How Statisticians Speak Risk

K.S. Redus
Redus and Associates LLC
189 Lafayette Drive, Suite C, Oak Ridge, TN 37830 USA

ABSTRACT

The foundation of statistics deals with (a) how to measure and collect data and (b) how to identify models using estimates of statistical parameters derived from the data. Risk is a term used by the statistical community and those that employ statistics to express the results of a statistically based study. Statistical risk is represented as a probability that, for example, a statistical model is sufficient to describe a data set; but, risk is also interpreted as a measure of worth of one alternative when compared to another. The common thread of any risk-based problem is the combination of (a) the chance an event will occur, with (b) the value of the event. This paper presents an introduction to, and some examples of, statistical risk-based decision making from a quantitative, visual, and linguistic sense. This should help in understanding areas of radioactive waste management that can be suitably expressed using statistical risk and vice-versa.

INTRODUCTION

Like most professions, the lexicon of statistics is sometimes confusing to the layperson. Terms like *value*, *chance*, and *risk* are the basis for radioactive waste management decision making. This paper discusses how these terms relate to each other in the world of statistical risk. First, we provide a background to the terms value, chance, and risk to illustrate how risk has evolved to its current usage, especially in decision making. We then present some examples of risk as used in contemporary statistics that are directed to radioactive waste management applications. We conclude with future directions and promising areas of research in the continually developing frontier of risk and risk-based decision making.

BACKGROUND

There are two components that describe risk: value and chance. *Value* is a basic ethical, social, or economic construct present in all civilizations. *Chance* was not considered to have any scientific merit until the 17th century since life was managed by the Fates, God, or the Gods. The scientific approach to risk – namely the combination of value and chance – has blossomed since then. In the 21st century, we are discovering areas of risk that were not imagined even a decade ago.

Confucius (5th century BC) espoused a notion of value as a state of being formed by the triad of heaven, earth, and humans bound by *ch'i* – the whole of human relationships are mutually beneficial if individuals place values for what is best for all, including nature, before themselves (1). Aristotle (4th century BC) argued that value is a behavior and philosophy that has as its principal concern the nature of human well-being (2). The pre-Spanish conquest Aztecs pragmatically viewed value as a sacred energy which served as the glue to keep the balance of the cosmos and its human inhabitants (3).

The economic meaning of value is equally common. Prince Raimondo Montecuccoli, Prince of the Holy Roman Empire in the 15th century, captured value in his prescient statement "to wage war, you need first of all money; second, you need money, and third, you also need money" (4). Christopher Columbus sought venture capital in an agreement with Queen Isabel and King Ferdinand of Spain to offer them a valuable return.

It was not until the Enlightenment in the 17th and 18th centuries that value began to be perceived as something other than an economic, religious, natural, or cosmic phenomenon (5). Value became linked with reason, namely, reason as defined using a logical inductive or deductive calculus. The Enlightenment notion of value is firmly entrenched in today's society, as evidenced by the simple fact that virtually every social, political, or scientific discipline has uniquely defined value as some form of tangible or intangible asset that produces some measurable good.

Chance was also formalized during the Enlightenment. Even then it was considered more superstition than science. The historical concept of chance – whether playing the Persian game of *as-zahr* (a type of dice game) or the conversation between Morocco and Portia in *The Merchant of Venice* ("men that hazard all do it in hope of fair advantages") (6), or combinatorial enumerations of the Chinese game *go* in the 10th century – is that "chance" was not considered to even exist; chance was due to a divine will, end of discussion.

In the mid-17th century a French monk named Blaise Pascal raised a significant question:

> Let us examine this point of view and declare: 'Either God exists, or He does not.' To which view shall we incline? Reason cannot decide for us one way or the other: we are separated by an infinite gulf. At the extremity of this infinite distance a game is in progress, where either heads or tails may turn up. What will you wager? According to reason you cannot bet either way; according to reason you can defend neither proposition (6, 7).

Pascal's attempt to deal with this question led to collaboration with the mathematician, Pierre de Fermat, to develop the basis of modern probability theory. Their work culminated in the *Port-Royal Logic*, a treatise describing chance as a series of events occurring with some frequency (8). Coupled with the efforts of Thomas Bayes (an employee of the Guinness brewery), Christiaan Huygens, Thomas Bernoulli, Pierre-Simon Laplace, Andrei Kolmogorov, and many others, modern probability – or the logic of chance – is virtually ubiquitous in modern economics, politics, science, physics, and engineering.

The etymology and historical basis of *risk* is equally diverse. One of the more credible explanations of risk is based on the classical Greek ριζα, a metaphor for "difficulty to avoid at sea." In Vulgar Latin, the word meant "root, and later "cliff." The Italian, *riscare*, meaning "to dare" is considered the modern derivation of risk (9).

Risk is frequently part of the jargon of contemporary statistics. Many statistical results infer "risk" as synonymous with "probability." We are familiar with the terms "false negative" errors and "false positive" errors. Error has a specific meaning: it is the chance of either rejecting or accepting a hypothesis. When risk is used in classical statistics and when there is no value associated with the hypothesis, risk relates to a specific hypothesis only in terms of probabilities or chances of rejecting or not rejecting the hypothesis.

Statisticians realized the confusion regarding "error" in the early-1930s – namely that risk quantification depends on both the linguistic and the mathematical form of the hypotheses being examined. Jerzy Neyman and E. S. Pearson put it this way:

> Is it more serious to convict an innocent man or to acquit a guilty? That will depend on the consequences of the error; is the punishment death or fine; what is the danger to the community of released criminals; what are the current ethical views on punishment? From the point of view of mathematical theory all that we can do is to show how the risk of errors may be controlled and minimised. The use of these statistical tools in any given case, in determining just how the balance should be struck, must be left to the investigator (10).

Understanding the Neyman-Pearson issue, Abraham Wald formally linked the notions of value and chance by employing what is known as weight functions (now commonly referred to as loss functions) to risk. Wald stated:

> The question as to how the form of the weight function should be determined is not a mathematical or statistical one. The statistician who wants to test certain hypotheses must first determine the relative importance of all possible errors, which will depend on the special purposes of his investigation (11).

