERENCES

EPA, *Verification of PCB Spill Cleanup By Sampling and Analysis*. Exposure Evaluation Division, Office of Toxic Substances, U.S. Environmental Protection Agency. August 1985.

EPA, *Data Quality Objectives for Remedial Response Activities: Development Process*. Office of Waste Programs Enforcement, Office of Solid Waste and Emergency Response, U.S. Environmental Protection Agency. EPA/540/G-87/003, March 1987.

EPA, *Guidance on Classification For Purposes of Disposal of Stored Natural Gas Pipe Which was not Part of a Pipe Removal Project Carried Out Under an EPA-Approved PCB Disposal Activity.* Chemical Regulation Branch, Exposure Evaluation Division, Office of Toxic Substances, Office of Pesticides and Toxic Substances, U.S. Environmental Protection Agency. Draft, February 1991.

EPA, *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities: Addendum to Interim Final Guidance*. Office of Solid Waste, Permits and State Programs Division, U.S. Environmental Protection Agency. June 1992.

EPA, *Guidance for Planning for Data Collection in Support of Environmental Decision Making Using the Data Quality Objectives Process*..(Interim Final). Quality Assurance Management Staff, U.S. Environmental Protection Agency. EPA QA/G-4, 1993.

Gilbert, R.O., & J.S. Simpson, *Statistical Methods for Evaluating the Attainment of Cleanup Standards, Volume 3: Reference-Based Standards for Soils and Solid Media.* PNL-7409 Vol. 3, Rev. 1, Pacific Northwest Laboratory, Richland, Washington.

Kiefer, J., *Introduction to Statistical Inference*. Springer-Verlag, New York, 1987.

Stuart, A., & J.K. Ord, *Kendall's Advanced Theory of Statistics*. Fifth Edition, in two volumes. Oxford University Press, New York, 1987.

6.0 APPENDICES

6.1 Appendix 1: The Draft Guidance

[A copy of the draft guidance appeared on pages 6-1 through 6-11, inclusive.]

6.2 Appendix 2: The Exponential Distribution

The subsections within this appendix address four technical objectives:

1. Clearly define the terminology
2. Sketch the theory behind the statistic "X BAR"
3. Derive the K-factors published in the Guidance
4. Compute power curves

6.2.1 Definition

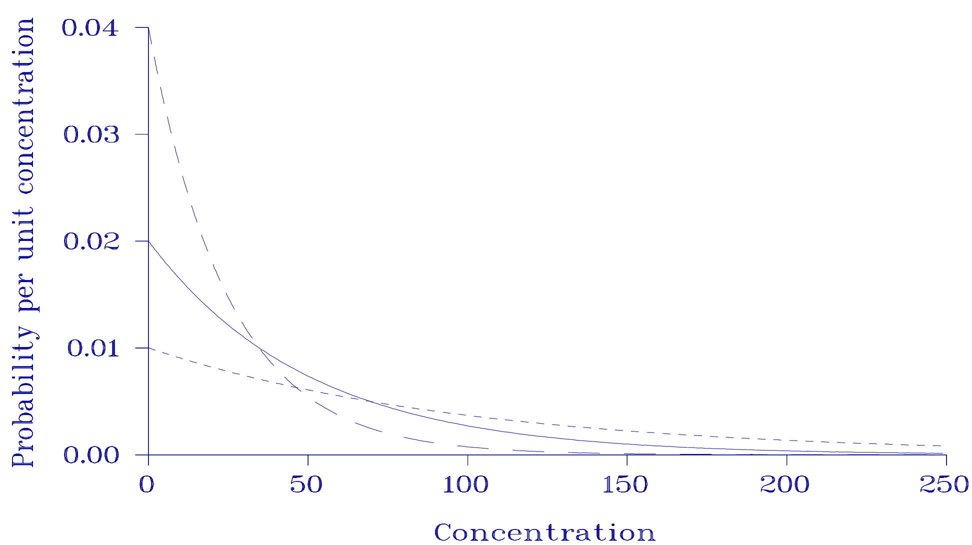
The "exponential distribution" is a one-parameter family of probability distributions, F_θ, where for any possible value x, $F_\theta(x)$ gives the probability of a measurement falling below x. The mathematical expression of F_θ is an exponential function:

[1] $F_\theta(x) = 1 - \mathbf{e}^{-x/\theta}$

The moment generating function of F_θ is $1/(1-u\theta)$, so that the kth moment (about zero) is $k!\theta^k$. Therefore the mean and standard deviation of F_θ are both equal to θ. The probability density function, f_θ, is the derivative of F_θ and therefore is equal to $\mathbf{e}^{-x/\theta}/\theta$.

