

RISK MANAGEMENT GUIDE



EG&G Idaho, Inc.

P.O. Box 1625
Idaho Falls, Idaho 83415

September 1982

UNITED STATES
DEPARTMENT OF ENERGY
Office of the Deputy Assistant
Secretary for Environment
Safety and Health

This document contains new concepts or the author(s) interpretation of new calculations and/or measurements; accordingly, EG&G Idaho, Inc. is required by the United States Government to include the following disclaimer:

DISCLAIMER

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

Available from:

System Safety Development Center
EG&G Idaho, Inc.
P. O. Box 1625
Idaho Falls, Idaho 83401

RISK MANAGEMENT GUIDE

Prepared by

Glen J. Briscoe

System Safety Development Center
EG&G Idaho, Inc.
Idaho Falls, Idaho 83415

September 1982

FOREWORD

This Guide presents an expansion of the risk discussion in the Management Oversight Risk Tree Analysis Manual.^{1,2} It was prepared as a textbook for use in Risk Analysis Workshops for Department of Energy personnel and for safety staffs of Department of Energy contractors.

The discussion includes the risk analysis of operational accidents and the role of risk analysis in line management and safety functions. Elementary probability, statistics, and risk theory are given. Practical applications for safety professionals and line managers are also given.

Line managers will be able to determine the necessary elements for a comprehensive risk management or loss control program. In addition, safety professionals will be able to apply basic risk evaluation techniques to new or existing systems, ranging from a single operation or process to an entire project or company.

Engineering analysis techniques (such as fault tree analysis or consequence analysis) and the processes of integrating risk with other organizational factors leading to managerial decisions are outside the scope of this Guide.

CONTENTS

FOREWORD	ii
GLOSSARY	v
1. INTRODUCTION	1
2. SUMMARY	3
3. BACKGROUND	6
3.1 Understanding Risk	10
3.2 Risk Perception	15
4. RISK MANAGEMENT	19
5. REPORT TO MANAGEMENT	30
6. ANALYTICAL METHODS FOR RISK QUANTIFICATION	33
6.1 Actuarial Risk Assessment	33
6.2 Example Problem	34
6.3 Subjective Risk Estimate	36
6.4 Survey Methods	37
6.5 Insurance Risk	39
6.6 Life Shortening Effects	39
6.7 Trend Analysis	41
6.8 Log-Normal Distribution	45
6.9 Extreme Value Analysis	55
6.10 Fault Tree Analysis and Other Hazard Identification and Evaluation Techniques	59
7. CONSEQUENCE ANALYSIS	61
7.1 Direct and Indirect Accident Costs	64
7.2 Direct Accident Costs	67
7.2.1 First Aid, On-Site Medical	67
7.2.2 Off-Site Medical	67
7.2.3 Workdays Lost	67

7.2.4	Fatality Costs	68
7.2.5	Property Damage	70
7.3	Indirect Costs	71
7.3.1	Injured Worker Time	71
7.3.2	Co-Worker Time	71
7.3.3	Supervisor Time	72
7.3.4	General Losses	72
8.	RISK ASSESSMENT OF EXISTING SYSTEMS	75
8.1	General	75
8.2	Risk Identification and Ranking	75
8.3	Theoretical Risk	80
9.	RISK ASSESSMENT OF NEW SYSTEMS	89
9.1	Resource Allocation	90
10.	REFERENCES	105
11.	BIBLIOGRAPHY	107
APPENDIX A--RISK IDENTIFICATION TREE	A-1	
APPENDIX B--USE OF RISK PROJECTION TECHNIQUES IN INVESTIGATION OF ACCIDENTS AND INCIDENTS	B-2	
APPENDIX C--PROBABILITY AND STATISTICS PRIMER	C-1	
APPENDIX D--PLOTTING METHODS, GOODNESS-OF-FIT TESTING, AND CONFIDENCE LIMITS FOR LOG-NORMAL AND EXTREME VALUE DATA	D-1	
APPENDIX E--RISK ASSESSMENT EXAMPLES	E-1	

RISK MANAGEMENT GUIDE

1. INTRODUCTION

Frequently, management allocates significant resources to correct specific hazards without first obtaining sufficient information to determine whether more hazardous conditions are being neglected, or whether the corrective costs are justified by the benefit or the reduction in risk. In addition, management frequently has little or no information of how risk compares to the actual value of a given program, and thus must make many safety-related decisions without sufficient information.

The Management Oversight Risk Tree (MORT) methodology provides a system for identifying management oversights and specific risks. Once risks have been identified, it is then management's responsibility to provide required resources to reduce or eliminate specific risks and to assume the residual risks.

Risk assessment estimates of future losses and the effectiveness of additional controls provides management information to make sound decisions regarding risk. Indeed, knowledge of risk permits the responsible person to decide whether a danger can be accepted, must be reduced, or eliminated by application of additional protective measures, or whether the operation must be cancelled.

As such, risk management and assessment is basic to a system approach to safety management.

Specifically, risk assessment permits or provides:

1. Probability estimates of large or catastrophic accidents.
2. Addition of such loss estimates actuarial predictions of loss to provide a more complete risk estimate.
3. Making safety programs more cost effective by concentrating on high risk areas.

4. Optimization of the combined cost of safety programs and the cost of accidents which present at a given level of control. This includes selection of the list of the various alternatives regarding specific hazards and control measures.
5. Evaluation of the effects of codes, standards, and regulations and the need for relaxation or additional controls.
6. Consideration of various types of risk on a consistent basis minimizing the effects of emotions, fears, and personalities with regard to such related subjects as low probability, high consequence events, environmental and health issues, and immediate versus latent effects.

Various types and degrees of danger are thus treated objectively with biases minimized.

Thus, the role of risk assessment is to provide the necessary information to make decisions regarding the cost effective commitment of resources to accident prevention and reduction. Risk assessment can also be used to determine if a proposed action is acceptable in those situations where it is impractical to eliminate particular hazards. Obviously, those areas where the greatest gains can be made with the least effort should be given top priority. Such prioritization will effect the greatest safety with any given level of effort.

A limitation in this process is that estimates of future losses are necessarily based on probabilities, statistics, and even subjective judgment; and therefore can never be precise. The decision to allocate resources, thus, is always made in the face of uncertainty. The purpose of risk analysis is to reduce that uncertainty as much as practical by providing a framework for the incorporation of all available information regarding the costs and risks of various alternatives. This guide provides some methods for analyzing and presenting this data to management.

2. SUMMARY

Risk analysis is the scientific measurement of the degree of danger or hazard involved in any operation or activity. More precisely it is a product of the frequency and severity of unwanted or accidental events. Measurements of the frequency of unplanned events can never be precise and therefore involve various degrees of uncertainty. In addition, adverse consequences involve a great variety of primary adverse effects and many secondary effects. The tangible effects include degradation of the environment, latent health effects for both the public and employees, property damage, vehicle accidents, and many secondary effects such as reduced environmental values, programmatic delays, etc. As such, the assessment of risk is not simple and requires a wide range of knowledge. The very complexity and lack of understanding of risk leads to gross misconceptions. Many very low risks are perceived as extremely risky and vice versa. Scientific data collection, analysis, and preparation of results can do much to provide an understanding of risk and to provide management with an estimated probable cost of accidents in an operation or activity and the uncertainty in that estimate including the range of severity and probability.

With this information, management can make sound decisions related to allocation of safety resources. This systems approach to safety, or risk management includes the following steps:

1. Establishment of company policy, setting of acceptable or upper limits of risk, and setting goals for reduction of risk
2. Determination of risk through risk assessment and analysis of hazards
3. Allocation of resources to control the quantified risk below the upper limits and to achieve the risk goals
4. Acceptance of residual risk or losses which are expected to occur at the specified control level

5. Monitoring the operation and safety program for change to assure continuance of acceptable levels of safety.

The risk analysis collect and analyze risk data and prepare reports which permit the manager to fulfill his functions in the above risk management steps. To prepare useful reports, the exact purpose or expected use of information must be clearly understood and stated. Assumptions must be distinguished from facts. Not only the results but the analytical methods must be clear and concise.

A large number of analytical methods are available for the risk analyst. The simplest is the direct use of actuarial data (accident statistics).

Last year's losses are the simplest most direct estimate of next year's expected loss or risk. Basic probability and statistical methods can provide knowledge regarding the range and uncertainty of these future losses and add meaning to accident statistics normally presented to management.

In the absence of data, subjective estimates may be required or a survey conducted. Properly made, these provide risk information that is far superior to hunches or pure guessing. Collection, analysis, and use of these actuarial and subjective data are very similar to that of the insurance industry; long-term average losses must be estimated and precautions made for not only the average loss but also for the unusual year in which an extremely large loss occurs.

Predictability and identification of these large losses enhances the ability to prevent them. Such information can be gained through graphical analysis of the frequency-severity relationship of accidents. Two methods for doing this are the log-normal and extreme value analyses.

Not only do these methods permit prediction of large losses, but they also provide insights into safety management. A relatively large number of

midrange accidents compared to smaller accidents indicates either or both under reporting of small losses and inadequate systems for control of large losses.

The different types of losses present a risk assessment problem in that there are no standard common units in which to sum different types of risk. Either techniques which thinly disguise placing a dollar value on the environment, health, or on human life, or a direct dollar value must be assumed if comparisons between various types of risk and subsequent equitable allocation of resources are to be made.

In the assessment of loss of human life, the loss is greater for accidents which occur more frequently at younger ages and latent health effects which result in fatalities later in life. This difference can be accounted for by stating the risk in terms of years-of-life lost rather than by the number of premature fatalities.

Finally, a number of methods are available for summarizing the various kinds of loss in order to provide an overview of company risks. Neglect of one discipline or concentrating too much in another can thus be identified and rectified. Use of these methods will place safety programs in a sound objective basis and will provide the greatest amount of safety for a given budget line. Human life is far too valuable, injuries far too painful, property damage and delays far too costly to do otherwise.

3. BACKGROUND

Risk evaluation has its origins in probability theory and statistics. The first formulation of probability theory was made by Pascal in the 17th century in order to evaluate gambling risks. Today, games of chance, such as dice and roulette, are used as examples of probability theory. In 1713, about a half century later, Bernoulli developed what is called the Bernoulli theorem of binomial distribution. This theory is useful in dealing not only with games of chance but also with quality control, inspection, public opinion polling, genetics, etc.

Later, Poisson developed basic theory dealing with how often events occur. If more than one event can occur per trial, it determines the probability that "x" events will occur. For example, what is the probability of a given number of counts on a Geiger counter in a 15-s interval, the number of worms in a cubic foot of soil, or the number of accidents in a given period of time?

It appears that the first application of probability mathematics to accident frequency or risk evaluation was by Von Bortkiewicz¹ in the 19th century. He studied the records of soldiers dying from kicks of horses in 20 Prussian Army Corps over a period of 10 years. For these 200 sets of observations, he calculated the relative frequency with 0, 1, 2, 3, or 4 deaths would occur and compared the results to actual experience. In one instance there were four deaths even though the average was only 0.6 deaths. The calculations were in good agreement and Von Bortkiewicz concluded there was no evidence that in any one corps in any given year, soldiers were more careless or horses were more wild.

The lesson for the safety engineer is that if a "rash" of accidents occur, it is not easy to determine whether changes have occurred causing an increase in accident frequency or whether the rash is a rare, random situation such as when four soldiers were kicked to death in a single year in one corps.

Near the end of the 18th century, Gauss developed the theory of normal or Gaussian distribution. This theory deals with continuous rather than

the discrete distribution of the Bernoulli (binomial) and Poisson theories. For example, the earlier theories predict that an event will or will not happen (two possibilities); thus, the term binomial. The Gaussian theory approximates distributions of measurements in nature, industry, psychology, etc. For example, what fraction of the students in a classroom are in a given weight or height range, rather than simply dealing with how frequently an event will occur. This theory can also predict the probable number of accidents which will occur in a given time period.

Risk evaluation was next applied by the insurance industry. Until recently, their approach to risk evaluation has been strictly actuarial or statistical. (Based on past experience, what are the expected losses next year?) Their approach to the quantification of risk has been to develop increasingly complex and narrower classes of risk. Preferred risk premiums apply, for example, to buildings with fire protection systems, people who do not smoke, adults with no teenage drivers, etc. Where experience has been lacking to predict future losses, insurance companies have protected themselves by very large premiums and/or by limitations of liability. These are not viable options for the program manager, therefore he needs greater risk assessment capability.

The first national tabulation of work accidents and rates was published in Accident Facts by the National Safety Council in 1928. Safety engineers soon began statistical analysis of accidents. In the 1930s, Heinrich studied accident frequency and severity and concluded that for each 300 minor injuries there were 30 serious injuries and 1 fatality. While these statistics may represent the average throughout all industry, their use could be misleading and dangerous. For example, the estimation of the probability of a fatality based on these statistics and the number of injuries in an office may lead to undue concern and safety efforts. Obviously, we cannot predict the chance of a fatality based on paper cuts, fingers shut in drawers, etc. On the other hand, no high rise construction worker should take comfort in the fact there had been few minor injuries among his coworkers.

The first large attempt to analyze and control hazards was with the Manhattan project. Previously, new technology was developed with practically no safety considerations in the design or development stages. Steamboat explosions were common on the Mississippi River in the 19th century. In the 1930s, the automobile death rate per vehicle mile, even at the lower speeds, was nearly three times the current rate. Countless eyes were needlessly lost before the need for safety glasses was realized.

However, beginning with the Manhattan project, the nuclear industry introduced safety analysis reports, safe work permits, etc. Each phase of each project was routinely and systematically analyzed for hazards, and control measures were adopted prior to starting the actual work.

These original safety analyses were limited to identification of hazards and evaluation of maximum consequences (worst-case analysis). The safety analysis reports were primarily concerned with limiting the worst accident (the Maximum Hypothetical Accident, later called the Design Basis Accident) to a given consequence level. For example, the risk was considered acceptable if the off-site radiation dose from the maximum credible accident did not exceed specified limits. The risks of more frequent but smaller accidents were treated superficially or not at all. The identification of hazards usually resulted in control measures being applied without cost/benefit analysis (risk quantification).

In the 1950s, Gumbel² developed the extreme value theory which can be used to predict the frequency of maximum events. This theory was first applied to natural events such as maximum river flow, highest winds, etc. The theory was also used to determine the adequacy of dams and flood control projects, the necessary wind resistance capabilities of building structures, etc.

With the development of intercontinental missiles equipped with nuclear warheads, a major advance in risk evaluation was necessary. An unplanned or inadvertent release of a nuclear missile programmed for the destruction of a foreign city was beyond any previously conceived or actual accident. No previous experience was available to apply statistical theory. A search

for ways the accident could happen and appropriate counter measures (as was done in the nuclear industry) was necessary but inadequate. A systematic method for evaluating the probability for inadvertent missile launch was needed. As a result, fault tree theory was developed.

In fault tree analysis, a single event (such as the accidental release of a missile) is postulated. Then, different events which can lead to this accident are searched for and arranged in a diagram which resembles a "tree." This process is continued until individual component failure or initiating human error is reached. The tree arrangement permits sequence of events and failure relationships and consequences to be evaluated. Assignment of probabilities of initiating events in the fault tree permits the evaluation of probability propagation to the top event. As far as possible or practical, all possible paths leading to the top event are identified; and the propagation of consequence up through the tree from the multitude of individual component failures and human errors are analyzed by the use of probability theory. Thus, the likelihood of the top event (or accident) can be estimated. Of perhaps greater value is that the various chains of events which can lead to the top event are identified, and additional systems control can be applied where most needed.

The application of probability (frequency-severity) distributions to industrial accidents has been developed recently. Gumbel's extreme value analysis could have predicted that a large fire had a relatively high probability of occurring at Rocky Flats. This technique is currently being used to calculate the frequency of maximum accidents including fires. The log-normal distribution, a specialized case of the general Gaussian or normal distribution, has been used to plot the frequency and severity of accidents and to predict the frequency of large events. Such predictions are generally in good agreement with extreme value theory. They have the added advantage of including all accidents, not just the worst accident in each time period. As such, the log-normal distribution can be integrated to quantify the entire spectrum of accidents. For example, the log-normal plot is extrapolated to include the large events which may be underrepresented in the historical data. The integration then includes the entire spectrum of accidents. While the probability of the maximum or worst-case

accidents can be reduced to acceptable levels with fault tree analysis, this technique provides some assurance that the sum total costs of all accidents will be within tolerable levels.

Yet to be developed are standard values for different kinds of risk (life, property environment, etc.). Also in the formative stage are standards for risk acceptability and resource allocation. Currently the science and use of risk assessment and management is growing rapidly. Many companies now have a position of risk manager. The Federal Government now requires risk cost/benefit studies for proposed regulations to reduce hazards. Insurance companies are becoming aware that more sophisticated risk assessment techniques are needed. It is hoped that this report will provide assistance for DOE and its contractors who wish to begin or improve an existing risk assessment and/or management program.

3.1 Understanding Risk

Laplace wrote in 1814, "Strictly speaking, it may be said that nearly all of our knowledge is problematical." Thus, managers and safety officials (in fact everyone in all matters of life) make decisions based on evidence which is logically incomplete.

The amount and quality of evidence available to predict a given outcome determines the confidence or degree of assurance in the likely outcome and provides a measure of probability of given outcome. As evidence changes, our confidence in the outcome or our estimate of the probability of the outcome changes. Thus, probability is not an intrinsic characteristic or trait of a future event but only a measure of evidence for that event. Thus, consideration of probability whether quantified or intuitive plays a fundamental role in rational thought and conduct, and has been declared to be the guide of life.

Estimates of probability may be very precise, as in the probability of a five in a single throw of a die as 1/6, or very imprecise as in probability of a given return on a stock investment. In neither case does an estimation or probability influence the outcome. In every individual

trial, regardless of the probability and regardless of how accurately that probability is known, the proposed event will either occur or not occur.

An estimate on a subjective probability is a measurement of how strong the estimator feels about a situation. While this may vary from individual to individual, the uncertainty can be reduced by using a panel of experts and/or by averaging subjective estimates. Incorporation of such feelings (numerically) into a risk analysis is better than no analysis and also serves to document or record differences of opinion. Indeed, it provides a record of estimates for the risk evaluation, which can be changed, if desired; and new results can then be calculated. In such cases, the chief value of risk analysis may not be the final risk figures obtained, which are certain to be open to much criticism and questioning. The value will lie in revealing many, if not most, of the various possible damage causing mechanisms; and thereby provide better insights to effective control measures.

Thus probability can be defined as (a) a measure of subjective expectation, (b) a degree of confidence in an outcome whose numerical value can be estimated by logical reasoning, and (c) the relative frequency with which any event occurs in a class of events.

In a broad sense, risk refers to the uncertainty in any outcome. Risk management and assessment includes assembly, analysis, and use of knowledge in a systematic way to define and reduce the uncertainty in any outcome whether associated with danger to personnel and property or not.

This guide is limited to the narrow concept of risk which deals with the danger of loss from accidents. As explained in more detail later, risk is defined as the probability of loss multiplied by a measure of the consequence.

There is an element of danger in every human activity. Usually, people try to avoid danger and take all due precautions to preserve life and limb. Yet there is an element of intrigue and excitement in risk taking. The death defying high wire acts and other stunts where daredevils

deliberately flirt with death attract crowds and much public attention. In spite of the fact that risk is common and all live with it everyday, when it comes to evaluating and understanding risk, many feel there is a mystique about the unfamiliar subject of risk. Indeed, many are prone to say of a fatal accident that "his time had come." Nevertheless, the concept of risk is quite simple. The dictionary defines risk as "the chance (probability) of harm or adverse consequences" or as "the degree of exposure to loss or injury." These are the qualitative and quantitative definitions of risk used throughout this guide; with the term "risk" when used quantitatively being synonymous with "degree of risk." Risk, safety, and danger are analogous to the terms temperature, cold, and hot; temperature being a measure of how cold or how hot. Just so, risk is a measure of how safe or how dangerous. (Safety and danger are relative terms for loss potential but at opposite ends of the scale similar to cold and hot.) The degree of risk (how safe or how dangerous) is measured by the probability of a potential loss multiplied by the severity or cost of that potential loss. Thus, risk is the expected loss. If a person bets \$10 on the flip of a coin, his risk or expected loss is \$5 (\$10 times a 50% chance of losing).^a He also will win \$5 half of the time, so his risk will be equal to the gain from the gambling venture in this case.

Somewhat confusing is the fact that risk is sometimes defined and used to denote only one of the two risk parameters (either the probability or the amount of the potential loss). Another dictionary definition of risk is "the probability of loss." Frequently, the statement that a venture is risky means only that there is a high probability of loss. Another dictionary definition is "the amount the insurance company stands to lose." With this definition, risk in the previous coin toss example would be \$10 (he risked \$10 on the flip of a coin). A third, qualitative definition of risk is "exposure to a hazard": "He risked his life to save a child."

a. Since risk is a composite function of how often and how severe, frequency-severity distributions of accidents define a risk spectrum.

For our purposes, risk will be restricted to the primary definition, that of expected loss which equals the product of the probability and the consequence; and thus includes both aspects of risk.

The probability term indicates to what extent one can expect the loss to take place. Probability is stated as a number between zero and one. A value of one indicates total certainty, however, the loss in question must take place in the considered period of time. A probability of zero means that the event cannot take place. In nearly all cases where risk is discussed, the probability is neither one nor zero, but is at some intermediate level. This simple observation is very basic and very important. It means that there is nearly always a residual risk. Many fruitless discussions could be avoided if this concept were understood and accepted.

Germane to this concept is that probability or risk approaches zero asymptotically. That is, the time interval between events, being the inverse of probability, approaches infinity as the probability approaches zero. In other words, the time between low frequency events is unbounded. The other end of the scale is bounded, as the probability approaches one, the probable time for at least one event to occur approaches the considered time interval. This skewness of the probability distribution will result in the geometric mean of high and low probability estimates being low. Another difficulty is that few of us have very much practice in dealing with very low probabilities. We see numbers like 10^{-5} ($1/100,000$); the meaning of which is difficult to grasp.

The words "certainty" and "uncertainty" as they relate to probability and risk are also frequently a source of confusion. A probability of one means that certainty is absolute; the event will always occur. In this sense, a probability of zero could also denote certainty in a negative way--it is certain the event will never occur. Thus, a probability of 0.5 represents the maximum uncertainty--there is an equal chance the event will or will not occur.

This concept of certainty must not be confused with how well the probability value is known. In flipping a coin, the probability is known

to be precisely one-half (0.5). In most risk assessments, the probability value itself is not exactly known and must be assigned an uncertainty value. In the probability value of 0.9 ± 0.01 , the 0.9 denotes the degree of certainty that the event will occur, and the 0.01 represents the degree of certainty with which the probability value of 0.9 is known. This distinction is important and should be understood when referring to uncertainty.

The other term in the definition of risk, cost or severity, may be thought of as the degree of undesirability in the event which is of interest. The undesirable event usually involves loss of some value and can thus be measured in terms of

- Monetary value
- Loss of life or damage to well being
- Environmental damage

or even intangible values such as

- Loss of freedom
- Public reaction
- Employee morale.

Another factor to remember is that while these items have different degrees of undesirability, the degree itself is usually uncertain--we may expect a strong public reaction, but due to unforeseen circumstances it may be quite mild. This amorphous nature of risk analysis is not well understood and sometimes results in risk assessments being criticized or rejected. The fact is, that probability and risk theory is an exact science which deals with or measures uncertainty.

3.2 Risk Perception

Lack of knowledge, fear, the public media, and other factors influence our perception of risk. Since acceptance or opposition is necessarily based on how risk is perceived, it is important that the risk analyst understand risk perception. This understanding will also enable the analyst to make better subjective estimates.

A recent study³ in which members of the League of Women Voters were asked to estimate risks of various activities on products is quite revealing. The women were given a list of activities and products, then asked to rank them in order of risk and assign risk values to them. A value of 10 would be assigned to the least risky. For example, the annual number of deaths in the United States being the measure of risk, an activity causing 10 times as many deaths as the least risky activity would be assigned a value of 100.

Given in Table 1 are (a) selected perceived risk values from this exercise, (b) the number of deaths per year from either statistical tables or risk analyses, and (c) the ratio of the perceived risk to the actual risk normalized to a value of one for the smallest ratio. Since the league

TABLE 1. PERCEIVED RISK

Item	Risk as Perceived by League	Number Deaths	Perceived Risk Divided by Number of Deaths (Normalized)
Food coloring	31	1	16,400
Nuclear power	250	10	13,000
Football	37	20	975
Vaccination	17	10	900
Fire fighting	92	170	285
Commercial aviation	52	100	275
Handguns	220	1,000	116
Private aviation	114	1,100	54
Railroads	37	400	48
Bicycles	65	1,000	34
Motorcycles	176	4,000	23
Motor vehicles	247	50,000	3
Smoking	189	100,000	1

was asked to estimate risks in arbitrary units (not the number under or over), estimation of each risk cannot be determined. The ratio demonstrates only the extreme inconsistency of risk perception.

From Table 1, we can make the following observations:

1. The range of risk perceived by the league results in a ratio of only 15 to 1 (nuclear power is rated at 15 times riskier than vaccinations), whereas the actual ratio is 100,000, (smoking causes 100,000 times as many deaths as food coloring). Note, if we eliminate estimates and use only known statistical values the range is still 2500: motor vehicles (50,000) divided by football (20) equals 2500. This range is a factor 170 times the perceived range.
2. There is a strong inverse correlation between the actual number of deaths and the ratio of perceived to actual risk.
3. Activities involving relatively few people such as fire fighting and football have a high perceived to actual ratio.

From these observations, we conclude:

1. The public has little knowledge of actual risk values which are, in fact, fairly well known to statisticians and risk analysts.
2. Reading about risk distorts risk perception. For example, football and nuclear power which are much in the news are grossly overestimated.
3. Estimating a societal or average risk of an activity involving a small percentage of the population generally requires a detailed analysis to avoid overestimating the risk (football was overestimated).

4. There is a strong aversion to catastrophic risk. In a followup study⁴ students were asked to estimate the number of deaths in a normal year and in a disaster year, and the disaster year was overestimated.

In this last respect, nuclear power was in a class by itself. The 24 students, who were asked to describe the worst nuclear accident that would occur in their lifetime, expected few deaths in a normal year; but 25% of the students expected more than 100,000 deaths in a disaster year. The Rasmussen report⁵ states that an accident with 3300 prompt fatalities has a probability of 5×10^{-9} per reactor year. Assuming 100 reactors operating for 60 years, the probability would be 3×10^{-5} or once in 33,000 years. Yet 10 of the 24 students expected an accident of greater severity in their lifetime.

Without arguing the merits of the Rasmussen report, it is sufficient to alert the risk analyst to the phenomenon of risk aversion. Many believe if it can happen, it will happen. The risk analyst must deal with future risk versus current costs and must decide whether to value loss on a linear basis--as losses became catastrophic, this risk appears to be unacceptable to some regardless of how small the probability is estimated.

This guide does not recommend any particular discount rate for future losses in estimating risk. This guide does assign the same value for 100 lives lost in a single event as it does for 100 times the value of one life lost. It is recommended that these factors be fully considered and explicitly stated. To keep risk analysis simplified, these factors are not considered in examples and formulas presented in this guide.

The primary bias which must be considered by the person estimating probability is the tendency to underestimate high frequency and overestimate low frequency. The ordinary mind does not readily perceive the vast difference between 1 in 10^4 and 1 in 10^7 ! To make a better estimate, one should:

1. Relate probability estimates to known experience.

2. Divide a project or operation into subtasks and estimate the probability of the subtask.
3. Obtain estimates from a panel of experts. Group estimates tend to be better than individual estimates. Also, variance in the estimates of several persons is an indication of the degree of uncertainty in the probability.

4. RISK MANAGEMENT

Risk management is loss control exercised by sound management principles. Loss from a manager's point of view can be anything that increases cost of operation or reduces productivity. Risk management involves the understanding of potential adverse effects and the systematic application of controls to optimize productivity by minimizing losses.

The risk management function includes gathering and organizing the necessary risk information, recommendation, developing a system, using the information, and, perhaps, making recommendations.

The manager's function is to make decisions and allocate resources to accomplish a given task or mission. To proceed, the manager must control costs, schedules, and undesirable side effects. Effective control, in turn, requires planning and forecasting to eliminate those events which will cause failure. Four basic failure modes are:

1. Failure to produce a specified product
2. Failure to produce the product at an acceptable cost
3. Failure to produce the product within an acceptable schedule
4. Failure to produce the product with acceptable undesired outputs.

Acceptable, herein, means informal agreement within legal and ethical constraints. These failures are further developed in Figure 1, Mission Failure Mode Tree. Lower tiers of the tree indicate the specific failures under the four basic failure modes which will compromise success and, therefore, constitute the family of risks involved in the mission or project.

Examination of the tree indicates that a total coherent evaluation of risk includes the business or economic risks as well as those risks which

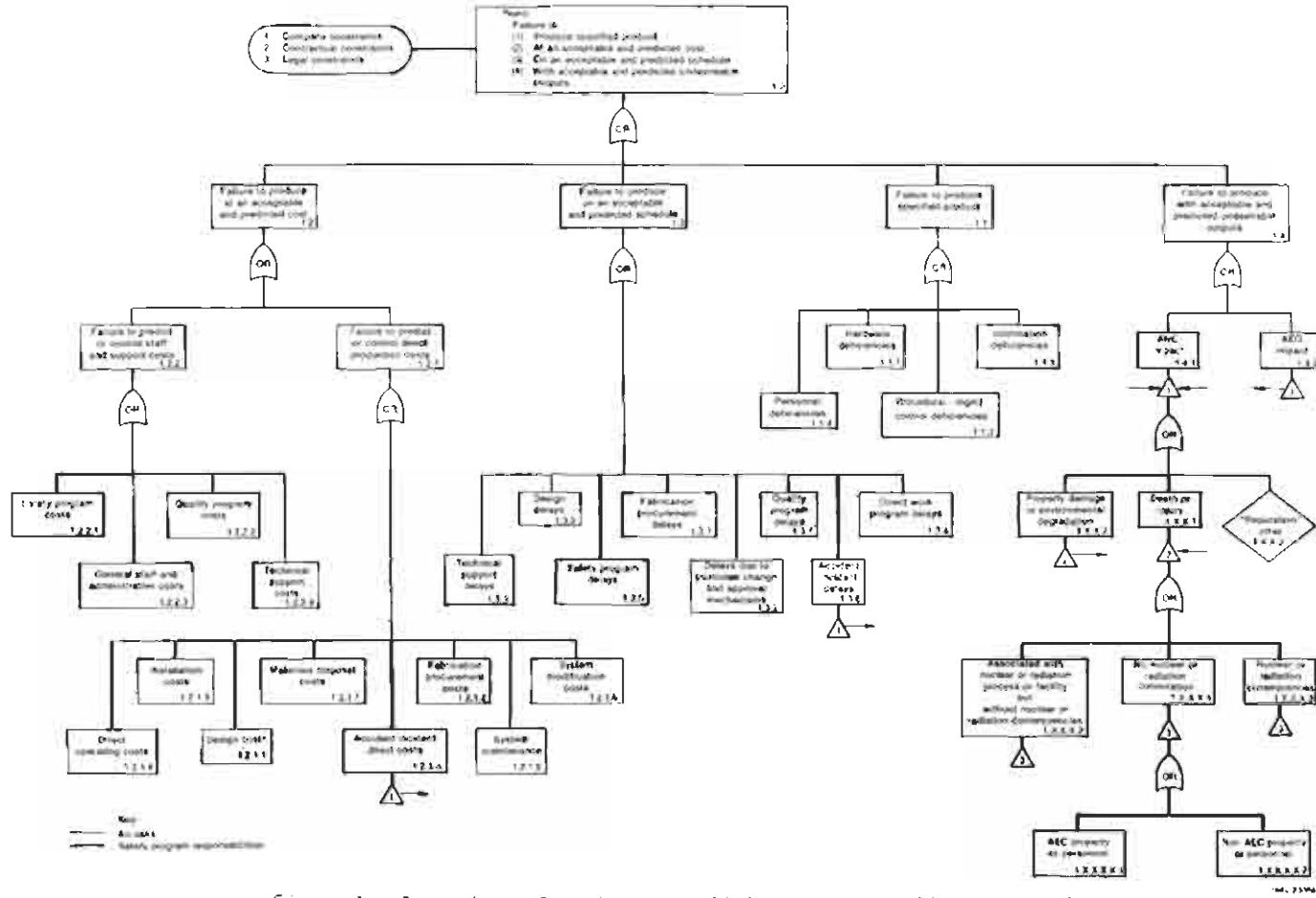


Figure 1. Comparison of various normalizing factors applied to annual direct property damage costs.

are essentially "safety" in nature (personnel, property, or environmental harm). These safety areas are those portions of the tree that are in bold-line.

Two points may be noted:

1. The "Safety Program" is found in three of the four major failure mode branches. The one branch, "failure to produce a specified product," could include property damage (accident cost), if quality control inadvertently broke down and permitted impurities or other imperfections in the final product (degrading its value).
2. The safety program is clearly an integral part of the total risk management program. As such, the safety program risk evaluation must be communicated to management in the programmatic and economic language of the project so that it can be combined with or considered in the same terms as other business risks. While only one branch is labeled, "Failure to produce at an acceptable and predictable cost," cost can be assigned to the other branches. Thus, the "cost" branch is labeled direct costs while the other three branches may be considered as indirect and/or intangible costs.

The tree can be considered in two ways; as a success tree, a failure, or a risk tree.

1. To convert to a risk failure mode tree to a suggestion tree, change all "or" gates to "and" gates and remove the word failure from each box. Thus, the total cost of the project is the sum of the direct support and production costs and the indirect costs of the other three branches. As is clearly illustrated, accident costs are an integral part of the costs to produce a product. The transfer symbol indicates that property damage, environmental harm, death, and injury are considered as direct production costs, the costs of undesirable outputs, or the cost of delays. If only direct accident costs are included in direct costs

(Block 1.2.1.8), and the indirect accident costs only are included under delays (1.3.6) and impact (1.4.1), there will be no duplication. If, however, as is usually the case for the risk analyst who is considering only accidents, the total costs of accidents are assessed as a unit for the various hazards (vehicle, inplant property, and personnel), then care must be taken to avoid duplication of risk.

The tree was not originally intended as a tool or format for compiling or tabulating risks, but rather as an illustration that to achieve success, management must identify and control the potential sources of failure. If labor costs, delivery schedules, quality control, etc., are not controlled, failure will result. A balance must be achieved between control costs and failure probabilities (or risk) to provide an optimum for success. Either excessive safety program costs or excessive accident costs can jeopardize success.

2. To complete the illustration, consider the tree as a failure tree, as drawn in Figure 1. The failure to control production or accident costs will produce a cost overrun. The risk for each element in the tree is the probability of control failure multiplied by the consequence. Evaluating the total tree then provides the probable total cost overrun (this is an exercise for an experienced fault tree analyst). This exercise is not necessarily recommended, but if assessments are based on most probable production, delay, product deficiency, and undesirable output costs, then the cost evaluated from the "success tree" will be the most probable cost.

Thus, probable accident costs as well as safety program costs must be included in project cost estimates if the risk of cost overruns and the risk of project failure are minimized. This concept of risk refers to business risk and deals with uncertainty of loss estimates. "Risk" as used elsewhere throughout this document does not include the nonaccident elements of business risk. It does include both the loss estimates and the uncertainty in

the estimates involving injury, exposure to harmful agents (health effects), property damage, programmatic delays, and adverse environmental and public impact. The major steps required to control these losses define the basic risk management process as follows:

1. Establish a company policy and set tolerable or acceptable risk levels; i.e., set an upper limit of risk beyond which people or property will not be exposed; and set goals for minimizing risk
2. Determine risk and allocate resources
3. Allocate resources
4. Accept reduced risks or apply additional controls to further reduce risk
5. Monitor operation and loss control program for change.

Since the conduct, control, and safety of operations are line functions, the responsibility for risk management rests with line management. Generally, Steps 2 and 4 (the hazards search and risk analysis, and the monitoring) will be delegated to a safety organization because they are not directly related to conduct of operations and require special expertise. Steps 1 through 3 (the specification, and the acceptance of risk and application of controls) require input from various groups both within and outside the company organization. Regardless of the company organization, it is important that each of the functions be defined and assigned to a specific department.

Each of the five steps are discussed below.

1. Establish Acceptable Risk Levels and Goals--Within the constraints of codes, standards, and regulations, there is some latitude for the manager to establish upper limits of risk. Also, the risk management process will identify either over or under regulation of hazards. In addition, the wise manager or safety professional

will not assume that compliance with codes, standards, and regulations is equivalent to adequate safety. Hazards must be systematically identified because no code or standard can ever apply to all conditions at all times.

The first and primary guide for establishing an acceptable risk level is that risk not be out of line with that which is commonly accepted. A second guide is that occupational risk should be small compared to mortality risk from disease. For reference, the following fatality rates are given:

Cause	Annual Deaths (United States)	Deaths/100,000 Population
All causes		
All ages	1,933,000	914
Age 40	--	400
Age 20	--	100
Natural	1,778,370	--
Heart disease	746,480	353
Cancer	--	--
All other natural causes	358,400	170
Accidents	673,490	--
All accidents	103,000	49
Vehicle accidents	52,000	25
Work accidents (USA)		
All occupations	13,200	13
Construction	2,600	52
Transportation	1,600	30
Manufacturing	1,800	7
All DOE and Contractor	9	6
Other	51,630	--
Suicide	26,430	13
Homicide	20,780	6
Other	4,420	2

The construction death rate is about one-half the natural death rate at age 20. There are high risk occupations with an

occupational death rate of several hundred deaths per year per 100,000 workers (or approaching the natural death rate at age 40). The ethics of permitting unequal death rates in different occupations and the impracticality of equalizing risks are outside the scope of this document. Our goal is ordinarily to minimize loss, but not at the expense of subjecting (sacrificing) any individual to extremely high risks.

One approach that has been suggested for establishing risk acceptance criteria is that, for involuntary risks to the public, the death rates should not exceed those from natural causes. As a guide the following fatality rates per 100,000 population are given.

Cause	Annual Deaths (United States)	Deaths/100,000 Population
All natural causes	1500	0.07
Excessive cold	634	0.03
Tornado, flood, earthquake	200	0.01
Lightning	100	0.005

The death rates for both public and occupational rates are presented only as information. These rates could be used as a suggested starting point for discussion and establishment of upper or acceptable levels of risks. The intent of establishing upper levels is that whatever resources are required to meet these goals should be expended. In any case, total losses should be small compared to net gain or profit expected from an activity.

In addition to establishing upper risk levels, goals should be established and plans formulated in order to minimize risk or cost of accidents.

The total accident cost is the cost of accidents plus the cost of preventing accidents. These total costs are minimized if large resources are not expended on small risks or inadequate resources are not allocated to large risks.

Also, goals can be humanitarian; that is, resources could be expended somewhat beyond that which returns economic dividends. The intangible benefits in improved employee morale and goodwill may justify a safety program beyond that which can be justified by tangible losses from accidents. While general goals may be established at the beginning of a project, they may be modified later if it becomes evident that some goals might be too difficult or if further gains might be realized.

Finally, several large corporations have outstanding safety programs that demonstrate that extremely low injury rates and property loss risk are compatible with efficiency and profitability.

2. Determine Risk--Since most of this guide deals with hazard identification and risk analyses, only general principles are discussed in this section. The following steps are applicable to any risk assessment.
 - a. Decide what questions need answering and exactly what the risk assessment is to accomplish. Do not obscure the analysis with irrelevancies.
 - b. Define the operation being analyzed. Unless the operation or hazard is bounded and properly documented, the analysis becomes infinite. The operation being analyzed may be as simple as a single critical crane lift or as complex as the entire life cycle of a major operation.
 - c. Identify hazards. A large number of techniques for identifying hazards exist in the literature. All involve classifying or placing hazards in various categories and systematically searching each class. A thorough and exhausting search can be made by using the Risk Identification Tree given in Appendix A. This method is too detailed and time consuming to apply to every hazard in a large operation or company.

To simplify, the usual hazards from normal industrial activities can be treated collectively and quantified using previous accident experience.

- d. Assess risk. Determine the potential consequence of each hazard and the probability of its occurrence. The usual risks of occupational injury, fire, property damage, and vehicle accident can be treated collectively and quantified using previous accident experience. Unusual or high consequence, low frequency events which cannot be quantified from statistical accident data should be determined individually and added to the statistical risk. Formulas, techniques, and methods for assessing the statistical risk estimates and assessing individual risks are given in the Analytical Methods section. Multiplying the probability of each potential loss by its consequence value will give the risk in units of expected loss. Thus, the units of risk are the number of fatalities, injuries, workdays lost, quantity of pollution released as well as dollar losses from property damage, medical expenses, etc. These various types of risk can be itemized, but to reach a single risk value requires risk evaluation.
 - e. Evaluate risk. Evaluating risk requires placing a degree of undesirability upon the various types of risk. If equivalencies between environment, safety, and health risk are established with management concurrence and used in all risk evaluations, much time could be saved; and environment, safety, and health issues can be treated consistently and objectively by arguing their relative merits in each proposal. In special situations, the equivalencies could be reexamined without necessarily compromising this system.
3. Allocate Resources--It is essential to allocate sufficient resources to a safety program and to line management to control risks within the upper limits established in Step 1. Additional

resources to meet goals established for minimizing risk can also be considered. One consideration is the cost savings in risk reduction gained from additional safety expenditures.

4. Accept Residual Risk--The manager rather than the risk analyst should make the final decision to accept the residual risk. However, the analyst should not submit a risk report to management until he is satisfied that all special or unique hazards have been identified and adequate controls to minimize cost and ensure that the success of the project or activity will not be jeopardized by accidents.

It is important that risk responsibility be carefully defined and formally documented. As a general rule, the same authority which sets standards and approves procedures may also bypass safety requirements. As an example, a foreman was asked if, in order to meet a schedule, he had authority to bypass a limit switch. His reply was "Yes," but when it was pointed out that limit switches were required by the safety manual which had been issued under the signature of the General Manager, the foreman changed his mind. In short, risk acceptance procedures are needed so that each foreman, supervisor, and employee clearly understands what level of risk he is authorized to accept.

5. Monitoring and Control Review of each phase of a project will help ensure that the entire lifecycle is carried out in accordance with the controls and limitations set forth. The operational controls and the required resources necessary to maintain risks within the established levels and to meet the minimum risk goals will have been identified. In Steps a, b, and c, highlighting these controls in a safety document⁶ for distribution to appropriate design, construction, installation, test, operation, maintenance, project, quality assurance, and safety groups will facilitate compliance.

Monitoring will provide assurance that these controls are implemented and maintained. To be most effective, the monitoring will begin at the conceptual design stage and follow through to operation and dismantling and/or decommissioning. (See Operational Readiness-SSDC-1).⁷

Design review, quality control, and safety inspections will help assure that no changes are made which would violate the safety documentation without prior review and approval by those who reviewed and approved the original safety documentation. This monitoring is a backup to the line manager who has first and prime responsibility for operating within the safety envelope.