Statistical decision theory was the result of statisticians pursing Wald’s challenge. In 1954, L. J. Savage wrote *The Foundations of Statistics*, and in 1962, Howard Raiffa and Robert Schlaifer published *Applied Statistical Decision Theory*. These two texts combined “personalism" and “behavioralism" (decision-orientation) to deal with Wald’s question. Savage’s theory of probability was a “subjective” interpretation of Wald's minimax theory (a key element of Wald’s idea of a weight function). Raiffa and Schlaifer introduced the idea of a likelihood kernel that related what was known prior to the investigation to what can be concluded following the investigation (12)

Recent years have produced a wealth of advances in statistics and all other sciences dealing with risk. Risk is measured in terms of loss functions, as the maximum entropy using principles from information theory, using the constructs of possibility theory and a linguistic calculus (fuzzy sets), and incorporated in evidence theory and belief or plausibility functions (13, 14, 15, 16, 17). Current available statistical software uses all these risk measures to estimate statistical parameters and identify models (18).

Modern statistical techniques that describe and estimate risk are not without fault. These approaches are only tools that attempt to combine the concept of value and the notion of chance in an attempt to estimate and communicate risk. Research in the measurement, estimation, and interpretation of risk continues to provide approximations to perceptions of value and chance-defined realities.

STATISTICAL RISK-BASED DECISION MAKING

There is a wide spectrum in which risk is applied in statistical studies. In some cases, the problem is to assess the probability that a statistical model is sufficient to describe a data set. In other cases, risk is interpreted as a measure of worth of one alternative when compared to another. Statistical risk-based problems examine both (a) the chance an event will occur and (b) the value of the event. The bottom line is that any problem that uses statistical risk approaches has a simple objective: *help make decisions by understanding the values, the chances, and the risks of the problem*.

Statistical risk-based decision making deals with making decisions using information about the value, the probability, and the risk associated with various alternatives or hypotheses. Risk is defined as the probabilistic expectation of a consequence associated with an event E (Eq. 1).

$$\text{Risk of E} = \text{Probability of E} \times \text{Consequence caused by E} \qquad \text{(Eq. 1)}$$

Value is expressed in terms of the consequence of the event E, for example, in terms of amounts of money, fatalities prevented (incurred), radiation doses averted (received), or functional forms of correlated consequences. Probability is interpreted as the chance, or the likelihood, the event will occur or not occur. There are many variations of this formula, and they all basically imply the same concept – *risk quantifies a consequence in a probabilistic manner*. There are several interesting – but rather disturbing – attributes of the above “risk equation” which are summarized in Table I.

Table I. Attributes of the Risk Equation

Sources of Uncertainty	Consequence	Time Scale	Behavioral Issues
▪ Law of large numbers ▪ Future events ▪ Ignorance ▪ Noise or measurement error ▪ Instrument limits ▪ Models / parameters ▪ One-shot events / unpredictability	▪ Modest loss (product arrives too late to sell) ▪ Major setback (lose company) ▪ Catastrophic (lose a city)	▪ Real-time ▪ < 1 year ▪ > 3 years	▪ Scenario definition ▪ Ambiguity aversion ▪ Prospect theory ▪ Behavioral probability ▪ Security ▪ Nature of decision ▪ Perception ▪ Expectation and Surprise ▪ Complexity

- Uncertainty is usually tackled by dealing with random variables. Random variables are characteristics that are not constant but have some attribute of variability. If the event is not random, the risk is simply the consequence of the event.

- In order for the risk equation to make any sense, the event E must occur repetitively in a series of trials so that the relative frequency of E is likely to approach the probability of E. This is the law of large numbers. If this is not the case, then the estimates of the probability of E are suspect, biased, or wrong.

- Is the event E decomposed into independent or conditional events that have unique meanings? In chemical risk assessments, E consists of conditional events of scenario definition, risk identification, and exposure and hazard assessments. If the probability calculus applied to these events is incorrect and if the events have not been reasonably decomposed, slight changes in any one component could result in unexplainable or meaningless changes in the aggregation.

- The risk equation does not consider events that may occur as one-shot events or events that occur in a non-repetitive fashion. An attack on the World Trade Center occurred twice over a decade, the latter attack was considered a one-shot event that had catastrophic local and global consequences considerably greater than the former.

- The risk equation does not capture unique behavioral qualities (19). When a consequence is defined, it is defined by people with an understanding of the system in which risk is examined, but people often tend to do what is comfortable rather than what is important. When a consequence is defined, it may tend to ignore information that is inconsistent with *a priori* beliefs resulting in a consequence definition that is a distorted view of reality.

- The risk equation fails to quantify biases and heuristics employed when identifying consequences. Such biases include "anchoring" on information that is readily available, vivid or recent, making insufficient adjustment from initial anchor points, overestimating what is already known, and underestimating what is unknown.

- Consequences may be incorrectly defined based on flawed reasoning. Consequences associated with complicated and complex problems may be overly simplified; a consequence may be based upon circular reasoning, and when confronted with a sure loss, a consequence may be artificially inflated (19).

Approaches Used to Examine Statistical Risk

Space limitations prevent an exhaustive examination of the universe of statistical risk approaches, and we limit our discussion to several areas significant to the radioactive waste management community.

- *Hypothesis formation* (also known as hypothesis testing) is the basis of statistical risk examination. A hypothesis test results in a binary decision – either a statistical parameter or relationship "is" or it "is not" – and there is a chance the decision that is made is correct or incorrect. The risk of a hypothesis decision is exactly that of Eq. 1, namely risk is the product of the probability of the correctness of the decision and the value (or the consequence) of the decision. Two probabilities are used: the false positive rate and the false negative rate.
- *Visualization and interval estimation* express risk in terms of portraying and bounding a statistical parameter. This approach, for example, addresses the question "what is the Tc-99 concentration in a building planned for clean up such that 5% of the time, this concentration will be exceeded?" When the value (or the consequence) of the bounding estimate is unity, risk is the probability for the bound of the estimate of the statistical parameter.
- *Determining number of samples* for environmental characterization purposes combine hypothesis formation with interval estimation to address, for example, the question "how many samples are needed – and what is the total expected cost of sampling analysis – to confirm or reject the hypothesis that a waste form meets specific waste acceptance criteria?" Risk is expressed as the cost of sampling for the number of samples.
- *Simulation approaches* are useful in large-scale risk problems that deal with hundreds or thousand of probability functions and consequence relations. For example, the Oak Ridge Environmental Management Waste Management Facility (EMWMF) uses simulation for waste volume forecasting and capacity management risk decisions. Risk is expressed in terms of life cycle probabilities of exceeding volumetric and radiological capacity for this Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA) disposal cell.