Figure 2

Exponential Probability Density Functions



This figure shows probability density functions for three members of the exponential family of distributions. The probability density function is the derivative of the cumulative density function F_θ. The values of θ shown are 25 (long dashes), 50 (solid), and 100 (short dashes). The curves differ by a change of scale only; their heights are adjusted to ensure that the total probability remains exactly 1.

6.2.2 Sampling Distribution of "X BAR:" The Gamma Distributions

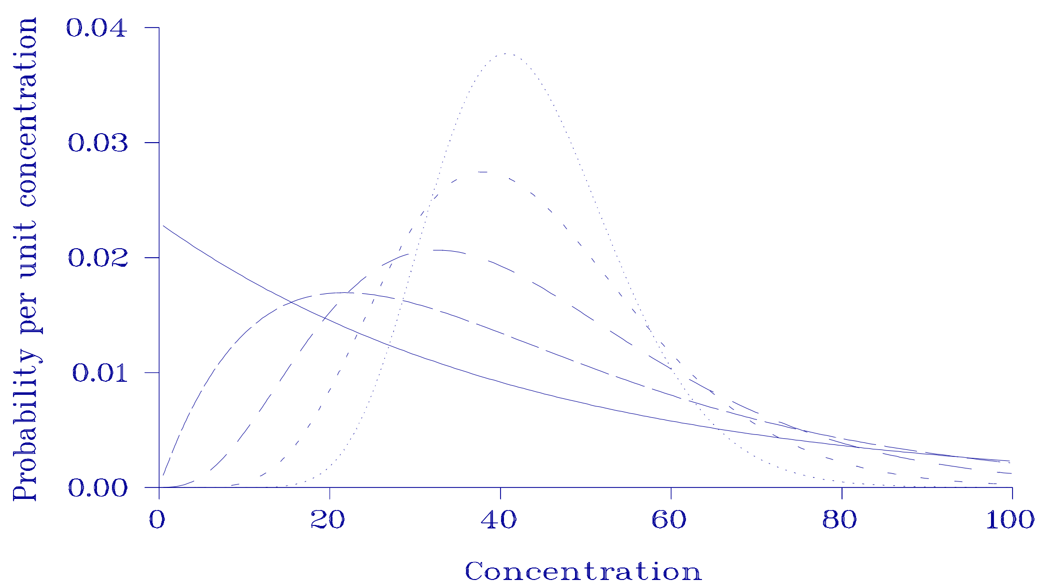
The statistic "X BAR" is the mean of a sample of N identically distributed exponential variables; or, equivalently, is 1/N times their sum. The sum of N exponential variates has a gamma distribution whose moment generating function is $1/(1-u\theta)^N$. We shall write $\Gamma_{\theta;N}$ for its cumulative distribution function and $G_{\theta;N}$ for its inverse, the "percentage point function:" for any percentage point p between 0 and 100%, $G_{\theta;N}(p)$ is the pth percentile of $\Gamma_{\theta;N}$. Note that $\Gamma_{\theta;1} = F_\theta$, so that $G_{\theta;1}$ is the percentage point function for the exponential distribution with parameter θ. A useful property of $\Gamma_{\theta;N}$ is that

[2] $$\Gamma_{\theta;N}(x) = \Gamma_{1;N}(x/\theta),$$

because θ is a scale parameter.

Figure 3

Gamma Probability Density Functions



This figure shows probability density functions for five members of the gamma family of distributions. The probability density function is the derivative of the cumulative density function $\Gamma_{\theta;N}$. The value of θ (the *scale parameter*) has been fixed at 100/ln(10) = 43.43 throughout; only the second parameter, N (the *shape parameter*), varies. Values of N=1,2,4,8, and 16 are shown. As N increases, the size of the dashes decreases, so that the solid curve corresponds to N=1 (an exponential distribution) and the dotted curve (having the highest peak) corresponds to N=16.