5. REPORT TO MANAGEMENT

The scope and depth of a risk assessment report depends upon the reason or purpose for doing the assessment. There are at least three separate purposes (types of risk assessments) each of which determine not only the scope of the assessment but also the content of information reported to management:

1. Safety Assurance--The first purpose is to assure management that a specific hazard presents no undue risk to a project or operation. Risks associated with normal or routine operations may be acceptable on the basis that qualified safety professionals have a good safety program. An unusual hazard may surface requiring a risk assessment. For example at one DOE site, the safety director became concerned about a proposed location of an office building near the end of an airport runway. To assure management the risk was acceptable, an assessment was made. Only one hazard was considered; that of an aircraft crash into the office building. Alternatives, such as a different site and additional measures to reduce risk, were not considered because the probability of a crash was assessed as very unlikely. Of course had the risk been unacceptable, the assessment would have been expanded to the second type discussed below. This risk assessment is included as an example in Appendix E.
2. Cost/Benefit Trade-Offs--This type of assessment evaluates the cost of risk reduction measures against the estimated reduction in risk. It answers the questions: Are further controls warranted? Which controls are most cost effective? For example, reactor reflector blocks must be shipped cross-country to a test reactor. Five pairs of five sections are to be shipped on a single truck. An accident damaging the blocks would delay reactor startup by one year. Shipping single pairs of dissimilar blocks on five separate trucks would reduce the probability of reactor shutdown because it would take two accidents rather than one to shut down the reactor. What is the risk associated with one

truck versus five trucks? Is the extra cost justified? This risk assessment is also included as an example in Appendix E.

3. Overview Risk Assessment--This type of assessment quantifies risks by hazard categories. Its purpose is to assess the total organization or industrial risk and place these risks in perspective. This information can be used by safety program directors and line managers to determine if safety is balanced, and to adjust resource allocations or regulations to achieve a more cost effective safety program. In this type of assessment all risks are quantified. Routine industrial safety risks are quantified and placed in perspective to unusual risks such as nuclear or toxic material risks. The risk reduction effects of additional resource to a particular area of risk may be estimated.

The risk report will be better received if it is clear and communicates results to management in terms of economic cost and programmatic impact. While adverse environmental and public health effects are important, the programmatic impact from public reaction or regulatory action of these effects should be communicated, if a reasonable estimate of such effects can be made.

While a specific outline is not suggested, the following elements should be included on a report to management.

State the results in concise terms summarizing the basic assumptions and method. State the purpose of the risk assessment and why it is needed. The scope of the risk assessment should be included. Define the system being analyzed. Accidents and adverse consequences have far reaching effects and may adversely affect other systems. The assessment will be endless or incomplete unless limited to a well defined system.

Describe the method and analytical model. Discuss the limitations if possible or give an upper and lower range of risk. All the calculations should not be included in the report. The equations should be given with sufficient data (on a reference to data) that the analysis could be repeated

by a reader. A clear distinction should be made between assumptions or estimates and hard data. If a reader disagrees with any assumption or estimate, it should be easy to insert different assumptions or estimates to determine the effect on the results. Clearly separate the probability and consequence factors so that different assumptions or estimates of either can easily be inserted into the analysis to test the effect on the risk assessment. Either provide confidence levels on upper and lower bounds with a best estimate.

Discuss factors which were not considered in the analyses because data were unavailable or for other reasons.

Place detail and data in an appendix in order to keep the body of the report concise and clear. Basic methods, assumptions, and results should be readily available to the manager without sorting through a mass of detail.

If the risk assessment is a cost/benefit trade-off study, those who bear the cost and the recipients of the benefits should be identified. Much of the argument generated by many risk reports arise from the fact that frequently those who reap the benefits are not the ones placed at risk.

Present the results in graphical or tabular form if feasible. Headings, labeling of axis, etc., should be self-explanatory. Too much data on a single graph will not be as easy to read as several graphs.

If practical, give both the probability of loss and the consequence with the resulting risk. A catastrophic loss with a low probability may be more important than an equivalent risk with a higher probability and lower consequence.

The report to management should clearly identify those factors having the greatest effect on risk, the risk should be clear and well defined with limitations spelled out and should be communicated in business language avoiding risk jargon.

6. ANALYTICAL METHODS FOR RISK QUANTIFICATION

Accident risk is the expected or probable loss per unit of time or unit of activity and is equal to the probability of loss multiplied by the magnitude of loss. For any operation, the risk is the sum of the individual risks for each potential loss.

$$R = \sum p_i c_i \quad (1)$$

where

R = risk

Σ = summation symbol meaning add each consequence multiplied by its probability

p_i = probability of i^{th} accident

c_i = cost of consequence of i^{th} accident.

Since there is an infinite number of both probabilities and consequences, an accurate quantification of risk requires consideration of the entire accident cost-frequency spectrum.

6.1 Actuarial Risk Assessment

An actuary is a person who computes insurance premiums or risks based on statistical data. Thus, actuarial risk assessments are based on accident experience. Data can be obtained from Accident Facts⁸ published by the National Safety Council, an Almanac, the U.S. Statistical Abstract, the Bureau of Labor Statistics, or numerous safety records and reports maintained either by individual companies or by National/International Agencies such as the National Transportation Safety Board. The following vehicle risk problem is following as an example of actuarial risk assessment.

6.2 Example Problem

What is the annual risk of driving to and from work for the average person?

Accident Facts, 1981 edition, Page 40, indicates that in 1980 there were 2.98×10^7 accidents and a total vehicle mileage of 1.511×10^{12} miles. The unit probability is:

$$P = \frac{2.98 \times 10^7 \text{ accidents}}{1.511 \times 10^{12} \text{ vehicle - mileage}} = 1.97 \times 10^{-5} \text{ accidents/vehicle - miles.}$$

The exposure is the number of miles driven: assuming 20 miles/day and 225 work-days/year gives 4500 miles/year, P_{annual} = the annual probability is:

$$1.97 \times 10^{-5} \frac{\text{accidents}}{\text{mile}} \times 4.5 \times 10^3 \frac{\text{miles}}{\text{year-person}} = 0.089 \frac{\text{accidents}}{\text{year-person}} .$$

Accident Facts, 1981 edition, Page 4, indicates a total monetary loss of $\$39.3 \times 10^9$. With 29.8×10^6 accidents, the average cost is \$1319 including wage loss, medical expense, insurance, and administrative costs as well as repair costs. Indirect losses associated with legal courts and cargo damage are not included.

Risk = probability x average cost

$$\text{Risk} = 0.089 \frac{\text{accidents}}{\text{year}} \times 1319 \frac{\text{dollars}}{\text{accident}} = \$117.39/\text{year}$$

$$\text{Risk} = 0.089 \times 1319 = \$117.39/\text{year.}$$

To calculate the risks of each consequence separately, multiply the costs for wages lost, medical expense, insurance administration, and property damage as given on Page 5 of Accident Facts by the probability, 0.089.

The fatality risk can be similarly calculated:

Accident Facts, Page 40, gives 52,600 deaths in 1980. The mileage given previously is 1.511×10^{12} . The probability, P, is:

$$P = \frac{5.26 \times 10^4 \text{ deaths/year}}{1.511 \times 10^{12} \text{ total mileage}} = 3.48 \times 10^{-8} \text{ deaths/mile.}$$

The exposure (mileage) is the same, 4500 miles/year. Thus the annual fatality probability for an individual, P_f is:

$$P_f = 4500 \frac{\text{miles}}{\text{year-person}} \times 3.48 \times 10^{-8} \frac{\text{deaths}}{\text{mile}} = 1.6 \times 10^{-4} \frac{\text{deaths}}{\text{year-person}}.$$

The probability becomes more meaningful if we convert the probability of death to expected days lost assuming an average of 35 years lost for each death.

$$\text{Risk} = 1.6 \times 10^{-4} \frac{\text{deaths}}{\text{year}} \times \frac{70}{2} \frac{\text{years}}{\text{death}} \times \frac{365 \text{ days}}{\text{year}} = 2.04 \text{ days.}$$

Thus, on an average, there are four days of human life lost for each person driving to and from work for one year.

The risk can also be stated in economic terms assuming \$600,000 for the value of a life, the risk is $\$600,000 \times 1.6 \times 10^{-4}$ or \$96/year. In terms of productivity loss, the risk is one-half if we assume a lifetime salary of \$600,000 (\$20,000/year for 30 years):

$$\frac{2.04 \text{ days lost/year}}{365 \text{ days/year}} \times \$20,000/\text{year} \times \frac{30 \text{ year}}{70 \text{ years in lifetime}} = \$48/\text{year}.$$

Note that the 4.1 days lost/year is divided by 365 days because the days of life lost are not necessarily work days.

Conditional probability is the probability of a consequence conditioned on the probability of a prerequisite event. For example, the calculated

annual probability of an accident in the above example was 0.089/year. The conditional probability of a fatality is the probability of fatality if an accident occurs and is the number of fatalities divided by the number of accidents. With 52,600 deaths and 29,800,000 accidents, this probability is 52,600 divided by 29,800,000 or 0.001765 fatalities/accident which is one fatality for each 56.65 accidents.

To find the probability of a fatality from the conditional probability, multiply the first event (the accident) by the conditional probability of a fatality should the accident occur: 0.089 accidents/year x 0.00176 fatality/accidents = 1.6×10^{-4} /year, the same value as calculated directly.

6.3 Subjective Risk Estimate

The following example illustrates a process by which a subjective estimate of the annual probability of an accident fatality can be made if data are not available.

In a community of 30,000 to 40,000, one reads about a fatality several times each year. The number of fatalities is certainly not 50/year and its surely more than 1/year. A value between these extremes will provide the best estimate. The arithmetic average is (50 + 1) divided by 2 or 25.5. However, this value is a factor of 2 below the maximum value but is a factor of 25 greater than the minimum value. The geometric average will be an equal factor above and below the minimum and maximum values and will have the least chance of large error (a large factor above or below the actual value). To obtain the geometric average, the logarithm of the maximum and minimum are averaged and converted back to the geometric average:

$$\begin{aligned} \ln 50 &= 3.91 \\ \ln 1 &= 0 \\ \ln 50 + \ln 1 &= 3.91 \\ 1/2 (\ln 50 + \ln 1) &= 1.96 \\ \text{antilog of } 1.96 &= 7.07. \end{aligned}$$

This value of 7 deaths/year is 1/7 the maximum and 7 times the minimum is likely to be within a factor of 2 or 3 of the actual value.

More than half of the accidents seem to occur evenings or weekends, so we estimate two for driving-to-work fatalities. We estimate 1/4 of the city population (10,000) drive-to-work. Two fatalities out of 10,000 workers equals a probability of 2×10^{-4} /year, in good agreement to the statistical value of 1.6×10^{-4} /year. A similar process of looking at each piece of a problem will yield estimates of risk which are preferable to pure guesses or hunches when logical decisions are needed. As new information becomes available the subjective estimates can be modified. A rigorous method for adjusting probabilities giving weight to both the subjective estimate and new information is given by Bayes Theorem which is treated in most statistics textbooks.

6.4 Survey Methods

Frequently, information needed for a risk assessment can be obtained by using an employee questionnaire. For example, dispensary records for an ERDA contractor indicate that an average of nine toe injuries occurred per year. Even though the contractor operated a shoe store and permitted employees to purchase safety shoes at cost, many employees chose not to wear them. To determine the worth of the safety shoe program investment, the risk differential between those who wore safety shoes and those who did not had to be established. An employee questionnaire was selected as the means for obtaining the required data and was completed by about 300 employees. The analysis of the inputs revealed the following.

About 58% of the shopworkers wore safety shoes with an annual injury rate of 0.003 toe injuries per shopworker. Shopworkers not wearing safety shoes had an injury rate of 0.04. Nonshopworkers not wearing safety shoes had 0.003 toe injuries per year per employee, while no toe injuries were reported for nonshopworkers wearing safety shoes, indicating that toe injury risk for nonshopworkers was small, whether wearing safety shoes or not.

Based on an average value of \$150/toe injury, the annual toe injury risk for each shopworker wearing safety shoes is \$0.45, and is \$6.00 for those not wearing safety shoes. For the company as a whole, the toe injury risk was estimated to be \$1500/year. If no employee wore safety shoes, the company risk would be \$3000; but the risk would only be \$200 if everyone wore safety shoes. Management can now weigh the costs of a safety shoe program against these risk values. Other factors, such as OSHA requirements, will also influence the decision, but with such a survey the manager can get a feel for the economic return from his safety shoe investment.

A survey will usually require the services of a research statistician. Anyone conducting a survey should follow these basic rules:

1. Identify the problem.
2. Survey the literature.
3. Discuss the feasibility of the proposed survey with management. Explain the potential benefits of the survey to enhance your chance of obtaining management cooperation.
4. Define in clear specific terms the survey question(s) to be answered.
5. Develop the questionnaire with an instruction sheet. Explain the purpose of the questionnaire. Keep the questions as simple as possible. Put the instructions on a separate sheet rather than on the back of the questionnaire.
6. Search carefully for ambiguities. Test the questionnaire on a few representative individuals and revise as appropriate.
7. Request line management distribution and collection of the questionnaire.
8. Evaluate the data and draw conclusions.

9. Prepare report to management. Explain assumptions and methods in sufficient detail to permit rapid reevaluation if assumptions are changed.

6.5 Insurance Risk

Consider insurance as it relates to risk. A company insuring 100 cars for property damage will estimate the average loss per year as follows. On an average, 10 cars out of 100 will sustain property damage during a year. The average loss has been \$500. Thus, we can expect the loss during the year to be $\$500 \times 10$ or \$5000. To pay the \$5000, the insurance company must collect the \$5000 from the 100 car owners or \$50 each. The \$50 is the measure of risk for each car owner. If not insured, the car owner has a large (90% chance) of no loss and a small (10% chance) of a \$500 loss (a \$50 risk). The insurance company accepts the risk of all 100 owners so its risk is the sum of the individual owner risks or \$5000. The company risk will vary less than that of the individual. The individual has a 90% chance of no loss and 10% chance of a \$500 loss. The company has a 10% chance of paying out \$2500 and a 90% chance of paying out \$7500. In an average year the insurance company's payout will equal the \$5000 taken in to cover the risk. In actual practice, the insurance company will require a larger premium than \$500 (\$50 per car owner) to cover administrative costs, profit, etc.

The variance becomes smaller as the number of accidents becomes larger. The standard deviation is a measure of this variation and is the square root of the number of accidents. In the illustration above, the number of accidents each year would be $10 \pm \sqrt{10}$ or 10 ± 3.2 . The 90% level is $1.645 \sqrt{N}$ or 5, so the 10 to 90% confidence range is 5 to 15 accidents or \$2500 to \$7500 at \$500/accident. The standard deviation for 10,000 cases is only 1% of 10,000, compared to 10% for 100 cases and 32% for 10 cases.

6.6 Life Shortening Effects

For any specific causes of death, the probability of dying can be related to the average decrease in life expectancy.²⁰ For specific

hazards with a small probability of death compared to the natural death rate, the reduced life expectancy can be estimated by the following method (automobile accidents are used as an example):

1. Calculate the lifetime probability of death from a vehicle accident:

In 1981, there were 55,000 vehicle accident deaths of a total 1,900,000 deaths. Since 2.9% (55,000 divided by 1,900,000) of all deaths result from vehicle accidents and each individual is sure to die, the lifetime probability of a vehicle accident is 0.029.

This estimate can also be calculated as follows:

$$\frac{55,000 \text{ vehicle deaths/years}}{220,000,000 \text{ U.S. population}} = 2.5 \times 10^{-4} / \text{year.}$$

Multiplying the annual probability by 70 years (the average life-time) gives 0.0175. This value is in reasonable agreement. This calculation assumes there are an equal number of people in each age group and an equal probability of vehicle fatality at each age. Thus, the higher value of 0.025 is more correct. (For most risk estimates this degree of difference is not large.)

2. Estimate the average years lost by assuming the fatality occurred midway through life; i.e., multiply the lifetime death probability by 35 to obtain expected number of years lost:

$$0.025 \times 35 = 0.875 \text{ years reduced life expectancy.}$$

For motor vehicles, this estimate of reduced life expectancy would be somewhat low because younger males have a higher probability of a motor vehicle fatality. A more accurate estimate could be made by calculating reduced life expectancy for specific age groups (male and female separately). For most risk assessments, such accuracy is not necessary, and the data are not usually available.

Stating fatality consequences in terms of reduced life expectancy, rather than in terms of lives lost, is more appropriate for hazards such as smoking or air pollution which kill through increased cancer or lung and heart disease. These effects are more pronounced in the later years of life, so that the number of deaths does not give a clear picture. Many who smoke heavily will die earlier than if they did not smoke. Thus, a published estimate of the number of deaths per year from smoking is not meaningful unless the basis for the estimate is also given. For example, were only deaths caused by diseases closely associated with cigarette smoking (lung cancer) counted? Was it a statistical estimate of the number of early or premature deaths based on a study of age of death differences between smokers and nonsmokers?

If the estimation is the number of equivalent lives lost based on 70 years per life, what was the average number of years lost per early death? It is obviously not an average of 35 years (midway through life as assumed for accidents).

One percent of the population dying 35 years early and 35% of the population dying 1 year early is different; but each result in the same equivalent number of "lives lost" or in the same reduced life expectancy. There is no a priori "correct" way to measure the risk of health effects or loss of life. It is important, however, that comparisons of different risks involving loss of life be made on a conservative basis, and that the methods and assumptions be specifically stated in the risk assessment. Departures from the norm and any assumptions should be described qualitatively and, as practical, quantitatively.

6.7 Trend Analysis

The measurement of safety performance is essential to evaluation of new techniques or methods of reducing risk, such as proposed in the preceding section of this guide. Discussion of a comprehensive safety information system, which would provide good indicators of future safety performance, is included in another SSDC document.¹⁹ Of concern here is only the analysis of property damage and injury data in order to determine whether a trend or change in accident rates is statistically significant.

In a period of inflation and/or company growth, total accident costs will also increase unless there is a compensating decrease in accident rates. Frequency and severity rates are indicators of performance rather than total loss. Also, cents loss per \$100 property valuation is a method of measuring performance. However, factors such as number of employees (employee-hours worked), inflation (dollar value), and vehicle miles can be applied to loss data in order to determine whether an increase or decrease in accident costs (or frequency, severity, property loss rates) is significant or whether it could have happened as a result of random variation.

For example, during 1975, an ERDA (now DOE) contractor experienced large increases in 1975 over 1974 in-plant property damage, vehicle damage, and the number of injuries. Were these increases significant, or could they have happened by chance? From 1971 through 1975, the number of employees (and hence employee-hours) increased by 20%, and the number of vehicle miles increased by 32%. In addition, inflation increased by 35%. In order to determine if there was a significant trend in loss experience, these data (inflation, employee-hours, and vehicle miles) were determined and compared for each year, 1971 through 1975.

The number of property damage incidents and the number of injuries were normalized to employee-hours, while vehicle accidents were normalized to mileage. The cost per incident of both vehicle and property damage was normalized to the cost of living index. Public liability and fires were not analyzed since the experience to date has been so limited that any attempt to predict a trend would be meaningless. The normalized vehicle and property damage costs were obtained by multiplying the normalized cost per incident by the normalized number of incidents.

The normalizing factors and the normalized accident data are given in Table 2. As can be seen, all normalized 1975 values are within or close to one standard deviation. (The number of vehicle incidents is low by 1.1, but this is not significant because this value would occur one out of four times by chance.) The overall data show a general random character with no discernible trend.

TABLE 2. SAFETY PERFORMANCE ANALYSIS

	Normalizing Factors						
	1971	1972	1973	1974	1975		
Inflation	1.0	1.03	1.10	1.22	1.34		
Man-hours x 10 ⁶	4.1	3.9	4.0	4.2	4.8		
Vehicle miles x 10 ⁶	6.3	7.5	6.7	6.4	8.3		
	Accident Data						
	Normalized					Actual	
	1971	1972	1973	1974	1975	1975	Mean 1
Property damage							
Number of incidents	41	56	33	39	40	47	42 ± 8
Cost per incident	489	1190	1241	425	1311	1757	931 ± 435
Total cost x 10 ³ \$	20	65	41	17	52	83	39 ± 20
Vehicle damage							
Number of incidents	54	38	46	35	31	41	41 ± 9
Cost per incident	170	161	97	97	154	209	132 ± 42
Total cost x 10 ³ \$	9.2	6.1	4.5	2.7	4.9	8.6	5.4 ± 1.7
Number of injuries	542	607	677	635	624	730	617 ± 30

The only apparent normalizing factors for vehicle accidents and personnel injuries are mileage and employee-hours worked. For property damage, however, other factors than employee-hours (for the number of incidents) and inflation (for the average cost per accident) can be used. The total budget and the total property value at risk could also be used as factors for property damage. In Figure 2, property value, company budget, and inflation times employee-hours for 1971 through 1976 were plotted.

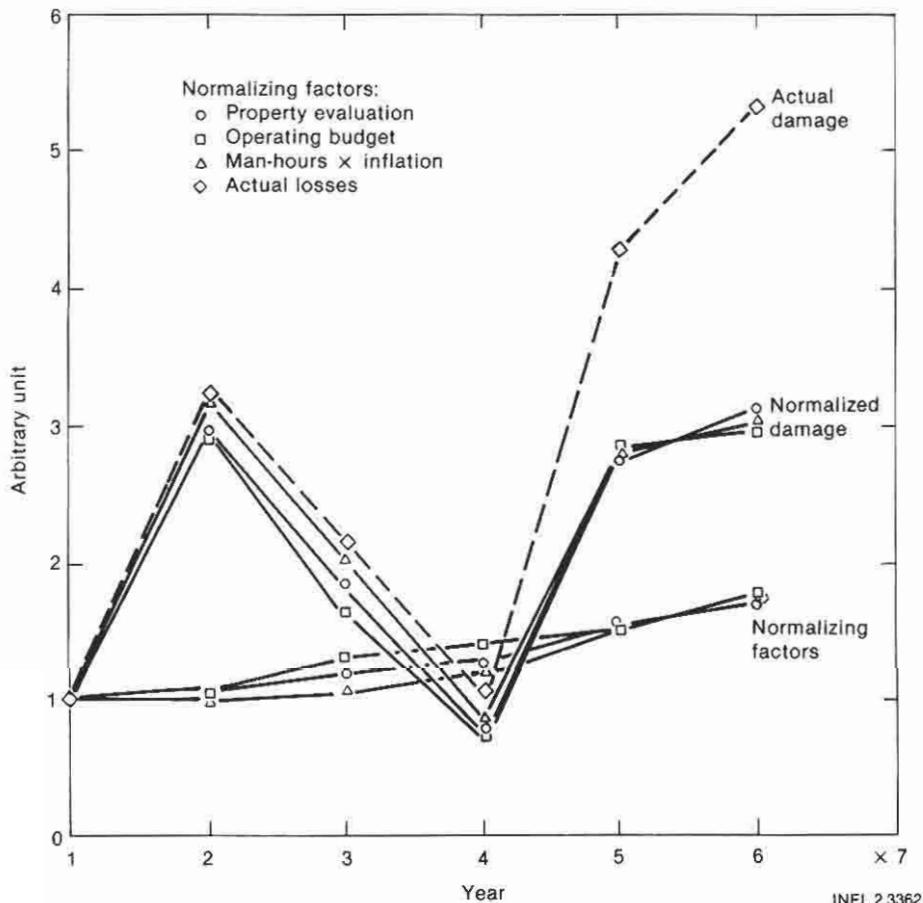


Figure 2. Comparison of various normalizing applied to annual direct property damage costs.

These three factors were applied to actual property loss, and the resulting normalized property damages were also plotted. In addition, all data normalized to a value of one for 1971 to permit direct comparison of the three factors yield remarkably similar results. (The budget and property value inherently include inflation. Employee-hours, however, must be multiplied by inflation. The reason for this is employee hours times inflation provides a constant measure of company property investment for a given plant, assuming wages are not rising faster or slower than property value.)

The various normalizing factors appear to provide consistent results. It is concluded that any of the above normalizing factors are acceptable, and the cents loss per \$100 property value is a good performance indicator.

6.8 Log-Normal Distribution

The simplest, first cut, method of assessing the average cost of an accident is to divide the total cost of accidents (of particular type) by the number of accidents.

The result, the total cost of last year's accidents, is an estimate of next year's risk; but only if the very costly or catastrophic accident is properly represented in the accidents of the previous year. This simple approach can lead to underestimating the risk. First, all of the consequences may not have been considered or represented in the experience data. With vehicles, which we are using for an example, the property damage may be adequately represented but the number of injuries or fatalities are likely to be few or none with a small data base of vehicle accidents. In such cases, the average number of fatalities can be estimated from statistical data: the injury and/or death rate per million vehicle miles for all DOE can be used to calculate the injury per death risk for a specific field office or contractor. However, due to peculiarities of a particular contractor's operation or climate conditions, such estimates must be made with caution.

The average cost can similarly be estimated from a larger data base. The average loss per event for a given contractor should be compared to the

average loss per event for all DOE contractors. If the difference is significant, then the risk analyst must determine whether the specific contractor's average loss per event is likely to be different or whether the single contractor's experience is inadequate to determine an average. A quick review of the data may reveal the answer. If all of the accidents have been minor (frequently the situation with few accidents), the long-term average accident cost is likely to be higher than the cost based on recent past experience. A quite unlikely, but possible, opposite situation is that only one accident occurred but the car was "totaled." The long-term average accident cost will be less. The second problem closely related to the first is that the frequency-severity relationships of the same type of accident may vary considerably from one contractor to another. That is, the probability of death based on the number of injuries may vary greatly. It is certainly not the same for a high rise steel worker and an office worker. The probability of a catastrophic fire may not be directly proportional to the number of waste basket fires. Factors such as types of construction, absence of sprinkler systems, etc., may have a greater bearing on the probability of a major fire and how it will affect the average cost of fires.

Assessing risk using the average cost of past accidents will usually result in a gross underestimation of risk. Accident patterns are random events with a log-normal or skewed distribution. For a linear-normal distribution, the average cost of an event is also the cost which occurs most frequently. For accidents, the average cost is usually larger than the most frequently occurring cost. This is because the very large or catastrophic accident may (and frequently does) constitute a significant portion of the risk event though none have occurred recently.

Analysis of the frequency-severity patterns of previous accidents using various types of probability graph paper can predict the probability of large accidents, determine the average cost of accidents, and provide a better risk picture.

The log-normal graph paper has a probability scale which converts the probability curve to a straight line. The consequence (cost) scale is

logarithmic. Plotting the cost and frequency of accidents on this graph paper usually approximates a straight line. The slope of the line is determined by the degree of skewness in the probability-cost (frequency-severity) curve. The slope (skewness) is determined by the relative frequency of large accidents compared to the frequency of smaller accidents. Thus, a steep slope indicates a high probability of a large accident. In addition because the plotted data are linear, extrapolations are fairly accurate. The extrapolated curve represents the entire potential frequency-severity distribution. The probability of an accident of any given severity and the average cost of all accidents can be determined, and the straight line on the log-normal graph paper can be transferred to log-log graph paper. The curve on log-log graph which can be used to provide additional insight to the frequency-severity and how it relates to risk.

The following exercise illustrates this procedure: Given in Table 3 are electrical property damage data over a 5-year period for a DOE contractor.

The 22 accidents total \$66,220, or an average of \$3,010 per accident. The 5-year average of 4.4 accidents/year indicates a mean annual direct property loss of \$13,200. However, one accident costs \$49,000 of the

TABLE 3. ELECTRICAL PROPERTY DAMAGE

Cost Range (\$)	Number of Accidents	Cumulative Number (N_i)	Percentage of Accidents Less than Cost $\{[N_i \div (N + 1)] \times 100\}$
10 to 100	9	9	39.0
100 to 500	8	17	74.0
500 to 1000	2	19	83.0
1000 to 5000	1	20	86.9
5000 to 10,000	1	21	91.3
10,000 to 50,000	1	22	95.6

N_i = The cumulative number of accidents through a given cost range.

N = The total number of accidents in the study.

\$66,000 total for all 22 accidents. The basic question is: What is the expected annual loss from electrical property damage? Also, did the \$49,000 accident result from unique causes or from causes which are typical of other electrical accidents, so that a control system weakness is indicated?

The analysis is accomplished as follows (refer to Table 3 and Figure 3):

1. Rank all accidents from the smallest to the largest in increasing order of dollar cost.
2. Select cost-range intervals in such a way that data points are approximately equally spaced on the probability paper.

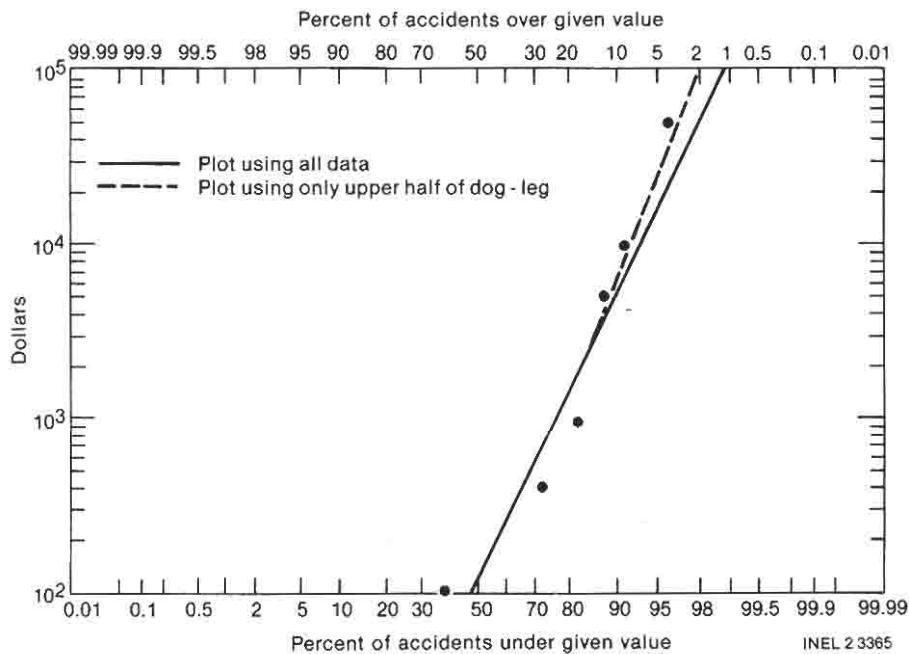


Figure 3. Log-normal plot of electrical property damage data from Table 1.

3. List the number of accidents for each cost range, as well as the cumulative number of accidents to that point.
4. Calculate the cumulative percentage for each cost range (as in Table 3) and plot the cost versus the cumulative percentage on log-normal paper (as in Figure 3). The cumulative percentage is calculated by dividing the cumulative number (N_i) in each cost range by the total number of events plus one, and multiplying the result by 100. (We divide by $N + 1$ because "N" points divide a line into $N + 1$ segments.)

The following guides explain how to interpret the resulting curve.

1. If the accident is an outlier, the accident is probably unique, and change analysis¹⁰ to determine the need for a specific fix for the unique accident is indicated. On the other hand, if the slope is steep (high probability of a large accident and the \$49,000 accident fits the rest of the data, as in this example) the \$49,000 is typical of the system. If the indicated frequency is unacceptable, systems analysis is required to prevent reoccurrence.
2. If the slope increases with cost (curves upward), the system may be diverging out of control. Careful attention to cause of increase is critical.
3. If the plotted line approximates a dog-leg with the two segments fairly straight, two different basic causes are probably involved. For example, the curve for small fires which are extinguished may not have the same slope as the curve for large fires beyond the control of fire fighting capabilities. Another example is that planned radiation exposures should not have the same characteristics as unplanned large exposures.
4. If the plotted line flattens out at the top (smaller slope), there may be a limit to the consequence. For example, the maximum

damage to a vehicle equals the value of the vehicle, and thus the single vehicle accident data will approach this value asymptotically as long as we consider only vehicle property damage.

5. If the plotted line flattens at the bottom, there is probably a reporting or natural minimum. For example, a curve of all accidents greater than \$50,000, as reported by DOE contractors, flattens out and approaches the \$50,000 consequence level asymptotically. Thus, in the \$50,000 to \$100,000 accident range, the frequency-severity slope is distorted.
6. The frequency of occurrence (or return period in extreme value language) for any given size accident can be calculated. For example, in Figure 3 about 1.3% of electrical accidents will cost more than \$100,000. Since 4.4 accidents have occurred each year, a greater than \$100,000 accident is expected about once every 17 or 18 years [$1/(0.013 \times 4.4)$]. Note that the data points (circles) in Figure 3 exhibit a slight dog-leg characteristic. If a straight line is fitted to the upper half of the data, the probability of a greater than \$100,000 accident increases to 2.3%, with a resulting expected frequency of every 7.5 to 11 years. [As will be seen in the example for extreme value analysis (discussed in the next section), which uses only the maximum accident in each of the 5 years, the extreme value analysis predicts a return period of 10 years. Thus, the upper half line is probably more accurate. However, one should not be concerned with this difference; the return periods are within a factor of 2, which indicates good agreement for the limited amount of data.]

Additional information from the data can be obtained by transferring the log-normal curve (Figure 3) to log-log paper (Figure 4). In making the transfer, use the "percent over" scale above Figure 3. Select the points to transfer from the curve, not the circled data points. For example, the first (x, y) point to transfer, (51%, \$100) on Figure 3 becomes 0.49 on the y axis and \$100 on the x -axis. The second point on Figure 3 ($x = 20\%$) $y = \$1000$ becomes 0.20 on y -axis and \$1000 on the x -axis in Figure 3.

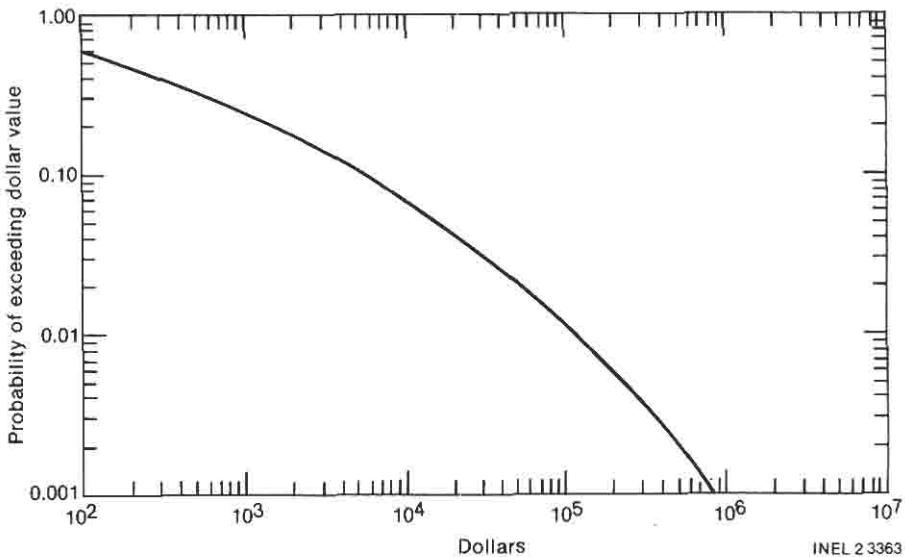


Figure 4. Log-log plot of log-normal curve given in Figure 3.

Once the curve is transferred to the log-log curve, much can be learned from visual inspection.

1. For example, in Figure 4 the slope becomes greater than one just beyond the $\$10^5$ consequence value. This is the cost range of maximum risk. (At a negative 45 degree slope minus one, the probability times the consequence is a maximum.) This line of balance concept is discussed in more detail in Appendix B.
2. The slope will approach infinity (vertical) as it nears the maximum potential consequence.

Additional examples of the log-normal and log-log data analyses are given in Appendix B.

The area under the log-log curve represents the cumulative risk. The risk can be grossly approximated as follows. The average cost of accidents in the range of \$100 to \$1000 appears to be about \$300. The probability of

an accident in this cost range (taken from the curve in Figure 4) given an accident is about 0.6 - 0.25 or 0.35. The risk is the consequence times the probability or about \$105. Repeating this process for each cost range and summing gives the average cost of an accident. This average is more accurate than the total cost of accidents divided by the number of accidents, because it gives proportionate representation to the large accident which occurs too infrequently to be accurately represented in experience data.

A more accurate estimate can be obtained by integrating the curve as follows (refer back to Figure 3).

1. Approximate the curve with a series of straight lines. In this case, a straight line for each decade introduces only a small error.
2. Integrate each line separately and sum the results. The general equation of a straight line on log-log paper is:

$$cf^A = C_i f_i^A$$

$$C = C_i f_i^A f^{-A}$$

where

C = cost of the accident

f = frequency or probability of the accident

A = the slope of the line which is a constant for a straight line.

NOTE: For any type of accident, for a given cost, there is a different probability of occurrence. Generally, as the cost increases, the probability decreases. We can choose to integrate over frequency or cost, depending on whether we wish to know the

risk ($\sum c_i f_i$) over a given frequency interval or cost interval. In the example below, the risk is calculated over the cost range of \$100 to \$1000.

A, the slope is calculated as follows:

$$c_1 f_1^A = c_2 f_2^A$$

$$c_2/c_1 = (f_1/f_2)^A$$

$$\ln(c_2/c_1) = A \ln(f_1/f_2)$$

$$A = \ln(c_2/c_1)/\ln(f_1/f_2)$$

The area under the curve is the risk and is equal to:

$$\text{Risk cost} = \text{Area} = \int_{f_1}^{f_2} Cdf = c_1 f_1^A \int_{f_1}^{f_2} f^{-A} df$$

$$\text{Area} = c_1 f_1^A \frac{f_1^{1-A} - f_2^{1-A}}{1-A} \Big|_{f_1}^{f_2} \quad A \neq 1$$

$$\text{Area} = c_1 f_1^A \frac{f_1^{1-A} - f_2^{1-A}}{1-A}$$

$$\text{Area} = c_1 f_1^A \ln f \Big|_{f_1}^{f_2} \quad A = 1$$

$$= c_1 f_1^A (\ln f_2 - \ln f_1) = c_1 f_1^A \ln f_2/f_1$$

In the example, the values of f and C for the cost range of \$100 to \$1000 are:

$$C_1 = \$100$$

$$f_1 = 0.6$$

$$C_2 = \$1000$$

$$f_2 = 0.25$$

and

$$A = \ln(C_2/C_1) \div \ln(f_1/f_2)$$

$$A = \ln(1000/100) \div \ln(0.6/0.25) = 2.3/0.88 = 2.63$$

$$\begin{aligned} \text{Area} &= C_1 f_1^{2.63} \left(\frac{f_1^{1-A} - f_2^{1-A}}{1 - A} \right) = 100 \times 0.6^{2.63} \frac{0.6^{-1.63} - 0.25^{-1.63}}{-1.63} \\ &= 26 \frac{2.30 - 9.50}{-1.63} = \$116 . \end{aligned}$$

The risk for accidents ranging in cost from \$100 to \$1000 is \$116.

Integrating each of the cost ranges in like manner gives the risk values tested in Table 4.

TABLE 4. ELECTRICAL PROPERTY DAMAGE RISK

Cost Range (\$)	Risk (\$)
10 ² to 10 ³	115
10 ³ to 10 ⁴	550
10 ⁴ to 10 ⁵	1640
10 ⁵ to 10 ⁶	2500
10 ⁶ to 10 ⁷	1300
Total	6105

The mean of actual losses over the 5-year period is \$66,219 divided by 22 or \$3010. The average cost of accidents, from the integration which includes the large accident which has not yet happened, is about \$6100. The \$6100 integrated value is the better estimate of the mean. The average number of accidents is 22 divided by 5 or 4.5/year. (The frequency is not affected by the size of the accident and can be calculated directly without adjustment.)

The integration was terminated at \$10,000,000 in Table 4. At this point, the slope is quite steep and the risk in the next decade would be only \$100 or \$200. In addition, the data have already been extrapolated from \$50,000, so the uncertainty is very large.

Judging from the tabulated risk values for each cost range, the cost range of maximum risk is \$100,000 to \$1,000,000. This is in agreement with the conclusion that the cost range of greatest risk is where the slope is equal to one, just beyond the \$100,000 consequence value.

For any accident type, if a sufficiently large number of accidents have been experienced, the integrated mean will be very nearly equal to the mean calculated directly from experience data. For example, a 5-year average cost of passenger vehicle property damage for one contractor was \$175. Going through this exercise for the more than 100 accidents experienced yielded an integrated mean of \$174.

In summary, plotting accident data and integrating the resulting frequency-consequence curves determine the cost range of maximum risk and provide approximate risk values, which are more accurate than simple projection of last year's losses. Appendix B provides additional detail for those interested in risk projection techniques.

6.9 Extreme Value Analysis

Extreme value analysis is discussed briefly in the MORT text,¹¹ the Accident/Incident Investigation Manual,¹² and in detail by Gumbel.² The maximum events taken from each of a large number of time intervals form

a special frequency-severity relationship which is somewhat similar to the log-normal distribution pattern. The scale on the extreme value graph paper converts this skewed, bell-shaped curve to a straight line on the x-axis. Also on the x-axis opposite the frequency (cumulative probability) scale is a scale which converts the frequency to "return periods." This return period is in the same time units as the time intervals from which the maximum events were taken. The y-axis represents the severity or cost. The average time between events (return period) can be read directly from the graph paper for maximum events of any given severity or cost. The value lies in the ability to extrapolate or extend the straight line to include longer time periods, and thus predict the occurrence of very large accidents.

There are two kinds of extreme value graph paper. On one paper, the cost scale is linear; on the other the cost scale is logarithmic. The reason is that for events resulting from multiple independent causes, the cost increases linearly with respect to the cumulative probability scale. For those events resulting from multiple interdependent causes, the cost increase exponentially or logarithmically. This indicates a possibility of a common cause, such as some factor in the management system, leading to the several causes of the large accident.

The extreme value equation is an empirical derivation of the frequency and severity of maximum events represented on the upper tail of the log-normal curve. The basic difference between extreme value and log-normal analysis is that all events are used in log-normal analysis, but only the maximum events are used in extreme value analysis. Extreme value analysis is generally preferred because: (a) maximum event data are frequently available when a record of events is incomplete; (b) it is quicker and easier to use in that less data is involved, and the return period can be read directly from the graph paper; (c) no judgment is required in plotting the data, whereas judgment is required in selecting cost intervals in log-normal analysis.

The log-normal and log-log analyses are required to determine the relative risk of small versus large accidents and to determine the average cost of potential accidents.

The maximum accident in each of the 5 years for the 22 electrical property damage accidents discussed earlier follows, and is plotted on extreme value paper (Figure 5).

<u>Year</u>	<u>Cost (\$)</u>
1	593
2	8,883
3	707
4	3,800
5	49,700

The data points are calculated as follows:

1. Select a time increment or period (in this case 1 year has been selected, but the time increment may be a day, week, or any other appropriate unit).