Hypothesis Formation

A hypothesis formation is a statistical algorithm that compares one statistical statement to another. The base hypothesis is referred to as the "null" hypothesis which is compared to an alternative hypothesis. "Null" implies nullification of the hypothesis in favor of an alternative hypothesis. There is a compelling reason to place the burden of proof on an alternative hypothesis. A base hypothesis is never really "accepted." It is rejected in favor of the alternative hypothesis or one must conclude there is insufficient evidence to reject it in favor of the alternative hypothesis. Many times, the word "acceptance" is simply used for communication convenience (20).

The notion of a "false positive" and a "false negative" springs from the concept of hypothesis formation. Hypothesis tests always involve a trade-off between the acceptable level of false positives and the acceptable level of false negatives. Table II provides the basic definition of these terms and examples of consequences are presented in Table III.

- *False positive* – if we decide to "reject the base hypothesis in favor of the alternative hypothesis" when the base hypothesis is "true," this is labeled a false positive. This situation occurs when we are observing some difference when in reality there is none. Statisticians refer to false positives as "Type I" errors or α-errors.
- *False negative* – if we decide to "not reject the base hypothesis in favor of the alternative hypothesis" when the base hypothesis is "false," this is labeled a false negative. This situation occurs when we fail to observe some difference when in reality there is. Statisticians refer to false negatives as "Type II" errors or β-errors.

A simple way to remember these terms is that (1) a false positive error occurs by rejecting the base hypothesis when it is true and (2) a false negative error occurs by "accepting" the base hypothesis when it is false. A common example of a false positive is "finding an innocent person guilty" and of a false negative is "finding a guilty person innocent."

The notion of sensitivity and specificity is a useful tool to understand the impact of false positives and false negatives in risk-based decisions as illustrated in Tables II and III.

- *Sensitivity* is the probability of correctly predicting an existing condition. Numerically speaking, sensitivity = 1 – False Positive = 1 – α (since the probability of incorrectly predicting the existence of a condition is a false positive result).
- *Specificity* is the probability that a test correctly predicts that a condition does not exist. Numerically speaking, specificity = 1 – False Negative = 1 – β (since the probability of incorrectly predicting the absence of a condition is a false negative result), Specificity is often called the power of a hypothesis test.

Table II. Notion of False Positive, False Negative, Sensitivity, and Specificity

<table>
<tr><td colspan="2" rowspan="2">Generalized Hypothesis Formation and Decision</td><td colspan="2">Decision</td></tr>
<tr><td>Reject Base Hypothesis in favor of Alternative Hypothesis</td><td>Do not reject Base Hypothesis in favor of Alternative Hypothesis</td></tr>
<tr><td rowspan="2">True Condition</td><td><u>Base hypothesis</u>
The condition exists</td><td>False Positive (α)
1 – Sensitivity</td><td>Correct decision
Sensitivity</td></tr>
<tr><td><u>Alternative hypothesis</u>
The condition is does not exist</td><td>Correct decision
Specificity (Power)</td><td>False Negative (β)
1 – Specificity</td></tr>
</table>

Table III. Examples of Consequences based on False Positives and False Negatives

<table>
<tr><td colspan="2" rowspan="2">Waste Acceptance Criteria (WAC) at Disposal Facility</td><td colspan="2">Decision</td></tr>
<tr><td>Waste meets WAC and can be disposed</td><td>Waste does not meet WAC and cannot be disposed</td></tr>
<tr><td rowspan="2">True Condition</td><td><u>Base hypothesis</u>
Waste does not meet WAC and cannot be disposed</td><td><u>False Positive</u>
Allow waste that does not meet WAC to be disposed</td><td>Correct decision
Sensitivity</td></tr>
<tr><td><u>Alternative hypothesis</u>
Waste meets WAC and can be disposed</td><td>Correct decision
Specificity</td><td><u>False Negative</u>
Reject waste that meets WAC and do not allow waste disposal</td></tr>
</table>

<table>
<tr><td colspan="2" rowspan="2">Computer Security</td><td colspan="2">Decision</td></tr>
<tr><td>User is an imposter</td><td>User is an authorized user</td></tr>
<tr><td rowspan="2">True Condition</td><td><u>Base hypothesis</u>
User is an authorized user</td><td><u>False Positive</u>
Classify an authorized user as an imposter</td><td>Correct decision
Sensitivity</td></tr>
<tr><td><u>Alternative hypothesis</u>
User is an imposter</td><td>Correct decision
Specificity</td><td><u>False Negative</u>
Classify an imposter as an authorized user</td></tr>
</table>

Medical Screening		Decision	
		Infant has a congenital disorder	Infant is healthy and does not have a congenital disorder
True Condition	Base hypothesis Infant is healthy and does not have a congenital disorder	False Positive Classify a healthy infant as one with a congenital disorder	Correct decision Sensitivity
	Alternative hypothesis Infant has a congenital disorder	Correct decision Specificity	False Negative Classify an infant with a congenital disorder as healthy

Regression Model		Decision	
		Regression coefficients are zero	Regression coefficients are not zero
True Condition	Base hypothesis Regression coefficients are not zero	False Positive The regression coefficients do not contribute to the model	Correct decision Sensitivity
	Alternative hypothesis Regression coefficients are zero	Correct decision Specificity	False Negative The regression coefficients contribute to the model

Consider how false positives and false negatives impact a waste management situation. A regulatory and cost consequence matrix for this example is illustrated in Table IV with the following observations.