Each distribution has the same mean value of 43.43. The first one, corresponding to N=1, is the exponential distribution for which the 90th percentile equals 100. The other curves give the sampling distributions for the means of N exponential variates. Evidently, once N exceeds 4, the probability that the mean exceeds 100 becomes vanishingly small.

WHY X BAR IS A USEFUL STATISTIC

The likelihood function for N independent and identically distributed exponential variates x_1, ..., x_N is the product of their probability density functions:

$$[3] \quad L(x_1, ..., x_N) = \mathbf{e}^{-x_1/\theta}/\theta * ... * \mathbf{e}^{-x_N/\theta}/\theta = \mathbf{e}^{-(x_1+...+x_N)/\theta}/\theta^N$$

The likelihood function depends on the x_i only through their sum, $x_1 + ... + x_N$, which is a fixed multiple N of their mean, $\overline{X}$ ("X BAR" in the Guidance). Therefore the mean is a minimal sufficient statistic and it contains all the statistical "information" about the sample.

6.2.3 Computing the K-Factors

In Case 1 of Section 6.3.1, the null hypothesis to test is

$$[4] \quad H = \{\theta \mid \text{the 90th percentile of } F_\theta < 100\} = \{\theta \mid \theta < 100/\ln(10)\}.$$

Because is a minimal sufficient statistic, and θ is a scale parameter, the test should be based on a fixed multiple K of . The confidence will equal the false positive rate when θ is exactly equal to 100/ln(10). This rate is the probability that K equals or exceeds 100, written as

$$[5] \quad P\{K \geq 100\}.$$

This probability is given by the inverse gamma distribution, as follows:

$$[6] \quad P\{K \geq 100\} = P\{KN \geq 100N\} = P\{x_1+...+x_N \geq 100N/K\}$$

$$= 1-\Gamma_{\theta;N}(100N/K),$$

because the sum of N exponential variates, $x_1+...+x_N$, follows a gamma distribution. Setting this rate to 10%, as determined by the required confidence of 90%, we can find K explicitly:

$$[7] \quad 0.10 = 1-\Gamma_{\theta;N}(100N/K) = 1-\Gamma_{1;N}(100N/K\theta) = 1-\Gamma_{1;N}(N\ln(10)/K)$$

$$[8] \quad \Gamma_{1;N}(N\ln(10)/K) = 0.90$$

$$[9] \quad N\ln(10)/K = G_{1;N}(0.90)$$

$$[10] \quad K = N\ln(10)/G_{1;N}(0.90)$$

More generally, by retracing the previous steps using an arbitrary confidence level $1-\alpha$ and coverage γ, the value of K will be $N\ln(1/(1-\gamma))/G_{1;N}(1-\alpha)$.

In Case 2, the null hypothesis is different:

$$[11] \quad H = \{\theta \mid \text{the 90th percentile of } F_\theta \geq 100\} = \{\theta \mid \theta \geq 100/\ln(10)\}.$$

The same sequence of steps followed for Case 1 yields

[12] $K = N\ln(1/(1-\gamma))/G_{1;N}(\alpha)$

Table 2 provides selected values for the two cases, computed using the Microsoft Excel 4.0 spreadsheet's GAMMAINV() function. The values for Case 2 include those provided by the Guidance (page 8), with which they agree within ±1 unit in the third significant figure, demonstrating that the theory provided here correctly explains the procedure specified in the Guidance.

Table 2
K-Factors For One-Sided Upper 90% Confidence, 90% Coverage
Exponential-Distribution Tolerance Limits

N	Case 1	Case 2	Ratio
1	1.000	21.854	21.854
2	1.184	8.659	7.314
3	1.298	6.268	4.829
4	1.379	5.279	3.829
5	1.440	4.733	3.286
6	1.490	4.383	2.943
7	1.530	4.138	2.704
8	1.565	3.956	2.528
9	1.595	3.815	2.392
10	1.621	3.701	2.283
12	1.665	3.529	2.120
15	1.716	3.353	1.954
20	1.778	3.170	1.783
25	1.823	3.055	1.676
30	1.857	2.974	1.601
35	1.885	2.913	1.546
40	1.907	2.866	1.503
45	1.927	2.828	1.468
50	1.943	2.796	1.439
100	2.037	2.634	1.293
200	2.109	2.529	1.199
400	2.163	2.459	1.137
∞	2.303	2.303	1.000

The "Ratio" column expresses the ratio of the Case 2 K-factor to the Case 1 K-factor. This ratio exceeds 2 for sample sizes less than 15. In other words, concentrations have to be smaller by a factor of 2 to meet the Case 2 test than to meet the Case 1 test. This illustrates the effect of adopting different null hypotheses.