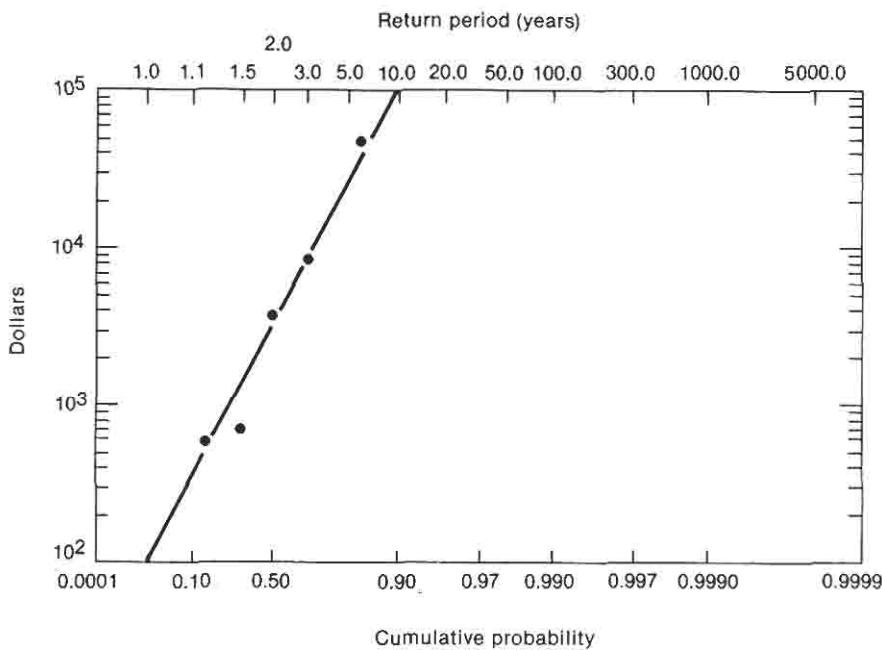


Figure 5. Logarithmic extreme--value electrical property damage (same accident data as plotted in Figures 3 and 4).

2. Select the maximum loss event for each time period and rank the selected events in order of increasing cost.
3. Select a vertical scale which will permit extrapolation to desired consequence levels (generally two to three times the maximum value).
4. Calculate the cumulative probability by dividing N_j by $N + 1$. Note that the five data points (one for each of five years) divide the cumulative probability scale into $N + 1$ or six intervals. The results data are given below:

<u>Cost (\$)</u>	<u>N_j</u>	<u>$(N_j/N + 1) \times 100$</u>
593	1	16.7
707	2	33.3
3,800	3	50.0
8,883	4	66.7
49,700	5	83.3

5. Plot the data on both logarithmic and linear extreme value papers and use the paper that gives the best straight-line fit. If a straight-line relationship does not occur in either case, analyze the data for homogeneity--first by scanning, then by formal change analysis.¹¹

If the plot approximates a straight line on linear graph paper, the accidents are likely to be independent and the prevention of extremely large accidents is well under control. If the plot approximates a straight line on logarithmic paper, the multiple causes can usually be traced back to a common source, or one cause influences another. Since a strong systems control program would eliminate common causes, review of the control or safety system may be in order. This is especially true if the slope of the logarithmic curve is steep so that the return period for a large accident is short.

Corrections for "number of employees" or company growth should not be made on the consequence scale. It is not reasonable to expect that "halving" the size of an activity would "halve" the severity of the most severe event. A more reasonable method for making such corrections is through the return period. That is to say, if one has data from an existing unit and adds an identical unit to the system, one would expect to halve the return period (the time when that consequence event is expected to recur). This is the same as normalizing the raw accident data to the number of employees.

From Figure 5, a return period of 10 years is found for a \$100,000 accident. This is considered to be good agreement for the 17-year period using all 22 accidents on log-normal paper. As discussed earlier, the log-normal curve exhibits a dog-leg characteristic. Fitting a straight line to only the upper segments will yield a return period of 10 years.

This illustrates that the upper segment of a dog-leg probability curve predicts more accurately the frequency of large consequence events, and that 10 years is the correct return period. However, until the reason for the dog-leg is identified, it should be assumed the return period lies in the range of 10 to 17 years. Other examples of extreme value projection are given in Appendix B.

6.10 Fault Tree Analysis and Other Hazard Identification and Evaluation Techniques

Prepared by P. L. Clemens, Sverdrup Technology Inc., Arnold Air Force Station, Tennessee, also available from the System Safety Development Center is a Compendium of Hazard Identification and Evaluation Techniques for System Safety Application. This document provides abstracts of 25 different techniques. Described are method, application, thoroughness, mastery required, and difficulty of application with comments provided.

Most of these techniques are useful in hazard identification and systems analysis but do not provide a measure of risk. One that does, fault tree analysis, identifies one undesirable event and the contributing elements (faults/conditions) that are required to precipitate the undesired

event. These events are arranged in a logic tree and the probability of the top (undesired) event is calculated using network paths through Boolean Logic gates. A Reliability and Fault Tree Analysis Guide, SSDC-22 is available.¹³

Fault Tree Analysis is time consuming and costly. A major effort is required to prepare and analyze a tree for a single event so that use of this technique is limited to specific high risk events. A limitation of this technique is that there are inadequate means to assure completeness. It is impossible to include or arrive at all possible combinations of events leading to the undesired events (or in other words all possible ways the undesired event can happen). Of great value is that the sensitivity of the top event to a particular component failure can be determined and redundancy or other means taken to make the component failure less critical. Also, fault tree analysis greatly increases the understanding of a system and how the various components interact, especially during failure.

7. CONSEQUENCE ANALYSIS

The identification of accident consequences and quantification of the associated risk serves two purposes. The identification permits control measures to be designed and applied. The risk quantification helps determine whether the risk is serious enough to warrant control measures. (The risk reduction should be balanced against the costs of additional control measures.) In the accident cost analysis, there are two dimensions that must be considered:

1. Types of consequences (property, environment, and human)
2. Indirect costs (lost time, administrative, legal, and replacement of services or products).

Both dimensions must be considered if the risk assessment is to be complete and various types of consequences are to be treated equally on a systematic basis.

The simplest way to treat the various types of consequences is to simply assign units to each of the identified consequences. Monetary units (dollars) have the distinct advantage of permitting a direct comparison of the dollar cost of risk reduction against the expected loss. These are terms a program manager can easily understand. A disadvantage is there may be a negative reaction when placing a dollar value on human life or environmental values.

It may be argued that assigning a cost for a fatality is not placing a value on life but is placing a standard value to help allocate safety resources efficiently so that loss of life may be minimized. Nevertheless, there will still be those who object. If dollar values are used, they should be stated in the risk document in such a way that the risk values can be easily revised based on different dollar values for intangible effects.

Merely assigning arbitrary units to intangible effects results in risk values expressed in arbitrary units which are useful only for comparing one

type of risk to another. The arbitrary units may be summed and then converted to dollars, but this is a thinly disguised method of assigning dollar values.

Another method of dealing with the problem is to classify consequences by ranges of severity such as A, B, or C events; with A being the worst accident to occur and C the least severe. The following values have been proposed:⁵

1. Loss of Life

- A event--30 fatalities
- B event--5 to 29 fatalities
- C event--1 to 4 fatalities

2. Environmental Pollution

- A event--10 million tons oil spilled
- B event--100,000 to 10 million tons oil spilled
- C event--1,000 to 100,000 tons oil spilled

3. Property Loss

- A event--2 billion dollars
- B event--20 million to 2 billion dollars
- C event--200 thousand to 20 million dollars.

By equating A, B, and C events, equivalent dollar values can be obtained for oil spills and fatalities (human life). In doing so, an inconsistency will be obvious. The dollar value per life is \$67 million for B events and \$200 thousand to \$4 million for C events.

To avoid assigning dollars to each of these three types of losses, it has been suggested¹⁴ risks be presented in all three dimensions (human, environment, and property) all the way to the decisionmakers.

The risks from these three categories can be ranked by Matrix analysis as pictured in Figure 6.

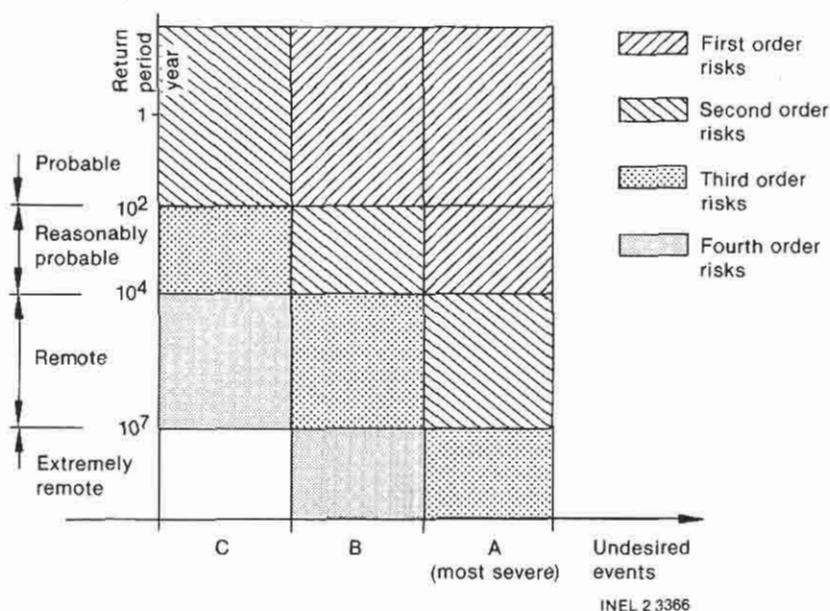


Figure 6. Matrix risk ranking.

The combination of frequency and severity thus determines the degree or ranking of risk. If desired, this method can be used with other scales. The method does provide priorities for resource allocation. Minor human risks are not given precedence over major economic risks, and use of the system also eliminates personal bias and prejudice. However, there are several disadvantages. The risks cannot be compared or summed as there are no common units. For example, the question "Is the risk small compared to the potential benefit?" cannot be answered directly in the case of human or environmental risk. Another disadvantage is that assigning a fixed amount of pollution to each severity classification does not adequately quantify the consequence. How valuable is the contaminated area? A spill in one area may have far greater consequences than a spill elsewhere. A final comment is that risk aversion is implicit in the event--severity classification. That is, 10 fatalities associated with one event have more than

10 times the consequence of one fatality; thus, relatively more should be done to protect groups than individuals. From the individual's viewpoint, it may make no difference whether others are killed with him; and thus, the risk should be linear. From an organizational viewpoint, risk aversion should be based on its impact on the organization. While many questions have been raised, no answers are provided in this document. These questions should be addressed in the formulation of a risk management program when establishing goals and acceptable levels of risk. It may be that flexibility is desired in that dollar values for environmental effects or human life are acceptable in some situations, but not others. Consideration of risk aversion and/or discounting future loss/benefits to present values may be desirable at times. In any event, the consequence and associated unit of risk should be explicitly stated in a risk assessment document.

7.1 Direct and Indirect Accident Costs

Direct and indirect costs should be assessed in light of the consequences discussed above. For a specific hazard, the various possible consequences and their costs could be evaluated on an individual basis. Costs for cropland destroyed could be based on the annual value of the crops or the market price of the land. Pollution could be measured by economic losses due to sickness or illness, plus a factor for human misery associated with illness.

Much time can be saved if only direct costs are considered and a multiplying factor is used for the indirect costs. This has an averaging effect which is not detrimental because while indirect costs may vary greatly, the probable value is, by definition, equal to the average indirect cost.

There is no logical basis for separating accident consequences into "direct" and "indirect" costs. However, a number of studies^{15,16,17} have been made to determine the "direct/indirect" or the "visible/hidden" accident cost ratio. Usually, the purpose is to determine the total to accident cost. For this same reason, we have labeled accident consequences which are normally reported as "direct" costs, although this is strictly a

convention used to arrive at the total risk without recording all of the consequences for every accident. (A more accurate description of direct/indirect costs might be "reported" and "not reported.") Each company should determine their own direct and indirect cost ratio, since both definitions and actual ratios will vary.

Accident losses normally reported are:

1. The amount of first aid, medical treatment, workdays lost, and fatality cases
2. The number of and dollar damage amount for vehicle accidents, property damage, and fires.

The balance of this chapter discusses methods for determining the direct and indirect accident costs starting from those losses normally reported. The actual data, calculations, and results of a DOE contractor experience are presented for illustrative purposes. The cost of each type of loss and the number of accidents will vary with individual organizations and with inflation. (Each organization should evaluate their own costs and numbers of accidents.) A summary of these data is given in Table 5. As can be seen, the direct cost is obtained by multiplying the average cost by the number of accidents. The values in Table 5 are not the same as would be obtained from adding the costs of all accidents because: (a) there are no "reported" cost values for injuries and fatalities, (b) fractional number of cases, such as 0.06 fatalities, are estimated (none had occurred), and (c) the average cost of property loss was estimated from log-normal analysis of each type of property loss accident, as explained earlier in this chapter.

The indirect cost multiple is the sum of the direct cost and the indirect cost divided by the direct cost. Multiplying the average direct cost per accident by this multiplier gives the total cost per accident, or (as in the table) multiplying the direct annual cost of a class of accidents gives the total annual cost of that class of accidents. These multipliers should not change with inflation and may be used in lieu of making indirect

TABLE 5. DIRECT AND INDIRECT ACCIDENT COSTS

Accident Category	Average Direct Cost/Accident (\$)	Number of Cases/Year	Direct Annual Dollar Cost (risk x 10 ³)	Indirect Cost Multiplier	Total Cost x 10 ³ (\$)
Injury					
First-aid, on-site medical	10	1,140	11.4	4.0	45.6
Off-site medical	60	98	5.9	7.5	44.3
Workdays lost	3,500	12	42.0	2.0	84.0
Fatality	100,000	<u>0.06</u>	<u>6.0^a</u>	2.0	<u>12.0</u>
Total injury	--	1,240.06	65.3	--	186.0
Property					
In-plant \$1000	6,000	8	48.0	2	96.0
In-plant \$1000	250	30	7.5	2	15.0
Vehicle	210	47	10.0	4	40.0
Fire	210	<u>6</u>	<u>1.3</u>	2	<u>2.6</u>
Total property	--	91	66.8	--	154.0
Total of injury and property	--	1,331	132.0	--	340.0

a. Theoretical value only; no death recorded.

cost studies for the accident classes considered in Table 5 for your own organization. They are believed to be conservative in that only those indirect factors which were identified and evaluated are included in Table 5.

7.2 Direct Accident Costs

7.2.1 First Aid, On-Site Medical

The cost for first aid and on-site medical treatment was estimated by the medical director to be \$10/case. (On-Site medical treatment costs are closer to first aid costs primarily because of travel time. In addition, off-site cases are paid for through workmen's compensation as they are treated by a "noncompany" doctor.)

7.2.2 Off-Site Medical

For the DOE contractor, off-site medical treatment paid through workmen's compensation costs \$60/case, averaged over a several year period.

7.2.3 Workdays Lost

Costs which appear in the company's financial records are the medical costs and the injured employee's wages paid through workmen's compensation. It is noted that the part of the employee's wages paid through sick leave benefits will vary widely from one state to the next. If sick leave benefits become exhausted, part of the loss may be borne by the injured employee. For this report, direct costs for lost time injuries are defined as medical costs plus wages paid to the employee from both sick leave and workmen's compensation. This was assumed to average \$1200 for medical costs and \$2300 for wages, or a total direct cost of \$3500/accident. (This assumption is based on the contractor's experience. Since this contractor has a severity rate much lower than the average DOE contractor, each contractor should determine their cost.)

7.2.4 Fatality Costs

The dollar value for a fatality is very subjective. A number of papers have been written on this subject.^{15,16,18} Estimates have been based on:

1. Future earnings
2. Propensity or willingness to accept risk in return for some benefit
3. Wage differential for high/low risk operations
4. Direct company costs, such as death benefits (insurance and replacement expenses).

All such analyses contain serious deficiencies. There is no consensus on whether the losses should be determined from a society, company, or individual viewpoint. In addition, it is the company which accepts the residual risk, but the majority of the losses are suffered by the injured employee and his dependents. The loss to society is usually considered fairly insignificant for a single individual, since the average person consumes much of what he produces. However, the economic cost to society resulting from the death of an employee with many dependents may be very large. Attempting to determine the dollar value of a human life from voluntary risk acceptance patterns is not practical, because individual perceptions of risk acceptability are extremely variable and it is difficult to determine whether the occupational risks are really understood by the worker (and, consequently, whether they are voluntary or involuntary).

Within a company, the costs for the loss of an employee also vary greatly. It would undoubtedly have a much greater effect on any company to lose a key scientist or executive than an average employee. For specific risk analysis where only management personnel are at risk, such as corporate aircraft travel hazards, a million dollars per life may not be too high. Yet, \$100,000 for the average employee is believed to be higher than the actual dollar costs to a contractor. Most of the studies referenced

previously arrive at a dollar value for human life in the range of \$100,000 to \$500,000. While it is evident that values in this range cannot be justified for the average employee solely on a company economic basis, most of those who have considered the problem believe that economics should not be the sole basis. Currently, there are large differences in explicit or implicit values applied by different regulatory agencies and individual analysts. The most valid basis appears to be a consensus of informed employers and workers who must bear the risk and pay for risk control. Unfortunately, such a consensus is not available. Nevertheless, if a consistent value is used by all DOE contractors, risk studies could be compared on a consistent, relative basis.

A fatality cost of \$200,000 was arbitrarily selected for use in this guide. (For direct/indirect cost analysis, \$100,000 is assigned to direct costs and \$100,000 to indirect costs.) Lifetime earnings could be used, which would result in, perhaps, \$500,000 for computational purposes.

While it may seem distasteful to place a dollar value on a human life, no other scale appears adequate to provide a direct comparison of the risks and the resources allocated to risk reduction. In addition, the use of almost any yardstick (even placing a dollar value on human life) seems preferable to the usual practice of allocating finite resources to lifesaving measures on a completely arbitrary and subjective basis.

The contractor being used as an example had never experienced an occupational fatality. The fractional number of fatal cases per year was derived as follows. The National Safety Council and company safety records give the following national and company vehicle accident data:

- Company vehicle mileage = 1.7×10^6 miles/year
- Company accident rate = $6.4/10^6$ motor vehicle miles
- National accident rate = $19/10^6$ motor vehicle miles
- National death rate = $1.3/10^8$ passenger miles

- Assumed number of passengers per vehicle = 2
- National rate of injuries per fatality = 37
- $P = 1.3 \times 10^{-8} \frac{\text{deaths}}{\text{mile}} \times 1.7 \times 10^{-6} \frac{\text{miles}}{\text{year}}$
 $\times \frac{6.4 \times 10^{-6}}{19 \times 10^{-6}} \frac{\text{company accident rate}}{\text{national accident rate}} \times 2 \frac{\text{passengers}}{\text{vehicle}}$
 $\times 1/2 \text{ seat belt factor} = 0.008 \text{ fatalities/year.}^a$

The probability of a lost time injury from a vehicle accident is:

$$P = 37 \text{ injuries/fatality} \times 0.008 \text{ fatalities/year} = 0.3 \text{ injuries/year.}$$

These theoretical probabilities should be added to the other theoretical probabilities, and the actual number of fatalities or lost time injuries should be related to other accidents to obtain the expected or average number per year for all company accident types.

7.2.5 Property Damage

Direct costs of property damage accidents are defined by DOE to include labor and material for replacement and cleanup costs. Depreciation adjustments are not made. Average direct costs can be derived by separating property damage and vehicle cases into proper categories, and by dividing the total cost in each category by the number of cases in that category. This gives a measurement only of accident costs experienced to date. It does not provide a risk measurement of the large accident which has not yet happened. Adjustments to include the large consequence potential should be made as discussed in detail in Section 6.

a. The 1/2 factor is for greater use of seat belts by company employees.

7.3 Indirect Costs

For each of the categories discussed under Direct Costs and in Table 5, indirect costs were evaluated. These values were used to derive the indirect cost multipliers as explained, which were evaluated separately for each category and also at different severity levels, because indirect cost factors are not a linear function of severity.

Indirect cost items have been identified. These are believed to include nearly all of the hidden costs for each of the accident categories listed in Table 5. However, not all of the items are applicable to all categories; for example, there are no new employee training costs for first aid cases. The items are discussed on the following pages.

7.3.1 Injured Worker Time

Productive time is lost by injured employee and is not reimbursed by workmen's compensation.

7.3.2 Co-Worker Time

1. Time is lost by co-workers at the scene, as well as when assisting the injured to dispensary or ambulance.
2. Time is lost through sympathy or curiosity, and work interruption at the time of injury and later from discussing the case, telling similar stories, swapping opinion of cause, grumbling, etc.
3. Incidental lost time results from cleanup, collecting donations to aid the employee and his family, review hearings, etc. The cost of other employee overtime required to accomplish the injured employee's work and the time spent by safety organization personnel on the accident should be included.

7.3.3 Supervisor Time

Supervisor time charged to the accident should include:

1. Assisting injured employee
2. Investigating accident cause, i.e., initial investigation, followup, research on prevention, etc.
3. Arranging for work continuance, getting new material, rescheduling
4. Selecting and training new employee; including obtaining applicants, evaluating candidates, and training new employee or transferred employee
5. Preparing accident reports
6. Participating in hearings or court proceedings.

7.3.4 General Losses

1. Production time is lost due to upset, shock, or diverted interest of workers, slowdown of others, discussion by others--"did you hear . . ."--(applies to employees of other units not included in Item 3, on the previous page)
2. Losses result from work stoppage of machines, vehicles, plants, facilities, etc., and can be either temporary or long term and effect related equipment and schedules
3. The injured employee's effectiveness is often reduced after his return to work, from work restrictions, reduced efficiency, physical handicaps, crutches, splints, etc.
4. Loss of business and goodwill, adverse publicity, problems in obtaining new hires, etc., are common general losses

5. Legal expenses arise from compensation hearings, liability claims handling, etc., that involve contractor legal services, rather than the insurance carrier legal expense that appears in direct costs
6. Costs can increase for insurance reserves and tax multipliers which are, respectively, small annual percentages of the gross incurred losses, and taxes based upon the dollar value of losses, that are tied up in reserves
7. Replacement of services or products loss during downtime and/or penalties administered under penalty contract clauses for DOE contractors
8. Miscellaneous additional items should be included which may be unique to particular operations and are appropriate to specific accident cases.

An average value for each of the above items can be estimated for each of the categories in Table 5. The sum of these values for each category is the total indirect cost. The sum of the indirect and direct cost, divided by the direct cost gives the indirect cost multipliers listed in Table 5. This indirect cost multiplier, as used in Table 5, gives the total annual cost of accidents when multiplied by the annual direct cost of accidents. (The multiplier could first be applied to the direct cost of each accident, and then the total cost of each accident multiplied by the annual number of accidents.)

Evaluating the individual items for each category is not simple. The results in Table 5 are our best estimates based on the results of:

1. An opinion survey of 35 supervisors
2. Field evaluations of accidents as they occur
3. Legal and financial records.

The opinion survey gave inflated results. For example, in 1976, 35 supervisors, when given the items and a description of a first aid case to evaluate, estimated losses ranging from \$75 to \$1490. On the other hand, costs for 11 first aid cases were evaluated jointly by safety engineers and supervisors as they occurred. Their estimates ranged from \$5 to \$80 with an average of \$30 per case. In general, the indirect costs do not appear to escalate linearly with severity. This confirms the need for evaluating indirect costs separately for different severity levels.

Items which are determined directly, such as legal expense or insurance costs, should not be included in an opinion survey, but should be taken directly from appropriate records. In addition, personnel making the evaluation should make estimates in hours for lost time accompanied by descriptive information, such as "reactor shutdown for 2 hours." Conversion to dollar values should then be made by one person with adequate cost and salary information. Estimates of general losses should be submitted directly in dollar values.

Once the evaluation has been made, the indirect cost multiplier can be used for several years, although periodic refinement is recommended. It is emphasized that the indirect multipliers in Table 5 need further field evaluation and should not be accepted as more than gross approximations. However, Table 5 does provide one way to compile total accident costs for any company.

8. RISK ASSESSMENT OF EXISTING SYSTEMS

8.1 General

A risk assessment of the total operations for an existing company serves several purposes.

1. Oversight are identified. In one case, a risk assessment resulted in adding a needed electrical engineer to the safety staff.
2. Assurance is provided that adequate safety control is being maintained.
3. The various risks are placed in perspective for management. Resources can be adjusted for cost effectiveness.
4. Areas where regulations are either less than or more than adequate can be identified. The single minded application of regulations frequently result in the wasting of resources where little real risk exists. Waivers, exemptions, or changes in the regulations should be granted wherever it is clear that little benefit is derived. Risk assessment provides continual feedback to administrators to make regulations more effective and cost efficient.

8.2 Risk Identification and Ranking

A systematic search for all risks greatly reduces the number of hazards which will be neglected because of management oversight. Based on the premise that all accidents result from an unplanned and unwanted transfer of energy, the Risk Identification Tree was developed by Dr. R. J. Nertney of the EG&G Idaho System Safety Development Center and is presented in Appendix A.

Another method for identifying hazards is to use a process called a Reported Significant Observation (RSO) study. An RSO study is an

information-gathering method which use employee-participants to describe situations they have personally witnessed, involving good and bad practices and safe and unsafe conditions. This information is utilized in the risk assessment process to help monitor the presence of hazards, and thereby help eliminate them and prevent their existence in future operations and designs. One such RSO study identifying high risk energy types was compared to actual AEC fatalities by energy type. The excellent agreement demonstrates that RSO studies can be highly accurate in predicting areas of higher than average risk.

This constitutes a rudimentary identification and ranking of risks. The risks can be further quantified from actuarial data and theoretical estimates. The general method in principle is quite simple: determine the average annual cost of accidents, as in the previous section, and add to these costs the theoretical risks not represented in the actual loss data. Specific steps which will accomplish this are:

1. Choose an accident classification. Although other classifications might be used, we suggest classifying by type of energy as given in Table 6. The reasons for choosing energy are:
 - a. Each type of energy creates a homogeneous class of accidents which increase the accuracy of the frequency-severity analysis. A heterogeneous group of accidents may completely mask the potential for a large accident in one energy group if the energy group has few accidents.
 - b. Energy sources are easily identified so that oversights are minimized. A thorough hazards identification may include a search of energy sources within each facility or department.
 - c. The energy types can be grouped by safety discipline (industrial, nuclear, traffic, fire, industrial hygiene, etc.) so that safety resource allocation can be compared to and balanced with risks for a more cost effective safety program.

TABLE 6. STATISTICAL ESTIMATE OF ANNUAL RESIDUAL RISK FOR A TYPICAL DOE OPERATION (in dollars x 10³)

<u>Energy Type</u>	<u>Injury Death</u>	<u>Property Damage</u>	<u>Tangible Losses</u>	<u>Intangible Effects</u>	<u>Total Losses</u>	<u>Percent</u>
Criticality ^a (in-reactor)	0.2	8.0	8.2	83.0	91.2	18.0
MGH ^b (falls)	68.0	2.5	70.5	7.0	77.5	15.1
Electrical	19.0	50.0	69.0	6.9	75.9	14.8
Kinetic/linear (in-plant)	45.0	18.0	63.0	6.3	69.3	13.5
Vehicle	10.0	41.0	51.0	5.1	56.1	11.0
Radiation	0.3	3.0	3.3	33.0	36.3	7.1
Fire	4.5	13.0	17.5	17.5	35.0	6.8
Cranes and lifts	11.0	7.0	18.0	1.8	19.8	3.9
Rotational	3.5	11.5	15.0	1.5	16.5	3.2
Toxic/pathogenic	6.0	--	6.0	6.0	12.0	2.3
PVC-KDD ^c (stored energy, pressure-volume springs, Young's constant-distance pressure, etc.)	6.0	0.5	6.5	0.7	7.2	1.4
Corrosives	5.0	--	5.0	0.5	5.5	1.1
Thermal	4.5	--	4.5	0.5	5.0	1.0
Explosive pyrophoric	3.0	--	3.0	0.3	3.3	0.6
Criticality ^a (out-of-reactor)	0.1	--	0.1	1.0	1.1	0.2
Total	186.1	154.5	340.6	171.1	511.7	100.0

a. Theoretical estimate only.

b. Potential energy.

c. Pressure--volume.

d. Young's constant--distance.

2. Assign each accident which occurred during the past several years to one of the classifications. Injury, property damage, and vehicle accidents should be classified separately (note that all vehicle accidents can be considered a special case of linear-kinetic energy). If injury logs contain insufficient information, the energy classification may be quite subjective.
3. Review separately, the injury, vehicle, and property damage accidents in each energy classification.
 - a. If there are sufficient data, do a frequency-severity analysis, and use the integrated value as the average cost of an accident. Multiply this average cost of an accident by the average number of accidents per year to obtain the average or probable loss per year. This expected loss is the most likely cost for any given year and is the risk for this category of accidents.

If it is assumed that the severity distribution of injuries is the same for each energy source, the cost estimates for injuries can be made on the basis of the number of injuries in each energy category.

- b. If there have been few or no accidents in a given category, make a theoretical risk estimate. Multiply this probability by the estimated cost to obtain the risk for this category. Examples are given following Item 5 below.
4. Enter the "cost" or risk for each energy category by injury and property damage as in Table 6. Sum the value to obtain the total tangible annual risk for the existing company or system.
5. Add intangible accident costs. Value factors should be added to various energy risk categories. These are based on subjective judgment of the acceptability of different kinds of risk. As such, adding value factors should be considered a form of risk.

evaluation rather than risk analysis or quantification. These intangible costs should not include any consequences included in direct or indirect costs. Intangible costs are primarily nonphysical in nature.

For the study in Table 6, an intangible value factor of 10 for radiation categories was selected because of the extreme public reaction (and the resulting reflection on the nuclear industry) which usually results from an accident involving radiation. Since a fire may or may not involve radioactive materials, the intangible loss from fires was assumed to be equal to the tangible loss (a value factor of 1).

A value factor of 1 was also assumed for the toxic/pathogenic category. This is both prudent and reasonable, since latent ill effects of toxic exposure are not likely to be immediately detected. Intangible costs for all other energy categories were assumed to be 1/10 of tangible costs. These value factors are arbitrary but were selected only after discussion with DOE and contractor safety personnel. The intangible costs derived from these factors are given in Table 6.

It is recognized that all intangible risk factors have not been included. One notable example is that a single, large consequence event is usually more undesirable than many small consequence events. Witness the general public acceptance of 50,000 annual vehicle deaths compared to the concern for a maximum potential nuclear reactor accident. In addition, no attempt has been made to place dollar values on peace of mind, employee morale, human suffering, etc. All factors included in an assessment should be explicitly stated. Presentation of risk information should be in a form which permits ready identification and selection of appropriate factors for inclusion in the final risk values. However, the limited development of intangible value factors is considered acceptable for the purposes of this report. The total losses given in Table 6 are not a dollar value of all losses, nor indeed can a dollar value be truly placed on esthetic, moral, or life values. As such, it is emphasized that these risk values are not absolute or complete risk values. They do place various kinds of risks in perspective to assure that all risks are managed and accepted in a rational,

systematic manner. In addition, the risk values enable management to compare resources allocated to various risks in a scientific manner.

8.3 Theoretical Risk

To complete the risk picture so that comparisons in Table 6 are valid theoretical derivation of risk is needed where actuarial data (experience) is lacking. A reasonable approximation of risk is better than no information. Examples of theoretical derivations and discussions of some of the energy categories in Table 6 are given below.

1. Radiation Risk--The radiation risk was subdivided into "criticality (in-reactor)," "criticality (out-of-reactor)," and "radiation" to correspond with the reactor safety, criticality safety, and health physics disciplines. Fatality and injury costs from accidental exposure are assessed in the same way as injury from other types of energy. (The number of injuries or fatalities is multiplied by the same average cost of injuries or fatalities as given for other conventional forms of risk.) These are \$100,000 for a fatality and \$3500 for the average workday lost injury. These values are multiplied by the indirect factor of 2 to obtain the total tangible radiation accident costs. To account for the intangible factors these values are multiplied by 10 to obtain the total expected loss.

The annual probability of a radiation fatality or injury for a typical contractor was calculated from the accident data¹⁷ given in Table 7.

These 41 accidental exposures occurred over a 32-year period. The annual probability for a typical contractor with 2.5% of the work force would be these values divided by 32 and multiplied by 0.025. Credit should also be taken for improved safety performance. Of the 41 exposures, 34 occurred during the first 15 years and only 7 during the last 15 years. Using a conservative factor of 2 for time improvement and the consequence values of \$200,000

TABLE 7. LOST TIME INJURIES FROM RADIATION ACCIDENTS
(32 years)

Severity Level	Radiation Source					Total
	Criticality (in-reactor)	Criticality (out-of-reactor)	Weapons Testing	Other Radiation		
Fatality	2 ^a	1	0	0	3	
Permanent disability	0	0	0	4	4	
Temporary clinical effects	2	8	4	7	21	
No observable effects	10	0	1	2	13	
Total	14	9	5	13	41	

a. Three fatalities at SL-1 are excluded.

for fatality and permanent disability and \$7,000 for observable injury, the reactor ^{a,b} risks are then calculated as follows:

$$\text{Fatality risk} = \frac{2 \text{ fatalities} \times \$200,000/\text{fatality}}{32 \text{ years}}$$

$$\times \frac{0.025 \text{ fraction of DOE work force}}{2 \text{ (improvement)}} = \$160/\text{year}$$

$$\text{Injury risk} = \frac{2 \text{ injuries} \times \$7000/\text{injury}}{32 \text{ years}}$$

$$\times \frac{0.025 \text{ fraction of DOE work force}}{2 \text{ (improvement)}} = \$6/\text{year}$$

$$\text{Total contractor personnel risk} = \$166/\text{year.}$$

a. Reactor is defined as an assembly where approach to criticality is planned.

b. See Pages 43 and 44 of SSDC 11, 1977.

These 41 accidental exposures occurred over a 32-year period. The annual probability for a typical contractor with 2.5% of the work force would be these values divided by 32 and multiplied by 0.025. Credit should also be taken for improved safety performance. Of the 41 exposures, 34 occurred during the first 15 years and only 7 during the last 15 years. Using a conservative factor of 2 for time improvement and the consequence values of \$200,000 for fatality and permanent disability and \$7,000 for observable injury, the risks are then calculated as follows:

a. Reactor:^a

$$\text{Fatality risk} = \frac{2 \times \$200,000}{32} \times \frac{0.025}{2} = \$160/\text{year}$$

$$\text{Injury risk} = \frac{2 \times \$7000}{32} \times \frac{0.025}{2} = \$6/\text{year}$$

$$\text{Total personnel risk} = \$166/\text{year}.$$

b. Criticality (out-of-reactor):

$$\text{Fatality risk} = \frac{1 \text{ fatality} \times \$200,000}{32 \text{ years fatality}}$$

$$\times \frac{0.025 \text{ fraction of work force}}{2 \text{ (improvement factor)}} = \$80/\text{year}$$

$$\text{Injury risk} = \frac{8 \text{ injuries} \times \$7000}{32 \text{ years injury}}$$

$$\times \frac{0.025 \text{ fraction of work force}}{2 \text{ (improvement factor)}} = \$22/\text{year}$$

$$\text{Total contractor personnel risk} = \$102/\text{year}.$$

a. Reactor is defined as an assembly where approach to criticality is planned.

c. Other radiation:

$$\text{Fatality and permanent disability risk} = \frac{4 \text{ case} \times \$200,000}{32 \text{ years case}}$$

$$\times \frac{0.025 \text{ fraction of work force}}{2 \text{ (improvement factor)}}$$

$$= \$312/\text{year}$$

$$\text{Injury risk} = \frac{7 \text{ injuries} \times \$7000}{32 \text{ years injury}}$$

$$\times \frac{0.025 \text{ fraction of work force}}{2 \text{ (improvement factor)}} = \$19/\text{year}$$

$$\text{Total contractor personnel risk} = \$331/\text{year.}$$

The above injury values are given in Column 1 of Table 6.

The Rasmussen study⁵ gives the probability of a fatality from reactor operations as one in 5×10^7 reactor years. For a contractor with four test reactors, the risk would be:

$$\text{Fatality risk} = \frac{4 \text{ reactors}}{5 \times 10^7 \text{ reactor-years/fatality}}$$

$$\times \frac{\$200,000}{\text{fatality}} = \$0.02/\text{year} .$$

This value is much lower than the \$166/year calculated previously, but the reactor fatalities resulted from critical experiments while Rasmussen analyzed the fatality risk of power reactors. On the other hand, no reactor fatality (other than SL-1) has occurred in the past 20 years. Therefore, the current fatality risk may be lower than the calculated \$166/year because of the significant improvement in reactor safety.

For this study, \$200 is assumed to be the annual reactor fatality risk and \$100 is assumed for the typical contractor out-of-reactor or criticality risk for four reactors and 3000 employees. However, the actual risk may vary greatly depending upon the type of operation. For example, the criticality risk for an R&D contractor who does not process fissile material in liquid form would be much smaller than that for contractors operating processing plants. This is based on the fact that no criticality incident involving unmoderated material, except where attempting to achieve or examine criticality, has ever occurred in the nuclear industry. The reason for this is that a comparatively large mass of fissile material must be arranged in a highly reactive configuration, and a moderating/reflecting material must be added to attain criticality. It is true that spent or recycled fuel elements may be stored and handled in water, but since fuel elements are handled only one at a time with a long handling tool, the probability of an inadvertent criticality, through the arrangement of a critical number of elements, is extremely low.

Property damage from a radiation accident consists primarily of decontamination costs or loss of contaminated items, where the cleaning costs exceed the property value. Routine decontamination costs which are expected and are a part of a planned operation should not be included.

Property damage from out-of-reactor criticality is insignificant, and no value is listed in the table. Criticality incidents have been caused by improper manual handling and usually result in one or more fatalities. The decontamination costs are small in comparison to these costs.

The property damage from "other" radiation accidents was estimated at \$3000 per year for a typical contractor, based on an examination of radiation incidents reported by one contractor over a 5-year period. The actual risk value may be somewhat higher, since there is evidence that some minor spills are decontaminated without reporting the cleaning costs as property damage.

Two methods were used to reach a gross approximation of property damage risks from the nuclear energy generated in a reactor core. Probabilities of various failures and subsequent consequences were estimated, is one method. For each of the analyzed accidents, the probability times the consequence sums to a risk of \$500/reactor-year.

Another approach is to assume that the property damage risk from test reactors is similar to that of reactors used for the commercial generation of electrical power, as given in the Rasmussen Reactor Safety Study. This gives an order of magnitude risk estimate for test reactors. (A separate risk assessment would be more accurate, but the cost puts it outside the scope of this Guide.) Integrating the property damage frequency-severity curve, given in the Reactor Safety Study, yields a property risk of about \$5000/reactor-year. Of this total, about \$4000 results from large accidents in the >\$100,000,000 range. Since test reactors are smaller and are more isolated from the populated areas, this \$4000/year risk is beyond the upper limit and can be excluded. This leaves a risk of about \$1000/reactor-year, or about \$4000/year for four test reactors. Doubling this value for indirect costs yields a risk of \$8000/year, as given in Table 6.

These estimates, by various means, all agree within a factor of 2 or 3 which gives assurance that the risk is indeed very small compared to other costs of operating reactors.

2. Fire Risk--Fire risk should be based on an individual company's experience adjusted for the large or catastrophic fires. If sufficient experience data are available, the large fire adjustment can be made using statistical projections (log-normal extreme value analysis). If fire experience is very limited, risk can be estimated from DOE-wide experience data or from insurance rates. (Fire loss averaged over all industry is 30 to 40% of the insurance rate minus the total premiums paid.)

The AEC (DOE) 28-year cumulative fire loss ratio (1947 to 1974) was 1.7¢ per \$100 property valuation.¹⁷ Without the Rocky Flats fire, the ratio would have been less than a third of this value. In the 5 years after the Rocky Flats fire (1970 to 1975), the ERDA-wide fire loss ratio has been 0.093¢ per \$100 value (if adjusted for catastrophic loss, the loss rate would undoubtedly be higher). However, improvement is evident from the chart given on Page 19 of WASH-1192,¹⁷ and probably results from extensive fire protection improvements and increased concern for fire loss. There is, however, statistical evidence that the fire loss ratio has been increasing since the Rocky Flats fire, indicating a possible decreasing concern as time passes. In addition, the fire loss ratio is variable, being considerably higher at some sites than others. Contractor risk estimates can be made from local, DOE, or national experience, making adjustments as appropriate.

3. Toxic Pathogenic--Except for massive or serious acute exposures, losses from toxic, carcinogenic, or pathogenic effects usually are not identified. Since many months, or even years, may pass prior to the onset of serious effects, an illness or injury caused by exposure to toxic materials may never be related to specific exposure incidents. In addition, the toxicity of many substances has not been recognized because of the long, latent periods before biological damage appears.

Of the 120 AEC occupational fatalities from 1959 through 1975, one was caused by solvent vapors and three by asphyxiation in confined spaces or inert atmospheres. Assigning asphyxiation to the toxic category, 3.3% of the fatalities occur in this category. Although there are few toxic injuries, the injury/death risk is assumed to be the same for toxic materials as it is for other energy categories. Thus, 3.3% of the total injury/fatality risk is assigned to toxic/pathogenic in Table 6.

4. Electrical Risk--Electrical risk includes hazards of downtime from power failure, property damage from electric system failures, and

personnel injuries from contact with electricity. One should be cautious not to overlook the large consequence potential inherent in electrical systems. Failures have potential for long downtimes and/or large fires or other property damage. For example, for electric property damage risk Table 4 gives a value twice the average loss over a 5-year period.

In addition, the injury risk may be high even though few electric shock injuries have occurred. An electrician may not report a shock if he is not injured, which under slightly different circumstances may be fatal. To estimate this fatality risk, multiply the number of injuries by the conditional probability of fatality if an injury occurs. An example is given below:

In the past 6 months, Company Z had 6 injuries from electrical sources; 2 of which were 600 volts.

The following from "Accident Facts" shows derivation of return period and risk for electric fatality:

Statistical data published by the National Safety Council in Accident Facts indicate:

18.5% of injuries from >600 volts are fatal

1.6% of injuries from <600 volts are fatal.

Fatality frequency for >600 volts:

$$\frac{2}{6} \text{ accidents/month} \times 0.185 \frac{\text{fatalities}}{\text{injury}}$$

$$\times \frac{12 \text{ months}}{\text{year}} = 0.74 \frac{\text{fatalities}}{\text{year}} .$$

Fatality frequency for <600 volts:

$$4/6 \frac{\text{accidents}}{\text{month}} \times 0.016 \frac{\text{fatalities}}{\text{injury}}$$

$$\times \frac{12 \text{ months}}{\text{year}} = 0.13 \frac{\text{fatalities}}{\text{year}} .$$

$$\text{All voltages} = 0.74 + 0.13 = 0.87 \frac{\text{fatalities}}{\text{year}} .$$

$$\text{Return period} = \frac{1}{\text{frequency}} = 1/0.87 = 1.15 \text{ years} ..$$

Assuming \$200,000 for a fatality, the annual risk is:

$$R = 200,000 \times 0.87 = \$174,000/\text{year}.$$

NOTE: This is an example and this risk value is not included in Table 6.