Table IV. WAC at Disposal Facility Regulatory and Cost Consequence Matrix

Waste Acceptance Criteria (WAC) at Disposal Facility		Decision	
		Waste meets WAC and can be disposed	Waste does not meet WAC and cannot be disposed
True Condition	Base hypothesis Waste does not meet WAC and cannot be disposed	False Positive *Disposal facility violates license agreement that results in significant regulatory impacts, potential closure for an indefinite time, mitigation costs > $50 million, and at least four years of ~ $1 million annual standby cost* > - $54 million	Sensitivity *Disposal facility does not violate license agreement and maintains $2 million annual operating cost* -$2.00 million
	Alternative hypothesis Waste meets WAC and can be disposed	Specificity *Disposal facility does not violate license agreement and maintains $2 million annual operating cost* *Disposal facility gains revenue of $50K from generator* - $1.95 million	False Negative *Disposal facility does not violate license agreement and maintains $2 million annual operating cost* *Disposal facility loses revenue of $50K from generator* -$2.05 million

- The problem is to identify the risk associated with waste meeting or not meeting Waste Acceptance Criteria (WAC) at a disposal facility. The consequence matrix illustrates (1) the importance of sensitivity – namely being able to correctly conclude the waste does not meet the WAC and cannot be disposed and (2) the importance of specificity – namely being able to correctly conclude the waste meets the WAC and can be disposed at the disposal facility.

- The risk question is how to minimize the false positives, or, equivalently, how to maximize the sensitivity of correctly predicting waste will not meet the WAC. Using the notation of Table II and the form of Eq. 1, the risk associated with a false positive decision is at least -$54 million x α million. If α is 5%, the expected cost and regulatory risk is -$54 million x 0.05 = -$2.7 million. This risk is greater than each of the consequences in the consequence matrix. Knowing which waste will not meet the WAC for disposal is exactly the reason why disposal facilities have WAC and stringent WAC approval processes.

Visualization and Interval Estimation

One of the first tasks in a statistical risk-based analysis is to address the question *what does the data say*? To accomplish this, visualization of the data and the estimation and calculation of intervals that bound true values of the statistical parameters are performed.

Two techniques used for data visualization are a box plot and an empirical cumulative distribution (ECDF) plot. A box plot portrays the data in terms of the 25th and the 75th quantiles, the interquartile range, and the median sample value. Box plots can rapidly identify potential data outliers. An ECDF plot examines the cumulative distribution function of the data, and a box plot summarizes the distribution of the data points. Both are accomplished for one or more groups of data as required. An ECDF plot orders the data and calculates the cumulative frequency of occurrence of the data points. (18)

For example, suppose we have Tc-99 concentration (pCi/g) measurements for three buildings denoted as N-334, N-335, and N-336. Visualizing the data using a box plot and an ECDF plot, Figure 1, the statistician would say:

- The box plot illustrates the spread of the Tc-99 data for each of the three buildings. It appears the Tc-99 concentrations in building N-335 are greater than that of buildings N-334 and N-336. The building N-334 box hardly overlaps the building N-335 box, so there are probably different concentrations. The building N-335 and the building N-336 boxes do seem to overlap and they may not be that different at all. The "high" outlier for building N-336 may be a measurement risk and should be examined.

- The ECDF plot illustrates similar information as the box plot. The 0.90 CumProb level indicates the 90th percentile value of Tc-99 for the buildings. The 90th percentile value of Tc-99 concentration in building N-334 is very different than buildings N-335 or N-336. The 90th percentile value of Tc-99 concentration in building N-335 seems close to that of building N-336, but the slope of the two ECDFs looks very different. The variation of Tc-99 concentrations may be different from building N-335 to building N-336.

Three forms of interval estimation can be employed to supplement the visualization.

- A confidence interval gives an estimated range of values for a statistical parameter (the mean, variance, proportion, etc.) or function (a regression model or a correlation model). This range has some probability of including the true value of the parameter or function, and it is calculated from the sample data.

- A tolerance interval is an interval with an upper and a lower bound that one can claim contains at least a specified proportion of a data set with a specified degree of confidence

- A credible interval is also a probability interval about a statistical parameter that supplements prior knowledge of the bounds for the statistical parameter with additional data.

Of course, there are several key assumptions that must be validated before interval estimation is performed. These assumptions include determining the underlying probability function of the data and the appropriate method to calculate the quantity to add to (or subtract from) the statistical parameter in order to determine the lower or upper bound of the parameter.

Using Eq.1 and assuming there is no consequence value for the event "intervals of the true Tc-99 mean concentration for Building N-334," risk is expressed as the probabilities of the event:

- Confidence interval: there is a 90% probability the true mean Tc-99 concentration is between 4 pCi/g and 7 pCi/g when the mean Tc-99 data for Building N-334 follows a normal distribution.

- Tolerance interval: there is a 90% probability that 90% of the Tc-99 concentration values for building N-334 range between 0 pCi/g and 10.9 pCi/g.

- Credible interval: if prior information indicates the mean Tc-99 concentration for building N-334 is uniformly distributed between 5 pCi/g and 25 pCi/g, and if sampling results indicate the mean concentration is normally distributed with a mean of 5 pCi/g and standard deviation of 1 pCi/g, then there is a 90% probability credible interval for the mean Tc-99 in building N-334 is between 2 and 18 pCi/g.