6.2.4 Assessing the Power

The power of a hypothesis test is the ability to correctly identify true departures from the null hypothesis. In Case 1, this will be the probability that K will exceed the target of 100 μg/100cm^2. It depends on the exponential parameter θ. One can compute this probability as follows:

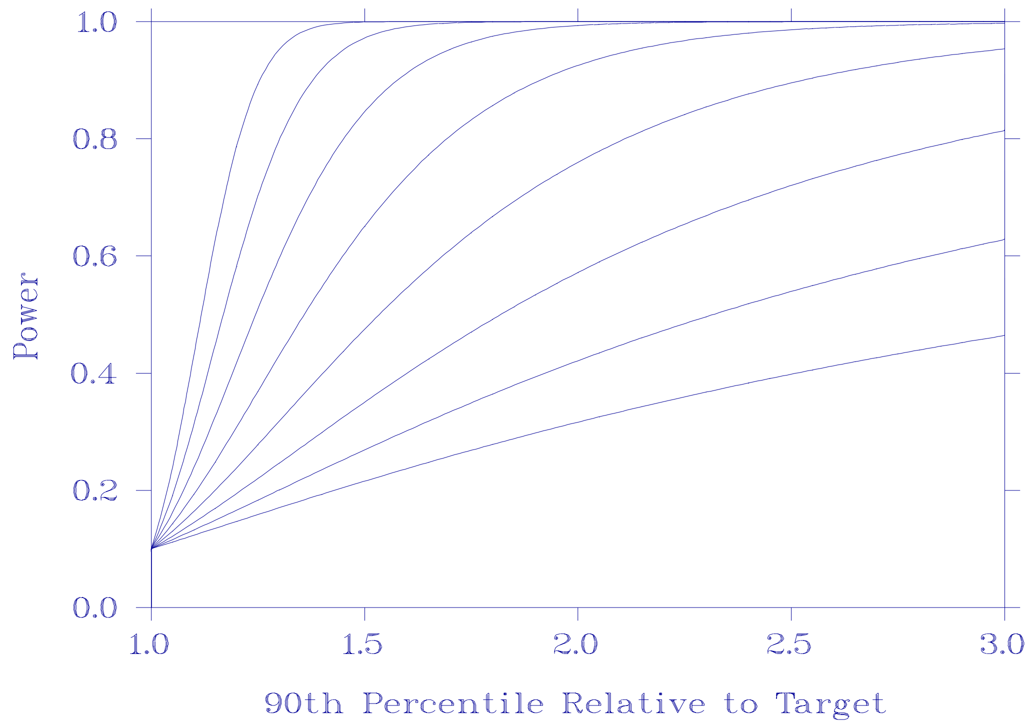
$$
\begin{aligned}
[13]\quad P\{K \geq 100\} &= P\{N \geq 100N/K\} = 1 - \Gamma_{\theta;N}(100N/K) = 1 - \Gamma_{\theta/100;N}(N/K) \\
&= 1 - \Gamma_{1;N}(100N/[\theta\ K]) \\
&= 1 - \Gamma_{1;N}(100N/[\theta\ N\ln(1/(1-\gamma))/G_{1;N}(1-\alpha)]) \\
&= 1 - \Gamma_{1;N}(\ G_{1;N}(1-\alpha) / [\ \ln(1/(1-\gamma))\ (\theta/100)\ ]\ ),
\end{aligned}
$$

where again 1-α is the confidence level and γ is the coverage of the tolerance limit. The parameter θ/100 is the value of the true mean concentration relative to the target level; the expression ln(1/(1-γ))(θ/100) is the true γ percentile concentration relative to the target level. Calling this value x, the power function becomes

$$
[14]\quad P\{K \geq 100\} = 1 - \Gamma_{1;N}(\ G_{1;N}(1-\alpha) / x\ )
$$

Evidently the power depends on the sample size N. Given N, the only free parameter is the relative value of the 90th percentile of concentrations. Figure 5 displays the power function for several typical sample sizes. Values on the horizontal axis correspond to ln(1/(1-γ))(θ /100).