9. RISK ASSESSMENT OF NEW SYSTEMS

If a proposed facility and its operation involve no unique or unusual hazards, the life cycle risk can be estimated from appropriate DOE-wide or national property loss and injury incidence rates.

The following steps are necessary.

1. Divide the project into stages appropriate for risk estimation such as:
 - a. Planning, design, and review
 - b. Construction
 - c. Operation
 - d. Decommissioning and dismantling.
2. Assess the routine risk in each stage by:
 - a. Determining the employee-hours and property values at risk
 - b. Multiplying the employee-hours and property values by appropriate incidence rates; i.e., construction employee-hours and property values times construction incidence rates
 - c. Considering any environmental or public effects for each stage.
3. Assess the nonroutine risk in each stage by:
 - a. Considering the breakin period and training of new employees during the first year. (Because personnel are unfamiliar with procedures and equipment, mistakes are more likely.) Data show that even experienced bus drivers have an increased

accident probability when assigned to a new route, and consequently, first year employees have a higher injury rate. In the absence of data, a factor of 2 or 3 over average accidents rates is suggested.

- b. Consider the latter stages of plant life. Will risk increase because of complacency and carelessness engendered by familiarity? Will funds be provided to replace old equipment or will recommended service life likely be exceeded? No guideline is given here because the increased risk varies greatly depending upon hazards associated with breakdown of aging equipment and because data of a general nature are unavailable. However, the risk can increase enormously as equipment ages.
 - c. Screening energy sources at each stage and assessing the risk for unusual hazards such as nuclear, vehicle (for operations with large transportation needs), thermal (for solar reflectors), etc.
4. Sum the various risks to obtain the total. While outside the scope of this risk manual, risk assessment is only a part of an adequate safety analysis report which should, among other things, document the control procedures and safety staff assumed in the risk assessment. There is ample evidence that a professional safety staff and comprehensive review system will reduce risks by factors of 5 to 10 over that of a project with only one safety engineer for 5 to 10 thousand workers.

9.1 Resource Allocation

For optimum efficiency, the effort applied to risk control should be consistent with the degree of magnitude of risk. The curve in Figure 7 pictures qualitatively the relationships between safety investment costs and accident costs. As shown in Figure 7, safety investment is represented by a straight line, while accident costs are represented by the decreasing

curve which approaches a minimum to the right. The total cost to a company is the sum of the safety investment and the residual accident costs, which still persist at the current level of investment. This total cost value is represented by the top curve. Zero investment in safety results in maximum accident costs. Initial safety efforts correct the most obvious and easily corrected hazards and yield the greatest dividends in reducing accident costs, as shown by the rapidly decreasing curve. Each incremental investment yields successively smaller dividends. As seen in the figure, the minimum total cost occurs where the decreasing slope of accident costs equals the constant slope of the safety investment curve. This minimum will usually occur at or very near the intersection of the two cost curves, where the safety investment is equal to the residual accident cost. It should be noted that this phenomenon is not very sensitive to accident and safety costs. If a large investment is required to produce a relatively smaller reduction in accident costs, the investment curve would have a higher positive slope and would intersect the accident cost curve where its

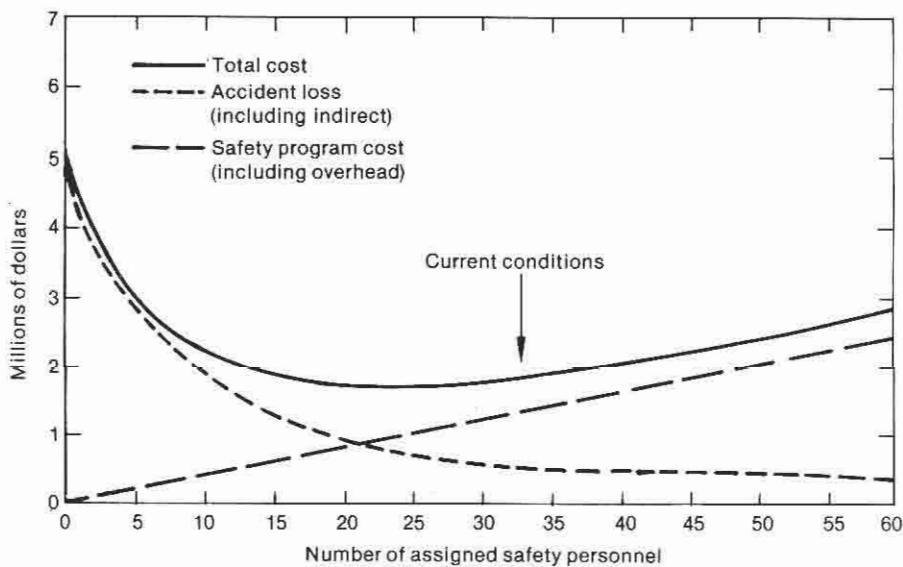


Figure 7. Total accident costs as a function of safety investment and residual accident costs.

negative slope is also greater. Likewise, a low investment yielding large dividends results in an intersection farther to the right where the accident cost curve is also relatively flat.

Thus, it is reasonably certain that total costs are at a minimum when resources devoted to safety are about equal to residual accident costs. It should also be noted that the total cost curve is relatively flat over a fairly wide range of safety investment. This means that a company can invest resources in safety that are significantly larger than the residual accident costs with little total increased cost. It also means that gross estimates are often adequate to determine whether insufficient or excess resources are being invested in safety.

The actual dollar values given in Figure 7 were derived as follows. The safety program costs were based on a cost of \$40,000 per professional safety engineer or radiological engineer. (The \$40,000 includes all overhead expenses and indirect personnel salaries averaged over 33 direct safety personnel.) The accident loss curve was approximated from only one firm point: the current accident loss of \$500,000 given in Figure 7, and the current number of 33 safety professionals assigned to operational activities for a typical contractor. Since the average severity, frequency, injury, and vehicle accident rates given by the National Safety Council range from 5 to 15 times greater than a typical DOE contractor, it was assumed that if the company had only one instead of three safety engineers, the residual risk of accident losses would be 10 times as high (or five million dollars per year). This gives a second point: one engineer versus five million dollars. An approximation of the slope of the accident cost curve at the current level of 33 engineers was obtained from a hazard sensitivity study. The consensus given in the Delphi study (discussed later in this section) is that at the current level of safety staffing, the addition or deletion of one man would eventually change the risk by about \$10,000/year. Since two points and the slope at one point are known and the curve is approximately exponential, the complete curve can be drawn with reasonable accuracy. By exponential we mean that the first person corrects the easiest most obvious hazards, and the second person corrects items the first person missed. With succeeding persons, we reach a point of

diminishing returns; thus, slope is steeper at first and flattens out. It is believed that the curve is accurate within a factor of 2 on the left, with the accuracy increasing to about 20% in the range of 20 to 40 safety personnel.

It is impractical to construct a similar resource accident cost curve for each energy type. However, it is obvious that total risk control resources are allocated at optimum efficiency if the resources and specific energy residual risks are relatively uniform, rather than large sums being spent on small risks, and vice versa.

To measure the degree of balance or uniformity in a safety program, determine the percentage of resources spent on each type of risk and divide by the degree of risk. As an illustration, this exercise for a contractor follows.

The annual resources in man-years allocated to each risk in Table 8 were determined. Some of the categories given in Table 8, such as criticality and fire, were easily and accurately estimated. Others, such as corrosive and thermal, are only educated guesses. The estimates can be greatly improved if safety professionals working in multienergy categories keep a log of their time spent in each category. The log needs to be kept only long enough to establish the division of time and at periodic intervals to monitor changes. A table similar to Table 8 can then be constructed. Local decisions must be made regarding which personnel should logically be included in the safety manpower allocation. Noticeably absent in Table 8, for example, are salaries and other costs for the fire department which is DOE-operated, rather than contractor-operated, at this particular location.

Assuming the man-years spent on each category are proportional to the safety program costs for each category, resource percentages were calculated directly from the man-years.

There are, no doubt many ways to estimate resources allocated to safety. Total salaries and overhead costs could be used but these are

TABLE 8. CURRENT SAFETY RESOURCE ALLOCATION MAN-YEARS

Energy Type	Operational Safety			Engineering				Total Man-Years
	Site 1	Site 2	Site 3	Radiation	Industrial Safety	Industrial Hygiene	Nuclear	
MGH	0.4	0.37	0.52	--	0.2	--	--	1.49
Criticality (in-reactor)	--	0.04	0.03	--	--	--	2.0	4.3
Kinetic/linear (in-plant)	--	1.14	0.91	--	0.2	--	--	6.37
Electrical	0.5	0.16	0.17	--	1.15	--	--	2.25
Vehicle	0.1	0.02	0.02	--	1.2	--	--	1.985
Radiation	8.0	10.8	5.97	5.0	--	0.44	0.4	--
Fire	0.5	0.44	0.3	--	0.1	2.8	1.3	--
Cranes and lifts	0.2	0.37	0.33	--	0.2	--	--	5.44
Rotational	0.7	--	--	--	0.1	--	--	1.1
PV-KD (stored, energy, springs, pressure, etc.)	0.4	0.18	0.11	--	0.3	--	--	0.8
Corrosives	--	0.18	0.2	--	0.1	--	0.1	--
Thermal	0.2	0.14	1.1	--	0.1	--	--	0.58
								1.54

TABLE 8. (continued)

Energy Type	Operational Safety				Engineering					Total Man-Years
	Site 1	Site 2	Site 3	Radiation	Industrial Safety	Fire	Industrial Hygiene	Nuclear	Other	
Explosive pyrophoric	0.1	0.04	0.04	--	0.4	--	--	--	--	0.58
Criticality (out-of-reactor)	0.5	0.04	0.25	--	--	--	--	2.0	0.5	3.065
Other	0.1	--	--	--	--	0.9	--	--	--	1.0
Toxic/pathogenic	0.3	0.08	0.26	--	--	1.3	1.5	--	--	3.44
Total man-years	12.0	14.0	9.99	5.0	3.05	5.44	3.3	9.0	5.9	62.68

reasonably proportional to the manpower spent in the field. Costs which may not be proportional are personnel safety equipment and increased capital equipment and operational costs required to meet safety requirements.

Specific cost/benefit analysis is often more appropriate than generalization of these items as safety investments. The cost of meeting minimum codes and standards should be included only in cost/benefit analyses during evaluation of proposed standards, but not in overall safety investment/accident cost trade-off studies.

The percentage of risk for each energy category (calculated from risks in Table 8), the resource percentage and the resource to risk rates are calculated and tabulated as in Table 9. The percentage resource divided by the percentage risk gives a ratio that is indicative of the relative safety resource/risk relationship. A ratio of <1 means that less than average attention is being given to safety, and a ratio >1 means more than average attention to safety. These ratios are given for total risks (including intangible effects) and company risks only (tangible direct and indirect losses). In Table 9, the energy types are listed, with those receiving the least relative attention at the top and those receiving more at the bottom. The total resource/risk comparison is also given in bargraph form in Figure 8. The multiplier of 10 applied to nuclear risks (included in the bargraph) changes these ratios significantly. The total risk ratios are the ones of direct interest, since implicit in the multiplier of 10 is that it is worthwhile to spend 10 times as much to prevent a nuclear accident, as to prevent a nonnuclear incident having the same direct dollar loss.

There are no data available from which to determine directly the sensitivity of the residual risk level to changes in the level of safety effort allocated to each energy type; however, an approximation can be obtained from opinions of safety specialists. One method of obtaining a consensus is a Delphi study. This technique permits interchange of thought without the study participants meeting as a group. The steps in a Delphi study are as follows:

TABLE 9. RELATIVE RESOURCE/RISK RANKING

Energy Type	Total Risk			Tangible Risk Only	
	Annual Resource (%)	Annual Risk (%)	Ratio ^a	Annual Risk (%)	Ratio ^a
MGH (falls)	2.2	15.1	0.14	20.7	0.11
Electrical	3.2	14.8	0.21	20.2	0.16
Vehicle	2.5	10.9	0.23	15.0	0.17
Kinetic/linear (in-plant)	3.2	13.5	0.24	18.5	0.17
Kinetic energy (rotational)	1.2	3.2	0.38	4.4	0.27
MGH (cranes and lifts)	1.6	3.9	0.41	5.3	0.30
Nuclear (reactor transient)	11.8	18.0	0.66	2.4	4.92
Corrosives	1.0	1.1	0.91	1.5	0.67
PV-KD (springs, pressure, etc.)	1.5	1.4	1.07	1.9	0.79
Fire	9.5	6.8	1.40	5.1	1.86
Explosive pyrophoric	1.0	0.6	1.67	0.9	1.11
Thermal	2.2	1.0	2.2	1.3	1.69
Toxic/pathogenic	6.2	2.3	2.7	1.8	3.44
Radiation	45.6	7.1	6.42	1.0	45.6
Nuclear (criticality out-of-reactor)	5.5	0.4	3.75	0.1	55.0
Total	100	100	--	100	--

a. Percent resource divided by percent risk.

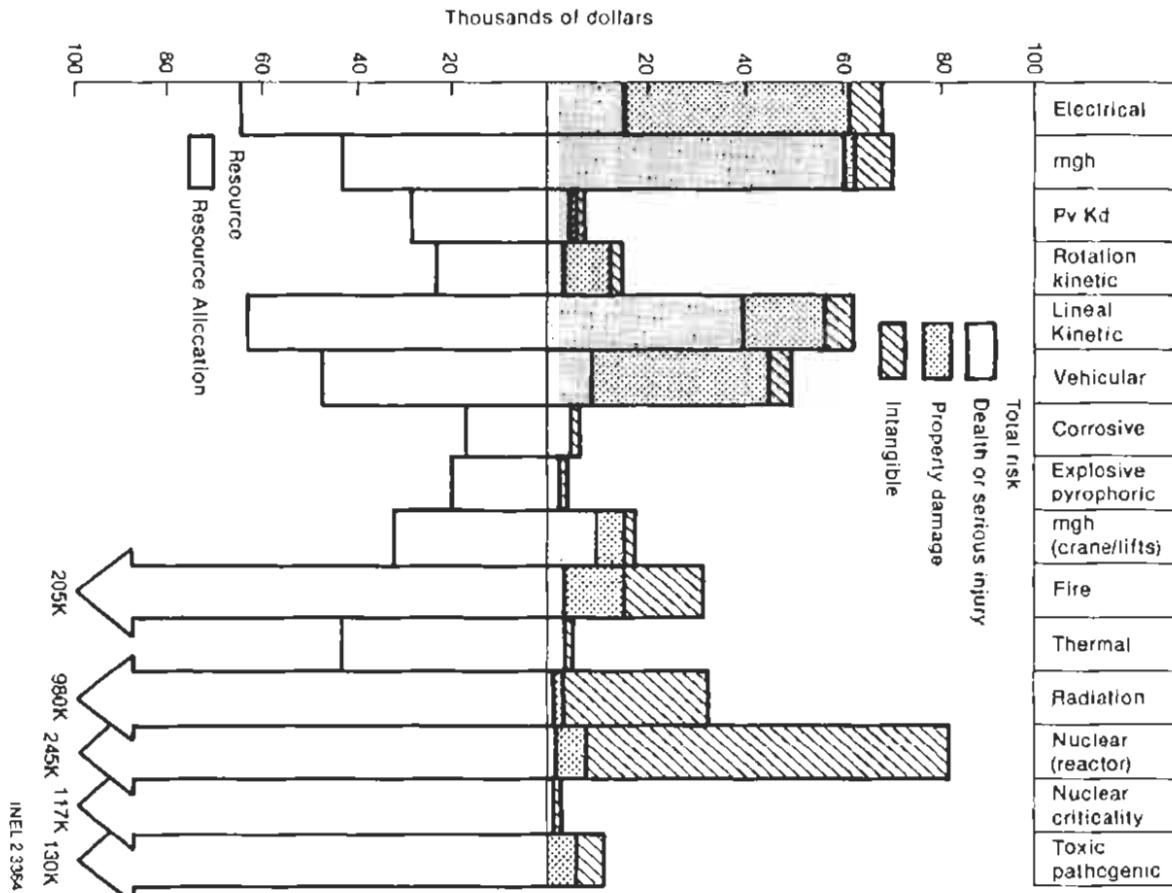


Figure 8. Comparison of safety investment (resources allocated to accident prevention) to residual risks by energy type.

1. Select group of qualified or recognized professionals in the field being studied.
2. Round 1 prediction: solicit predictions from participants.
3. Determine:
 - a. Median prediction
 - b. Innerquartile range (IQR = middle 50%). The middle 50% excludes the lower 25% and the upper 25%.
4. Round 2 prediction:
 - a. Feedback median and IQR to participants
 - b. Participants change prediction, reevaluate if they desire
 - c. Participants state their reasons if new prediction lies outside of IQR.
5. Round 3 prediction:
 - a. Feed back median, IQR, and concise summary of reasons for extreme positions (outside of IQR)
 - b. Participants change predictions if they desire
 - c. Participants state reasons for extreme position (outside of IQR).
6. Repeat Round 3 steps as necessary in order to achieve a reasonable consensus or a steady state in the results.

7. Report:

- a. Median (or medians if bimodal)
- b. Innerquartile range (IQR)
- c. Arguments for extreme position (or two positions if bimodal).

For the hazard sensitivity study, each of 12 safety specialists were given the current risk and the number of men currently allocated to each energy category. They were asked to estimate the eventual change in risk level for a 1/2 man-year level of effort added to or subtracted from the current level.

The results of each of the three rounds are given in Table 10. The median estimate and the IQR are given for each category. The median, rather than the mean, is used because one "wild" guess by an individual has no effect on the median, but could change the mean significantly. As can be seen, the range of estimates decreased, but in most cases there was little or no change in the median. It is recommended that anyone doing a Delphi study ask those who state reasons for differing from the group opinion to be very explicit. To demonstrate the consistency of hazard sensitivity estimates, each individual's estimates for +1/2 and -1/2 man-year (third round results) were added and are given in Table 11. Although the feedback and request for explanations result in a tendency to conform to majority opinion, there were not significant changes from the first round predictions. This indicates a single round of opinion gathering may have been adequate in this case.

From the Delphi study, the net change in residual risks, which results from increases and decreases in control efforts for specific energy categories, can be calculated. The data in Table 10 were used to calculate the effects of proposed changes. Table 12 gives the resulting recommendations for change in resource allocation. As can be seen, with no increases in the number of safety personnel, the estimated risk can be reduced by approximately \$38,000/year.

TABLE 10. HAZARD CONTROL SENSITIVITY--COMPANY LOSSES ONLY

Energy Type	Current Resource Number of Men	Current Risk (10 ³ \$)	Change in Risk ^a (10 ³ \$)					
			First Round		Second Round		Third Round	
			+1/2 man	-1/2 man	+1/2 man	-1/2 man	+1/2 man	-1/2 man
Electrical	2.0	75.0	10/5-10	20/10-20	10/5-10	20/15-20	10/5-10	15/10-20
MGH cranes and lifts	1.1	22.0	2/2-3	4/2-5	2/2-5	5/1-5	3/3-4	3/3-5
Other (MGH)	1.5	92.0	7/5-8	10/10-10	8/5-10	10/5-10	5/5-8	5/4-5
Kinetic/linear (in-plant)	2.3	78.0	8/5-8	5/5-10	8/5-10	5/5-10	8/5-8	6/5-8
Criticality (in-reactor)	2.0	8.3	1/0-1	1/1-1	1/0-1	1/0-2	1/0-1	1/1-1
Vehicle	0.4	51.0	10/8-10	10/10-15	10/5-10	10/2-20	10/9-10	10/10-15
Fire	3.7	150.0	10/10-10	15/10-30	10/10-20	10/10-40	10/10-10	15/10-20
Corrosive	0.6	7.5	1/1-2	3/2-5	1/1-2	2/1-5	1/1-2	3/2-5
Rotational	0.8	16.5	3/3-4	3/2-5	3/1-4	2/2-5	3/3-4	3/2-5
Radiation	29.6	3.5	0/0-0	0/0-0	0/0-0	0/0-0	0/0-0	0/0-0
Toxic	2.5	10.0	0.5/1-1.5	1.5/1-3	0.5/0.5-1	1.5/1-3	1/0.5-1	1.5/1-1.5
Pv-KD	1.0	8.5	1/1-1	2/1-2	1/0-2	2/0-2	1/1-1	2/1-2

TABLE 10. (continued)

Energy Type	Current Resource Number of Men	Current Risk ($10^3 \$$)	Change in Risk ^a ($10^3 \$$)					
			First Round		Second Round		Third Round	
			+1/2 man	-1/2 man	+1/2 man	-1/2 man	+1/2 man	-1/2 man
Thermal	1.5	6.0	1/1-1	1/1-2	1/0-1	1/0-2	1/1-1	1/1-2
Explosive pyrophoric	0.6	4.0	1/1-2	2/2-4	1/0-2	2/1-5	1/1-2	2/2-4
Criticality (out-of-reactor)	2.5	0.2	0/0-0	0/0-0	0/0-0	0/0-0	0/0-0	0/0-0

a. Median/range.

TABLE 11. CHANGE IN RISK/ONE-MAN EFFORT
(\$10³)

Energy Type	Individual Estimates by Safety Specialists												Median
	1	2	3	4	5	6	7	8	9	10	11	12	
Electrical	30	45	25	35	15	25	10	15	20	25	23	15	24
Cranes and lifts	7	2	7	7	8	6	20	8	13	15	3	3	7
Other (MGH)	15	15	15	18	18	17	18	14	15	15	0	8	15
Kinetic/linear (in-plant)	13	10	10	16	13	15	13	10	18	15	24	11	13
Criticality (in-reactor)	1	2	2	1	2	1	2	1	0	3	0	0	1
Vehicle	15	20	20	35	20	23	20	20	7	30	15	25	20
Fire	30	20	25	30	20	40	20	20	20	35	6	35	23
Corrosive	7	6	2	7	3	4	3	7	6	7	3	2	5
Rotational	8	3	5	8	4	7	5	8	5	9	7	5	6
Radiation	0	--	--	0	0	0	0	1	0	0	0	0	0
Toxic	10	2	4	4	1	2	2	2	5	3	1	2	2
PV-KD	--	--	3	3	3	3	3	1	4	3	3	1	3
Thermal	3	0.5	2	3	2	7	2	1	2	3	2	1	2
Explosive pyrophoric	6	25	3	4	5	5	10	7	5	5	3	1	5
Criticality (out-of-reactor)	0	--	0	0	0	0	0	0	0	0	0	0	0

TABLE 12. RESIDUAL RISK REDUCTION FROM CHANGE IN RESOURCES

<u>Energy Type</u>	<u>Proposed Change in Man-Years</u>	<u>Change in Risk (10^3)</u>
Electrical	+1	-20
Vehicle	+1/2	-10
Fire	<u>+1/2</u>	<u>-10</u>
Subtotal	+2	-40
Criticality	-1/2	0
Radiation	-1/2	0
Reactor	-1/2	+1
Thermal	-1/4	+0.5
PV-KD	<u>-1/4</u>	<u>+0.5</u>
Subtotal	-2	+2
Total	0	-38

It is admitted that this reduction has not been supported by hard data, but is based on, at least, educated guesses of risk sensitivity to resource allocation by safety specialists. Nevertheless, these techniques, as described, offers a systematic method of allocating resources to safety. As such, it is a significant improvement over subjective and/or intuitive management response to the particular safety discipline which agrees best for increased resource or to pressures from elsewhere.

10. REFERENCES

1. I. B. Campbell, Accident Statistics and Significance, Occasional Paper, No. 34, Director of Safety Accident Compensation Commission, Wellington, New Zealand, 1980.
2. J. Gumbel, Statistical Theory of Extreme Values and Some Practical Applications, National Bureau of Standards Applied Mathematics Series-33, 1957.
3. B. Fischhoff et al., How Safe is Safe Enough? A Psychometric Study of Attitudes Towards Technological Risks and Benefits, Decision Research, 1201 Oak Street, Eugene, Oregon, 97401, 1976.
4. P. Slovic et al., Images of Disaster-Perception and Acceptance of Risks From Nuclear Power, Decision Research, 1201 Oak Street, Eugene, Oregon, 97401, 1979.
5. "Reactor Safety Study," An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants, WASH-1400, October 1975.
6. G. J. Briscoe et al., Applications of MORT to Review of Safety Analysis, SSDC-17, July 1979.
7. R. J. Nertney, J. L. Clark, and R. W. Eicher, Occupancy-Use Readiness Manual--Safety Considerations, ERDA-76-45-T, SSDC-1, September 1975.
8. Accident Facts, National Safety Council, Chicago, Illinois, 1981.
9. B. L. Cahen and I-Sing Lee, "Catalog of Risk," Journal of Health Physics, June 1979.
10. M. G. Bullock, "Change Control and Analysis," SSDC-21, March 1981.
11. W. G. Johnson, MORT--The Management Oversight and Risk Tree, SAN 821-2, February 12, 1973.
12. Accident/Incident Investigation Manual, ERDA-76-20, August 1975.
13. P. A. Crosetti, "Reliability and Fault Tree Analysis Guide," SSDC-22, February 1982.
14. Risk Assessment: A Study of Risk Levels Within Norwegian Offshore Petroleum Activities, Report No. 26-27/2, The Royal Norwegian Council for Scientific and Industrial Research, November 1979.
15. "Are Government Programs Worth the Price?," Business Week, June 30, 1975, pp. 114-115.
16. R. Wilson, "The Costs of Safety," New Scientist, 68, 973, October 20, 1975, pp. 274-275.

17. Operational Accidents and Radiation Exposure Experience 1943-1975,
WASH-1192, 1975.
18. "On the Evaluation of Human Life in Accident Studies," Accident Analysis and Prevention, Pergamon Press, 5, 1973, pp. 287-293.

11. BIBLIOGRAPHY

M. G. Bullock, Safety Information System Guide, ERDA 76-45/9, SSDC-9, March 1977.

N. W. Knox and R. W. Eicher, MORT User's Manual, ERDA-76-45-4, SSDC-4, March 1976.

Fault Tree for Safety, Support Systems Engineering, Boeing Military Aircraft Product Development, November 1968.

W. D. Rowe, Draft, An Anatomy of Risk, Environmental Protection Agency, March 1975.

B. L. Cohen and I-Ling Lee, "Catalog of Risks," Journal of Health Physics, June 1979.

APPENDIX A
RISK IDENTIFICATION TREE



APPENDIX A

RISK IDENTIFICATION TREE

A logic structure in the form of risk identification tree can be used to make a thorough search for all hazards. It will minimize the number of hazards which may otherwise be overlooked.

Block 1.0 of the Risk Identification Tree defines the objective, i.e., bringing to the attention of management the "Residual Operational Risks" remaining after the risk analysis has been completed, and corrective action has been taken to eliminate and control major risks. Subordinate blocks (1.0 through 5.0) define the necessary and sufficient conditions to achieve fulfillment of the objective stated in Block 1.0. These conditions are:

1. All energy sources must be identified
2. All potential targets of uncontrolled energy release must be identified for each energy source
3. All control mechanisms and barriers to energy release must be identified for each energy source
4. An analysis must be performed in each case to determine failure modes and effects, in order to identify the residual risks.

The balance of the tree provides a guide for identifying all energy sources. The two lower tiers on Page A-4 identify the various forms of energy. The transfer symbols relate to tabulations on Pages A-5 through 12 of specific risk situations. The tabulations are general in nature, but are traceable to specific hardware, locations, and organizational units.

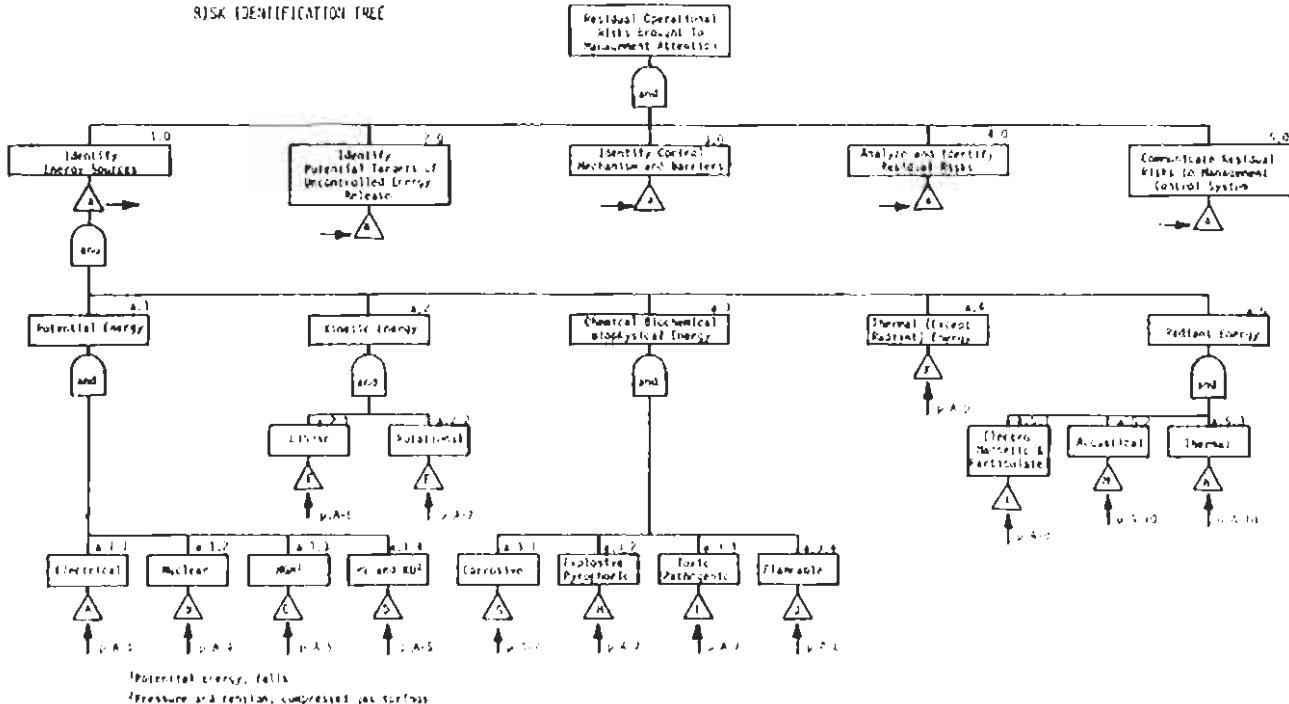
Identifying of all energy sources and tracing them to specific hardware has the primary benefit of preventing oversight of specific hazards. The safety analyses associated with Conditions 2, 3, and 4, above, are time-consuming. Therefore, the high risk energy sources should be considered first and the analytical effort scaled to the degree of risk. The selection

and scaling should be made by safety specialists. Although the selection of high risk hazards does not quantify the risk, it will help to prevent oversight of high risk areas.

The tree provides a logical sequence for a structured search of hazards which include the following steps:

1. Identify each energy source. An energy source is any material or condition which could result in a release of energy. Examples are combustibles, toxic substances, corrosive materials, electric or radiant energy, moving objects or machinery and objects which could fall or drop. A list of energy types are given in Table 6 in the main text. A list with many examples are given in the Safety Analysis Guide SSDC.
2. Identify all potential targets of uncontrolled energy release for each energy source equipment, facilities, employers, public, and environment should all be considered.
3. Identify all control mechanisms and procedures for each energy source and target. These include time, space, and physical guards or barriers.
4. Perform an analysis to determine failure modes and effects to identify the residual risks. (This only identifies the risks; probability estimates are necessary to quantify the risk.)

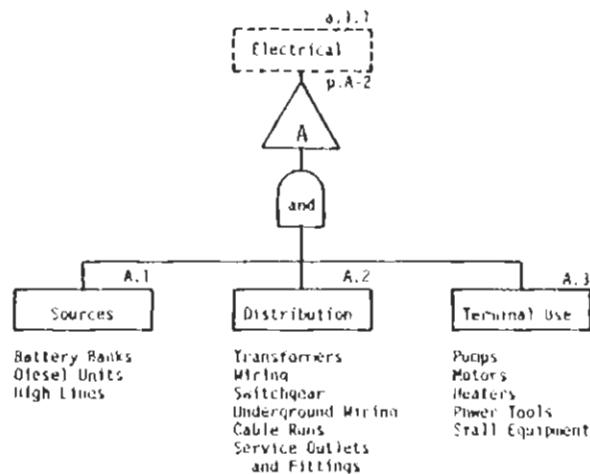
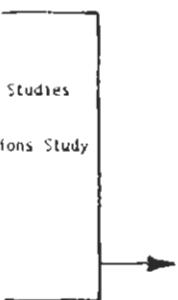
An understanding of this method will alert a person to hazards in making plant inspections or in just walking through the plant even though a structured search is not conducted. A thorough search using the tree will minimize hazard oversight. Since the procedure is time consuming, high energy sources could be considered (Steps 2, 3, and 4) first.



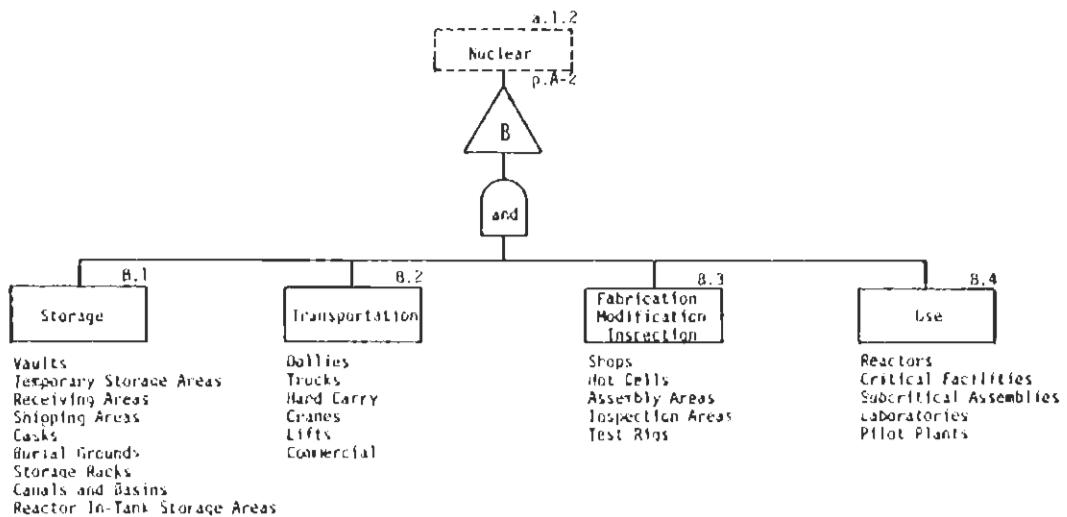
TYPICAL LOWER TIER STRUCTURE
(ENERGY SOURCES GEOGRAPHICALLY FIXED)

Inputs

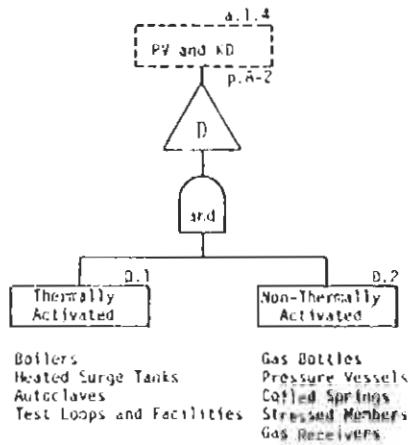
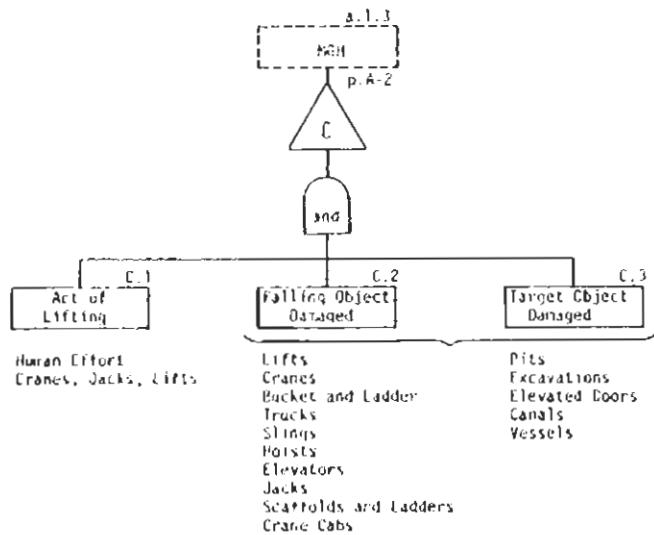
- Field Monitoring Programs
- Statistical Data Stores
- Reported Significant Observation Studies
- Area Safety Manager Reports
- Fire Safety and Operation Conditions Study
- Priority Problem Lists
- ROT Incident Reports
- OSHA Inspections
- Design Studies and Analyses
- Safety Engineer Lists

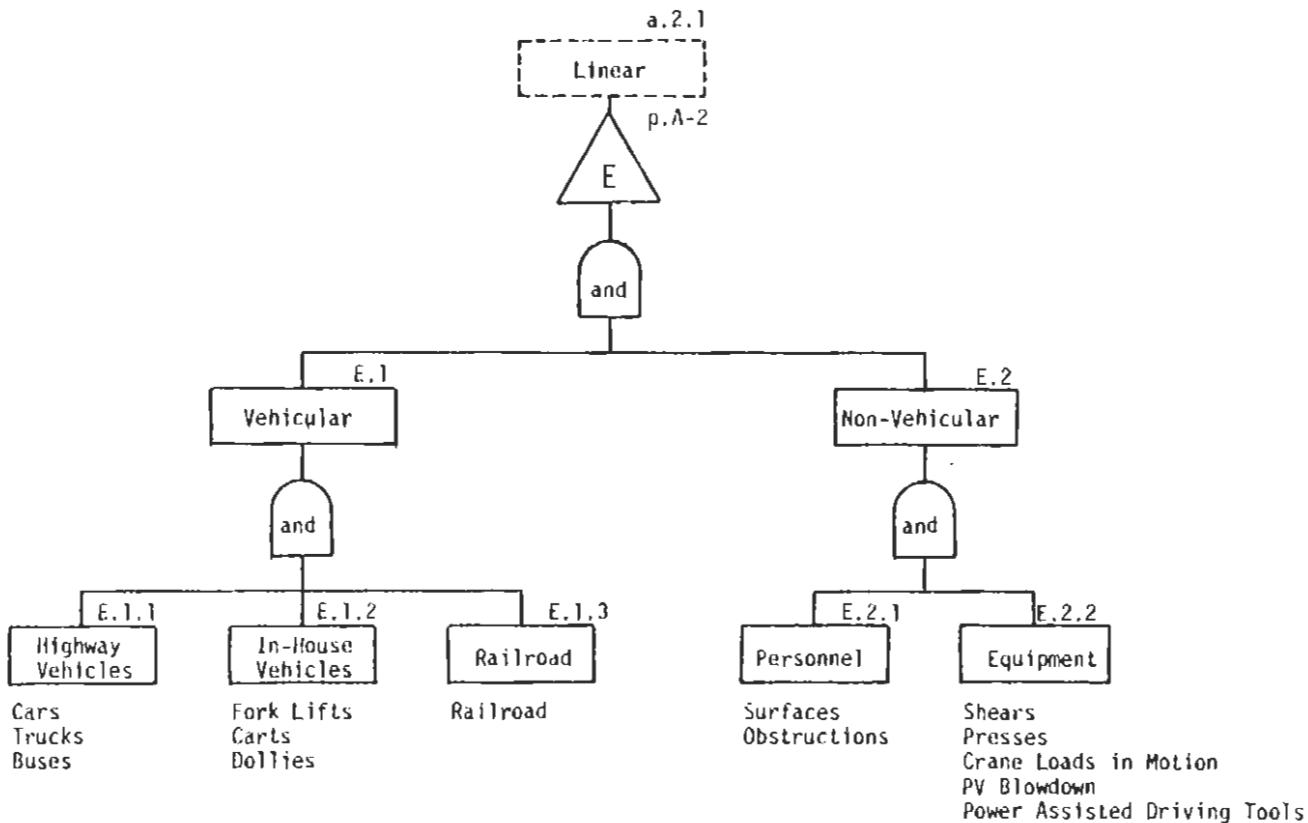


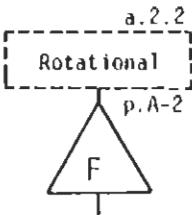
TYPICAL LOWER TIER STRUCTURE
(ENERGY SOURCES GEOGRAPHICALLY MOBILE)



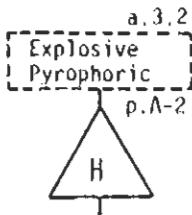
A-8



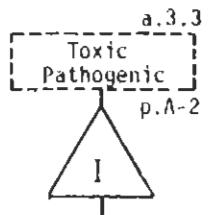




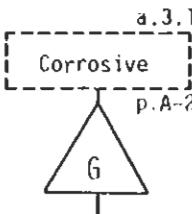
Centrifuges
Motors
Pumps
Cooling Tower Fans
Cafeteria Equipment
Laundry Equipment
Gears
Shop Equipment (Grinders,
Saws, Brushes, etc.)
Floor Polishers



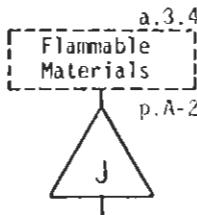
Caps
Primer Cord
Dynamite
Powder Metallurgy
Dusts
Hydrogen (Incl. Battery
Banks and Water Decomp.)
Gases - Other
Gd Nitrate
Electric Squibbs
Peroxides - Superoxides



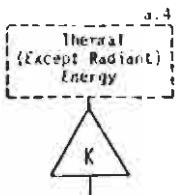
Acetone
Flourides
Co
Lead
Ammonia
Asbestos
Trichloroethylene
Dusts and Particulates
Pesticides - Herbicides -
Insecticides
Bacteria
Be
Chlorine
Decon Solutions
Sandblast
Metal Plating
Asphyxiation - Drowning



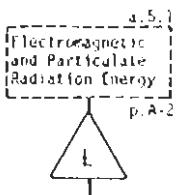
Acids
Caustics
"Natural" Chemicals
(Soil, Air, Water)
Decon Solutions



Packing Material
Rags
Gasoline (Storage and In Vehicles)
Lube Oil
Coolant Oil
Paint Solvent
Diesel Fuel
Buildings and Contents
Trailers and Contents
Grease
Hydrogen (Incl. Battery Banks)
Gases - Other
Spray Paint
Solvent Vats



Convection
Heavy Metal Weld Preheat
Exposed Steam Pipes
Electric Heaters
Fire Boxes
Lead Melt Pot
Electrical Wiring and Equipment
Furnaces



p.A-2



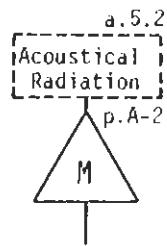
Canals
Plug Storage
Storage Areas
Storage Buildings

"Sources"
Rattle and Scrap
Contamination
Irrad. Experimental
and Reactor Equip.