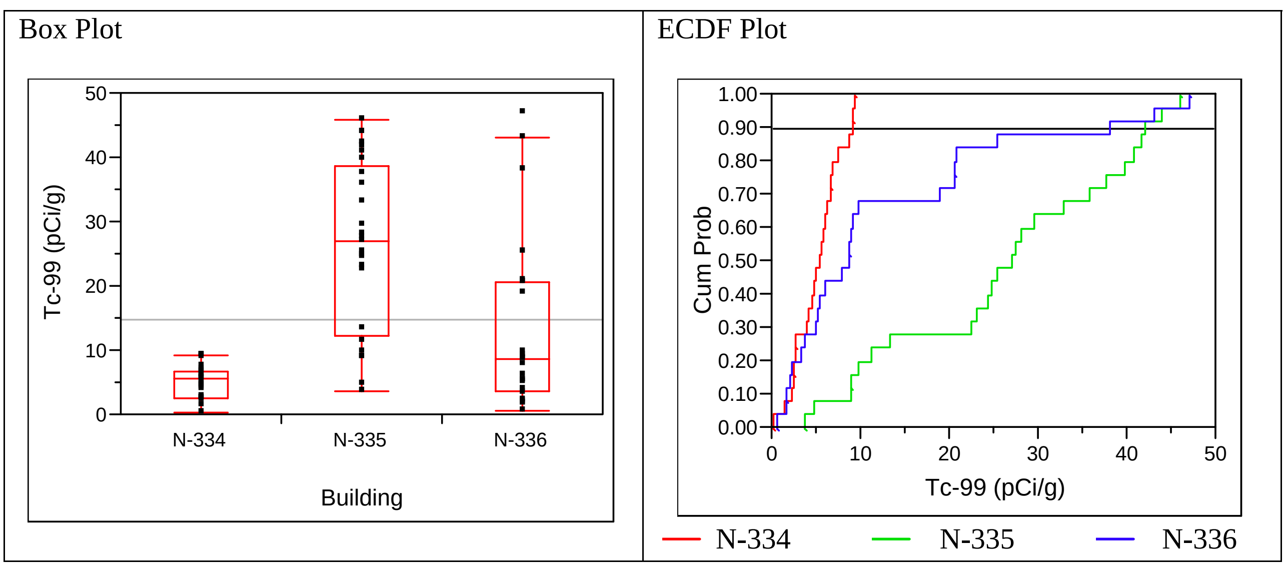

Fig 1 Box Plot and ECDF Plot used to compare Tc-99 (pCi/g) concentration for three buildings

How can we use Eq. 1 and the notion of a consequence matrix to examine risk associated with schedule and cost risk to remediate Tc-99 in the three buildings? Define two events, Schedule Duration (measured in months required for remediation) and Cost (measured in dollars ($) per month required for remediation). Table V elaborates these event definitions for a Base Case and two sensitivities using a relationship between the schedule duration or the cost and the upper 90% confidence bound for the mean Tc-99 concentration (denoted as X in the table below) as measured in pCi/g

Table V. Event and Consequence Definition for Schedule Duration and Cost Events for Decontamination of Three Buildings

	Schedule Duration (months)	Cost ($/month)
Base Case	6 if X $\leq$ 20 pCi/g 9 if 20 pCi/g < X $\leq$ 35 pCi/g 12 if X > 35 pCi/g	$10,000
Sensitivity #1	6 if X $\leq$ 20 pCi/g 9 if 20 pCi/g < X $\leq$ 35 pCi/g 12 if X > 35 pCi/g	$10,000
Sensitivity #2	6 if X$\leq$ 20 pCi/g 9 if 20 pCi/g < X $\leq$ 50 pCi/g 12 if X > 50 pCi/g	$10,000 if X $\leq$ 40 pCi/g $20,000 if X > 40 pCi/g

Risk results are presented in Table VI for the Base Case, Sensitivity #1, and Sensitivity #1 when the X is 6 pCi/g for Building N-334, 36 pCi/g for Building N-335, and 45 pCi/g for Building N-336. Use Eq. 1 to determine the expected schedule risk and the expected cost risk. For example, in the Base Case, X is 36 pCi/g for Building N-336; and, using Table V, the schedule duration is 12 months. Since there is a 10% chance of X being exceeded, the expected schedule risk is 0.10 x 12 months = 1.2 months. At $10,000 per month cost, the expected cost risk is 1.2 x $10,000 = $12,000. The largest schedule and cost risk is for Buildings N-335 and N-336. The ordering of risk between the three buildings remains essentially the same.

Table VI. Risk Results for Decontamination of Three Buildings

Case	Building	Decontamination Schedule (months)	Expected Schedule Risk	Expected Cost Risk
Base Case	N-334	6	0.6	$6,000
	N-335	12	1.2	$12,000
	N-336	12	1.2	$12,000
Sensitivity #1	N-334	6	0.6	$6,000
	N-335	9	0.9	$12,000
	N-336	9	0.9	$12,000
Sensitivity #2	N-334	6	0.6	$6,000
	N-335	9	0.9	$9,000
	N-336	9	0.9	$18,000

Determining the Number of Samples

Statisticians refer to risk when addressing the question “how many samples are needed?” The answer requires that (1) a base and an alternative hypothesis must be stated, (2) α and β must be defined, (3) the relative standard deviation or the variability of the statistical parameter in question must be quantified, and (4) the underlying probability function defining the statistical parameter must be known.

Figure 2 illustrates the situation of how many samples are needed to measure Tc-99 concentrations and confirm whether we meet – or do not meet – a disposal facility WAC threshold of 150 pCi/g. We are in the situation in which (1) the true mean Tc-99 concentration is unknown which requires the hypotheses must be defined in terms of the WAC threshold, (2) the false positive and false negative rates are defined per stakeholder and regulatory requirements, (3) the true relative standard deviation (RSD, which is the ratio of the standard deviation to the mean) is unknown and must be treated as a hypothesis sensitivity, and (4) while the underlying probability density function describing the mean Tc-99 is unknown it is assumed to be highly skewed to the right (for example, a lognormal distribution).

Define several hypotheses for various RSD, as stated below, under a fixed false positive value of 0.05, a sensitivity of 0.95, and unknown false negative rate. Let us assume the probability of occurrence of the hypotheses are Probability (H_1 is true) = 0.40, Probability (H_2 is true) = 0.35, and Probability (H_3 is true) = 0.25. The sum of the probabilities must equal one.

- H_1: True Tc-99 Mean = 90% of threshold = 0.90 x 150 pCi/g = 135 pCi/g
- H_2: True Tc-99 Mean = 80% of threshold = 0.80 x 150 pCi/g = 120 pCi/g
- H_3: True Tc-99 Mean = 50% of threshold = 0.50 x 150 pCi/g = 75 pCi/g

Figure 2 illustrates the relation of the required number of samples to these hypotheses. As can be seen, the closer the true Tc-99 mean is to the threshold and the larger the variability (as expressed by the RSD) about the true mean, the more samples that are required.