Figure 4
Power Curves For Case 1



This figure shows portions of the power curves corresponding to sample sizes N=1,2,4,8,16,32,64, and 128, using the Case 1 (smaller) K-factors in Table 2. The power increases uniformly with increasing N. Thus, for example, the probability of correctly deciding to dispose of pipe whose 90th percentile of contamination is 3 times the target level (300 µg/100cm^2) on the basis of a single sample is approximately 45%, while it is virtually certain that using 128 samples will correctly identify any elevation of the 90th percentile above 1.5 times the target (150 µg/100cm^2).

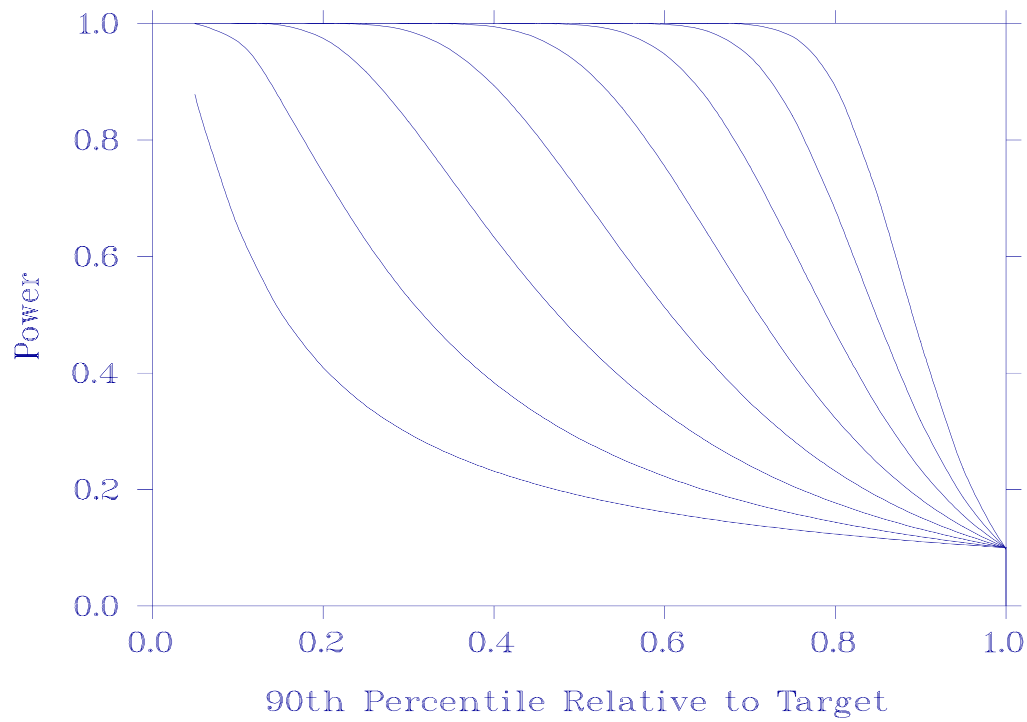
In Case 2 the power expresses the probability that Kµ will fall below the target of 100 g/100cm^2:

[15] $$P\{K < 100\} = P\{N < 100N/K\} = \Gamma_{\theta;N}(100N/K) = \Gamma_{\theta/100;N}(N/K)$$
$$= \Gamma_{1;N}(100N/[\theta\ K])$$
$$= \Gamma_{1;N}(100N/[\theta\ N\ln(1/(1-\gamma))/G_{1;N}(\alpha)])$$
$$= \Gamma_{1;N}(\ G_{1;N}(\alpha)\ /\ [\ \ln(1/(1-\gamma))\ (\theta/100)\ ]\ ).$$

Figure 5 displays the power function for several typical sample sizes.

Figure 5

Power Curves For Case 2



This figure shows the power curves corresponding to sample sizes N=1,2,4,8,16,32,64, and 128, using the Case 2 (larger) K-factors in Table 2 and the Guidance. The power increases uniformly with increasing N. Thus, for example, the probability of correctly deciding, on the basis of 16 samples, not to dispose of pipe whose 90th percentile of contamination is 50 percent of the target level (50 µg/100cm^2) is approximately 90%.