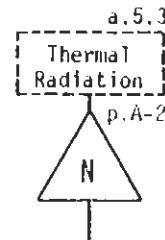
"Sources"
Waste and Scrap
Contamination
Irrad. Experimental
and Reactor Equip.

Electric Furnace
Blacklight (e.g., Magniflux)
Laser
Medical X-Ray
Radiography Equipment
Welding
Electric Arc - Other (High Current CKTS)
Electron Beam

A-13



Equipment Noise
Ultrasonic Cleaners



Furnaces
Boilers
Steam Lines
Lab and Pilot Plant
Equipment

APPENDIX B
USE OF RISK PROJECTION TECHNIQUES
IN INVESTIGATION OF ACCIDENTS AND INCIDENTS

APPENDIX B
USE OF RISK PROJECTION TECHNIQUES
IN INVESTIGATION OF ACCIDENTS AND INCIDENTS

1. Introduction

An important factor in investigation of accidents or incidents is that of relating the accident or incident under investigation to normal behavior of the organization experiencing the event.

"Item fixes" (remedial actions based on the specifics of the accident) and "system fixes" (remedial actions having to do with the overall control system within which the accident or incident occurred) form an essential part of reaction to any accident or incident. In order to frame effective remedial recommendations, it is, however, necessary to understand the degree to which the accident is typical of organizational behavior. When related organizational operating experience exists, there are two statistical methods for relating the consequences of a severe accident to normal organizational behavior. The first method involves study of past frequency-severity data and derivation of the probability (expected frequency) of the event under study. The second method makes use of only the more serious events experienced in the past to derive the likelihood that the accident under study represents "normal" behavior of the organizational control system.

These two basic methods will be discussed and compared. In both cases, the discussion will be based on standard forecasting methodology, i.e., the predicting of future performance based on past performance. In the context of an accident or incident, the questions to be answered are: "Would an event this serious have been expected in terms of normal system behavior?" If the answer is affirmative, emphasis must be placed on correcting the entire control system within which the accident occurred. If the answer is negative, one must determine the unique characteristics of the particular accident or incident which allowed escape from normal levels of control and seek to eliminate similar escape in the future.

1.1 Frequency-Severity Distributions

As we have indicated, frequency-severity distributions utilize the entire spectrum of experience data, ranging from events of trivial consequence to the most serious events experienced by the organization and system under study.

Figure B-1 illustrates the basic frequency-severity matrix in simplified (logarithmic) form. The "line-of-balance" indicates a situation in

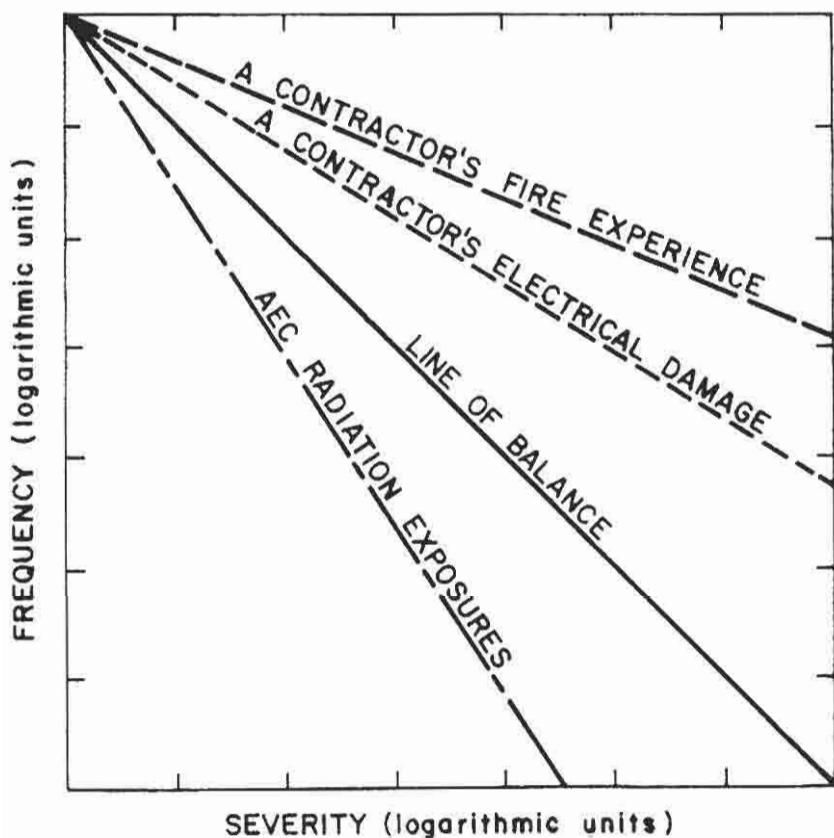


Figure B-1. Accident frequency-consequence relationship illustrating line of balance.

which losses are balanced in the sense that a single \$1000 loss is controlled with the same effectiveness as one thousand \$1.00 losses. Such balance is not necessarily "good" or "bad" but forms a convenient frame of reference in evaluating the nature of the control system.

As may be seen, the AEC radiation exposure distribution indicates that the system is relatively permissive in permitting low level exposures, but is more restrictive in the upper exposure levels. This result is not surprising if one studies the nature of the system constraints and controls utilized by AEC during the period under study.

The two contractor distributions relating to the fire and electrical losses represent a different situation in that the system was more permissive in case of the severe consequences. This situation has high potential for resulting in a "fools paradise" situation. This occurs because the high frequency events having less severe consequences are often relatively well controlled. When the infrequent events having more serious consequences occur, they may be too easily rationalized as unusual and isolated events. This, in fact, occurred to some degree in both contractor organizations, whose performance is indicated in Figure B-1, and relatively severe accidents appeared as "surprises" due to inadequate information and reaction to the less severe precursor accidents.

Once an accident has occurred and the consequences are evaluated, one may directly enter frequency-severity distributions of the type indicated in Figure B-1 to determine the expected frequency of such an accident. An alternative to this procedure involves use of curves of the type shown in Figure B-2. The curves in Figure B-2 are obtained by integrating the frequency-severity curves to obtain the probability of exceeding a specified consequence level. Figure B-2 represents such a study performed a number of years ago relating to radioactive shipments.

The Figure B-2 curves indicate the probability of exceeding a given dollar loss on a per shipment and on a per year basis for a given operation (identified as the "ITS" operation). Since the derivation of predicted loss for the operation under study was prepared largely on the basis of theoretical data (due to lack of actual accident data), AEC data, which

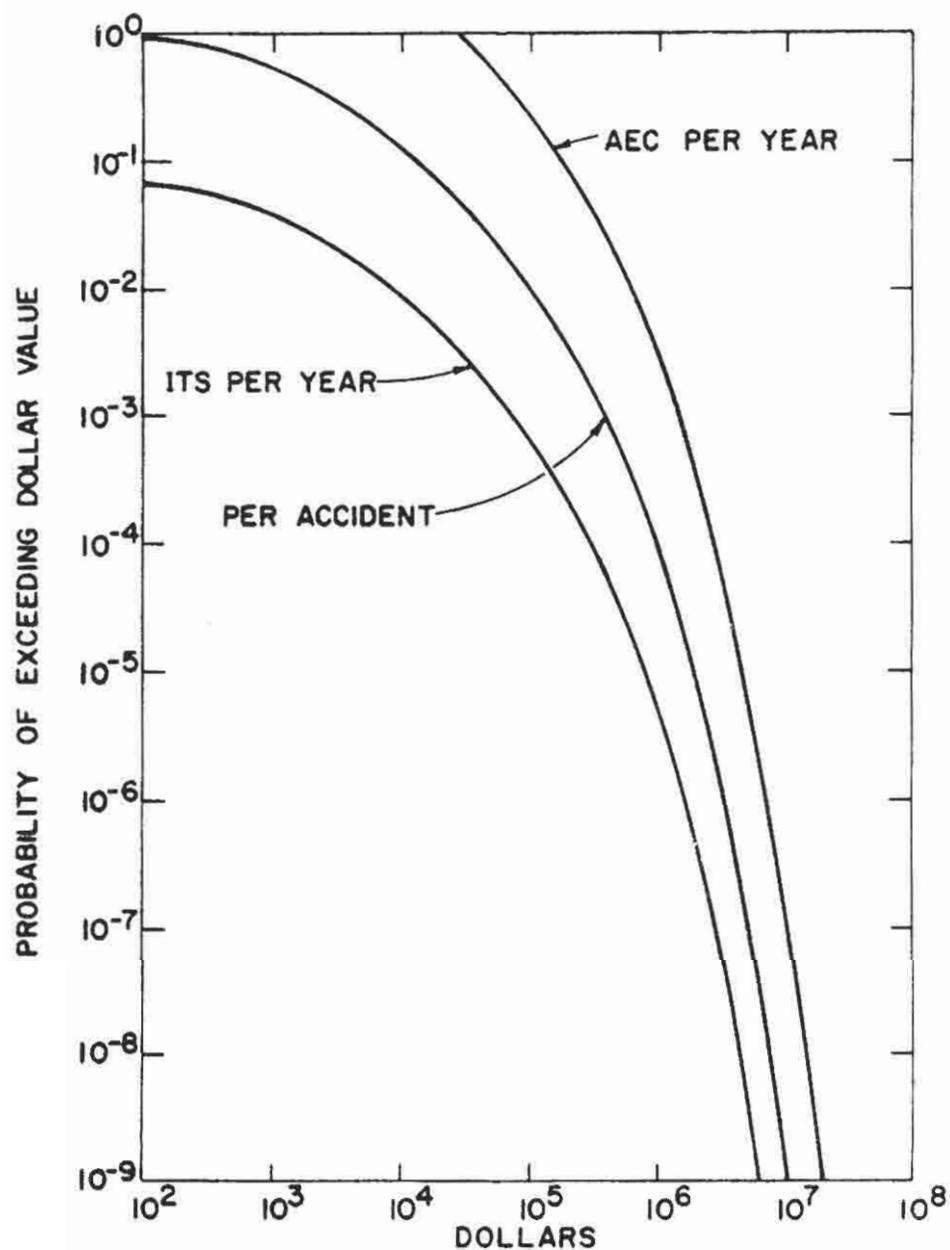


Figure B-2. Risks of shipping radioactive materials by truck.

includes many more such shipments, were used to construct the "AEC" curve. This is done to provide an envelope and to evaluate the reasonableness of predictions relating to the specific "ITS" operation.

1.2 Extreme Value Analysis

As indicated earlier, application of frequency-severity distributions require use of all data generated by the organization ranging from low frequency-severe consequence to high frequency-low consequence levels. Furthermore, these data may be distributed in a variety of ways which require sophisticated statistical analysis.

This leads to a number of difficulties in use of frequency-severity distributions:

1. Large quantities of data must often be processed (the AEC radiation exposure curve in Figure B-1 represents analysis of more than one million pieces of data).
2. Low severity data are often not readily available or may not exist at all (due to reporting thresholds).
3. The great mass of the data will ordinarily lie in the low severity range. If one does not use extreme care in analysis, one can arrive at a situation in which "the tail wags the dog" in the sense that a large mass of low consequence data predominate in predicting the severe consequence events (which we are interested in).
4. One must use extreme care in selecting, using, and testing statistical distribution functions in order to avoid invalid conclusions.

This raises the question as to whether other predictive and analytical methods exist which:

1. Require a smaller quantity of more readily available data.

2. Are self-testing as to applicability.

The extreme value analysis of Gumbel⁷ represents such a method. This method is described in MORT and in Gumbel's excellent monograph referenced in MORT. We will not go into the detailed theory or specifics of application here. In brief, the method involves:

1. Selecting a period of homogeneous operation prior to the event under study.
2. Breaking the period of time down into appropriate intervals.
3. Obtaining the most severe consequence event for each of the time intervals.
4. Plotting the worst-case events on the special extreme value paper in accordance with the rules provided in the MORT manual.
5. Testing for applicability. (Does the data plot as a straight line?)
6. Determining how the accident event relates to the normal system behavior as indicated by the remaining data points.
7. Structuring the investigation and recommendations in accordance with Item 6, above.

Figures B-3, -4, and -5 indicate application of extreme value analysis to AEC property loss data. Figure B-3 indicates the results using raw loss data. In this case, only a small portion of the more recent data fit the straight line distribution. Examination of the data reveals that the dollar value of AEC property placed at risk, as well as the property losses, grew at a high rate during the period under study, i.e., the situation changed rapidly.

This suggests use of loss ratios (cents loss per \$100 property owned by AEC). Loss ratio as a function of time is indicated in Figure B-4. If one now plots the loss ratio figures on extreme value paper, it may be seen

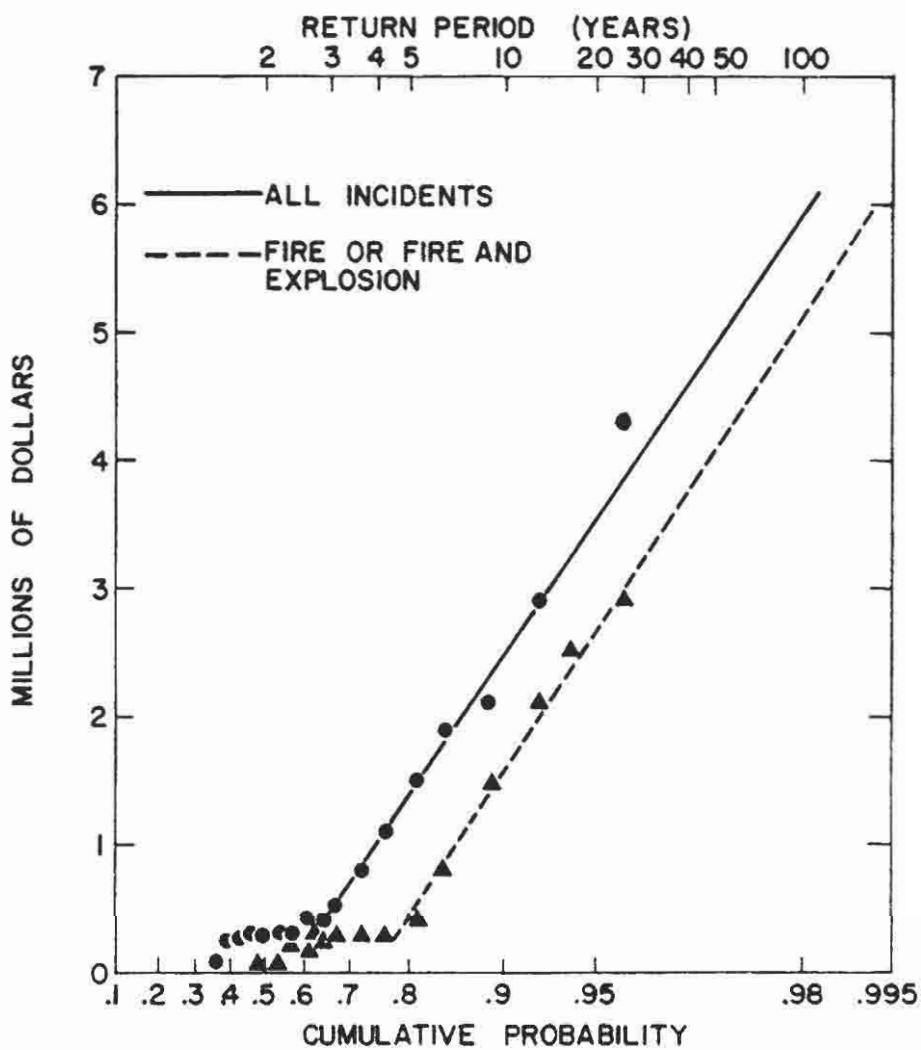


Figure B-3. AEC incidents from 1943 to 1967 (property damage or loss).

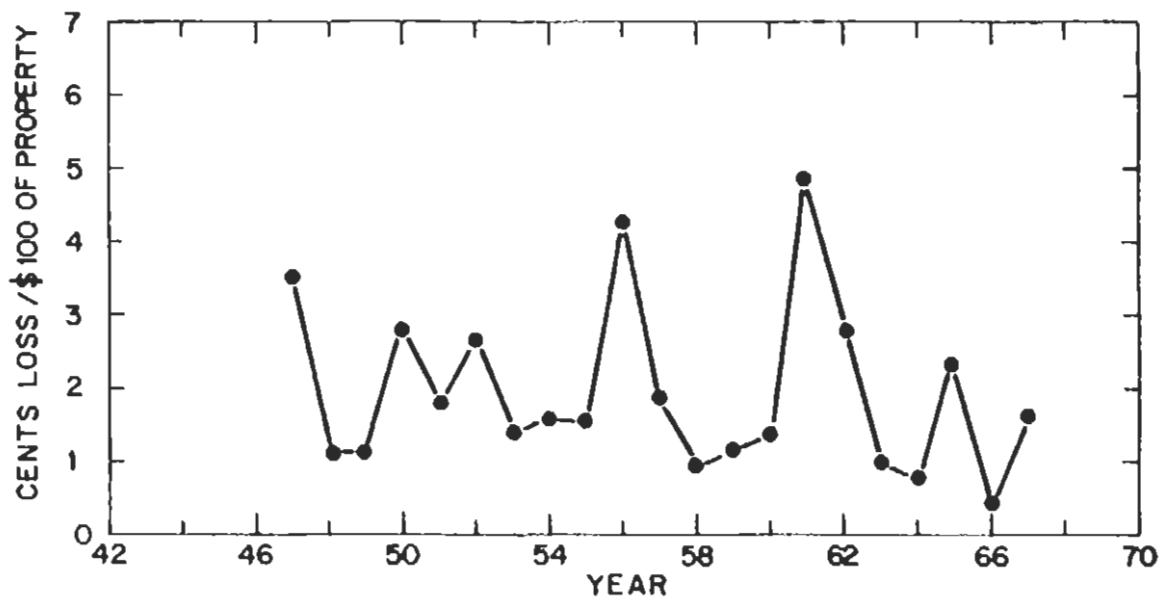


Figure B-4. Property damage ratio vs. time from 1947 to 1967.

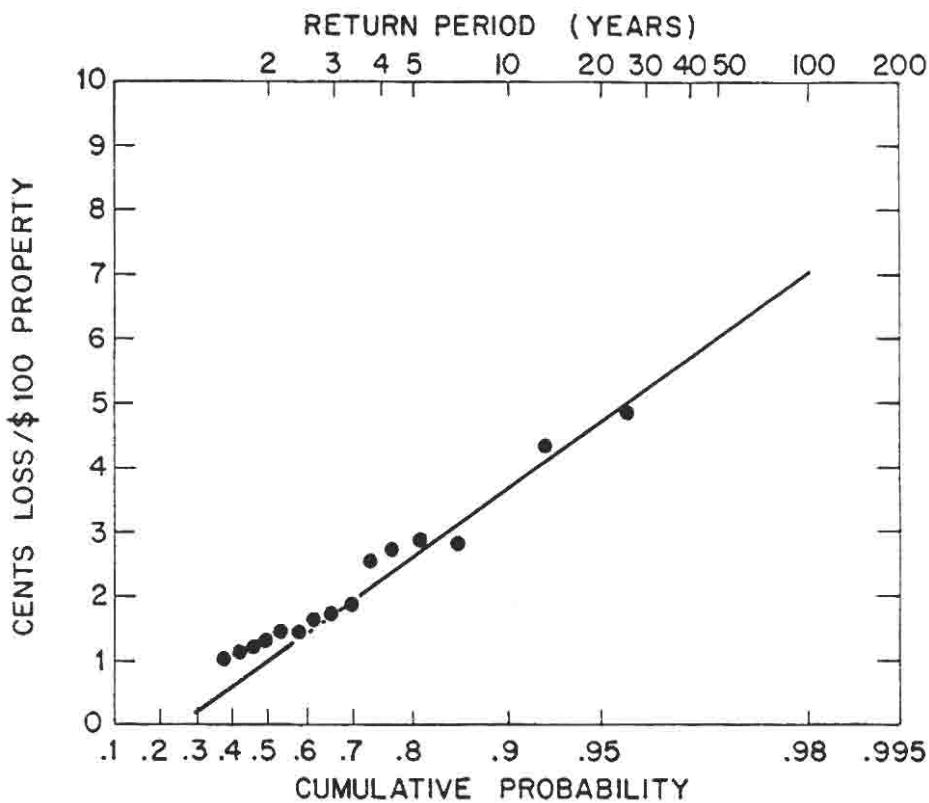


Figure B-5. Property damage ratio from 1967 (cents/\$100 property).

that the required linear relationship is achieved and the expected "most probable maximum" loss ratio over any period of observation may be obtained from the return period scale on the top of the extreme value paper (Figure B-5). Figure B-6 represents an example of the extreme value method for single event property loss for a single contractor.

2. Extreme Value Analysis Exercises

2.1 Case One

2.1.1 Situation. You are investigating accidents in two plants identified as Plants A and B. These accidents involve 6000 unit losses in

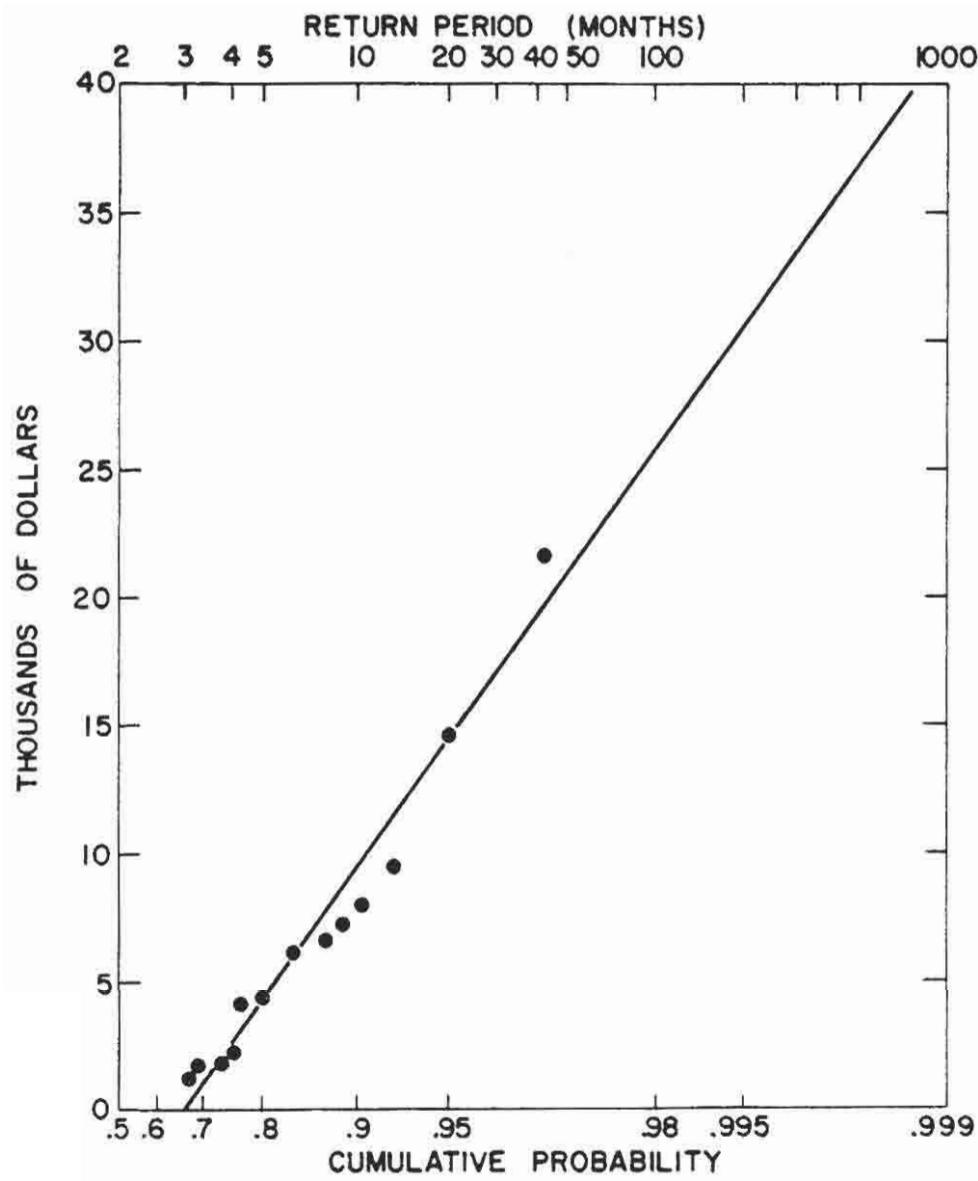


Figure B-6 Property loss for single events from 01/68 thru 02/71
(38 months).

both plants. You request and obtain historical data, given in Table B-1, indicating the maximum similar losses in Plants A and B for each year over the past 9 years. Investigation indicates the operating modus operandi for the two plants has been essentially constant for "the past 10 years," and that effective maintenance programs have kept the plants "in good shape."

TABLE B-1. CASE ONE DATA FOR EXTREME VALUE EXERCISE

Year	Maximum Single Event Loss	
	Plant A	Plant B
1972	6000 ^a	6000 ^a
1971	3600	1600
1970	2400	600
1969	1200	2100
1968	600	500
1967	1600	1000
1966	4200	1500
1965	<500	1000
1964	2600	1900

a. "Accident" under current investigation.

2.1.2 Exercise. The data as prepared for plotting are given in Table B-2. Plot the data for the two plants on extreme value paper, as in Figure B-7.

2.1.3 Questions

1. Is extreme value projection valid for these two plants? How do you know this?
2. In terms of extreme value analysis, how would you expect the course of the investigation and the nature of the recommendations to differ for Plants A and B?

TABLE B-2. CASE ONE DATA PREPARED FOR PLOTTING
(Figure B-7)

Cumulative Probability	Maximum Single Event Loss	
	Plant A	Plant B
0.90	6000	6000
0.80	4200	2100
0.70	3600	1900
0.60	2600	1600
0.50	2400	1500
0.40	1600	1000
0.30	1200	1000
0.20	600	600
0.10	<500	<500

3. What is the significance of the information that the operating methods and maintenance have been essentially constant for the past 10 years? How does the extreme value analysis validate or fail to validate this information?
4. What difficulties might have been experienced if one had utilized frequency-severity data rather than extreme value analysis in this case?

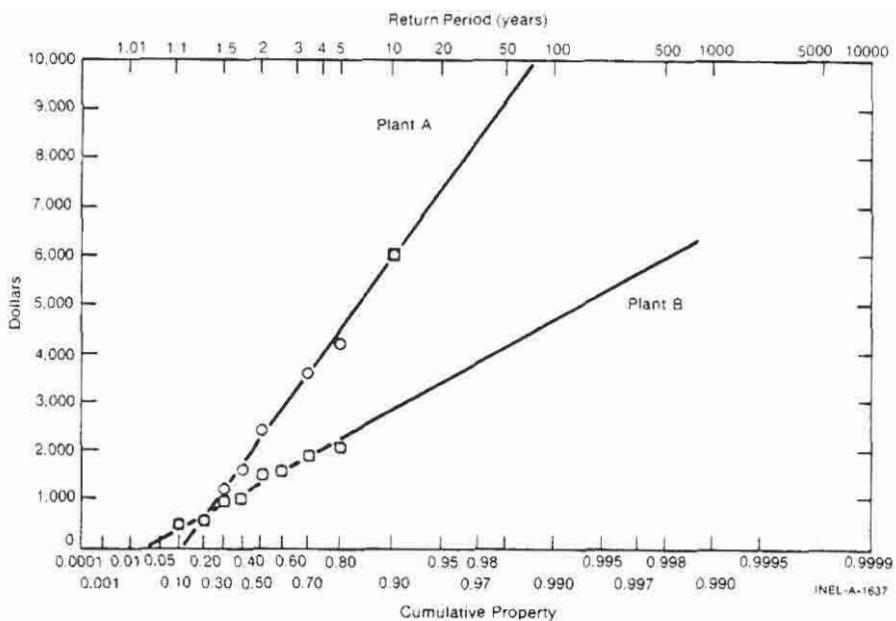


Figure B-7. Case 1 data plotted on linear scale extreme value paper.

2.1.4 Answers

1. Yes, extreme value projection is valid because most of the data fit a straight line in both cases.
2. Accident control system weakness is indicated for Plant A by the steep slope of the Plant A curve. The relatively flat slope of the Plant B curve indicates a relatively good control system. In addition, for Plant B, the \$6000 accident is an outlier indicating that the causes for this "norm." Thus, investigation for the Plant A \$6000 accident should spend more time on general management oversights and omissions, while the investigation for Plant B should spend relatively more time on change analysis and specific fix.

3. There have been no major changes which would change or distort the accident frequency-severity distribution. The fact that the data do fit a straight line validates that no external influence has perturbed the system.
4. It is usually more difficult to obtain consistent data on all accidents over a 10-year period, that is to get information about the largest accident. In addition, a much larger quantity of data must be analyzed.

2.2 Case Two

2.2.1 Situation. You are investigating fatal accidents in two contractor organizations identified as Contractors X and Y. Neither organization had experienced a fatality during the past 10 years. You request and receive disabling injury data for the two plants in terms of the maximum disabling injury during each 6-month period for the past 10 years. When the data are ranked according to magnitude and related to cumulative probability in the usual manner, the results indicated in Table B-3 are obtained.

2.2.2 Exercise. Plot the data in Table B-3 on the linear scale extreme value paper, as in Figure B-8, and on the log scale extreme value paper, as in Figure B-9, for both contractors.

2.2.3 Questions

1. Which representation is more appropriate for Contractor X, the linear or the log scale?
2. The statistician advises us that the log representation is more appropriate for situations in which the limits do not exist on how bad things can get at the high severity end of the curve. In this light, what can we say about basic energy-safety controls in the Contractor X organization?

TABLE B-3. MAXIMUM DAYS CHARGED FOR DISABLING INJURY DURING EACH SIX-MONTH PERIOD

Cumulative Probability	Contractor	
	X	Y
0.95	3700	190
0.90	1500	180
0.85	900	180
0.80	900	150
0.75	600	140
0.70	450	125
0.65	450	110
0.60	350	85
0.55	300	78
0.50	170	74
0.45	140	66
0.40	120	65
0.35	120	50
0.30	120	49
0.25	90	48
0.20	40	47
0.15	11	41
0.10	11	29
0.05	None reported	27

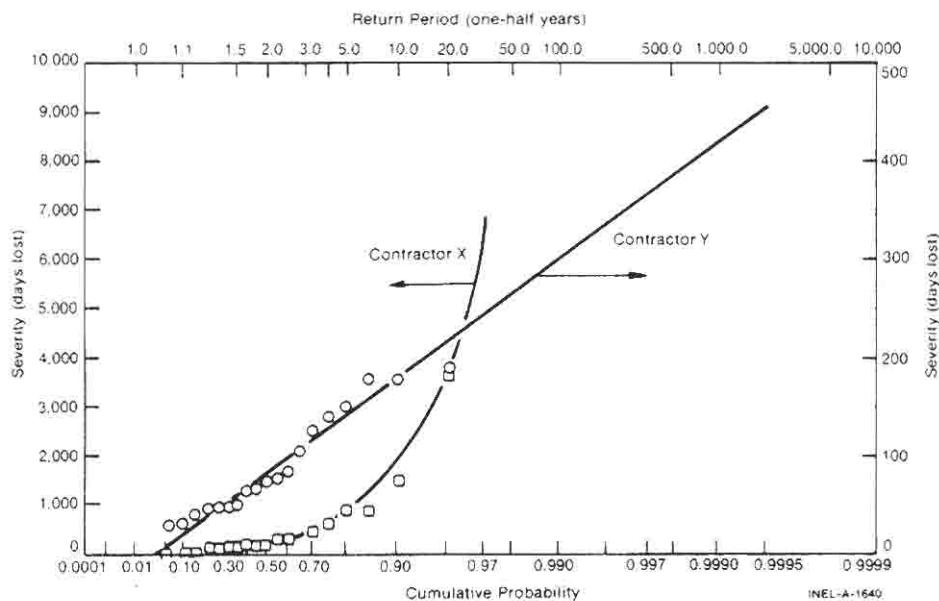


Figure B-8. Case 2 data plotted on linear scale extreme value paper.

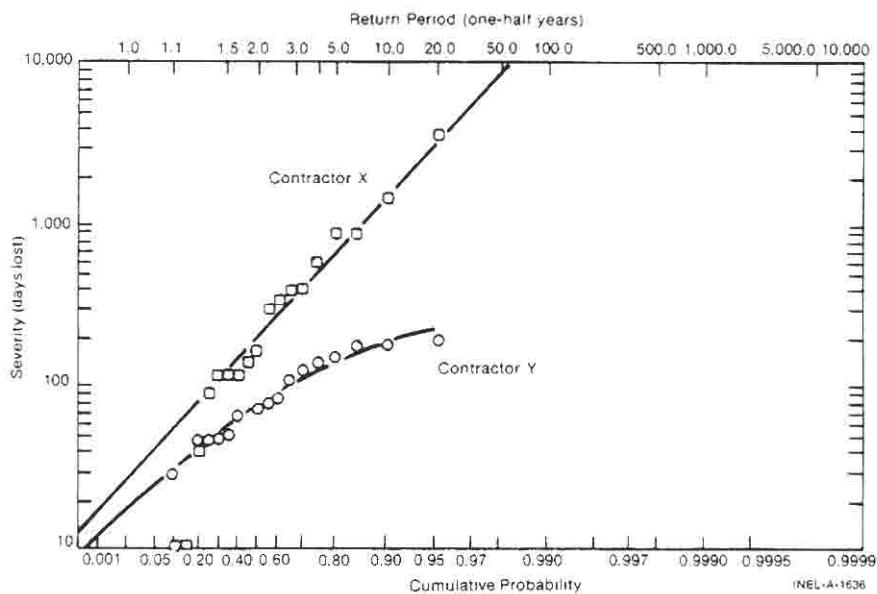


Figure B-9. Case 2 data plotted on logarithmic scale extreme value paper.

3. Both Contractors X and Y management advise us that no changes should be made in their basic safety controls and programs as a result of the deaths. How would you respond to these arguments?
4. Using the convention that a death is "equivalent" to 6000 man-days loss, how often would one anticipate a fatality in each of these contractor organizations, if no changes occur in their basic safety controls?

Answers

1. Logarithmic scale--Contractor X data approximates a straight line on the log scale in Figure B-9.
2. The basic energy-safety controls are probably inadequate or nonexistent for Contractor X.
3. The extreme value projection for Contractor X indicates a high frequency for very serious accidents, and thus a high probability of an accident so severe as to cause death. Unless the safety program is reoriented toward high severity injuries, more deaths will occur. I would agree with Contractor Y, since the extreme value projection indicates a low frequency of the fatality being repeated.
4. For Contractor X, approximately 16 years. For Contractor Y, the extreme value projection indicates that control of severe accidents has been so good as to virtually eliminate the possibility of a fatality from those types of accidents causing injuries. The probability is so low as to make prediction unreliable.

3. Log-normal Frequency-Severity Exercise

3.1 Case Three

3.1.1 Situation. You are investigating two types of property damage accidents, A and B. For Type A, you obtain cost-frequency data and construct the following, Table B-4.

From the data in Table B-4, you plot the log-normal distribution given in Figure B-10.

For Type B events, you request and obtain the data in Table B-5.

The data for both Types A and B events represent 5 years of actual experience. The maximum Type A event is \$49,700; the maximum Type B event is \$4140.

3.1.2 Exercise. Calculate the necessary data and plot the log-normal curve for Type B property damage, as also given in Figure B-10. (Draw a straight line ignoring any point which appears to be an outlier.)

3.1.3 Questions

1. Does the \$4140 incident represent normal behavior for Type B events?

TABLE B-4. TYPE A PROPERTY DAMAGE

Cost Range (\$)	Events	Accumulative Events	Less Than Cost $[N_i \div (N_i + 1) \times 100]$ (%)
10 to 100	9	9	39
101 to 1000	10	19	83
1001 to 10,000	2	21	91
10,001 to 50,000	1	22	96

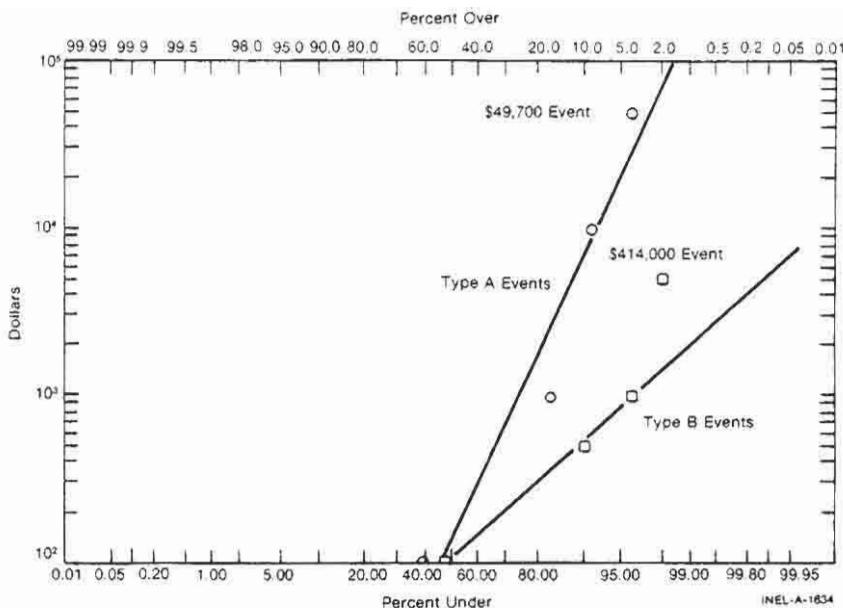


Figure B-10. Log-normal frequency-severity exercise. Property damage data plotted on log-normal paper.

TABLE B-5. TYPE B PROPERTY DAMAGE

<u>Cost Range (\$)</u>	<u>Events</u>
0 to 25	6
26 to 100	19
101 to 500	21
501 to 1000	3
1000 to 5000	1

2. Does the \$49,700 incident represent normal behavior for Type A events?

3. Verify your answer by calculating how often in years a >\$50,000 Type A and a >\$4000 Type B event would occur. (Hint: From the

percent over scale, determine what fraction of the events exceed the cost value in question. From this and the number of events per year, the frequency in years can be calculated.)

4. With regard to the \$4140 incident, should one be more concerned with the control system or specific condition requiring correction? With regard to the \$49,700 incident?
5. From the data, which cost ranges represent the greatest risk for Types A and B events? (risk = expected loss = consequence \times frequency - the relative risks may be approximated by the average cost in each range multiplied by the number of events in that range.)
6. When should log-normal be used rather than extreme value?

3.1.4 Answers

1. No, the \$4140 Type B event is an outlier.
2. Yes, the \$49,700 Type A event lies close to the curve and is part of the log-normal population.
3. The 22 Type A accidents in the 5-year period are equivalent to 100 accidents in 22.7 years. From Figure B-10, the \$50,000 severity level occurs at approximately 2.8 "percent over," which indicates that for each 100 accidents, 2.8 accidents each greater than \$50,000 will occur each 22.7 years or one \$50,000 accident each 8 years. (Or, 2.8% of 4.4 accidents/year is 0.123 accidents greater than \$50,000/year: $1/0.123 = 8$.) The 50 Type B accidents in the 5-year period are equivalent to 100 accidents in 10 years. From Figure B-10, approximately 0.20 accidents greater than \$4000 will occur for each 100 accidents in the 10-year interval or 1 accident greater than \$4000 each 50 years. (Or, 0.2% of 50 accidents/year is 2×10^{-2} accidents/year, or 50 years/accident for those greater than \$4000.)

4. For the \$4140 incident, the specific condition requiring correction should be sufficient. The control system applicable to Type B events should be investigated in addition to correcting the specific condition. (MORT analysis should be done in each case; it is more urgent for the Type A events.)
5. For Type A events, the top severity range is most important (one \$50,000 event is worse than two \$10,000 events, etc.). For Type B events, the \$100 to \$500 range represents the greatest loss.
6. Log-normal should be used when the number of events in each time period is small, or when additional information beyond predicting the return period for large events is desired.

4. Log-Log Frequency-Severity Exercise

4.1 Case Four

4.1.1 Situation. You desire further analysis of the Types A and B data given in the log-normal exercise. You transfer the Type A log-normal curve (accumulative frequency vs. severity) to log-log paper, using the "percent over" and "cost greater than" values. (This curve is shown in Figure B-11.)

4.1.2 Exercise. Transfer the Type B log-normal curve (accumulative frequency vs. severity) to log-log paper.

4.1.3 Questions

1. Compare the two log-log curves for Types A and B events. Which curve has the greatest slope? (Note that the risk is increasing where the slope is <1 and decreasing where the slope >1 .)
2. What are the peculiar advantages of log-normal and log-log curve?

4.1.4 Answers

1. The greater negative slope for Type B events indicates a low risk for large events.
2. The log-normal curve can be more accurately extrapolated since it is usually a straight line. The frequency-severity relationship or line-of-balance can be determined from visual inspection of the log-log curve.

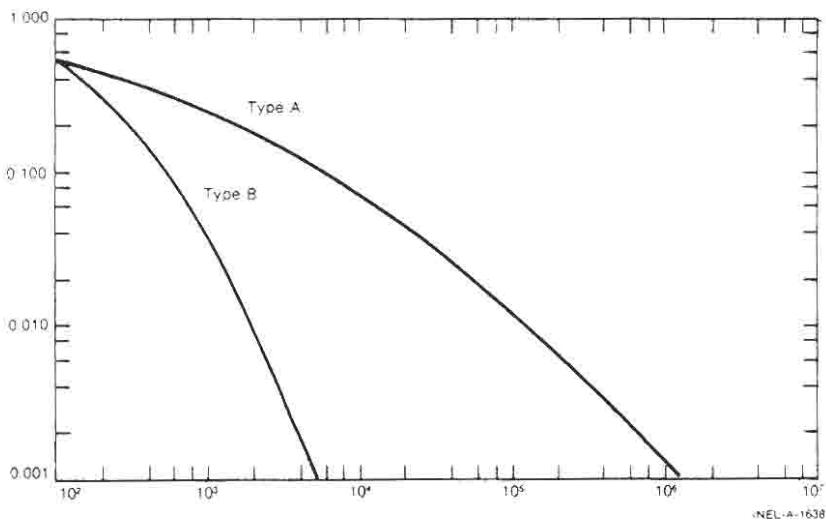


Figure B-11. Log-log frequency-severity exercise. Property damage data plotted on log-log paper.