Let the unit cost for each sample be $1500. Apply Eq. 1 to determine the risk for each hypothesis for a specific RSD value. For example, 80 samples are required for H_2 when the RSD is 125% of the threshold. The probability of H_2 is 0.35 and the sampling cost is 80 x $1500 = $120,000. From Eq. 1, the risk is 0.35 x $120,000 = $42,000. See Table VII for a summary of results. Suppose the risk threshold is $100,000; then the feasible risk-based sampling decisions are those situations with risk less than $100,000 as italicized in Table VII.

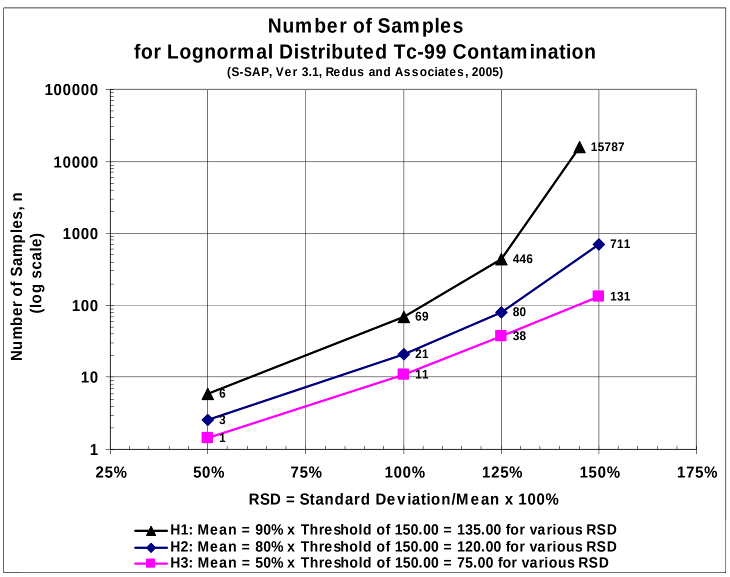

Fig 2. Number of samples for H_1, H_2, and H_3 (21).

Table VII. Risk Matrix for Number of Samples under H_1, H_2, and H_3

RSD	H_1 Risk	H_2 Risk	H_3 Risk
50%	*$3,534*	*$1,366*	*$530*
100%	*$41,469*	*$10,815*	*$4,102*
125%	$267,586	*$41,975*	*$14,207*
145%	$9,471,993	$373,362	*$49,151*

Simulation Approaches

Simulation approaches combine many probability functions and many consequence relations to efficiently deal with large-scale risk problems. Often there are only several primary risk measures, but there may be thousands of probability functions and consequence relations that are inputs to the risk equation. The theory is easily summarized but the implementation is rigorous. The conditional relationship of events is defined using a decision analysis technique known as influence diagrams that portray the logical sequence of events. Probability distributions for each event are needed to include credible estimates of parameters. Consequences and associated uncertainties are incorporated into the influence diagrams, and the consequences are defined in a probabilistic manner (22).

A real world problem that uses a simulation approach for risk-based decision making is disposition of waste at the EMWMF, a CERCLA landfill that can receive low-level radioactive waste in the form of contaminated soils and building demolition waste, located in Oak Ridge, TN (23, 24). The risk problem is two-fold.

- The first question is *what is the expected volume in the EMWMF from the start of operations in 2002 until (a) the end the current Accelerated Cleanup Program (ACP) in 2008 and (b) the end of the EMWMF life-cycle in 2015?*
- The second question is *how much radiological risk capacity is used by the end of the ACP?*

The magnitude of the problem is rather significant.

- The design basis volumetric capacity of the EMWMF is 1.7 million yd^3 and the volumetric risk measure is the volume value with a 5% chance it will be exceeded. The radiological risk measure is a volume weighted sum of fractions (VWSF) such that if the VWSF is less than one, radiological capacity is sufficient. If radiological capacity is exceeded, the CERCLA Record of Decision – and the requisite human health and environmental risk constraints – are violated.
- The sensitivity requirement (re: Table IV) is a 95% probability that waste not approved for disposal indeed does not meet the EMWMF WAC. The specificity requirement is an 80% probability that waste approved for disposal does meet the EMWMF WAC.
- From the third quarter of FY2002 (when the EMWMF opened) through the first quarter of FY2007, over 100 generators have disposed 46,000 shipments comprising 400,000 yd^3 of waste and clean fill at the EMWMF.
- Between the second quarter of 2007 and the end of the ACP, approximately 1.2 million yd^3 of soil and debris waste is planned for disposal.
- The WAC for the EMMWF consists of over 100 radionuclides, metals, and volatile and semi-volatile organic chemicals.
- Generators deal with uncertain budgets during any quarter of any fiscal year which, in turn, affects the amount of waste disposed at the EMWMF and the schedule of disposition.
- The ACP scope requires demolition of numerous gaseous diffusion process buildings and remediation of waste storage sites. The sufficiency of characterization, the waste volumes, and the types of waste generated over the course of a specific project exhibit high variability.
- Best management landfill practices at the EMWMF use waste compaction that effectively reduces the volume disposed by a variable amount. Because the waste is heterogeneous, the compaction efficiency and supplemental clean fill requirements are highly variable.

Waste Acceptance Criteria Forecasting Analysis System (WACFACS) is the tool used to address both questions of the problem (25, 26). WACFACS is a validated decision support system that has been used since the opening of the EMWMF in FY2002. WACFACS employs Monte Carlo simulation techniques and explicitly addresses operational uncertainties and variabilities, namely uncertainties in terms of (1) generator schedule and scope, (2) variability in waste volumes and (3) variability in waste stream characterization results obtained from Data Quality Assessment (DQA).

In order to address volumetric capacity, WACFACS uses a module that propagates the uncertainties for the following volumetric parameters:

- *In-place and as-generated volume estimates* are the uncertainties in the volumes as estimated by the generator. Such uncertainties are captured by the Confidence in Volume Value [a] (CIVV) as provided by the generator. The distributions of in-place and as-generated volumes are modeled using a beta distribution that bounds the point estimate of the volumes by the minimum and the maximum in-place volume and uses the point estimate as the most likely value for the in-place or as-generated volume.
- *Uncertainty as-generated measured tonnage* is the measurement uncertainty in the as-generated measured tonnage as measured by the EMWMF Operations Subcontractor. The distribution of as-generated measured tonnage is modeled using a uniform distribution that bounds the point estimate of in-place volume by the minimum and the maximum as-generated measured tonnage.
- *Uncertainty in density estimates* are the uncertainties associated with the various waste streams (soil, concrete with rebar, building and construction debris, sediment, etc.) that constitute the waste lot disposed at the EMWMF by a project. The distribution of density – for each waste stream – is modeled using a beta distribution bounded by the minimum and the maximum density and uses the point estimate as the most likely value for the density.
- *Uncertainty in soil volumes required for fill* are the uncertainties associated with the amount of fill material required to mitigate void spaces associated with disposed waste (especially debris). Fill material is needed to surround wastes within the cell to fill voids between waste lots much like mortar is used to piece together bricks. It is also used to replace air within debris wastes that otherwise are not removed by compaction or careful placement. The distribution of soil volume required for fill is modeled for each waste stream using a beta distribution bounded by the minimum and the maximum soil volume required for fill and uses the point estimate as the most likely value for the soil volume required for fill.

As illustrated in Figure 3, WACFACS capitalizes on the variability and the variabilities and uncertainties present in the waste constituents and volume data to examine the radiological capacity. These uncertainties are propagated to compute a Sum of Fractions (SOF), the VWSF, and the 90%-Upper Confidence Limit (UCL-90) for the VWSF.

Given the input conditions and the modeling assumptions, the WACFACS model identifies the upper bound risk value for the volumetric capacity as illustrated in Fig. 4. The current EMWMF is authorized to contain 1.7M yd^3, and the results substantiate that condition. Examining the probability function of total capacity, we discern there is a 5% risk that the total volume placed in the EMWMF at the end of FY2008 exceeds 1.2 million yd^3 (70% capacity) and 1.5 million yd^3 by the end of 2015. Sensitivities identify two post ACP projects (White Wing Scrap Yard Soil and Bethel Valley Soils) substantially contribute to the capacity followed by three ACP projects (East Tennessee Technology Park (ETTP) Scrap Removal, K-29/K-31/K-33 D&D Debris and Concrete, and Main Plant Area Facilities D&D Debris and Concrete).

WACFACS also determines the radiological risk capacity by the end the ACP. The expected VWSF is 0.7 and the UCL-90 of the VWSF is 0.8. The UCL-90 implies there is a 10% risk that more than 80% of the radiological capacity is consumed by the end of the ACP in FY2008. Sensitivities reveal that the K-25 D&D Project Metal and Concrete and the Zone 1 Remedial Action Soils are the two key drivers affecting the 80% radiological capacity.

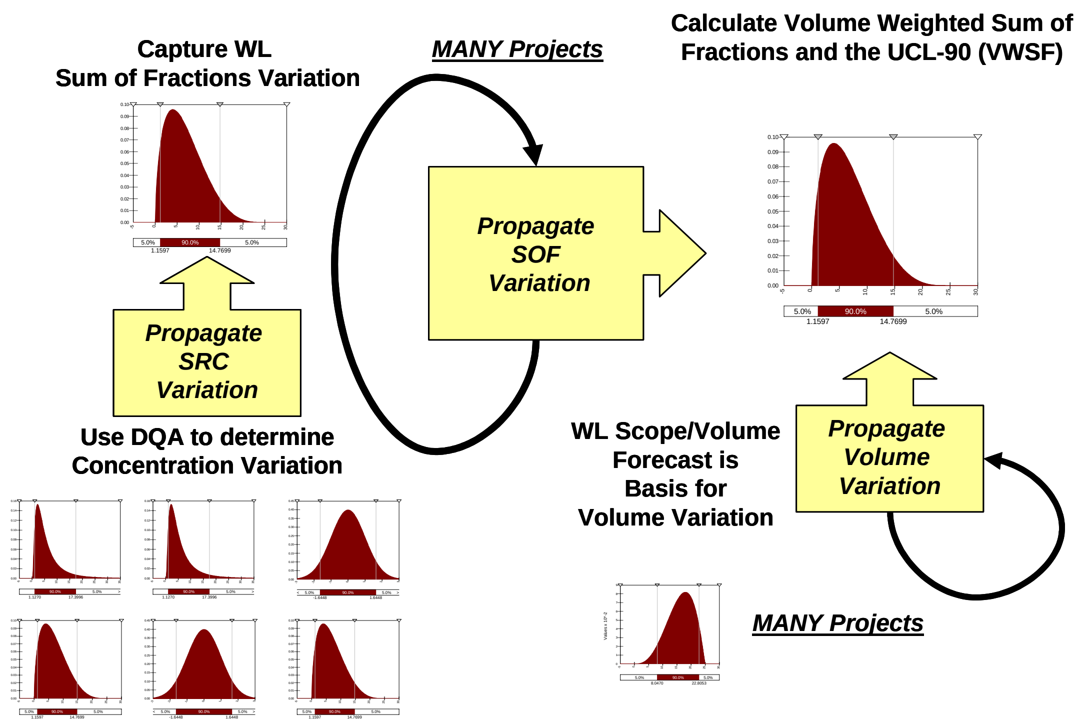

Fig 3. WACFACS process (25).