APPENDIX C

PROBABILITY AND STATISTICS PRIMER

This appendix provides a basic introduction to probability theory and statistical distribution. Some additional detail is provided in the body of the report. Many textbooks and references are available for further study. Recommended are "Introduction to Statistical Analysis" by W. J. Dixon and F. J. Massey and "Statistical Analysis" by Bennet and Franklin (particularly Chapter II for trend analysis and tests for randomness), TEAM Associates, P.O. Box 25 Tamworth, NH 03886, can provide graph paper and methods for their use.

The theory of probability deals with the chance occurrence of random events. Random, as defined in the dictionary, means lacking a specific pattern or causal relationship, haphazard. Chance refers to the nature of unpredictable events. An example is that of nine black balls and one white ball in a black bag. What is the chance of picking a white ball? The position of the balls and the selection are completely uncontrolled; thus each ball has an equal chance of being picked. The probability is simple--1 chance in 10 or 0.1. However, if the experiment is repeated 100 times, the white ball would not necessarily be picked 10 times. The best we can predict is that it would be picked $10 \pm X$ times, the value of X depending upon probability theory. It is possible that the white ball could be picked any number of times from 0 to 100 times. The likelihood of any number in this range can be calculated. Of course, the chance of the white ball being picked 100 times is virtually zero, it being 10^{-100} (a decimal fraction preceded by 100 zeros).

Chance events are not determined by luck, but are events which result from activities from which more than one outcome is possible. Chance also refers to an unplanned event which results from a combination or interaction of conditions and/or activities which are not sufficiently monitored to permit a prediction of the exact time and place of occurrence. As such, accidents are chance events which can be analyzed using probability and statistics. A person who habitually crosses a road without looking will eventually be hit by a vehicle, but without knowing

the exact time of crossing or passing of vehicles, we cannot predict which crossing will result in the accident. With sufficient crossing and traffic information, we could, however, deduce the probability of being hit by a vehicle.

Probability refers to the chance or likelihood of a specific event occurring given an opportunity for its occurrence. Probability values range from zero (impossible to occur) to 1.0 (certain to occur). A probability of 0.5 means that the occurrence or nonoccurrence is equally likely. While probability refers to the likelihood of a specific event given a single opportunity for its occurrence, statistics deal with the number of times an event will occur given many opportunities. Statistics also deal with the variation in the "probable" numbers of events. For example, if we toss a coin 10 times, the probable number of heads is 5. This means that if we repeated the 10 tosses many times, 5 heads would occur most frequently. The frequency of 4 or 6 heads would be less frequent, 3 or 7 heads even less frequent, and so on. The relative frequency of the number of heads falls into a well-defined pattern, called a statistical distribution of a variable. The variable is the number of heads in the experimental 10 tosses.

These two concepts of (a) probability of an event and (b) the statistical distribution of variables are fundamental to risk analyses. A few basic laws of probability and a discussion of statistical distributions are presented in this section.

1. Probability

Mathematically, probability is defined as the number of times an event will produce a given result divided by the total number of events. The probability value can be deduced or inferred. For example, with a throw of a single die, there are six possible outcomes or numbers. Any particular number (say a 5) has 1 chance in 6. Thus, the probability of a 5 is 1/6. This is a deduced probability and is exact if the die is not loaded and is in perfect symmetry.

If we cannot determine the equality and number of outcomes through reasoning, then we can, by trial, determine the approximate probability. If we throw the die 36 times, and the 5 comes up only 5 instead of 6 times, we would still conclude (because of the small number of trials) that the probability of a 5 is $1/6$, not $5/36$. On the other hand, if we suspected that the die was loaded, we might conclude that the probability of a 5 was about $5/36$, but could be $6/36$ or even $4/36$. Our best estimate is $5/36$. Now, if the die were thrown 36,000 times, and a 5 comes up 5,000 times, we conclude that the die is indeed loaded, and the probability of a 5 is very close to $5,000/36,000$ (or $5/36$) and very likely not $6/36$ or $4/36$. This probability of $5/36$ is inferred from observation. Statistics is the mathematics of inference.

The domain of statistical inference includes both estimating probability and its uncertainty from previous experience (how many times it has occurred in the past), and estimating how many times something will happen based on a deduced probability. The accuracy of these estimates is also determined from statistics. Probability can be relatively exact if deduced from known conditions. A coin, die, roulette wheel, all have very exact probabilities. But even though the probability of an outcome is known very precisely, the results of a small number of trials cannot be precisely predicted.

The most common method of estimating the probability of an accident is from previous experience. If a contractor has experienced 5 injuries for each 100 employees, the average employee injury probability is 0.05. This value multiplied by the number of employees will estimate the number of injuries for the next year. In probability language this estimated number is called the "expected" number and is referred to as "expectation." An average probability value may be grossly misleading if applied to an individual, who may have a relatively higher or lower risk than average.

There are no hard guidelines for avoiding errors when applying a general average to a specific situation or vice versa. However, the population from which a probability is estimated should be as similar as practical to the population to which it is applied. Even though small

groups within the population vary significantly, if the two populations have the same small group distribution, the variations within the two populations will average out.

For example, to estimate a vehicle accident probability:

1. Use U.S. Motor Vehicle Accident Statistics from the National Safety Council's "Accident Facts" for the average probability of a U.S. citizen
2. Use DOE vehicle statistics for the average probability of a DOE employee
3. Use the contractor experience for the average contractor employee
4. However, for a DOE bus driver, use national professional bus driver statistics if DOE bus statistics are unavailable.

Conditional probability is the probability of a consequence conditioned upon a prerequisite event. For example, what is the probability of an injury if a vehicle accident occurs? This conditional probability is obtained by dividing the number of injuries by the number of accidents. To obtain the probability of an injury per mile of travel, multiply the accident probability by the conditional probability:

$$\text{Probability of injury/mile} = \text{accidents/mile} \times \frac{\text{injuries}}{\text{accidents}}$$

The probability of an injury/mile can be obtained directly by dividing the number of injuries by the number of miles, but the example illustrates the concept of conditional probability.

Using this concept one can objectively estimate the probability of a fatality even though no fatalities have occurred, using the ratio of injuries to fatalities based on experience. Extreme care should be used in applying one type of experience to another because these ratios (or conditional probabilities) may vary widely.

The injury to fatality ratio for types of DOE and U.S. activities are given below:

Activity	Number of Lost Workday Injuries per Fatality
All U.S. industry	208 (170 per disabling injury)
U.S. construction industry	? (96 per disabling injury)
All DOE	148 (1977 through 1980)
DOE construction	384 (based on 3 deaths)
DOE services	192 (based on 2 deaths)

The average of these values, excluding the U.S. construction industry is 233 injuries/fatality. Notice that all of the ratios are within a factor of 2 of this value. This may not be true of office workers (one extreme) or parachute jumpers (other extreme). Nevertheless, estimates of this type, if used judiciously, can provide reasonable (within a factor of 2 or 3) estimates of probability. Calculating the probability from more than one source--such as total injuries, lost workday cases, days away, as well as from different types of industry--will give a range of values from which the uncertainty can be estimated.

Rules for probability calculations are:

1. The probability of an event not occurring is one minus the probability of the event occurring.

For example:

If the probability of a fatal vehicle accident is 1.6×10^{-4} /year, the probability of no fatal accident is $1 - 1.6 \times 10^{-4} = 0.99984$.

2. The probability of n independent events all occurring is the product of the probabilities of each event.

For example:

The probability of two individuals both dying in a vehicle accident in any 1 year is $(1.6 \times 10^{-4})^2$ or 2.56×10^{-8} . The probability of both living is 0.99984×0.99984 or 0.99968. Further, the probability of one individual not being killed in n years is 0.99984^n .

3. The probability of at least one event occurring is one minus the probability of no event. (The probabilities for all possible outcomes always total one.)

For example:

The probability of at least one of the two individuals having a fatal accident is one minus no fatal accident or $1 - 0.99968$ or 3.2×10^{-4} .

4. Small probabilities (<0.10) may be added with little error to determine the probability of either (any one) event occurring.

For example:

Small probabilities may be added (<0.1) with little error. The probability of either of two individuals having an accident is $1.6 \times 10^{-4} + 1.6 \times 10^{-4}$ or 3.2×10^{-4} .

2. Statistics

Common statistical terms and definitions are:

Distribution The frequency or manner in which observations of different values are distributed over the range of values. These values can be numbers, frequency, size, cost, severity, etc.

Frequency distribution	The relative frequency with which a variable quantity or variate assumes particular values.
Mode	The most common or frequent observation or value.
Mean	The arithmetic average of all observations or values.
Median	The point at which half the observations or values lie above and half below.
Variance	A measure of dispersion or variation in observations values. It is the summation of the squared difference between the mean and each value in the distribution, divided by the number of observations or values.
Standard deviation	The square root of the variance. It is a standard measure of dispersion: in the normal distribution for example, it is 68, 95, and 99.7% of all values occur between ± 1 , 2, or 3 standard deviations from the mean, respectively. Statistical tables provide values from which we can determine the fraction of observations lying within a specified deviation.
Skewed	The distribution is skewed if it is not balanced or symmetric around the mean. Most distributions found in nature, including accidents, are skewed. The reason is simple. In a symmetric distribution the mean is half way between the smallest and largest possible values. In nature, the largest value is usually more than twice the average, and frequently many times the average. Another example: with an average of only three accidents per year, zero is the fewest possible while more than six is possible. The smaller the average value relative to the maximum possible value, the greater the degree of skewness.

Confidence	The chance of likelihood that a specified value is <u>part of the population</u> (if a specific value lies outside \pm three standard deviations from the mean, then we have 99.7% confidence that this observation is different than the population since 99.7% of the population lies in that range).
Range	A measurement of the difference between two observations. The entire range is the difference between the smallest and largest value. We also speak of the inner two quartile range which includes 50% of all values (excluding the smallest 25% and the largest 25%).
Cumulative frequency (probability)	The frequency or probability which includes (or accumulates) all observations above or below a specified value.
Extreme value	The largest observation during a given period of observation.
Return period	The average space or time interval between a given observation. (A \$10,000 accident will occur every 10 years.) This value is the reciprocal or inverse of the frequency, and is equal to $1/(1 - p)$, where p is the cumulative probability.
Density function	The value of the y-axis on a probability distribution curve (see Figure C-2 later in this Appendix). It is a measure of frequency of stated values on the x-axis.
Histogram	Pictorial representation of a distribution; a bar graph.
Most probable maximum value	The return period measures how often an event equal to or greater than a specified value will occur. This specified value is the probable maximum value.

Probability paper A special graph paper in which the x-axis and y-axis are scaled in such a way as to convert a distribution curve into a straight line. The purpose is twofold--a straight line function is (a) easier to fit and (b) easier to extrapolate. The test to determine the type of distribution is: what type of probability paper results in a straight line fit. If there is a good straight line fit of the data, we can be reasonably sure that the experience can be represented by that type of distribution.

Accidents occur as a result of unplanned combinations of events and, as such, are statistical in nature. By statistical, we mean the exact cost or time of a specific accident cannot be predicted. A small sample of accidents appear to be random with no specific pattern of frequency or cost. However, as the sample size increases (experience is accumulated), a pattern begins to emerge in that we can begin to estimate the average number of accidents which will occur in a given time period and the relative frequency with which accidents occur within a given cost range (assuming there are no significant changes in the major factors which cause these accidents).

This pattern is called a statistical distribution and can be defined using a histogram or a probability curve. Figure C-1 depicts a histogram and probability curve, giving the fraction of years (probability) in which 0, 1, 2, . . . 10 accidents would occur given an average of 5 accidents/year. The smooth line approximating the histogram is a probability curve.

The shapes of various probability curves are determined by the fundamental processes such as counting, addition, multiplication, exponentiation, or combinations of these processes. Each process creates a statistical distribution, the probability curve having a unique identifiable shape peculiar to the process from which it originates.

These curves are converted to straight lines by use of probability paper with special scales. A special type of probability paper is required for each type of distribution. Plotting the data on various types of

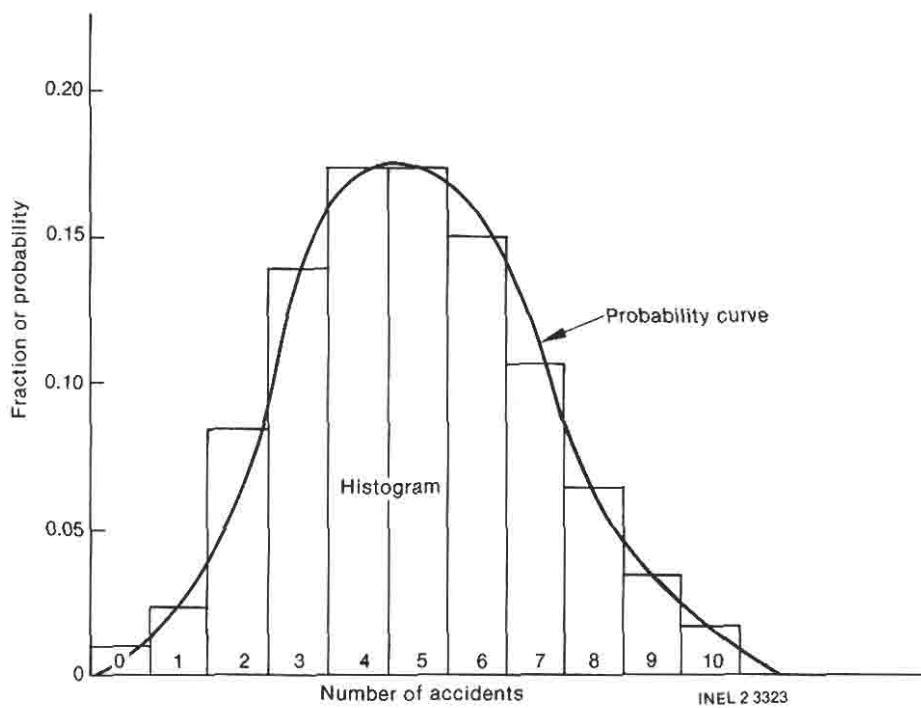


Figure C-1. Example of a statistical distribution.

probability paper will determine what type of distribution the data represent. If the plotted curve approximates a straight line on normal probability paper, then the distribution (curve) is normal; if it fits a straight line on log-normal paper, then the distribution is log-normal, etc.

Since the distribution is a straight line, it is easier to accurately extrapolate the curve. From this extrapolation one can determine the probability and cost of rare events, such as a very large accident.

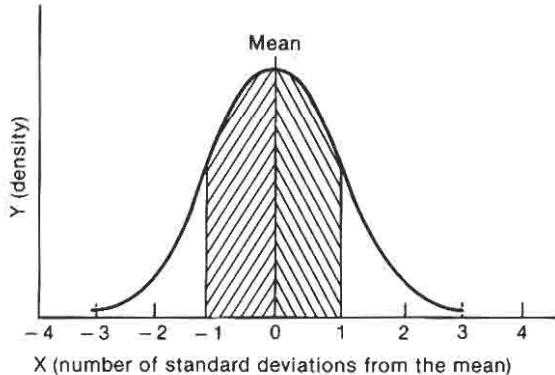
Several distributions useful in risk analysis will be discussed next. The first discussion is given in greater detail to explain basic concepts common to all distributions.

3. The Normal Distribution

The normal distribution process is linear (additive). Examples are the machined weights, sizes, and tolerances which all follow the normal distribution. Counting statistics (the number of accidents per year, the number of apples on a tree, the number of radioactive disintegration per minute, etc.,) with numbers larger than 20 are also approximated by the normal distribution.

A very important property of distribution, the Central Limit Theorem, states that "means" or average values are normally distributed regardless of the population distribution from which they are taken. This means that even though the cost or severity of accidents are log-normal distributed, average costs from different divisions or companies are normally distributed. These averages as well as counting numbers are represented by the normal distribution.

The normal distribution is a symmetrical bell-shaped curve extending infinitely in both the negative and positive directions on the x -axis as shown in Figure C-2. The x -axis is labelled in units of standard deviation which are the same for every normal distribution and in real values which are different for each distribution. The real values, in this case, represent the number of expected accidents where the mean is 100 accidents.



INEL 23324

Figure C-2. The normal distribution.

Not all bell-shaped symmetrical curves are normal distribution curves. The words "normal distribution" refer to the fact that the area under the curve is distributed in a specified manner which will be discussed later. This area represents the relative frequency with which variables fall within two points on the x-axis. The points on the x-axis are labelled in units of standard deviation beginning with zero at the center with negative values extending to the left and positive values to the right. These units of standard deviation are also called "Z scores." The height of the curve is measured by points along the y-axis which measures the "probability density." That is, the higher the curve, the greater the area and, hence, the greater the probability between two points on the x-axis. For example, between zero and one standard deviation, the probability is 0.34; between one standard deviation and two standard deviations, the probability is 0.135. The total area under the curve is one square unit or a probability of one. Thus the probability (area) of 0.34 between 0 and 1 standard deviation means that 34% of the variables in a normal distribution fall between 0 and 1 standard deviation. These probability values (in terms of standard deviations) are given in Table C-1. The mean is denoted by μ and the standard deviation by σ . By using standard deviations and a conversion table instead of labeling the x-axis directly in probability units, one set of values (one table) is applicable to distribution of any average and any standard deviation.

The standard deviation is calculated from a set of observations using the following equation:

$$\text{Standard Deviation} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}} \text{ (or, square root of the variance)}$$

where

\bar{x} = the average value of the variable also denoted as μ

x_i = the value each observation or variable

n = the number of observations or variables.

TABLE C-1. ORDINATES, AND AREAS BETWEEN $-z$ AND $+z$, OF THE NORMAL CURVE

<u>z</u>	<u>x</u>	<u>Ordinate</u>	<u>Area</u>	<u>z</u>	<u>x</u>	<u>Ordinate</u>	<u>Area</u>
0	μ	0.399	0.0000	± 1.50	$\mu \pm 1.50 \sigma$	0.1295	0.8664
± 0.05	$\mu \pm 0.05 \sigma$	0.398	0.0399	± 1.55	$\mu \pm 1.55 \sigma$	0.1200	0.8789
± 0.10	$\mu \pm 0.10 \sigma$	0.397	0.0797	± 1.60	$\mu \pm 1.60 \sigma$	0.1109	0.8904
± 0.15	$\mu \pm 0.15 \sigma$	0.394	0.1192	± 1.65	$\mu \pm 1.65 \sigma$	0.1023	0.9011
± 0.20	$\mu \pm 0.20 \sigma$	0.391	0.1585	± 1.70	$\mu \pm 1.70 \sigma$	0.0940	0.9109
± 0.25	$\mu \pm 0.25 \sigma$	0.387	0.1974	± 1.75	$\mu \pm 1.75 \sigma$	0.0863	0.9099
± 0.30	$\mu \pm 0.30 \sigma$	0.381	0.2538	± 1.80	$\mu \pm 1.80 \sigma$	0.0790	0.9281
± 0.35	$\mu \pm 0.35 \sigma$	0.375	0.2737	± 1.85	$\mu \pm 1.85 \sigma$	0.0721	0.9357
± 0.40	$\mu \pm 0.40 \sigma$	0.368	0.3108	± 1.90	$\mu \pm 1.90 \sigma$	0.0656	0.9426
± 0.45	$\mu \pm 0.45 \sigma$	0.361	0.3473	± 1.95	$\mu \pm 1.95 \sigma$	0.0596	0.9488
± 0.50	$\mu \pm 0.50 \sigma$	0.352	0.3829	± 2.00	$\mu \pm 2.00 \sigma$	0.0540	0.9545
± 0.55	$\mu \pm 0.55 \sigma$	0.343	0.4177	± 2.05	$\mu \pm 2.05 \sigma$	0.0488	0.9596
± 0.60	$\mu \pm 0.60 \sigma$	0.333	0.4515	± 2.10	$\mu \pm 2.10 \sigma$	0.0440	0.9643
± 0.65	$\mu \pm 0.65 \sigma$	0.323	0.4843	± 2.15	$\mu \pm 2.15 \sigma$	0.0396	0.9684
± 0.70	$\mu \pm 0.70 \sigma$	0.312	0.5161	± 2.20	$\mu \pm 2.20 \sigma$	0.0355	0.9722
± 0.75	$\mu \pm 0.75 \sigma$	0.301	0.5467	± 2.25	$\mu \pm 2.25 \sigma$	0.0317	0.9756
± 0.80	$\mu \pm 0.80 \sigma$	0.290	0.5763	± 2.30	$\mu \pm 2.30 \sigma$	0.0283	0.9786
± 0.85	$\mu \pm 0.85 \sigma$	0.278	0.6047	± 2.35	$\mu \pm 2.35 \sigma$	0.0252	0.9812
± 0.90	$\mu \pm 0.90 \sigma$	0.266	0.6319	± 2.40	$\mu \pm 2.40 \sigma$	0.0224	0.9836
± 0.95	$\mu \pm 0.95 \sigma$	0.254	0.6579	± 2.45	$\mu \pm 2.45 \sigma$	0.0198	0.9857
± 1.00	$\mu \pm 1.00 \sigma$	0.242	0.6827	± 2.50	$\mu \pm 2.50 \sigma$	0.0175	0.9876
± 1.05	$\mu \pm 1.05 \sigma$	0.230	0.7063	± 2.55	$\mu \pm 2.55 \sigma$	0.0154	0.9892
± 1.10	$\mu \pm 1.10 \sigma$	0.218	0.7287	± 2.60	$\mu \pm 2.60 \sigma$	0.0136	0.9907
± 1.15	$\mu \pm 1.15 \sigma$	0.206	0.7499	± 2.65	$\mu \pm 2.65 \sigma$	0.0119	0.9920

TABLE C-1. (continued)

<u>z</u>	<u>x</u>	<u>Ordinate</u>	<u>Area</u>	<u>z</u>	<u>x</u>	<u>Ordinate</u>	<u>Area</u>
± 1.20	$\mu \pm 1.20 \sigma$	0.194	0.7699	± 2.70	$\mu \pm 2.70 \sigma$	0.0104	0.9931
± 1.25	$\mu \pm 1.25 \sigma$	0.183	0.7887	± 2.75	$\mu \pm 2.75 \sigma$	0.0091	0.9940
± 1.30	$\mu \pm 1.30 \sigma$	0.171	0.8064	± 2.80	$\mu \pm 2.80 \sigma$	0.0079	0.9949
± 1.35	$\mu \pm 1.35 \sigma$	0.160	0.8230	± 2.85	$\mu \pm 2.85 \sigma$	0.0069	0.9956
± 1.40	$\mu \pm 1.40 \sigma$	0.150	0.8385	± 2.90	$\mu \pm 2.90 \sigma$	0.0060	0.9963
± 1.45	$\mu \pm 1.45 \sigma$	0.139	0.8529	± 2.95	$\mu \pm 2.95 \sigma$	0.0051	0.9968
± 1.50	$\mu \pm 1.50 \sigma$	0.130	0.8664	± 3.00	$\mu \pm 3.00 \sigma$	0.0044	0.9973 ^a
--	--	--	--	± 4.00	$\mu \pm 4.00 \sigma$	0.0001	0.99994
--	--	--	--	± 5.00	$\mu \pm 5.00 \sigma$	0.000001	0.9999994

a. Equals 99.7% for \pm standard deviations.

NOTE: If only the probability of falling either above or below the mean is wanted, divide the given probability area by two.

For counting statistics where the variable is merely the number of events, the standard deviation is simply the square root of the average number. For instance if the average number of accidents per year is 100, the standard deviation is 10 accidents.

Multiplying or dividing each observation by a constant, will multiply or divide the standard deviation by the same constant.

Adding or subtracting each observation by a constant will add or subtract from the mean but will not change the standard deviation.

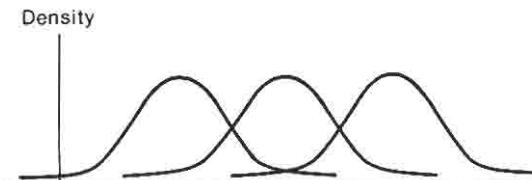
To add standard deviations, take the square root of the sum of the squares of each standard deviation. For instance, if Contractor A and B have averaged 100 ± 10 and 49 ± 7 accidents/year, the expected number of accidents for both is: $49 \pm 7 + 100 \pm 10 = 149 \pm 7^2 + 10^2 = 149 \pm 12.2$.

Figure C-3 depicts two sets of different normal curves. Each curve has an area of one. In the upper set each of the three curves has the same shape, standard deviation, and variance; but each has a different mean. In the lower set, each of the three curves has the same mean but a different shape, standard deviation, and variance. The curve with more area out near the edges has the larger standard deviation, but the same percentage of the area falls within ± 1 standard deviation, for each of the curves. Thus, only one table is needed to convert any value for any of the six curves to a probability value or a percentile ranking.

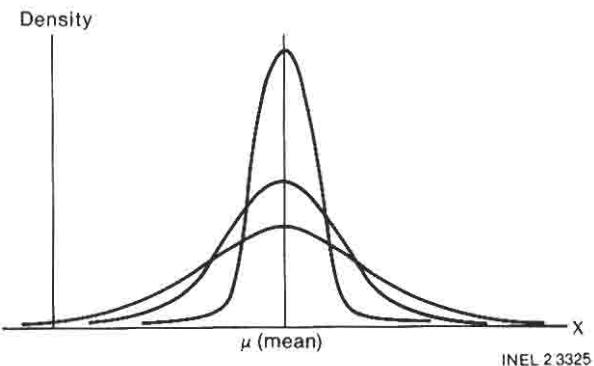
To illustrate, this procedure, find the probability of a deviation from the mean > 15 in a population having a mean of 100 and a standard deviation of 10. (This is a counting statistic applicable to the number of injuries or accidents.)

4. Solution

1. Divide the deviation (15) by the standard deviation (10) which equals 1.5 standard deviations.



(a) Normal distributions with the same variances different means.



(b) Normal distributions with the same means different variances.

Figure C-3. Normal distributions.

2. In Table C-1, Column Z find a score of ± 1.5 and go across to the corresponding value of 0.8664 under the "Area" column.

The value of 0.8664 is the probability that a variable will fall within ± 1.5 standard deviations.

3. Subtract 0.8664 from 1.0 to obtain 0.1336, the probability of exceeding ± 1.5 standard deviations.

In the above example 13.4% of the values fall outside the range of 85 to 115. Any number < 85 and any number > 115 , represents a significant deviation from 100. Any value within the range of 85 to 115 does not represent a significant deviation.

The degree of sureness or certainty is called the significance level. It is an arbitrarily selected value used to test an assumption or a hypothesis. This significance level is sometimes called a confidence level. The confidence level or confidence interval refers to the mean when it is not known but estimated from sampling or statistical measurement. For example, in the above sample problem, if the mean is estimated from a sample population of 1000 (10 measurements of 100 each) the mean is known to an accuracy of 100 ± 3.2 standard deviations calculated as follows:

$$\frac{1000}{10} \pm \sqrt{\frac{100}{10}} = 100 \pm 3.2.$$

In either case, the procedure described above is used to determine the percent of values with any range of values. The significance level is the level or percent of time a single observation will fall within a population range. The confidence level is the level or percent of a large number of samples giving a mean or average within the stated range.

The procedure in statistical testing is as follows:

1. State a hypothesis (make an assumption).

Example: A value of 85 represents a significant change from an average of 100.

2. Select a significance level.

Example: We want to be 90% sure we do not reject a good statistic.

3. Test the hypothesis by determining the probability that the variable is part of the population.

Example: A deviation of 15 (100 - 85) will occur 86.6% of the time as determined from Table C-1.

4. Reject the hypothesis since 86.6% is <90%. We conclude that a value of 85 does not represent a significant change. [However, if the value had been 80, with all else the same, we would have accepted the hypothesis, since a deviation of 20 (100 - 80) will occur 95.45% of the time and this is more than 90%. Here we would have concluded that a value of 80 does represent a significant change.]

In the above examples, the testing has been against exceeding a deviation in terms of absolute value (greater than a positive or negative deviation of 15). This testing is said to be two-sided since the probability includes the area on both sides of the curve around the mean. One-sided testing refers to testing for only a decrease (or an increase). There are no one-sided probability tables since these values are just one-half of the two-sided value.

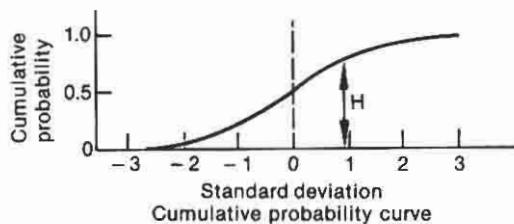
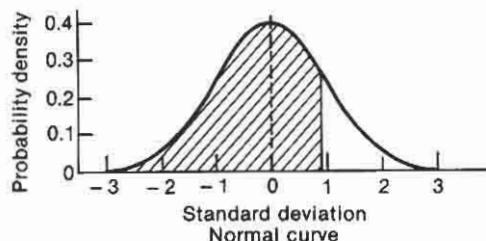
For example, the probability of <85 accidents is one-half the 0.13 value or 0.065. The complementary value is 0.935 and corresponds to a confidence level that 85 accidents or less represents a significant reduction.

This may appear contradictory to the previous conclusion that a deviation of 15 is not significant, but it is not because significance is determined by an arbitrary preselected confidence level. One can correctly choose either a two- or one-sided confidence level at whatever degree of confidence is desired, depending on whether the testing is being done for improvement or merely change (good or bad) as long as the choosing is done with an understanding of one- and two-sided values and confidence levels.

5. The Cumulative Normal Distribution

This distribution is merely a different representation of the same statistical pattern described by the normal distribution. Rather than giving the relative frequency with which variables fall on either side of the mean as in the normal distribution, the cumulative distribution gives the relative frequency with which observations will fall below (to the left of) any specified value. In the upper (normal) distribution curve in

Figure C-4, this relative frequency is the shaded area under the distribution curve from minus infinity to the specified value. In the lower (cumulative normal distribution), the relative frequency is indicated by the height of the curve rather than the area under the curve. (The height of H is numerically and conceptually identical to the shaded area under the normal curve.)



INEL 23327

Figure C-4. Cumulative probability for the normal curve compared to the cumulative probability curve.

The values of the cumulative normal distribution are given in Table C-2 with the specified values measured in units of standard deviation.

The only difference between Table C-2 and C-1 is that Table C-1 gives the probability of a value falling within a specified range above and below the mean, whereas Table C-2 gives the probability of a value falling below a specified value.

To illustrate, consider the normal distribution example in which the occurrence of 85 accidents was tested against an expected or average of 100 accidents.

TABLE C-2. AREAS BELOW Z (TO THE LEFT) OF THE NORMAL CURVE

<u>z</u>	<u>x</u>	Area	<u>z</u>	<u>x</u>	Area
-3.0	$\mu - 3.0 \sigma$	0.0013	0.1	$\mu + 0.1 \sigma$	0.5398
-2.9	$\mu - 2.9 \sigma$	0.0019	0.2	$\mu + 0.2 \sigma$	0.5793
-2.8	$\mu - 2.8 \sigma$	0.0026	0.3	$\mu + 0.3 \sigma$	0.6179
-2.7	$\mu - 2.7 \sigma$	0.0035	0.4	$\mu + 0.4 \sigma$	0.6554
-2.6	$\mu - 2.6 \sigma$	0.0047	0.5	$\mu + 0.5 \sigma$	0.6915
-2.5	$\mu - 2.5 \sigma$	0.0062	0.6	$\mu + 0.6 \sigma$	0.7257
-2.4	$\mu - 2.4 \sigma$	0.0082	0.7	$\mu + 0.7 \sigma$	0.7580
-2.3	$\mu - 2.3 \sigma$	0.0107	0.8	$\mu + 0.8 \sigma$	0.7881
-2.2	$\mu - 2.2 \sigma$	0.0139	0.9	$\mu + 0.9 \sigma$	0.8159
-2.1	$\mu - 2.1 \sigma$	0.0179	1.0	$\mu + 1.0 \sigma$	0.8413
-2.0	$\mu - 2.0 \sigma$	0.0228	1.1	$\mu + 1.1 \sigma$	0.8643
-1.9	$\mu - 1.9 \sigma$	0.0287	1.2	$\mu + 1.2 \sigma$	0.8849
-1.8	$\mu - 1.8 \sigma$	0.0359	1.3	$\mu + 1.3 \sigma$	0.9032
-1.7	$\mu - 1.7 \sigma$	0.0446	1.4	$\mu + 1.4 \sigma$	0.9192
-1.6	$\mu - 1.6 \sigma$	0.0548	1.5	$\mu + 1.5 \sigma$	0.9332
-1.5	$\mu - 1.5 \sigma$	0.0668	1.6	$\mu + 1.6 \sigma$	0.9452
-1.4	$\mu - 1.4 \sigma$	0.0808	1.7	$\mu + 1.7 \sigma$	0.9554
-1.3	$\mu - 1.3 \sigma$	0.0968	1.8	$\mu + 1.8 \sigma$	0.9641
-1.2	$\mu - 1.2 \sigma$	0.1151	1.9	$\mu + 1.9 \sigma$	0.9713
-1.1	$\mu - 1.1 \sigma$	0.1357	2.0	$\mu + 2.0 \sigma$	0.9772
-1.0	$\mu - 1.0 \sigma$	0.1587	2.1	$\mu + 2.1 \sigma$	0.9821
-0.9	$\mu - 0.9 \sigma$	0.1841	2.2	$\mu + 2.2 \sigma$	0.9861
-0.8	$\mu - 0.8 \sigma$	0.2119	2.3	$\mu + 2.3 \sigma$	0.9893
-0.7	$\mu - 0.7 \sigma$	0.2420	2.4	$\mu + 2.4 \sigma$	0.9918
-0.6	$\mu - 0.6 \sigma$	0.2741	2.5	$\mu + 2.5 \sigma$	0.9938
-0.5	$\mu - 0.5 \sigma$	0.3085	2.6	$\mu + 2.6 \sigma$	0.9953
-0.4	$\mu - 0.4 \sigma$	0.3446	2.7	$\mu + 2.7 \sigma$	0.9965
-0.3	$\mu - 0.3 \sigma$	0.3821	2.8	$\mu + 2.8 \sigma$	0.9974
-0.2	$\mu - 0.2 \sigma$	0.4207	2.9	$\mu + 2.9 \sigma$	0.9981
-0.1	$\mu - 0.01 \sigma$	0.4602	3.0	$\mu + 3.0 \sigma$	0.9987
0	μ	0.5000	--	--	--
-3.090	$\mu - 3.090 \sigma$	0.001	+3.090	$\mu + 3.090 \sigma$	0.999
-2.576	$\mu - 2.576 \sigma$	0.005	+2.576	$\mu + 2.576 \sigma$	0.995
-2.326	$\mu - 2.326 \sigma$	0.010	+2.326	$\mu + 2.326 \sigma$	0.990
-1.960	$\mu - 1.960 \sigma$	0.025	+1.960	$\mu + 1.960 \sigma$	0.975
-1.645	$\mu - 1.645 \sigma$	0.050	+1.645	$\mu + 1.645 \sigma$	0.950
-1.282	$\mu - 1.282 \sigma$	0.100	+1.282	$\mu + 1.282 \sigma$	0.900
-1.036	$\mu - 1.036 \sigma$	0.150	+1.036	$\mu + 1.036 \sigma$	0.850
-0.842	$\mu - 0.842 \sigma$	0.200	+0.842	$\mu + 0.842 \sigma$	0.800
-0.674	$\mu - 0.674 \sigma$	0.250	+0.674	$\mu + 0.674 \sigma$	0.750
-0.524	$\mu - 0.524 \sigma$	0.300	+0.524	$\mu + 0.524 \sigma$	0.700
-0.385	$\mu - 0.385 \sigma$	0.350	+0.385	$\mu + 0.385 \sigma$	0.650
-0.253	$\mu - 0.253 \sigma$	0.400	+0.253	$\mu + 0.253 \sigma$	0.600
-0.126	$\mu - 0.126 \sigma$	0.450	+0.126	$\mu + 0.126 \sigma$	0.550
0	μ	0.500	--	--	--

The Z score (units of standard deviation) in Table C-2 is the number of standard deviations and is equal to:

$$z = \frac{\bar{x} - \mu}{\sigma} = \frac{0.85 - 100}{10} = -1.5$$

where

\bar{x} = specified value

μ = mean

σ = standard deviation.

In Table C-2, a z score of -1.5 corresponds to an area probability of 0.0668, indicating that 85 accidents or less would occur <6.68% of the time, the same value as obtained by using the normal distribution table (1), and dividing by two to obtain one-sided values.

6. Normal-Probability Paper

By changing the vertical scale on the graph of the cumulative-normal-distribution curve, it is possible to have the cumulative-normal-distribution curve take on the shape of a straight line. This can be visualized if we think of the curve as plotted on an elastic sheet and the sheet stretched in an appropriate fashion. Figure C-5 shows the cumulative-normal-distribution curve and indicates the necessary stretching. To the left of the mean the cumulative curve should be pulled down, and on the right of the mean it should be pulled up until it coincides with the dotted line.

This, of course, means that the scale on the vertical axis is changed. Special paper, scaled appropriately, can be purchased. It is called normal-probability paper. In Figure C-6 is shown a sheet of this type of graph paper. The horizontal axis is marked for a mean of 100 and a standard

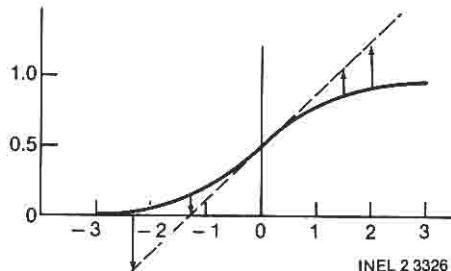


Figure C-5. Conversion of cumulative curve to a straight line.

deviation of 10 units. The line drawn represents the cumulative normal distribution with a mean (μ) = 100 and the standard deviation (σ) = 10. (The mean of a set of variables is always at the 50th percentile but may not agree exactly with the mean calculated by summing the variables and dividing by N , the number of variables. For statistical purposes, the graphical mean is preferred to the arithmetic mean.)

The probability of a variable falling below a certain value is called a percentile and can be read directly from the graph. For example, a value of <85 (reading the right hand scale has a probability of 0.066 on a 6.6% percentile rating. This is the same value as determined previously from both the normal and cumulative normal Tables C-1 and C-2. The reason the percentiles can be read directly from probability paper but a statistical table is needed to convert the standard deviation on the bell-shaped curve to percentiles is that the slope of the line on probability paper (rather than shape of the curve) determines the size of the standard deviation.

As stated, the mean (which has a value of 100) can be read directly from the 50th percentile and thus the standard deviation can be determined by subtracting the 16th percentile from the 50th percentile. Thus values from both Tables C-1 and C-2 can be determined directly from the normal probability paper.

To illustrate the use of normal probability paper, the number of lost work cases (LWCs) and the number of total recordable cases (TRCs) at one of

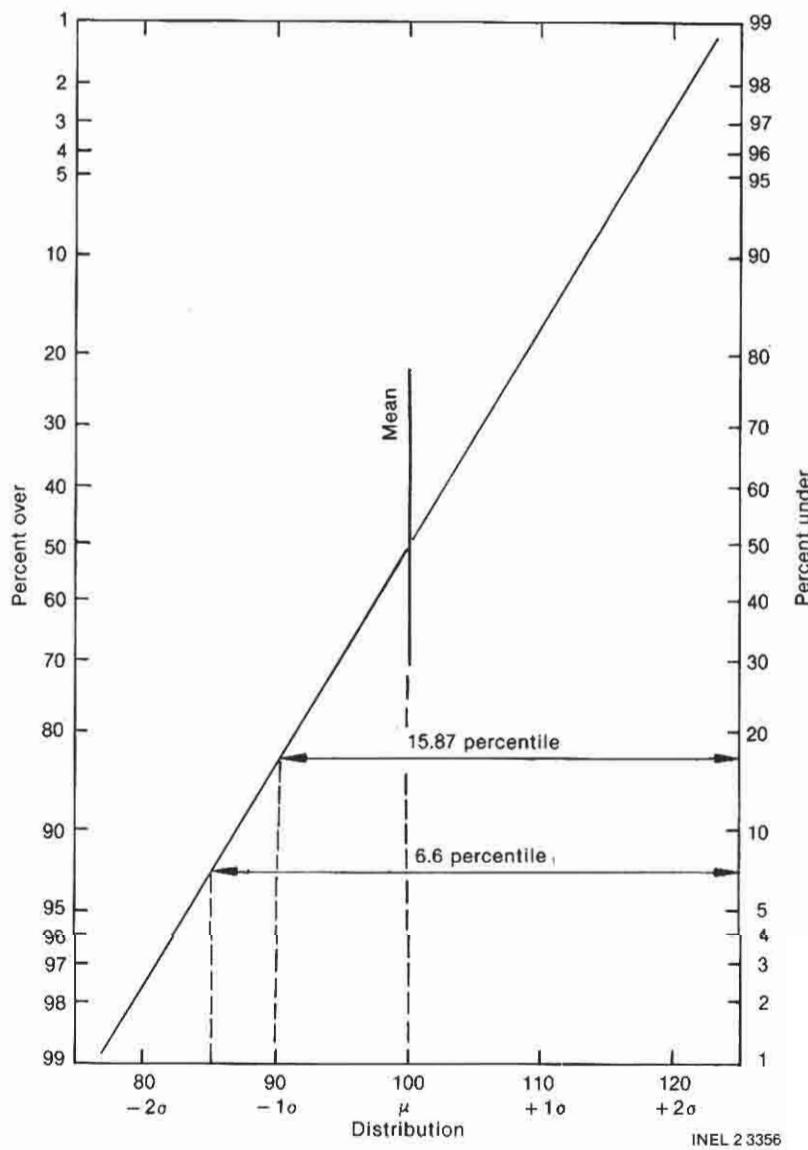


Figure C-6. Normal probability paper.

the DOE field operations are plotted on Figure C-7. The points for plotting were obtained from the following data:

<u>Year</u>	<u>TRCs</u>	<u>LWCs</u>
1976	673	295
1977	568	305
1978	546	287
1979	<u>505</u>	<u>309</u>
Average	573	299

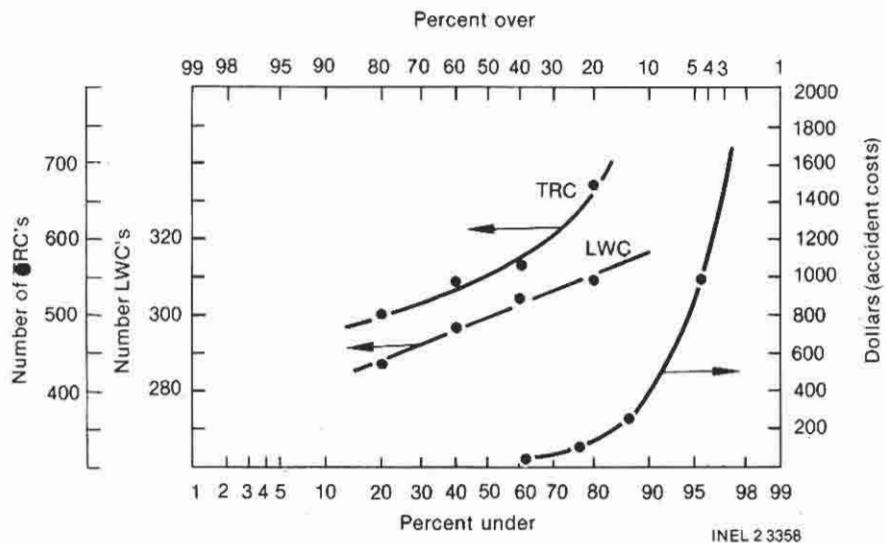


Figure C-7. Cumulative probability.