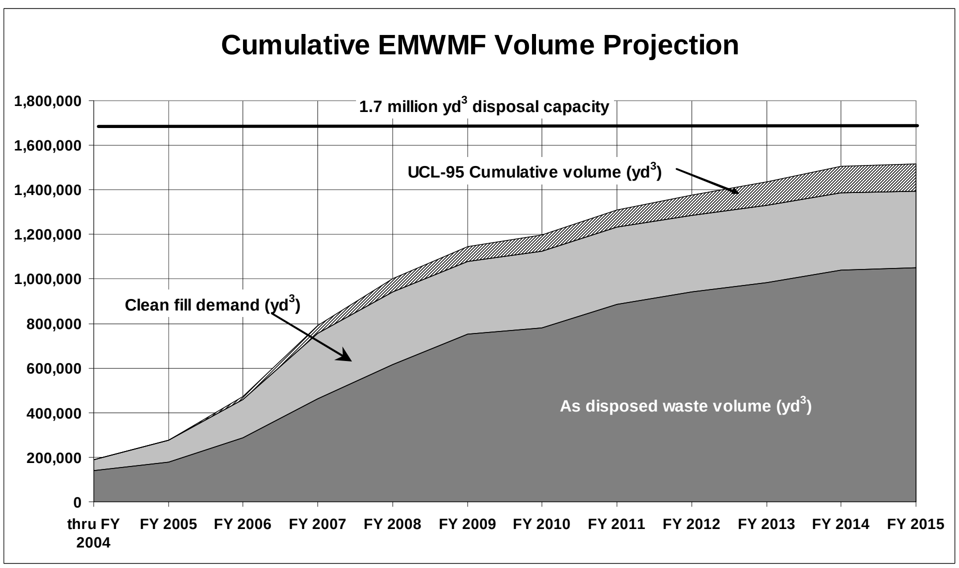

Fig 4. Cumulative EWMWF utilization (24).

CONCLUSION

While it may seem that the way statisticians speak risk is confusing, it is based on the concept that risk expresses expected consequence. Regardless of the statistical or numerical technique used to calculate risk, the scientific principles of having a clearly defined problem statement, understanding the constraints, and using validated data remain in place. Future directions and promising areas of research in the continually developing frontier of risk and risk-based decision making are exciting and include such topics as integration of evidence theory with stakeholder biases, nano-economics that examine neurophysiological changes in the brain and body chemistry when individuals are confronted with risky decisions, and efforts to better understand "one-shot, high consequence" events in national security decision making. While tools and techniques may become more sophisticated, risk still remains a difficulty to avoid in the sea of radioactive waste management decisions.

REFERENCES

1. M.E. Tucker, *Introduction to Confucianism*, *Confucianism and Ecology: Potential and Limits*, The Forum on Religion and Ecology. http://environment.harvard.edu/religion/religion/confucianism/, (acc. December 2006).

2. R. Kraut., *Aristotle's Ethics*, Stanford Encyclopedia of Philosophy, Metaphysics Research Lab, CSLI, Stanford University. http://plato.stanford.edu/entries/aristotle-ethics/, (acc. December 2006).

3. J. Maffie., *Aztec Philosophy*, The Internet Encyclopedia of Philosophy. http://www.iep.utm.edu/a/aztec.htm, (acc. December 2006).

4. S. Wolfram, A New Kind of Science, Wolfram Media, page 1082 (2002).

5. P. Brians, *The Enlightenment*, in Paul Brians' Study Guides. http://www.wsu.edu/~brians/hum_303/enlightenment.html, (acc. December 2006).

6. J.M. Ross, *Pascal's Legacy*, European Molecular Biology Organization Reports, 5(Suppl 1), (2004).

7. *Pascal's Wager*, Stanford Encyclopedia of Philosophy, as discussed in *Pascal's Pensées*, translated by W. F. Trotter (1910). http://www.stat.ucla.edu/history/, (acc. December 2006).

8. A. Hájek And C. Hoefer, *Chance*, Encyclopedia of Philosophy, 2nd edition, ed. Donald Borchert, Detroit: Macmillan Reference (2006).

9. R. Skjong, Etymology of Risk, http://www.dnv.com/research/transport/skj.asp, (acc. December 2006).

10. J. Neyman and E.S Pearson, 1933. *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, Philosophical Transactions of the Royal Society A, 231: 289-337 (1933).

11. A. Wald, *Contributions to the Theory of Statistical Estimation and Testing Hypotheses*, Annals of Mathematical Statistics, 10, (1939).

12. D. Mccloskey, *Econowannabes*, in University of Iowa Economic Sociology Editorial Series, http://www.gsm.uci.edu/econsoc/essays.html (Acc. December 2006)

13. M.G. Kendall and A. Stuart, The Advanced Theory of Statistics, Volume 2, Third Edition, Hafner Publishing Company, New York, (1973).

14. C.E. Shannon, *The Mathematical Theory of Communication*, The Bell System Technical Journal, 27(3, 4) (1948).

15. L.A. Zadeh, *Fuzzy Sets*. Information and Control, 8(3):338–353, 1965.

16. A. P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, The Annals of Mathematical Statistics, 38(2) (1967).

17. G. Shafer, A Mathematical Theory of Evidence. Princeton University Press, Princeton, NJ (1976).

18. Sas, http://sas.com/software/, (acc. December 2006).

19. Merkofer, L., http://www.prioritysystem.com/papers.html, (acc. December 2006).

20. R. R. Sokal and F. J. Rohlf, Biometry, 3rd ed. Freeman and Company, New York, (2000).

21. Redus and Associates, S-SAP, Version 3.1, Statistical Sampling Software, (2005).

22. K. S. Redus, *A Bayesian Approach for Data Fusion in Waste Characterization*, Proceedings of the Fifth Biennial International Meeting on Nuclear and Hazardous Waste Management, Spectrum '94, Atlanta, GA, (1994).

23. U.S. Department of Energy, Attainment Plan for Risk/Toxicity-Based Waste Acceptance Criteria at the Oak Ridge Reservation, Oak Ridge, Tennessee, DOE/OR/01-1909&D3, Office of Environmental Management, (2001).

24. U.S. Department of Energy, Environmental Management Waste Management Facility (EMWMF) 2006 Capacity Assurance Remedial Action Report, DOE/OR/01-2303&D1, Office of Environmental Management, (2006).

25. K. S. Redus, G. J. Hampshire, J. E. Patterson, and A. B. Perkins, *How to Deal with Waste Acceptance Uncertainty Using the Waste Acceptance Criteria Forecasting and Analysis Capability System (WACFACS)*, Proceedings of WM'02, (2002).

26. K. S. Redus, G. J. Hampshire, J. E. Patterson, and A. B. Perkins, *Waste Acceptance Decisions and Uncertainty Analysis at the Oak Ridge Environmental Management Waste Management Facility*," Proceedings of WM'03, (2003).