The number of injury cases was adjusted (normalized) to offset changes in the number of workhours, i.e., the number in the table is the actual number multiplied by the average workhours per year and divided by the workhours for the particular year.

To arrange the data for plotting, rank the number of injuries each year and determine the cumulative percentage as follows:

<u>ni</u>	<u>LWCs</u>	<u>Cumulative ni/N + 1 x 100 (%)</u>	<u>TRCs</u>	<u>Cumulative ni/N + 1 x 100 (%)</u>
1	287	20	505	20
2	297	40	546	40
3	305	60	568	60
4	309	80	673	80

$$N + 1 = 5$$

N = Total number of observations = 4, in this example ($N \pm 1 = 5$)

ni = Individual observations.

For distributions other than counting statistics, the data may be arranged as follows:

<u>Cost Range (\$)</u>	<u>Number of Accident</u>	<u>Cumulative Number of Accidents</u>	<u>Percentile <math>N \times 100/N + 1 (%)</math></u>
0-50	29	29	62
50-100	7	36	77
100-250	4	40	86
250-500	3	43	92
500-1000	1	44	95
1000-2000	2	46	98.9

$$N + 1 = 47$$

NOTE: There are also more accurate (and more complex) methods of determining the percentile value. See TEAM probability papers, TEAM, Box 25, Tamworth, NH 03886.

The cumulative number of accidents is the total number of accidents costing less than the maximum value in the corresponding cost range. The maximum value in the cost range is plotted against the percentile. This data, also plotted in Figure C-7, is labelled "percentile of accident costs." It is obviously not a normal distribution.

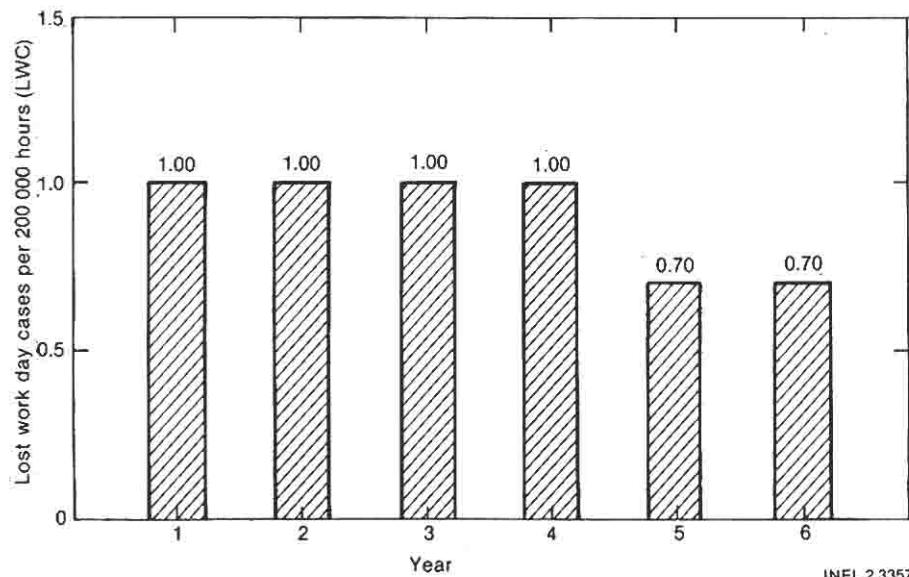
The following observations can be made from the TRC and LWC curves in Figure C-7. The LWCs appear to be normally distributed because the data

fit a straight line. The standard deviation is about 15 cases (obtained by reading the number of cases at the 16th percentile or at the 84th percentile and taking the difference from the 50th percentile). The counting standard deviation is the square root of the annual number (299), which equals 17. The good agreement indicates the degree of variability is normal so that there is no trend as cyclical influences changing the LWC accident frequency.

On the other hand, the TRC data do not fit a straight line. The 50th percentile value is 550 as opposed to the arithmetic average of 573 cases. The standard deviation from the curve between the 16th and 50th percentiles is about 60 compared to the "square root" standard deviation of 24. We conclude the TRC data are not normally distributed, indicating an increasing or decreasing trend. The larger standard deviation indicates a large variation, which is always the case if a trend exists. Cyclical variations with no trend will also produce a large standard deviation, but will fit a straight line. A bar graph representation of these data is given in Figures C-8 and C-9. (The data on the normal probability paper correspond to the LWC and TRC incidence rate bar graphs for only the years 1976 through 1979.) The LWC rates for these years are constant within one decimal place. The second decimal place is insignificant compared to normal statistical variation and is therefore not used. By contrast, the TRC bar graphs show a significant downward trend. Thus, normal probability paper is a special graphic representation with percentiles (cumulative percentages indicated right on the graph).

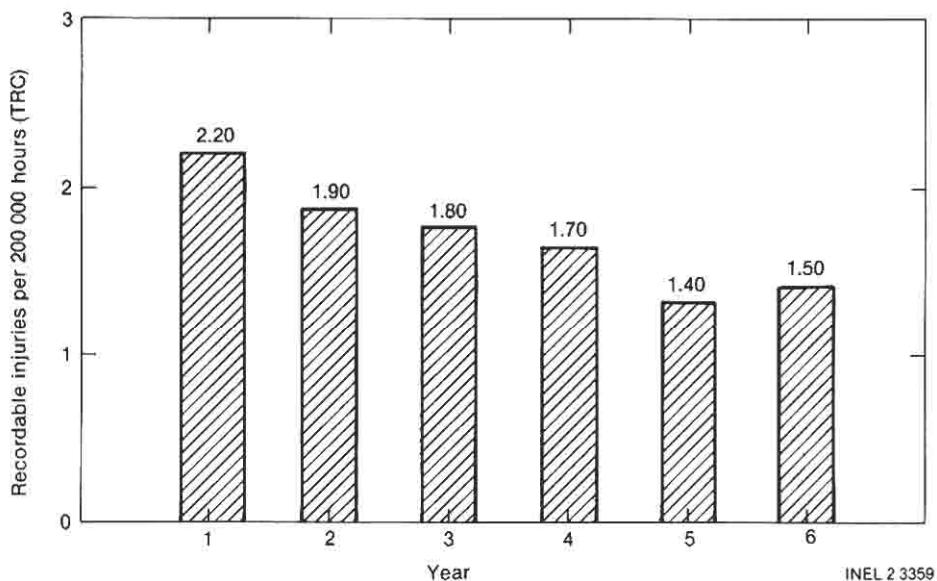
The vehicle accident cost data also do not fit a straight line and demonstrate graphically that accident cost data do not fit the normal distribution. Log-normal or extreme value paper should be used for plotting accident cost versus frequency as discussed later in this appendix.

While a curve on probability paper indicates a trend, two rules of thumb can be used to determine a trend on the bar graphs: (a) The overall deviation is large compared to deviations between adjacent years. For example, in Figure C-9 the large difference between 1976 and 1980 indicates a trend. The increase in 1981 over 1980 is too small to indicate a reversal



INEL 23357

Figure C-8. Lost work day case incidence rates.



INEL 23359

Figure C-9. Total recordable cases incidence rates.

in the trend; (b) A persistent change in one direction for three or more consecutive years indicates a trend. [The probability of a decrease is 0.5 (as much chance of decreasing as increasing) so that three such changes is $(0.5)^3$ or 0.125, indicating an 87.5% confidence level that there is a decreasing trend.]

These concepts also apply to curve fitting. Is a curve (or straight line) clearly identified by the data points, or is the scatter too great? A subjective feel for goodness of fit is usually adequate for tentative conclusions and/or identification of direction for further safety investigation. Those who desire confidence limit tests for goodness of fit should consult a textbook or a statistician.

As stated earlier, the normal distribution has limited application to risk analyses. This extended discussion provides an understanding of distributions and probability papers which are applicable to the log-normal and extreme value distributions.

7. Log-Normal Distribution

Processes in which the effects are multiplied fit the log-normal distribution. If the scale which measures the variable is logarithmic rather than linear, the curve representing this distribution has the symmetrical bell-shape of the normal curve. (An equivalent explanation is the logarithm of the variable is substituted for the variable to produce the same bell-shaped curve.)

On a linear scale, the log-normal curve is skewed with the peak to the left of center and a long tail to the right. Examples of this distribution are corrosion, gaseous diffusion, personal income, growth, and accident severity.

An example of log-normal probability paper is given in Figure C-10. It is identical to normal probability with the exception that the scale measuring the variable is logarithmic rather than linear. On both papers,

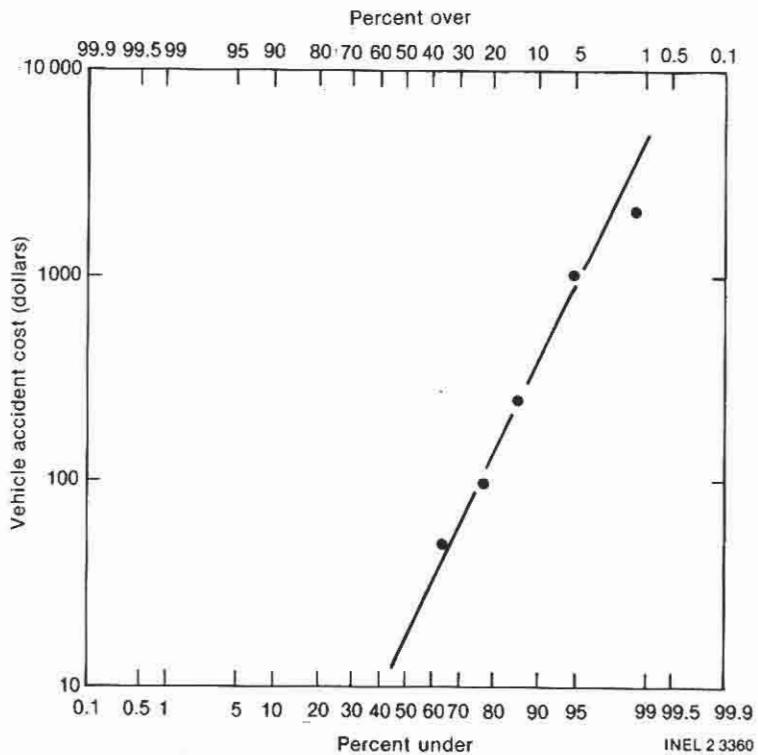


Figure C-10. Vehicle accident cost.

the percent scale is symmetrical with the 50th percentile at the center. Arranging the data for plotting is the same as for the normal probability paper.

The vehicle accident cost data plotted on normal probability paper in Figure C-7, is plotted on log-normal paper in Figure C-10. The data are taken from 46 vehicle accidents experienced by a DOE contractor. The data below the \$1000 level fit a straight line very well. Knowing that the maximum loss of a vehicle is limited by the value of the vehicles, we can surmise that this distribution is constrained by an inherent limitation and we would not expect the upper cost range to follow the logarithmic normal distribution. This procedure can be used even though the vehicle accident data may not be a true or unlimited log-normal distribution.

In fact, the multiple causes of accidents do have a multiplying effect and in most cases fit the log-normal distribution very well. As such, a flattening of the curve at the upper cost range is indicative of inherent or physical limitations to the size of the type of accident being analyzed. This flattening may not occur even though a limitation exists if the data is not sufficient to include accidents near the maximum possible value.

8. Extreme Value Distribution

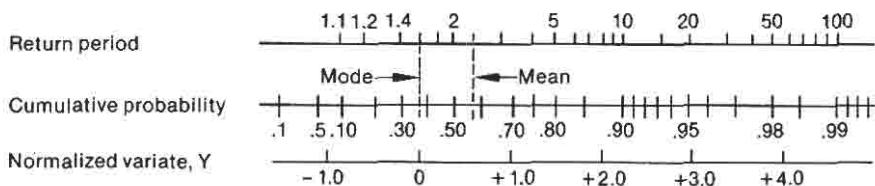
The extreme value model was developed from observations of maximum values from sets of observations, i.e., the tallest student in each class, the highest stream flow each year, the largest accident each year, etc. The curve representing this distribution is bell-shaped and skewed to the left similar to the log-normal wave. The degree of skewness or lack of symmetry is more extreme than the log-normal distribution. As such, the percentile scale is not symmetrical; the 50th percentile is to left of center.

The layout and design features of extreme value probability paper are different from those of normal and log-normal probability papers. Figure C-11 shows the three scales used on extreme value probability paper. The uniform scale of the normalized unit variable, y , is shown at the bottom. The scale for cumulative percentage is shown in the middle. For comparison, the extreme value distribution plotted on a linear scale is shown below. The spacing of the cumulative percentage scale is made to correspond to the shape of the extreme value distribution, which is highly compressed for low values and increasingly more spread out for high values. The peak of the distribution at the mode occurs at the 36.79% point which is much less than the 50% point of the normal and log-normal distributions.

The top scale, which is usually located along the top edge of the grid, is the return period scale. This is a nonuniform scale which increases in value from left to right. The value on this scale is equal to $1/(1 - p)$, where p is the cumulative probability. Due to the statistical behavior of the extreme value distribution, the return period has a unique

$$\text{Return period} = \frac{1}{(1-\text{cumulative probability})}$$

Scales on extreme value paper



Extreme value distribution

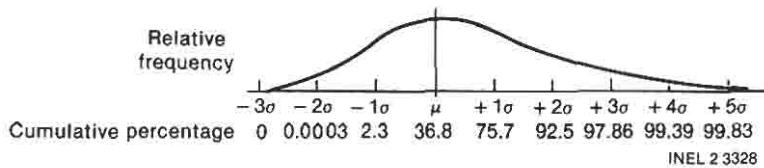


Figure C-11. Example of how the return period relates to the cumulative probability.

interpretation. The return period represents an estimate of probable sample sizes required for the largest observed value to equal a specified size. For example, the maximum accident observed in a 10-year period being \$50,000 is equivalent to stating that a \$50,000 accident has a 10-year return period.

The scale on the y-axis measuring the variable may be either linear as in Figure C-12, or logarithmic as in Figure C-13. In processes where the multiple effects are independent, the data fit the linear extreme value paper. In processes where there are multiple effects of interdependent, related causes, the data fit the logarithmic extreme value paper. In practice, accident data is plotted on both types of paper as a test to determine whether the extreme value distribution for the accidents under consideration is linear or logarithmic (does it fit a straight line?) and

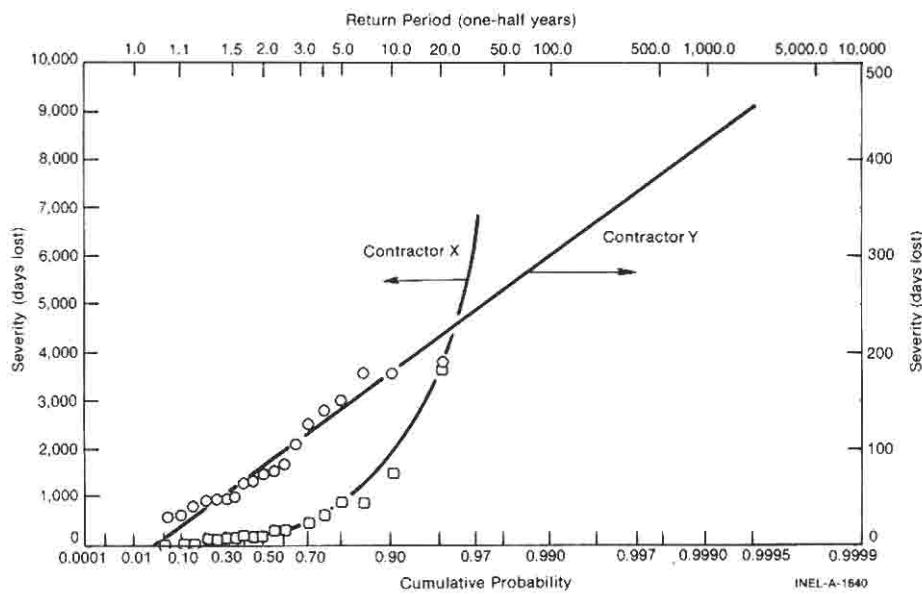


Figure C-12. Data plotted on linear scale extreme value paper.

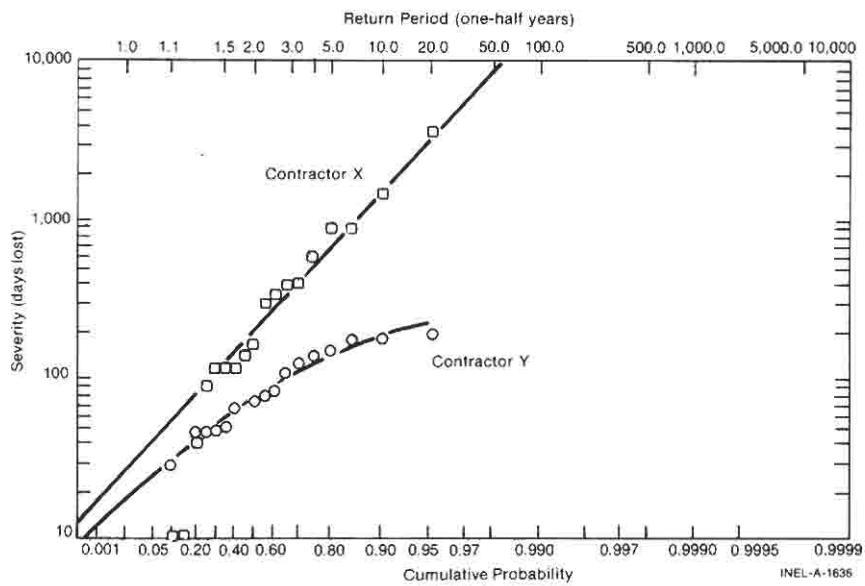


Figure C-13. Data plotted on logarithmic extreme value paper.

hence whether the multiple accident causes are independent or related by some system weakness in the control program. For example, the accident curves for Contractor X and Y are both plotted on linear and logarithmic extreme value paper in Figures C-12 and C-13. As can be seen, the severity for Contractor X increases more rapidly than does the severity and the curve for Contractor Y indicating a much higher potential for a very severe accident. In addition, the fact that Contractor Y's data fit a straight line on linear paper indicates that the multiple causation of accidents are relatively independent. On the other hand, the interdependency indicated by the straight line fit of Contractor X's data suggest common contributing factors to accident causation which could be corrected by strengthening the safety program rather than looking for a specific fix.

9. Binomial Distribution

The binomial distribution occurs in problems in which we take samples from a large population with specified "success" or "failure" probabilities and want to evaluate the chances of obtaining a certain number of successes in the sample. As such, it is a counting statistic. It has many applications in quality control, reliability, consumer sampling, and many other fields.

The variation or a distribution of the number of occurrences (accidents) such as is calculated from this fundamental law of probability. The binomial equation is:

$$B(r,n) = \frac{n!}{r!(n-r)!} p^r q^{(n-r)}$$

where

B = the probability of exactly r success in n trials

r = number of successes (occurrences)

n = n of trials (total possible number)

- p = probability of success (occurrence)
q = probability of not occurring ($1 - p$)
! = factorial, which is $(n \times n - 1 \times n - 2 \times \dots \times 1)$.

For large values of n, it approximates the normal distribution (which is easier to use). For small values of n, the Poisson distribution is easier to use. The binomial equation, however, is rigorously derived from probability combination theory using a differential equation series, and thus is not an approximation but accurately calculates the exact probability of r success in n trials. It should be used when accuracy is important or when redundancy is used to reduce risk in those situations where one or two failures can be tolerated, but a larger number of failures would have serious consequences.

10. Poisson Distribution

The Poisson distribution is also a counting statistic and approximates the binomial distribution when the average number of events is five or less. It is an approximate model for the number of elements per unit time or space.

The Poisson distribution has many applications in quality control, reliability, queuing theory, medical and biological statistics, and many other fields. In particular, this distribution may be an appropriate model for the number of defects in a piece of material, the number of insurance claims in a given period, the number of incoming calls per minute on a switchboard during a particular time of day, the number of bacteria in a given culture, the number of alpha particles emitted from a radioactive source in a specified time interval, the number of customer arrivals in a store at a given time of day, and the number of accidents in a short time period. The Poisson equation is:

$$P = \frac{e^{-n} n^x}{x!}$$

where

P = the probability of exactly x events given an average of n events

n = average number of events

x = the number of events for which the probability is being calculated.

The Poisson equation requires only a knowledge of the average number of events to determine the probability of a specified number of events. (If the average number of accidents per year is five, what is the probability of eight accidents/year?) Thus its use is not limited (as is the binomial equation) to situations where the probability of success or failure and the number of trials known. Note that the average number (n) is equal to the number of trials multiplied by the probability (Np) and this product can be substituted for n in the Poisson equation.

In addition, as in all counting statistics as explained in the discussion of the normal distribution, the square root of the average number of events is equal to the standard deviation.

The Poisson equation is recommended to determine whether a deviation from a small number of events is statistically significant.

For example, is an individual with several accidents a high risk? Is the number of accidents in a given department or time period statistically significant? Use of this analysis will determine the probability of a series of accidents in a short time period and eliminate arguments as to their significance between the line manager and safety engineer.

APPENDIX D
PLOTTING METHODS, GOODNESS-OF-FIT TESTING, AND
CONFIDENCE LIMITS FOR LOG-NORMAL AND
EXTREME VALUE DATA

This page intentionally blank

APPENDIX D
PLOTTING METHODS, GOODNESS-OF-FIT TESTING, AND
CONFIDENCE LIMITS FOR LOG-NORMAL AND
EXTREME VALUE DATA

1. General

The following methods for preparing data, plotting, and analyzing log-normal and extreme value distributions is given in a cookbook fashion with no mathematical or statistical derivations or justifications. The user is urged to consult the references at the end of this appendix for detailed theory and derivations. Much of the information contained herein was derived from these references.

The discussion to follow assumes some familiarity with statistical processes and use of probability graphing papers, where distributed data can be represented by a straight line on the graph. The beginner should read Appendix C, Probability and Statistics. Some emperical methods and simple tests are described to make a highly complex analysis process simple and easy to use by an engineer or scientist with only limited statistical analysis background.

2. Log-Normal Distributions

Random variations which lead to log-normal distributions are due to combinations of random effects which combine by relationships of multiplication and/or division. These distributions occur naturally for processes involving ratios, proportions, and rates. Safety severity data is generally log-normal in nature and as such can often be represented on a log-normal probability plot as a straight line. Conversely, data which can be well represented by a straight line on a log-normal probability plot can be said to come from a log-normal distribution. If the data can be represented by a log-normal distribution, certain characteristics of the system from which the data is taken can be implied in a manner similar to systems which produce normal (Gaussian) distributed data.

2.1 Preparing and Plotting Data on Log-Normal Paper

Data are best prepared for plotting by constructing a table. An example is shown in Table D-1 where the raw data points are given in Column A. The data are arranged for plotting as follows:

1. Arrange the data in Column A from smallest to largest as shown in Column B. If some data points are zero, include them in the

TABLE D-1. EXAMPLE OF LOG-NORMAL DATA ARRANGEMENT FOR PLOTTING

Column A Raw Data	Column B Ordered Data	Column C Rank Order	Column D ^a Plotting Position	Column E Percent Under (DX100)
85	15	1	0.025	2.5
15	25	2	0.066	6.6
170	50	3	0.107	10.7
270	55	4	0.149	14.9
110	60	5	0.190	19.0
55	60	6	0.231	23.1
205	70	7	0.273	27.3
155	70	8	0.314	31.4
60	85	9	0.355	35.5
780	90	10	0.397	39.7
190	110	11	0.438	43.8
70	145	12	0.479	47.9
25	155	13	0.521	52.1
145	160	14	0.562	56.2
200	160	15	0.603	60.3
70	170	16	0.645	64.5
50	190	17	0.686	68.6
220	200	18	0.727	72.7
160	205	19	0.769	76.9
60	220	20	0.810	81.0
470	225	21	0.851	85.1
160	270	22	0.893	89.3
225	470	23	0.934	93.4
90	780	24	0.975	97.5

- a. Plotting positions from formula in Table D-2.

arrangement, because even though they cannot be plotted on the log paper they figure in the plotting position for the remaining points.

2. Rank order the Column B data as shown in Column C.
3. For sample size less than 20, determine the plotting position for each ranked data point by entering Table D-2 with the sample size n (number of data points) and read the plotting position for each rank, Column C (from Table D-1). Enter the plotting positions into Column D. For sample size greater than 20 calculate the plotting positions from the relation shown on Table D-2. For example, the second plotting position of 0.066 units in Column D of Table D-1 is calculated as follows for rank order 2 and $n = 24$:

$$x_i = (i - 0.4)/(n + 0.2)$$

$$x_2 = (2 - 0.4)/(24 + 0.2) = 0.066.$$

Note that this method determines a plotting position for every value (event) in the sample in contrast to the method in the text which groups the data points into ranges. Also in the text, the percentile plotting positions are simplified using $n/n + 1$ rather than $(i - 0.4)/(n + 0.2)$. This more precise method is not necessary, particularly if the number of data points are >20 . In addition, the confidence limits and goodness-of-fit tests described later in this appendix can be applied to log-normal plots derived from either procedure.

4. Since the data have been arranged in descending order, plot the data points from Column B on the log-normal paper against the percentage plotting positions, Column E, as shown in Figure D-1, using the "percent under" axis.
5. "Best fit" a straight line through the data points using the median regression method given in Section 5.

TABLE D-2. PLOTTING POSITIONS FOR LOG-NORMAL ORDERED DATA BY SAMPLE SIZE n

Rank Number (i)	Sample Size (n)																				Rank Number (i)
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	0.500	0.273	0.188	0.141	0.115	0.097	0.083	0.073	0.065	0.059	0.054	0.049	0.045	0.042	0.039	0.037	0.035	0.033	0.031	0.030	1
2		0.727	0.500	0.301	0.208	0.258	0.222	0.195	0.173	0.157	0.143	0.131	0.121	0.111	0.105	0.099	0.093	0.088	0.083	0.079	2
3			0.813	0.619	0.500	0.419	0.361	0.317	0.283	0.255	0.232	0.213	0.197	0.183	0.171	0.160	0.151	0.143	0.135	0.129	3
4				0.857	0.692	0.581	0.500	0.439	0.391	0.353	0.321	0.295	0.273	0.254	0.237	0.222	0.209	0.198	0.188	0.178	4
5					0.885	0.742	0.639	0.561	0.500	0.451	0.411	0.371	0.348	0.324	0.303	0.284	0.267	0.253	0.240	0.228	5
6						0.903	0.778	0.683	0.609	0.549	0.500	0.459	0.428	0.394	0.368	0.346	0.326	0.308	0.292	0.277	6
7							0.917	0.905	0.717	0.647	0.589	0.541	0.503	0.465	0.434	0.407	0.384	0.361	0.344	0.327	7
8								0.927	0.826	0.745	0.679	0.623	0.576	0.535	0.500	0.469	0.442	0.419	0.396	0.376	8
9									0.935	0.843	0.768	0.705	0.652	0.604	0.566	0.531	0.500	0.473	0.448	0.426	9
10										0.941	0.857	0.787	0.727	0.676	0.626	0.593	0.558	0.527	0.500	0.475	10
11											0.946	0.859	0.803	0.746	0.697	0.654	0.618	0.582	0.552	0.525	11
12												0.951	0.079	0.817	0.763	0.716	0.674	0.637	0.604	0.574	12
13													0.955	0.887	0.823	0.778	0.733	0.692	0.656	0.624	13
14														0.958	0.895	0.840	0.791	0.747	0.708	0.673	14
15															0.961	0.901	0.849	0.802	0.760	0.723	15
16																0.963	0.907	0.857	0.813	0.772	16
17																	0.965	0.912	0.865	0.822	17
18																		0.967	0.917	0.871	18
19																			0.969	0.921	19
20																				0.970	20

NOTE: For log-normal data $x_i + (i - 0.4)/(n + 0.2)$ can be used to calculate plotting positions for sample sizes greater than 20.

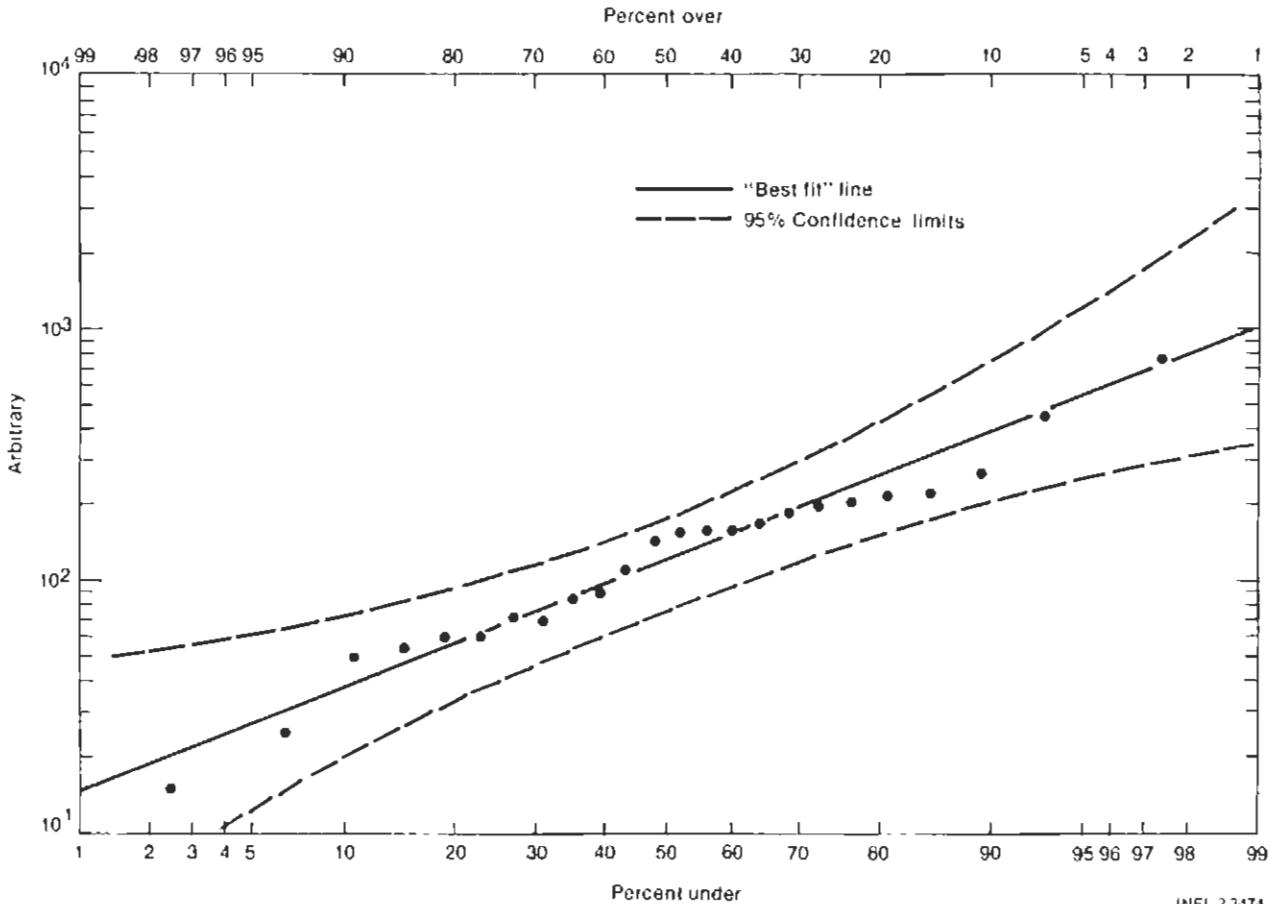


Figure D-1. Example of log-normal probability plot with confidence limits.
Data taken from Table D-1 and confidence limits from Table D-8.

6. Test the data for randomness and homogeneity by applying the run test given in Section 6.
7. If the data meets the run test, calculate and plot the confidence limits as shown in Figure D-1, by the method given in Section 7.1.
8. Make probabilistic estimates from the log-normal distribution represented by the data:
 - a. The median of the sample data is the value at 50% (the intersection of the fitted line and the 50% line).
 - b. The geometric dispersion is the ratio of the value at the 84% line to the median.
 - c. The percent of expected values under or over a specified value can be read directly from the fitted line by entering the graph at the appropriate percentage axis (percent under or percent over). Likewise, the expected percentage of events falling between two specified values can be obtained by entering the graph at the specified values on the fitted line and subtracting the two percentages read at these points.

3. Type I (Linear) Extreme Value Distributions

Type I (linear) extreme value distributions are the result of several independent causes, and has found use in studying such things as breaking loads, meteorological and geophysical phenomena (floods, tornados, earthquakes, etc.), chemical and electrical properties, stock market extremes, and economic data. Type I extreme value distributions result from systems which can be represented by algebraic polynomials and solutions to differential equations with constant coefficients. Some safety data distributions can be nicely represented with Type I extreme value.

Extreme value data is generated by selecting the largest (or the smallest) events in consecutive time periods, e.g., the largest accident loss experienced during consecutive year periods. The distribution of largest values is a right-skewed distribution; the distribution of smallest values is a left-skewed distribution.

One of the most useful pieces of information falling out of extreme value analysis is the "expected return period" for an event equal to or exceeding a given size.

3.1 Preparing and Plotting Type I Extreme Value Data

Data are best prepared for plotting by constructing a table. An example is shown in Table D-3 where the cost of the largest accident of a given type occurring in each six-month-period for eight consecutive years is tabulated in Column A. The data are arranged for plotting in the following steps:

1. Arrange the data in Column A from smallest to largest as shown in Column B.
2. Rank order Column B data as shown in Column C.
3. Determine the plotting position for each ranked data point by entering Table D-4 with the sample size n (number of data points) and read the plotting position for each rank in Column C. Enter the plotting positions into Column D. For sample size greater than 20, calculate the plotting positions from the relation shown on Table D-4.
4. Plot the data points from Column B at the plotting positions, Column D, on Type I (linear) extreme value graph paper (Figure D-2). The plotting position on the graph paper are on the cumulative probability axis.
5. "Best fit" a straight line through the data points using the median regression method given in Section 5.

TABLE D-3. EXAMPLE OF DATA ARRANGEMENT FOR TYPE I EXTREME VALUE PLOTTING

<u>Column A Raw Data</u>	<u>Column B Ordered Data</u>	<u>Column C Rank Order</u>	<u>Column D^a Plotting Position</u>
4800	3000	1	0.035
6200	4600	2	0.097
7200	4800	3	0.159
3000	5200	4	0.221
5200	5200	5	0.283
9300	5800	6	0.345
7300	6200	7	0.407
5200	6400	8	0.469
8000	6500	9	0.531
4600	7200	10	0.583
7200	7200	11	0.655
8300	7300	12	0.717
5800	7400	13	0.779
6400	8000	14	0.841
7400	8300	15	0.903
6500	9300	16	0.965

a. Plotting positions from Table D-4 with $n = 16$.

6. Test the data for randomness and homogeneity by applying the run test given in Section 6.
7. If the data meets the run test, calculate and plot the confidence limits, as shown on Figure D-2, by the method given in Section 7.2.
8. Estimate the return period, using the return period axis on the graph paper, for events exceeding a value of interest.

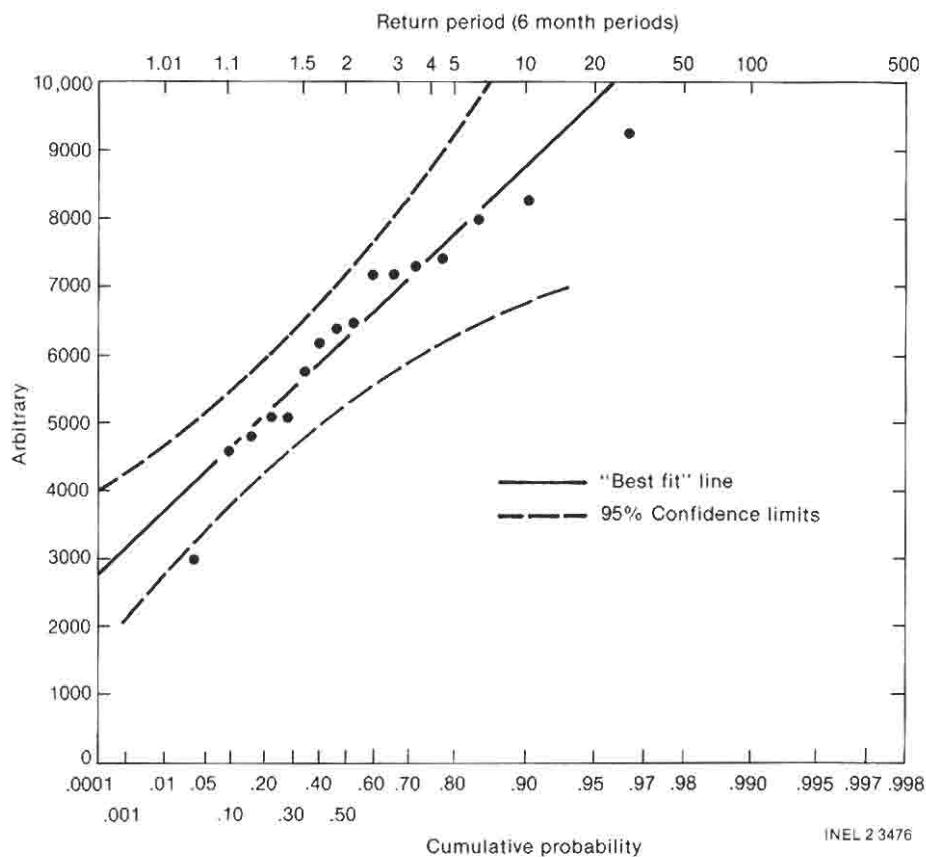
4. Type II (Logarithmic) Extreme Value Distributions

Type II extreme value distributions result from interdependent effects of several related causes and has found use in studying such phenomena as solid diffusion, chemical kinetics and particle breakage. Type II extreme value distributions result from systems which can be represented by

D-4. PLOTTING POSITIONS FOR TYPE I EXTREME VALUE ORDERED DATA BY SAMPLE SIZE n

Rank Number (i)	Sample Size (n)																				Rank Number (i)												
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20													
1	0.590	0.264	0.179	0.137	0.109	0.092	0.079	0.069	0.061	0.055	0.050	0.046	0.043	0.040	0.037	0.035	0.033	0.031	0.029	0.028	1												
2	0.776	0.599	0.379	0.105	0.255	0.219	0.192	0.171	0.154	0.140	0.129	0.119	0.110	0.103	0.097	0.091	0.086	0.082	0.078	0.078	2												
3		0.821	0.621	0.501	0.418	0.360	0.315	0.281	0.253	0.230	0.211	0.195	0.181	0.169	0.159	0.150	0.141	0.134	0.127	0.127	3												
4			0.963	0.695	0.582	0.509	0.438	0.390	0.352	0.320	0.294	0.271	0.252	0.235	0.221	0.209	0.196	0.186	0.177	0.177	4												
5				0.891	0.745	0.640	0.562	0.500	0.451	0.410	0.376	0.348	0.323	0.302	0.283	0.266	0.252	0.238	0.227	0.227	5												
6					0.918	0.781	0.685	0.610	0.549	0.500	0.459	0.429	0.394	0.368	0.345	0.325	0.307	0.291	0.276	0.276	6												
7						0.921	0.809	0.719	0.648	0.590	0.541	0.500	0.465	0.439	0.407	0.383	0.362	0.343	0.326	0.316	0.316	7											
8							0.931	0.829	0.747	0.680	0.623	0.576	0.535	0.500	0.469	0.442	0.417	0.395	0.376	0.356	0.336	8											
9								0.939	0.846	0.770	0.705	0.652	0.606	0.566	0.531	0.500	0.477	0.448	0.425	0.405	0.385	9											
10									0.945	0.660	0.789	0.729	0.677	0.632	0.593	0.558	0.528	0.503	0.475	0.455	0.435	0.415	10										
11										0.950	0.871	0.805	0.748	0.698	0.655	0.617	0.581	0.552	0.525	0.505	0.485	0.465	0.445	11									
12											0.954	0.881	0.819	0.765	0.717	0.675	0.638	0.605	0.575	0.555	0.535	0.515	0.495	0.475	12								
13												0.967	0.890	0.831	0.779	0.734	0.693	0.657	0.624	0.593	0.562	0.532	0.502	0.472	0.442	13							
14													0.960	0.897	0.841	0.792	0.748	0.709	0.674	0.643	0.613	0.583	0.553	0.523	0.493	0.463	14						
15														0.963	0.903	0.850	0.807	0.762	0.724	0.693	0.663	0.633	0.603	0.573	0.543	0.513	0.483	15					
16															0.965	0.909	0.859	0.814	0.773	0.743	0.713	0.683	0.653	0.623	0.593	0.563	0.533	0.503	16				
17																0.957	0.914	0.866	0.833	0.793	0.763	0.733	0.703	0.673	0.643	0.613	0.583	0.553	0.523	17			
18																	0.969	0.918	0.874	0.841	0.798	0.768	0.738	0.708	0.678	0.648	0.618	0.588	0.558	0.528	18		
19																		0.971	0.922	0.877	0.844	0.799	0.769	0.739	0.709	0.679	0.649	0.619	0.589	0.559	0.529	19	
20																			0.972	0.923	0.878	0.845	0.799	0.769	0.739	0.709	0.679	0.649	0.619	0.589	0.559	0.529	20

NOTE: For Type I Extreme value data $x_i = (i - 0.4)/(n + 0.12)$.



INEL 2 3476

Figure D-2. Example of Type I extreme value plot with confidence limits. Data taken from Table D-3 and confidence limit factors taken from Table D-10.

solutions of general differential equations or mathematical models involving multiplication of exponentials. Some safety data distributions can be represented with Type II extreme value (Table D-5). Often accident losses of a single, e.g., electrical type, can be represented by Type I extreme value distribution, however, when the losses from several types (i.e., electrical, mechanical, and nuclear) are combined into one distribution it generally results in a Type II extreme value distribution.

Data for Type II extreme value plotting is generated in the same manner as for Type I discussed in Section 3. Type II data must be all positive

TABLE D-5. EXAMPLE OF DATA ARRANGEMENT FOR TYPE II EXTREME VALUE PLOTTING

<u>Column A Raw Data</u>	<u>Column B Ordered Data</u>	<u>Column C Rank Order</u>	<u>Column D^a Plotting Position</u>
4.6×10^4	1.7×10^4	1	0.0625
5.0×10^4	2.0×10^4	2	0.1875
1.8×10^5	4.1×10^4	3	0.3125
1.7×10^4	4.6×10^4	4	0.4375
4.8×10^4	4.8×10^4	5	0.5625
4.1×10^4	5.0×10^4	6	0.6875
5.8×10^4	5.8×10^4	7	0.8125
2.0×10^4	1.8×10^5	8	0.9475

a. Plotting positions from Table D-6 with $n = 8$.

values, and although values of zero cannot be plotted, they should not be omitted because they are essential in determining the plotting positions for the remaining data points.

4.1 Preparing and Plotting Type II Extreme Value Data

To prepare and plot Type II extreme value data, use the same steps as described for Type I in Section 3.1 with the following exceptions:

1. The plotting positions in Step 3 are determined from Table D-6.
2. The plot in Step 4 is made on a Type II extreme value graph paper (has a logarithmic scale) as shown in Figure D-3.
3. The confidence limits, Step 7, are determined by the method given in Section 7.3.

TABLE D-6. PLOTTING POSITIONS FOR TYPE II EXTREME VALUE ORDERED DATA BY SAMPLE SIZE n

Rank Number (i)	Sample Size (n)																				Rank Number (i)
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	0.500	0.250	0.167	0.125	0.100	0.0933	0.071	0.0625	0.056	0.050	0.045	0.042	0.038	0.036	0.033	0.031	0.029	0.020	0.026	0.025	1
2		0.750	0.500	0.375	0.300	0.250	0.214	0.1875	0.167	0.150	0.136	0.125	0.115	0.107	0.100	0.094	0.089	0.079	0.075	0.075	2
3			0.823	0.625	0.500	0.417	0.357	0.3125	0.270	0.250	0.227	0.208	0.192	0.179	0.167	0.156	0.146	0.139	0.132	0.125	3
4				0.875	0.701	0.583	0.500	0.4375	0.369	0.350	0.318	0.292	0.269	0.250	0.233	0.219	0.206	0.194	0.184	0.175	4
5					0.900	0.750	0.643	0.5625	0.500	0.450	0.400	0.375	0.346	0.321	0.300	0.281	0.265	0.257	0.237	0.225	5
6						0.917	0.786	0.6875	0.611	0.550	0.500	0.458	0.423	0.393	0.367	0.332	0.321	0.304	0.269	0.275	6
7							0.929	0.8125	0.727	0.650	0.591	0.542	0.500	0.464	0.433	0.405	0.382	0.361	0.342	0.325	7
8								0.9475	0.833	0.750	0.662	0.625	0.577	0.536	0.500	0.469	0.441	0.417	0.395	0.375	8
9									0.944	0.850	0.773	0.708	0.654	0.607	0.567	0.531	0.500	0.472	0.447	0.425	9
10										0.950	0.864	0.792	0.731	0.679	0.633	0.594	0.559	0.526	0.503	0.475	10
11											0.955	0.875	0.808	0.750	0.700	0.656	0.618	0.581	0.553	0.525	11
12												0.959	0.885	0.821	0.767	0.719	0.677	0.639	0.605	0.575	12
13													0.962	0.893	0.833	0.781	0.735	0.694	0.658	0.625	13
14														0.964	0.900	0.853	0.794	0.750	0.711	0.675	14
15															0.967	0.905	0.854	0.805	0.763	0.725	15
16																0.969	0.912	0.861	0.816	0.775	16
17																	0.971	0.917	0.868	0.825	17
18																		0.972	0.921	0.875	18
19																			0.974	0.925	19
20																				0.975	20

NOTE: For Type II extreme value data $x_i = (i - 0.5)/n$.

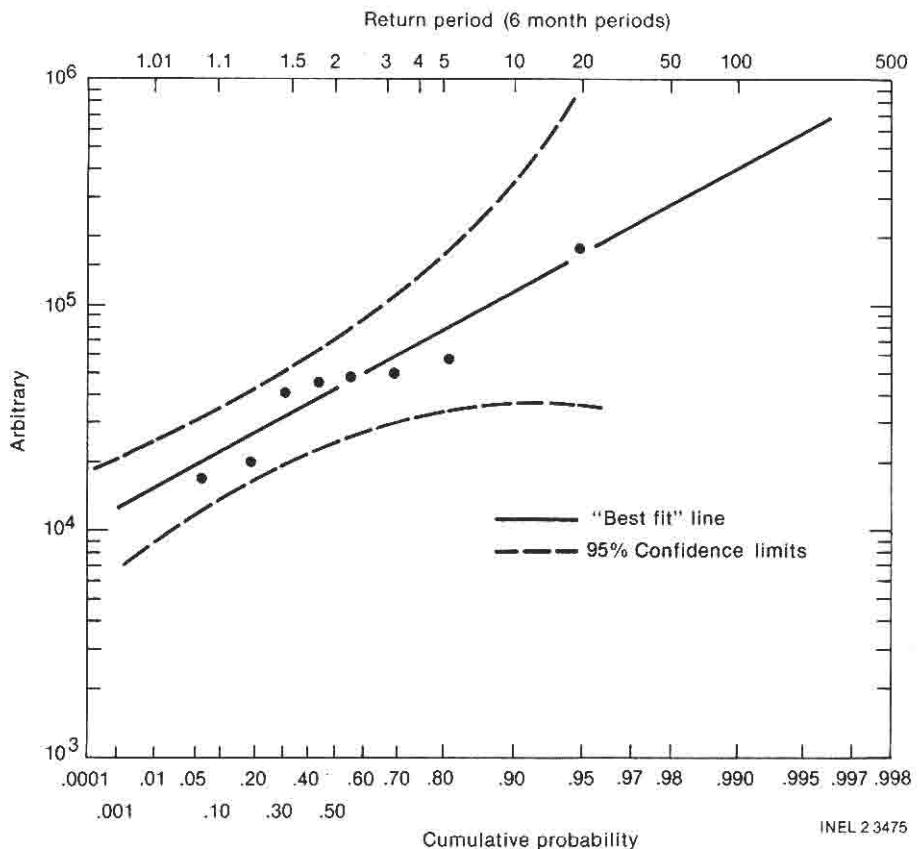


Figure D-3. Example of Type II extreme value plot with confidence limits. Data taken from Table D-4 and confidence limits from Table D-5.

5. "Best Fitting" A Curve Through Statistical Data Using A Median Regression Method

The following method of producing a "best fit" straight line through data points on probability graph paper has been found to be generally satisfactory. The method, although not exact, is quite reproducible and the same results can be obtained by two or more people working with the same data points.

1. Divide the data on the probability plot into two sets; one set includes the data to the right of 0.50 probability; the other set includes data to the left of 0.50 probability. If the total number of plotted data points is even, each set is distinct and no overlap occurs. If the total number is odd, a point will occur on the 0.50 probability line. It is usually best to treat this point it as if it belongs to both sets.
2. Using a clear plastic straight edge, place the edge on the extreme left point. Using this point as a pivot, divide the right set into two parts so that half of the points are above the straight edge and half are below.
3. Make a small pencil mark where the straight edge crosses the upper or right hand axis of the paper. Using this point as a pivot, divide the left set of data points into two parts so that half of the points in the set are above the straight edge and half are below.
4. Make a small pencil mark where the straight edge crosses the lower axis of the paper using this point as a pivot, readjust the straight edge through the right set of points if necessary to divide the points into two even parts, half above the straight edge and half below.
5. Iterate through Steps 3 and 4 until satisfied a good median regression fit has been obtained. To check the accuracy with which the right and left data sets are divided, count the number of points above and below the straight edge. The total number of points counted on one side the straight edge should not be more than two different from the points counted on the other side of the straight edge. When finished, the right and left sets of data points should each have half the points above and half below the line straight edge.
6. A line drawn along the straight edge is a "best fit" median regression line.

6. Testing for Goodness-of-Fit and Homogeneity of Data

If the number of points above and the number below the "best fit" line through the data on a probability plot are about the same, (i.e., the number of points above and below the line should not differ by more than two, else the line should be refitted) then apply the run test as follows:

1. A run is a group of consecutive points on the plot which lie on one side of the "best fit" line. If a point lies on the fitted line, this counts as the end of a run. Determine the number of runs above and below the "best fit" line for the plot. For example in Figure D-3, there are four runs.
2. Refer to Table D-7 which gives the expected number of runs for a given sample size and determine if the number of runs based on the total number of data points plotted are within the indicated range. For example, in Figure D-3, there are four runs in a sample of eight points. The range of runs for eight points as given in Table D-7 is three to seven; indicating valid data.
3. For sample sizes <10 the limits on runs from Table D-7 are essentially irrelevant. However, for larger sample sizes, if the number of runs are too few, there is good reason to believe the data are not homogeneous, i.e., they did not come from a stable or consistent population, assuming the data is plotted on the correct probability paper. If one is not sure the data is plotted on the correct probability paper, then place a straight edge across the extreme left and extreme right points on the plot. If all of the remaining points fall on one side or the other of the straight edge, it is likely that the wrong probability paper was chosen. Next, check the next extreme left and next extreme right points in the same manner with the straight edge. If all of the remaining points are still on the same side of the straight edge, it is highly likely that the wrong probability paper was used. If instead some points now fall on both sides of the straight edge, the problem may be caused by incomplete data.

TABLE D-7. APPROXIMATE 95% CONFIDENCE LIMITS FOR THE NUMBER OF RUNS ABOVE AND BELOW THE MEDIAN REGRESSION LINE

<u>Sample Size</u>	<u>Limits</u>
5	2-5
6	2-6
7	2-6
8	3-7
9	3-7
10	3-8
11	4-9
12	4-9
13	4-10
14	5-11
15	5-11
16	5-12
17	6-12
18	6-13
19	6-14
20	7-14
22	7-16
24	7-18
26	8-19
28	9-20
30	10-21
32	11-22
34	11-24
36	12-25
38	13-26
40	14-27
42	15-28
44	15-29
46	17-30
48	18-31
50	18-33

If there are too many runs, the sample is not random.

Evidence of invalid data as determined by these simple tests usually requires investigation and corrective action to establish conditions conducive to producing subsequent samples of data which will yield credible estimates.

7. Confidence Limits Calculations

After a data set has been plotted on probability plotting paper and a median regression "best fit" line has been drawn through the data as described in Section 5, confidence limits can be derived and plotted. Examples are shown in Figure D-1 (log-normal plot), Figure D-2 [Type I (linear) extreme value plot] and Figure D-3 [Type II (logarithmic) extreme value plot]. Theoretically, confidence limits can be calculated to any degree of confidence, but to simplify the application, tables of data are presented for calculating only the 95% and 80% confidence level. For most data plotting applications, these two levels are satisfactory and use of the appropriate table facilitates rapid plotting of the confidence level lines on the graph. The confidence interval for a line fitted to a set of data points, which are a sample from the population, should be visualized as a region within which the "true" line for the population may lie.

Alternately, a confidence interval may be considered as the region within which the specified percentage of additional samples will result in fitted lines within the confidence band. The location of the confidence intervals depend entirely on the fitted line. The amount of scatter in the points about the fitted line have no influence whatsoever on the location of the confidence interval. That is, the deviation of confidence limits assumes the data is representative of the distribution defined by the fitted line. As such, the confidence interval provide a test of data validity. For a 95% confidence interval, 5% of the data points may lie outside the interval; for 30% confidence, 20% of the points may lie outside the interval. A greater than expected percentage of points falling outside the confidence interval suggest the data are not valid; i.e., the points falling outside are not part of the distribution or the entire sample may be suspect. The source or cause of these points should be examined for differences from the other points and perhaps the data replotted after removing the "suspect" points.

Normally, a subjective estimate of the goodness-of-fit is adequate. That is, a good fit inspires a subjective degree of confidence while wide scatter instills less confidence. For those who desire a statistical test of whether the line fits all of the data, this method provides such a test.

7.1 Log-Normal Confidence Limits

Log normal confidence limits can be tabulated and plotted as shown in Table D-8 and Figure D-1. To construct a confidence limit table:

1. At the selected percentiles shown in Column A of Table D-8, tabulate as shown in Column B the values Y_p from the "best fit" line through the plotted data. The values used in this example are taken from Figure D-1.
2. Obtain the appropriate confidence factors from Table D-9, by interpolating for the sample size, if necessary, and enter the factors in Column C. The sample size n is the number of data points used to obtain the "best fit" line on the graph. In the example, Figure D-1, n is equal to 24.

TABLE D-8. EXAMPLE OF A TABLE FOR DETERMINING CONFIDENCE LIMITS ON A LOG-NORMAL PLOT

Column A Preselected Percentile	Column B ^a Value (Y_p)	Column C ^b Confidence Factors (95%)	Column D ^c $g_{x,y}$	Confidence Limits	
				Columns B x D Upper	Columns B/D Lower
5	27	0.831	2.141	57.8	12.6
10	37	0.701	1.901	70.3	19.5
30	75	0.482	1.555	116.6	48.2
50	120	0.425	1.476	177	81.3
70	195	0.482	1.555	303	125
90	390	0.701	1.901	741	205
95	550	0.831	2.141	1178	257

a. Values read from "best fit" line, Figure D-1.

b. Confidence factors interpolated from Table D-9.

c. $g_{x,y}$ is the ratio of $Y_{0.84}/Y_{0.50}$ read from "best fit" line on Figure D-1 where $Y_{0.84}$ and $Y_{0.50}$ are read at the 84 and 50 percentiles, respectively. Column D is generated by raising $g_{x,y}$ to the confidence factor powers from Column C.

D-9. CONFIDENCE FACTORS FOR THE NORMAL AND LOGARITHMIC NORMAL DISTRIBUTIONS

Sample Size (n)	Confidence Factors Preselected Percentiles									
	80%					95%				
	50	30 and 70	10 and 90	5 and 95	1 and 99	50	30 and 70	10 and 90	5 and 95	1 and 99
5	0.732	0.849	1.280	1.534	2.041	1.422	1.648	2.405	2.977	3.962
6	0.626	0.722	1.079	1.290	1.713	1.135	1.309	1.957	2.339	3.106
7	0.559	0.643	0.955	1.140	1.513	0.971	1.116	1.659	1.980	2.627
8	0.509	0.584	0.864	1.030	1.364	0.866	0.993	1.470	1.752	2.322
9	0.473	0.542	0.799	0.952	1.260	0.787	0.900	1.328	1.582	2.094
10	0.413	0.506	0.744	0.886	1.172	0.730	0.834	1.228	1.462	1.934
12	0.395	0.451	0.661	0.786	1.039	0.644	0.734	1.076	1.280	1.691
14	0.362	0.412	0.603	0.716	0.946	0.583	0.663	0.970	1.153	1.522
16	0.336	0.382	0.558	0.663	0.875	0.535	0.608	0.889	1.055	1.392
18	0.315	0.358	0.522	0.620	0.817	0.500	0.568	0.827	0.982	1.296
20	0.297	0.338	0.491	0.583	0.770	0.470	0.533	0.776	0.921	1.215
25	0.264	0.299	0.434	0.515	0.680	0.414	0.469	0.682	0.809	1.066
30	0.239	0.271	0.393	0.466	0.614	0.372	0.422	0.611	0.725	0.956
40	0.206	0.233	0.337	0.400	0.527	0.319	0.361	0.523	0.620	0.817
50	0.184	0.208	0.301	0.356	0.469	0.284	0.321	0.465	0.551	0.726
100	0.129	0.146	0.210	0.249	0.328	0.193	0.224	0.324	0.384	0.505

Intermediate values may be obtained by linear interpolation.

3. Determine the geometric dispersion $g_{y \cdot x}$ (the ratio of the value on the "best fit" line at the 84% percentile and the 50% percentile). For the example from Figure D-2, $g_{y \cdot x} = 300/120 = 2.50$. Raise $g_{x \cdot y}$ to the power of each factor in Column C and enter the results in Column D, (e.g., $g_{y \cdot x}^C$, Column D).
4. Multiply each value in Column B by the corresponding value in Column D to obtain points for plotting the upper confidence limit curve.
5. Divide each value in Column B, by the corresponding value in Column D to obtain the points for plotting the lower confidence limit curve.
6. At the appropriate percentiles from Column A, plot the points for the upper and lower confidence limit curves. Draw curves through the points using a French or a flexible curve.

7.2 Type I Extreme Value (Linear) Confidence Limits

Type I extreme value confidence limits can be tabulated and plotted as shown in Table D-10 and Figure D-2. To construct a confidence limit table:

1. At the selected percentiles shown in Column A of Table D-10, tabulate in Column B, the values Y_p from the "best fit" line through the plotted data on linear extreme value graphing paper. The Y_p values used in this example are taken from Figure D-2.
2. Obtain the appropriate confidence factors from Table D-11, by interpolating for the sample size if necessary, and enter the factors in Column C. The sample size n is the number of data points used to obtain the "best fit" line on the graph. In the example, Figure D-2, n is equal to 16.

TABLE D-10. EXAMPLE OF A TABLE FOR DETERMINING CONFIDENCE LIMITS ON A TYPE I EXTREME VALUE PLOT

Column A Preselected Percentiles	Column B ^a Value Y p	Column C ^b Confidence Factors (95%)	Confidence Limits		
			Column D ^c (1/ α)C	Columns B + D Upper	Columns B - D Lower
0.15	4850	0.615	826	5676	4024
0.30	5450	0.622	835	6285	4615
0.50	6200	0.707	950	7150	5250
0.70	7100	0.899	1207	8307	5893
0.85	8150	1.266	1700	9850	6450

- a. Values read from "best fit" line, Figure D-2
- b. Confidence factors from Table D-11 for 95% confidence.
- c. $1/\alpha = (Y_{0.85} - Y_{0.15})/2.457$ where $Y_{0.85}$ and $Y_{0.15}$ are the values read from the "best fit" line, Figure D-2, at the 85 and 15 percentiles, respectively. Column D is generated by multiplying $1/\alpha$ by each of the confidence factors for Column C.

3. Next, determine the Gumbel slope of the "best fit" line from the expression

$$\frac{1}{\alpha} = \frac{Y_{p.85} - Y_{p.15}}{2.457}$$

where

$Y_{p.85}$ = the value at the point on the "best fit" line where the cumulative probability is 0.85 and
 $Y_{p.15}$ is the value at the point where the cumulative probability is 0.15.

For example, from Figure D-2,

$$\frac{1}{\alpha} = \frac{8150 - 4850}{2.475} = 1343.$$

TABLE D-11. CONFIDENCE FACTORS FOR EXTREME VALUE DISTRIBUTIONS AT PRESELECTED PERCENTILE VALUES

Sample Size N	Confidence Factors Preselected Percentiles										
	95%					80%					
	0.15	0.30	0.50	0.70	0.85	0.15	0.30	0.50	0.70	0.85	
5	1.100	1.112	1.265	1.609	2.266	0.719	0.727	0.827	1.052	1.402	
6	1.004	1.015	1.154	1.468	2.068	0.656	0.664	0.755	0.960	1.352	
7	0.929	0.940	1.069	1.360	1.915	0.608	0.615	0.699	0.889	1.252	
8	0.869	0.879	1.000	1.272	1.791	0.568	0.575	0.654	0.832	1.171	
9	0.820	0.829	0.943	1.199	1.609	0.536	0.542	0.616	0.784	1.104	
10	0.778	0.786	0.894	1.137	1.602	0.508	0.514	0.585	0.744	1.048	
D-24	12	0.710	0.718	0.816	1.083	1.462	0.464	0.469	0.534	0.679	0.956
	14	0.657	0.665	0.756	0.961	1.354	0.430	0.435	0.494	0.629	0.885
	16	0.615	0.622	0.707	0.899	1.266	0.402	0.406	0.462	0.588	0.828
	18	0.580	0.586	0.667	0.848	1.194	0.379	0.383	0.436	0.554	0.781
	20	0.550	0.556	0.632	0.804	1.133	0.360	0.364	0.413	0.526	0.741
	25	0.492	0.497	0.566	0.719	1.013	0.322	0.325	0.370	0.470	0.663
	30	0.449	0.454	0.516	0.657	0.925	0.294	0.297	0.338	0.429	0.605
	35	0.416	0.420	0.478	0.608	0.856	0.272	0.275	0.312	0.398	0.560
	40	0.389	0.393	0.447	0.569	0.801	0.254	0.257	0.292	0.372	0.524
	45	0.367	0.371	0.422	0.536	0.755	0.240	0.242	0.276	0.351	0.494
	50	0.348	0.352	0.400	0.509	0.716	0.227	0.230	0.261	0.333	0.468
	100	0.246	0.249	0.283	0.360	0.507	0.161	0.163	0.185	0.235	0.331

Second Largest Value = 1.478; Largest Value = 2.236

Multiply $1/\alpha$ times each factor in Column C one at a time and enter the product in Column D.

4. Add the values in Column D to the respective values in Column B to obtain the points for plotting the upper confidence limit curve.
5. Subtract the values in Column D from the respective values in Column B to obtain the points for plotting the lower confidence limit curve.
6. At the appropriate percentiles from Column A, plot the points for the upper and lower confidence limit curves. Draw the curves through the points using a French curve or a flexible curve.

7.3 Type II Extreme Value (Logarithmic) Confidence Limits

Type II extreme value confidence limits can be tabulated and plotted as shown, in Table D-12 and Figure D-3. To construct the confidence limit table:

1. At the selected percentiles, shown in Column A of Table D-12, tabulate in Column B the values Y_p from "best fit" line through the plotted data points on logarithmic extreme value graphing paper. The Y_p values used in this example are taken from Figure D-3.
2. Obtain the appropriate confidence factors from Table D-11, by interpolating for the sample size if necessary, and enter the factors in Column C. The sample size n is the number of data points used to obtain the "best fit" line on the graph. In the example, Figure D-3, n is equal to 8.
3. Next, determine the geometric Gumbel slope of the "best fit" line from the expression

$$g' = \exp[\ln(Y_{0.85}/Y_{0.15})/2.457]$$

TABLE D-12.

Column A Preselected Percentiles	Column B ^a Value Y_p	Column C ^b Confidence Factor	Column D ^c g^c	Confidence Limits	
				Columns B x D Upper	Columns B/D Lower
0.15	2.5×10^4	0.869	1.57	3.93×10^4	1.59×10^4
0.30	3.1×10^4	0.879	1.58	4.90×10^4	1.96×10^4
0.50	4.2×10^4	1.000	1.68	7.06×10^4	2.50×10^4
0.70	6.0×10^5	1.272	1.94	1.16×10^5	3.09×10^4
0.85	9.0×10^4	1.791	2.54	2.29×10^5	3.54×10^4

a. Values read from "best fit" line at preselected percentiles, Figure D-3.

b. Confidence factors from Table D-11 for 95% confidence.

c. $g^c = \exp[\ln(Y_{0.85}/Y_{0.15})/2.457]$ where $Y_{0.85}$ and $Y_{0.15}$ are the values read from the "best fit" line, Figure D-3, at the 85 and 15 percentiles, respectively. Column D is generated by raising g^c to the confidence factor powers from Column C one at a time.

d. The upper and lower confidence limits at the preselected percentiles are obtained by multiplying and dividing Column B values by Column D values, respectively.

where $Y_{0.85}$ and $Y_{0.15}$ are the values picked from the "best fit" line at the cumulative probability 0.85 and the 0.15 points, respectively.

For example, from Figure D-3,

$$g^c = \exp[\ln(9.0 \times 10^4 / 2.5 \times 10^4) / 2.457]$$

$$g^c = 1.684.$$

Raise g^c to the power of each factor in Column C and enter the results in Column D.

4. Multiply the values in Column B by the values in Column D to obtain the points for plotting the upper confidence limit curve.
5. Divide the values in Column B by the values in Column D to obtain the points for plotting the lower confidence limit curve.
6. At the appropriate percentiles from Column A, plot the points for the upper and lower confidence limit curves. Draw curves through the points using a French curve or a flexible curve.

8. References

1. "Probability Charts for Decision Making," J. R. King, Industrial Press Inc., 200 Madison Avenue, New York, NY 10016, (Library of Congress Catalog Card Number 76-112966).
2. TEAM Easy Analysis Methods, Vol. 4, No. 1, 1977.
3. Ibid, Vol. 5 No. 1, 1978.
4. Ibid, Vol. 6 No. 1, 1979.
5. Ibid, Vol. 6 No. 2, 1979.
6. Ibid, Vol. 6 No. 3, 1979.
7. Ibid, Vol. 6 No. 4, 1979.
8. E. J. Gumbel, "Statistical Theory of Extreme Values and Some Practical Applications," February 1954, National Bureau of Standards Applied Mathematics Series, 33.

NOTE: TEAM Easy Analysis Methods publications are available from TEAM, Box 25, Tamworth, NH 03886, Telephone 603-323-8843.

APPENDIX E
RISK ASSESSMENT EXAMPLES

This page intentionally blank

APPENDIX E
RISK ASSESSMENT EXAMPLES

1. Logistics Probability and Risk Associated
With a Reactor Reflector^a

Apparently the decision to ship the latest replacement reflector on five separate trucks was based on an analysis made by an AEC statistician. Although we agree that the relative risk is reduced by this means, we are not convinced that the reduction in absolute risk is worth the increased cost.

Attached is a new study performed on the probabilities of cargo damaging accidents and associated risks of shipping the reflector from the manufacturer, American Beryllium Company, Sarasota, Florida. The results are briefly summarized for an optimized shipment (i.e., selection of the best trucks in the fleet; selection of the best drivers; selection of the best route, with an optimized driving schedule and truck spacing to avoid a cascading accident). Risk is the dollar value times the probability of loss.

1. The shipment could be made on one truck with a cargo damage probability of 7×10^{-4} . The loss of cargo risk is \$466. The loss of reactor operation risk is \$5250.
2. The shipment can be made by transporting the 10 reflector pieces in dissimilar pairs with a pair on each of five trucks. The cargo damage risk is essentially the same as for shipping all the packages on one truck. The risk from loss of reactor operation is reduced to \$37. This risk does not include a factor for programmatic disruption.
3. If no selection process is involved, to improve accident probabilities over the average for truck shipments, these risks could be expected to increase by at least a factor of 10.

a. This analysis is taken from a letter which has been disguised to conceal its origin.

In our opinion, if the risk involved in shipping via five trucks rather than one truck is weighed against the increased shipping cost, it is probably not worth the added cost to make a five-truck shipment. However, the customer's wishes and willingness to underwrite the cost of the incremental reduction (about \$5200) in risk must also be considered.

1.1 Statement of Problem

Ten beryllium reflector pieces, five of each type, are to be shipped via truck from the fabricator's shop to the reactor site. If an accident occurs, which prevents the receipt of four undamaged like sets of pieces, reactor operation could be delayed at considerable cost. It is proposed to ship the pieces on five separate trucks, in packages of two dissimilar pieces per package. How much is the probability of successful shipment of at least four sets of pieces improved by this method over shipping all of the pieces on one truck?

1.2 Assumptions

1. The probability of a cargo damaging accident is equal for all trucks. This assumes the trucks are in equal operating condition, driven by drivers of equal skill with equivalent driving records, over the same route.
2. All failure events (cargo damaging accident) and independent. (Trucks are spaced sufficiently to eliminate a cascading accident involving more than one truck.)

1.3 Solution

The binomial law states: If the probability of occurrence of an event in a single trial is p , then the probability that it will occur exactly r times in n independent trials is

$$p_r = {}_n C_r p^r (1-p)^{n-r}$$

where

$${}_n C_r = \frac{n!}{r!(n-r)!} .$$

This law can be extended to state: If the probability of the occurrence of an event on a single trial is p , then the probability that the event will occur at least r times in the course of n independent trials is

$$p_{\geq r} = p^n + {}_n C_1 p^{n-1} (1-p) + {}_n C_2 p^{n-2} (1-p)^2$$

$$+ \dots + {}_n C_{n-r} p^r (1-p)^{n-r} .$$

For our case the n independent trials are the five truck trips and the events r are cargo damaging accidents. Since two or more truck accidents are required to "fail the system" and delay reactor operation, the above equation reduces to

$$p_{\geq 2} = p^5 + 5p^4 (1-p) + 10 p^3 (1-p)^2 + 10 p^2 (1-p)^3.$$

This function was calculated for various values of p , the probability of a single truck having a cargo damaging accident (failure) and are tabulated below.

<u>p</u>	<u>$p_{\geq 2}$</u>
0.9	0.99954
0.8	0.99328
0.7	0.96922
0.6	0.91296
0.5	0.81250
0.4	0.66304
0.3	0.47178
0.2	0.26272
0.15	0.16479

<u>p</u>	<u>$p_{\geq 2}$</u>
0.10	0.08146
0.05	0.022593
0.01	0.00098015
0.005	0.00024751
0.001	0.000009980
0.0005	0.000002498
0.0001	0.000000100

Inspection of the table reveals facts which may not be intuitively obvious about this type logistic problem. If the probability of a truck accident is large, the probability of system failure (two or more accidents occurring in the five trials) is actually increased. Only when the single accident event probability is small is there an improvement in success probability.

This suggests that every effort should be made to reduce the value of p , the probability of a single truck accident while traveling the route.

This reduction is best accomplished by selecting the best trucks in the fleet, as demonstrated by inspection; selection of the best drivers with the best driving records; selection of the safest route with truck movement at controlled speed allowed only during the safest period of the day. It is expected that at least a factor of 10 improvement in safety could be realized over the average.

If the probability of a single truck accident can be reduced sufficiently, it then becomes important to look at the economics of shipping via five trucks or one truck. Is the probability of an accident so small, that the added cost of shipping the reflector on five trucks is justified?

Tri-State trucking statistics reveal that from 1964 to 1972, 3 million vehicle miles were traveled, and only one accident occurred which could be considered cargo damaging. It is assumed that the beryllium packages will be well constructed and properly tied down. Tri-State data should be characteristic of probabilities for well regulated shipments; i.e., the

extra mile for safety. (Reference 1 gives 3×10^{-7} /vm for moderate accident frequency and 8×10^{-9} /vm for severe accident frequency.)

Using the Tri-State frequency of 3×10^{-7} cargo damaging accidents per vehicle mile, the probability of the undesired event p is $(3 \times 10^{-7}) \times 2400$ miles (the shipment is to originate in Sarasota, Florida) or 7×10^{-4} . There is about one chance in 1400 of having a cargo damaging accident. By sending the packages on five trucks, the probability of at least two cargo damaging accidents is 5×10^{-6} , roughly a two order of magnitude improvement for system success over sending all the packages on one truck.

1.4 Risk

Risk is the loss due to an undesirable event times the probability of that event occurring. The risk to the cargo, where the cargo is \$666K worth of machined beryllium, is the same whether shipped via one truck or one-fifth of the cargo on each of five trucks. The risk, however, is distributed differently, since with five trucks there is nearly five times the probability of losing one cargo package, a small probability of losing two packages, and virtually no probability of losing more than two packages. These probabilities calculated using the binomial law and a value of 7×10^{-4} for p is tabulated below to illustrate this distribution.

Exactly 0 loss out of 5	0.9965
Exactly 1 loss out of 5	0.0034902
Exactly 2 losses out of 5	4.89×10^{-6}
Exactly 3 losses out of 5	3.43×10^{-9}
Exactly 4 losses out of 5	1.20×10^{-12}
Exactly 5 losses out of 5	1.68×10^{-16}

The total cargo loss risk for the beryllium is thus \$466.

The risk of keeping the reactor off-line due to loss of two or more packages based on a \$25K per day operating cost and a 10-month part replacement lead time is:

1. Five packages on one truck--\$25,000/day \times 300 days \times 7 \times 10^{-4} = \$5250
2. Five packages on five trucks--\$25,000/day \times 300 days \times 4.89 \times 10^{-6} = \$37 (requires two or more package losses).

These are actually the minimum risks, since no factor for programmatic disruption is included. The 10-month replacement time is based on the length of the present fabrication and test contract. It is assumed the beryllium is not destroyed in an accident, so the damaged material could be remelted, cast, and machined into the required reflectors (i.e., it is not necessary to find an additional source of high-grade beryllium).

2. Reference

Article in the NUCLEAR NEWS dated May 1973, entitled "Transportation Accidents: How Probable?", by William A. Brobst.

3. Risk From Aircraft to a New Support Building Addition

3.1 Summary

The proposed addition to the new support building, with a capacity of 500 personnel, will be located near the south end of the secondary runway. Approximately 20,000 landings or takeoffs per year, using the south end of this runway, result in flight paths near the proposed building. The probabilities of an aircraft impacting the building and the associated risks are given in Table E-1.

If the existing buildings are also considered, the values in Table E-1, with the exception of the individual risk value, are approximately doubled. The individual risk remains approximately the same. The best estimate values are approximately equivalent to a building impact of 1 in 12,000 years, a fatality risk of 1 in 1200 years, and an individual fatality risk of 1 in 600,000 years. The worst-case estimates are equivalent to a company

TABLE E-1. AIRCRAFT RISK TO THE ADDITION

	<u>Range</u>	<u>Best Estimate</u>
Annual building crash probability	9×10^{-6} to 4×10^{-4}	9×10^{-5}
Expected fatalities per year	9×10^{-5} to 4×10^{-3}	9×10^{-4}
Individual annual fatality risk	1.7×10^{-7} to 8×10^{-6}	1.7×10^{-6}
Annual dollar ^a risk	\$17 to \$800	\$170

a. \$200,000 per fatality assumed.

fatality of 1 in 250 years, and an individual risk of 1 in 100,000 years. For comparison, the estimated probability of a company fatality, for all operations, is 1 in 40 years, and the average citizen's fatality risk is 1 in 2000 years from all accidents.

3.2 Introduction

The new building addition is located near the flight path for landing or takeoff from the south end of the secondary runway at the local airport. A proposal to expand the existing facility to house an additional 500 personnel prompted an inquiry into the risk of an aircraft crashing into the building. This report is an assessment of that risk.

3.3 Description

The local airport has two runways. The primary one is instrumented and runs from northeast to southwest. The smaller, secondary runway begins at the northeast end of the primary runway, and extends nearly due south for about 1.6 km (1 mile). This runway is not instrumented and is normally used only by small aircraft.

The proposed addition will house about 500 personnel and will be two stories [about 8 m (24 or 25 ft)] high, with an area of 4500 m^2 ($49,000 \text{ ft}^2$). It will be located about 400 m (1300 ft) south and 200 m

(650 ft) east of the south end of the secondary runway. The perpendicular distance to the primary runway is about 1.4 km (7/8 mile).

In 1976, there were approximately 80,000 landings and takeoffs at the local airport, with about 8000 of these consisting of commercial aircraft landings or takeoffs. Nearly one-half of the small aircraft landings and takeoffs use the secondary runway, while nearly all of the large aircraft and the remainder of the small aircraft operations are from the primary runway.

3.4 Risk Analysis

An analytical method has been developed and used to determine the crash probability into major buildings surrounding the Los Angeles Airport.¹ Information given in this report (Reference 1) indicates that the probability decreases with distance from the end of the runway, being twice as large at 1 km as at 2 km. Approximately one-half of the crashes occur within 8 km (5 miles) of the runway. There is also an angular probability dependence, with the highest probability occurring at zero degrees. (The angle is defined by the intersection of the line formed by the runway and a line from the takeoff or intended landing point to the crash site.) However, specific data and equations were not given in this reference.

Sandia has used this method to calculate aircrash probabilities for two buildings located near the Albuquerque airport.³ This reference does include all of the information necessary to perform a similar analysis. Therefore, the equations and probability data given here were taken directly from the Sandia study.

3.4.1 Probability Equation. The annual probability that an aircraft will strike the proposed building addition is:

$$P_B = N A f(x) P_c \quad (1)$$

where

N = the number of operations/yr

A = the effective building area (km^2)

$f(x)$ = a distribution function for crash probability (per km)

P_c = probability of crash per km of flight.

A summation equation is given in Reference 2, which must be used if different types of aircraft, or different modes of operation, or more than one flight path are considered. Since we are considering only the south approach to the secondary runway and only general aviation (which has equal crash probabilities for takeoffs and landings), the summation is not needed. (The inflight mode is neglected because of the lower crash probability and lower number of flights near the building.)

3.4.2 Air Traffic. One-fourth of the 80,000 landings and takeoffs are assumed to use the south end of the secondary runway. All of these 20,000 operations are assumed to be general aviation aircraft. The other operations are neglected because they do not occur near the proposed facility. Therefore, $N = 20,000$ operations/year.

3.4.3 Effective Building Area. The effective building area is the roof area of 4500 m^2 ($49,000 \text{ ft}^2$), plus a "shadow" area of 1500 m^2 ($14,600 \text{ ft}^2$) defined by an assumed glide angle of 20 degrees, a building height of 8 m (25 ft), and a width of 70 m (230 ft). This gives a total area of 0.006 km^2 ($63,600 \text{ ft}^2$). A "skid" area, defined as the width of the building multiplied by the average skid length of 100 m (328 ft) or 0.007 km^2 (7200 ft^2) for general aviation aircraft is omitted because the kinetic energy would be rapidly dissipated so that the consequence of a light plane skidding into a building, on the average, would be much less than that from a direct strike above ground level. Therefore,
 $A = 0.006 \text{ km}^2$.

3.4.4 Impact Distribution. Crash probabilities are given for three different modes of flight: landing, takeoff, and inflight. (Landing and takeoff are defined as any operation within 8 km of the airport.) The inflight probabilities are considered constant for a given type of aircraft, but probabilities for the other two modes must be modified by a distribution function, which is the same for both modes and is given by:

$$f(x) = 1/2 \gamma e^{-\gamma x}$$

where

x = the perpendicular distance from the intended flight path to the new flight path to the new facility (0.2 km)

γ = 1/24/km

$f(x) = 0.48/\text{km}.$

3.4.5 Crash Probability. The crash probabilities per 10^5 km of flight, for different classes of general aviation, range from 0.095 to 0.26 for landing, and from 0.063 to 0.39 for takeoffs. However, these probabilities, averaged for all general aviation flights, are equal for takeoffs and landings, at 0.15 per 10^5 km. (Aircarrier crash probabilities are only 0.007 for takeoffs and 0.028 for landings.) Since the general aviation average is within approximately a factor of two of the specific values, it is used without attempting to determine the numbers of flights for the different types (charter, pleasure, etc.) within the general aviation category. Therefore, $P_c = 1.5 \times 10^{-6}/\text{km}.$

3.4.6 Probability of Building Impact. Using Equation (1) and the above input values, the probability of building impact is:

$$P_B = N A f(x) P_c = (20,000) (0.006) (0.48) (1.5 \times 10^{-6})$$

$$P_B = 8.6 \times 10^{-5} \text{ impacts/year.}$$

This is approximately equivalent to one impact in 12,000 years (or one in 6000 years if skidding into the building is included).

3.4.7 Consequences. Reference 1 gives a destruction coefficient of 0.06 for light aircraft impact into the average office building. This means that 6% of the persons in the average building would be killed. However, it appears that the percent of the occupants suffering a fatality would not be a constant 6%, but would be smaller for larger office buildings. A consensus of several engineers is that no more than several offices in the building would be destroyed, and that 10 fatalities is a reasonable estimate. Injuries and property damage are small by comparison. With an assumed value of \$200,000 per fatality, the consequence of a light aircraft striking the building is estimated at 10 fatalities or approximately \$2,000,000.

3.4.8 Risk. Risk can be thought of as the amount of insurance which will cover expected losses, and is defined as the consequence of an event multiplied by its probability of occurrence:

$$\begin{aligned} \text{Risk} &= P_B C = (8.6 \times 10^{-5}) (10 \text{ fatalities or } \$2,000,000) \\ &= 8.6 \times 10^{-4} \text{ fatalities/year or } \$172/\text{year.} \end{aligned}$$

The fatality rate is equivalent to about one fatality in 1200 years. With 500 personnel in the building, the annual individual risk is only one in 600,000. If the probability of skidding into the building is included, the risk estimate is increased, but not doubled, because the consequence would be much less than 10 fatalities. Also, if the existing buildings and personnel are included, the company risk is doubled, but the individual person's risk remains the same, since the total fatality risk is shared by twice as many persons. In comparison, a company accident fatality is expected once in 40 years, and an individual employee's chance per year of a fatal occupational accident is one in 120,000 years.

3.5 Discussion and Conclusion

The above risk values are considered best estimates. To determine the degree of uncertainty, a worst-case was calculated using the following conservative assumptions:

1. The air traffic is projected to increase by 50%.
2. A glide angle of 10 degrees rather than 20 degrees is assumed. This increases the shadow area by a factor of two and the building (target) area by 23%. Adding the skid area, but compensating for the less severe consequences of a skidding collision gives an estimated 50% increase in risk.
3. The flight path is 120 m (400 ft), rather than 200 m (650 ft) from the building. This increases the distribution function to 0.53 from 0.48, a 10% increase.
4. Aircraft from the primary runway are included. This does not double the risk because the distribution function is only 0.11 vs. 0.48 for the secondary runway. Adding 0.11 to 0.48 gives an increase of 23%. (Note that the building is not located near the airpath during takeoff or landing, but is about 1.5 km perpendicular to the center portion of the primary runway. The distribution function is thus probably strongly conservative in this instance. Also note that consideration of large aircraft leads to no change in risk estimate. Aircarrier crash probability is approximately 1/10 of that of general aviation, but the consequences are probably 10 times as severe.)
5. The mean takeoff and landing crash probabilities per 10^5 km for general aviation are: instructional--0.08; business--0.09; pleasure--0.32; aerial application--0.13; airtaxi--0.08. Therefore, a disproportionate number of pleasure craft using the secondary runway, compared to the national average would increase the crash probability above the 0.15 national average. Since it

is known that not all aircraft at the field are pleasure craft, the worst-case is estimated to be a 50% increase in risk.

When the above factors are compounded, the worst-case risk is approximately 4.5 times as great as the calculated "best estimate" value. On the other hand, the following conservatisms were not included in the "best estimate" calculation:

1. No credit was taken for shielding of the facility by other buildings. (There are several large buildings located between the runway and the new facility.)
2. No credit was taken for evasive action. (A pilot will attempt to avoid a structure.)
3. The office building is occupied only 8 hours per day, 5 days per week.
4. All operations using the secondary runway are under visual flight rules, while those operations under instrument flight rules likely have a higher crash probability.

A subjective estimate, considering the above conservative factors and optimistic (rather than worst-case) input parameters is that the risks are about a factor of 10 less than the best estimate values.

The resulting optimistic estimates, and the worst-case values, define the ranges of uncertainty in the risk. These ranges are given in Table E-1, together with the best estimate values.

4. References

1. D. K. Okrent, "Airplane Crash Risk to Ground Population," Hazard Prevention, 11, 3 (January/February 1975).
2. Betty Biringer, "Assessment of the Probabilities of Aircraft Impact with the Sandia Pulsed Reactor and Building 836, Sandia Laboratories, Albuquerque," SAND76-0366, November 1976.